Engaging diffractive ethnography to explore student and teacher perceptions of collaborative testing

Helen Bremert

A thesis submitted to fulfill requirements for the degree of Doctor of Philosophy, 2024

Sydney School of Education, Faculty of Arts and Social Sciences, The University of Sydney

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.
I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Acknowledgments

Debra, my incredible principal supervisor, no words can fully capture how grateful I am to have had your unwavering support through this long and arduous experience. Your insistence that I find my own way was both challenging and empowering, and I now realise how much it shaped this thesis and made it more meaningful. Introducing me to Karen Barad's book was a revelation, making me feel like I had found my intellectual place in the world. Your calm confidence, especially when you said, "I thought you'd get here," kept me going, especially through the pandemic and school closures. Your guidance has not only shaped this thesis but also helped me grow as a researcher. Thank you, for being such a vital part of this journey.

Claire, you came into this supervisory team just at the right time, and your presence has been invaluable to me. Your support, keen attention to detail, and ability to express exactly what I needed to know and when I needed it most, helped me get to this point. As I write this, I feel that thank you isn't enough to convey how happy I am that you joined the team, as I couldn't have reached this milestone without your guidance. Thank you again.

Jim, thank you for being at the start and end of this journey. Your encouragement, different perspectives, and words of wisdom helped focus this research. Thank you for your gracious support.

Nicole Brunker, you started this journey with me, and your support at the beginning of this thesis was encouraging and insightful. Thank you for being a valuable part of this experience.

I sincerely thank the teachers, Halo, Beta and Zeta, and all the other teachers and their students who participated in this study. Your encouragement, thoughtful insights, and the generous time you gave during the study, made a significant difference to this project. I truly appreciate all your contributions.

Karen and Pam, I owe a deep thanks. Karen, without your guidance, this thesis would not have been possible. Your willingness to trial 'new' methods was the motivation for my study. Pam, your support was what I needed, and I loved working with both of you. Thank you to the students I have taught over the years.

To Dr Penelope Hayes, thank you so much for proofreading my thesis. Your attention to detail in formatting and catching those small errors has been invaluable. I truly appreciate your help and service.

There are so many people who have kept me grounded over the years, far too many to name them all. Marion, Jenny, and Frances, words cannot express how thankful I am to have had you as a part of my life for so long. You have stood by me through the years, as I navigated this journey while moving and just being me, always offering your wisdom and unwavering support. I love you and am endlessly grateful for your presence and support when I needed it the most.

Phillip and Jill, my PhD partners at Sydney University, our discussions and the writing retreat were invaluable, and your insights made this journey so much more enjoyable. Moreover, to Landon, David, Adam, and everyone at my oasis, you kept me sane these last few months. Thank you for our conversations, the interest you showed, and the laughter we shared.

Campbell and Oliver, thank you for being in my life. I have always wanted to show you that the future is not set; you can be whoever you want to be at any stage in life. Life is meant to be lived well; never let anyone hold you back. Frodo, you were the perfect 'person' to have around during this last year. I would talk about the thesis, and you would purr, look at me, and walk away.

Alastair, I genuinely don't know where to begin. This has been an incredible journey, and I am so grateful that we have walked it together. Somehow, I have 'got it done,' but I know I couldn't have done it without you. Your unwavering support, through all the doubts and every step forward, means more to me than words can express. I feel so lucky that we found each other. Having you in my life makes everything better, and I can never thank you enough for being there through it all.

Authorship statement

Data collected as part of this project was used in the following publications and presentations during the period of my candidature:

Bremert, H. (July 2022). Exploring student and teacher perceptions of a novel assessment method to enhance learning and engage STEM students. *Network Gender and STEM Annual Conference, 2022*, Munich.

Bremert, H. (July 2023). Engaging diffractive ethnography to explore student and teacher perceptions of collaborative testing, to enhance learning and engage STEM students. *The International Academic Forum (IAFOR), The European Conference on Education (ECE), 2023*, London.

Bremert, H. (August 2023). Engaging diffractive ethnography to explore student and teacher perceptions of collaborative testing to enhance learning and engage STEM students. *European Conference on Education Research (ECER) Emerging Researchers Conference, 2023*, Glasgow.

Bremert, H. (September 2024). Engaging diffractive ethnography to explore student and teacher perceptions of collaborative testing. *British Education Research Association (BERA) Annual Conference, 2024*, Manchester.

Abstract

Engaging Diffractive Ethnography to Explore Student and Teacher Perceptions of Collaborative Testing

High-stakes testing has long been a pivotal element of educational policy and practice, essential for evaluating student performance, teacher effectiveness, and school quality both in Australia and globally. However, critiques of Australia's senior secondary assessment system indicate a misalignment with modern educational and workplace demands. Critics argue that the current education system overly emphasises discipline-based knowledge, thus undervaluing 21st-century skills such as critical thinking, communication and collaboration. This misalignment sets aside the holistic development of students and distorts teaching practices due to the pressure of high-stakes exams. In order to address these issues, a balanced approach is advocated that values both discipline-focused and general capabilities by integrating flexible and varied assessments, including teacher assessments, external exams, online assessments and collaborative testing. Such a diverse evaluation system would more accurately measure students' comprehensive abilities and potential. In response to these challenges, this study explores collaborative testing as a supplementary assessment method that emphasises 21st-century skills often overlooked in high-stakes environments.

Employing diffractive ethnography, aligned with Karen Barad's (2007) agential realism, this research delves into the complex dynamics of collaborative testing. This method is particularly suited for exploring how educational practices intra-twine with material conditions, providing a robust framework for assessing the intra-actions within collaborative environments. This approach is grounded in the principles of entanglements, phenomena, and the dynamic role of the apparatus in shaping educational realities: it analyses key elements such as the subject taught, gender distribution, ability, test length and school, alongside environmental factors like ambient temperature, noise levels, classroom setup and test timing. Diffractive ethnography views these components not just as a backdrop but as active participants in the educational process, influencing and being influenced by human actors and pedagogical frameworks, playing a critical role in educational landscapes. This approach facilitates a deeper understanding of how knowledge and educational realities are continually produced and transformed through ongoing intra-actions, offering a nuanced perspective of the collaborative testing environment.

Findings from this study suggest that collaborative testing enhances student outcomes by promoting active peer discussions, critical thinking and the development of 21st-century skills. Effective implementation necessitates certain conditions, such as positive interdependence among students, prior practice with structured collaborative activities, ample time for in-depth discussions and strategic student grouping. Moreover, external tests emerged as a significant barrier to implementing collaborative testing, highlighting a systemic constraint faced by teachers across the three sites. The insights gleaned demonstrate the benefits and challenges of collaborative testing and set the stage for discussing its broader implications for educational assessment practices.

The research contributes to the broader discourse on educational assessment by demonstrating the feasibility and benefits of collaborative testing in high schools. It also showcases the innovative application of diffractive ethnography to explore complex educational phenomena. This methodological approach not only reveals new insights into student and teacher experiences but also underscores the importance of the material environment in shaping learning processes.

Table of Contents

STATEMENT OF ORIGINALITY **2**

ACKNOWLEDGMENTS **3**

AUTHORSHIP STATEMENT **5**

ABSTRACT **6**

Engaging Diffractive Ethnography to Explore Student and Teacher Perceptions of Collaborative Testing 6

LIST OF ABBREVIATIONS **11**

APPENDICES **12**

LIST OF TABLES **12**

LIST OF FIGURES **13**

CHAPTER 1: INTRODUCTION **14**

Introduction to the Research Problem 15
Research Aims and Significance 16
Theoretical Foundations: Bridging Agential Realism and Materialism with Educational Practice 16
Methodological Approach 18
Thesis Outline 18
Conclusion 20

CHAPTER 2: LITERATURE REVIEW **21**

Part 1: Assessments and Their Role in Education 21
Managerialism of Education 23
Assessment in Australian Secondary Schools 26
Norm-Referenced and Criterion-Referenced Assessment 29
Standards-Referenced Assessment 30
Formative and Summative Assessment 31
Defining High-Stakes Testing 32
Impact of High-Stakes Testing on Student Learning 33
Surface and Deep Learning in the Context of High-Stakes Testing 35
Impact of High-Stakes Testing on Students' Anxiety and Well-being 37
Impact of High-Stakes Testing on Curriculum 38
Impact of High-Stakes Testing on Teachers' Pedagogy 41
High-Stakes Testing and Teachers' Wellbeing 43
Summary 44
Part 2: Collaboration and Collaborative Testing 45
Collaboration *45*
Collaborative Testing *47*
Collaborative Testing Overview *47*
Collaborative Testing and 21st Century Skills 52
Summary 53
Conclusion 54

CHAPTER 3: METHODOLOGY **55**

From Observer to Participant: Evolving Research Paradigms 55
Introduction to Onto-epistemology and Educational Implications 57
Agential Realism 58
Phenomena and Apparatus in Research Design 59
Phenomena *59*

Apparatus *60*
Material-discursive practices 61
Collaboration as a Material-Discursive Practice 62
Entanglement in Educational Practice 64
Ethnography and its Evolution Towards Diffractive Ethnography 64
Diffractive Ethnography 67
Components of the Diffraction 69
Transformative Nature of Data in Diffractive Methodology 71
Diffraction Methodology Applied 72
Conclusion 74

CHAPTER 4: RESEARCH DESIGN 75

Approach to Research 75
Embracing Diffraction over Reflexivity in Qualitative Research 76
Methodological Approach 77
Criteria for Selecting and Excluding Classes 78
Diverse Educational Contexts: Polaris High School and Ceres Independent High School 84
Engaging with the Material-Nonmaterial Elements 85
Evolving Research Methodologies: Integrating Fixed Methods with Dynamic Techniques 86
Collaborative Tests 87
Conversations 88
Audio Recording 91
Mind Maps 91
Observation and Journal Notes 92
Research Phases 93
Phases of Data Collection 95
Phase One (Early December 2022) *95*
Phase Two (Late February – April 2023) *95*
Phase Three (Immediately after Test Three) *95*
Phase Four (Ending of Testing Period) *95*
Discussion of Data and Methodology 96
Research Design, Strength and Limitations 97
Ethical Considerations 98
Chapter Structure and Methodological Approach 99
Conclusion 101

CHAPTER 5: HALO'S ENTANGLEMENT 102

Part 1: Setting the Stage with Collaborative Tests and Mind Map 103
Collaborative Test One *103*
Collaborative Test Two *110*
Collaborative Test Three *112*
Mind Map *114*
Halo-Researcher-Conversation-Collaboration-Environment 115
Part 2: Analysis of the Collaborative Testing Entanglement 117
Student Understanding *117*
Halo's Perceptions of Students' Understanding *120*
Collaboration – 21st Century Skills *121*
Halo's Perceptions of Students' 21st-Century Skills *122*
Part 3: Exploring the Material of the Classroom Environment 123
Part 4: Unique Observations and Broader Implications 126
Perceptions of Halo and Students on the Summative Collaborative Test *126*
Collaborative Testing and Accountability *127*
Collaborative Testing and Halo's Practise *128*
Conclusion 130

CHAPTER 6: BETA'S ENTANGLEMENT 131

Part 1: Setting the Stage with Collaborative Tests and Mind Map 132

Collaborative Test One *132*
Collaborative Test Two *135*
Collaborative Test Three *137*
Mind Map *139*
Beta-Researcher-Conversation-Collaboration-Environment 139
Part 2: Analysis of the Collaborative Testing Entanglement 141
Part 3: Exploring the Material of the Classroom Environment 145
Enhancing Critical Thinking Through Test Design *145*
Temperature Effects on Student Understanding *146*
Part 4: Unique Observations and Broader Implications 149
Strategic Group Formation in Collaborative Testing *149*
Messaging by Beta on Collaborative Testing *151*
Conclusion 154

CHAPTER 7: ZETA'S ENTANGLEMENT 155

Part 1: Setting the Stage with Collaborative Tests and Mind Map 156
Collaborative Test One 156
Collaborative Test Two *159*
Collaborative Test Three *161*
Mind Map *162*
Zeta-Researcher-Conversation-Collaboration-Environment 163
Part 2: Analysis of the Collaborative Testing Entanglement 165
Part 3: Exploring the Material of the Classroom Environment 170
Student Comfort *170*
Activating Prior Knowledge *171*
Part 4: Unique Observations and Broader Implications 172
Cognitive Fatigue *172*
Modelling Collaborative Testing *174*
Conclusion 175

CHAPTER 8: THE INTRA-ACTIONS OF COLLABORATIVE TESTING 176

Cross-Case Analysis 176
Peer Discussions and Argumentation in Collaborative Testing 177
Building Argumentation Skills Through Active Peer Discussions 180
Enriching Understanding Through Peer Discussions and Multiple Perspectives 183
Activating Prior Knowledge Through Peer Discussions in Collaborative Testing 184
Summary *185*
High-Stakes Testing and its Impact on Peer Discussions in Collaborative Testing 185
The Role of Group Dynamics in Shaping Peer Discussions During Collaborative Testing *187*
Impact of Group Formation on Peer Discussions During Collaborative Testing *188*
Addressing Gender Imbalances on Group Dynamics in Collaborative Testing *189*
Cultivating 21ST-Century Skills Through Collaborative Testing 190
Shaping Peer Discussions: The Material Environment's Role in Collaborative Testing 191
Summary *194*
Implementing Collaborative Testing: A Research-informed Guide for Teachers 195
Preparing Students for Collaborative Testing *195*
Designing Effective Collaborative Test Questions *196*
Creating a Supportive Learning Environment *196*
Optimising the Physical Classroom Environment *197*
Summary *197*
Conclusion 198

CHAPTER 9: CONCLUSION 199

Q1: How Did Students and Teachers Perceive Collaborative Testing? 200
Q2: How Does Collaborative Testing Impact Student Learning? 200
Q3: How Does Implementing Collaborative Testing Unfold in a Classroom? 201
The Material Impacts on Student Learning 202

Demonstrating the Impact of Collaborative Testing in High School Science Classes 202
The Influence of High-Stakes Testing on Collaborative Testing in High Schools 204
Innovative Methodological Contribution: Diffractive Ethnography in Educational Research 205
Implications and Possibilities for Future Research 206
Conclusion 208

REFERENCES 209

APPENDICES 254

Appendix A: University of Sydney Ethics Approval 254
Appendix B: Approval to Conduct Research in NSW Public Schools (SERAP) 256
Appendix C: Participant Information Sheets 257
Appendix D: Example of Test Questions 266

List of Abbreviations

ACARA	Australian Curriculum, Assessment and Reporting Authority
ACT	American College Test
AP	Advanced Placement
ATAR	Australian Tertiary Admission Rank
CF	Cognitive Fatigue
CGW	Collaborative Group Work
CRA	Criterion-Referenced Assessment
DET	Department of Education and Training
EBV	Estimated Breeding Values
GCSE	General Certificate of Secondary Education
GERM	Global Education Reform Movement
HOTQ	Higher-Order Thinking Questions
HSC	Higher School Certificate
HST	High-Stakes Testing
NAPLAN	National Assessment Program Literacy and Numeracy
NESA	New South Wales Education Standards Authority
NSW	New South Wales
PISA	Program for International Student Assessment
QCAA	Queensland Curriculum and Assessment Authority
QCE	Queensland Certificate of Education
SAT	Standards Admissions Test
SERAP	State Education Research Application Process
STEM	Science, Technology, Engineering, Mathematics
VA	Value-Added
VET	Vocational Education and Training

Appendices

Appendix A: University of Sydney Ethics Approval p. 257

Appendix B: Approval to conduct research in New South Wales (NSW) public schools (SERAP) p. 259

Appendix C: Participant information sheets p. 260

Appendix D: Example of Test Questions p. 269

List of Tables

Table 1	*School Pseudonym, Teacher Pseudonym, Selective, Community or Mixed Students, and Subject Taught*	p. 82
Table 2	*Collaborative Test Content for each Collaborative Test*	p. 84
Table 3	*The Setting for Pre-Collaborative Testing Conversations with Teachers*.	p. 90
Table 4	*The Setting for Post-Collaborative Testing Conversations with Teachers*	p. 91
Table 5	*Phases, Methodological Approaches, Techniques, and Emergent Data for Collaborative Testing*	p. 95

List of Figures

Figure 1	*Example of an HSC Syllabus Point from Module 5: Equilibrium and Acid Reactions*	p. 40
Figure 2	*Diagram of the Apparatus for each Collaborative Test*	p. 62
Figure 3	*Illustration of the Interference Pattern Resulting from the Diffraction of Light Through Two Slits.*	p. 69
Figure 4	*Diffraction of the Intra-Action for …Collaborative Test.*	p. 71
Figure 5	*Classroom Setup for the Three Collaborative Tests*	p. 86
Figure 6	*Student-Led Discussion Questions*	p. 91
Figure 7	*Individual Student Mind Map Questions.*	p. 93
Figure 8	*Illustration of the Phases and Techniques for Collaborative Testing*	p. 96
Figure 9	*Diffraction of the Entanglement for Halo, Collaborative Test One*	p. 104
Figure 10	*Photo of Halo's Classroom at Polaris High School*	p. 106
Figure 11	*Diffraction of the Entanglement for Halo, Collaborative Test Two*	p. 111
Figure 12	*Diffraction of the Entanglement for Halo, Collaborative Test Three*	p. 113
Figure 13	*Diffraction of the Entanglement for Halo, Collaborative Test Mind Map*	p. 115
Figure 14	*Diffraction of the Entanglement for Beta, Collaborative Test One*	p. 133
Figure 15	*Beta's Classroom Layout, Polaris High School*	p. 134
Figure 16	*Diffraction of the Entanglement for Beta, Collaborative Test Two*	p. 136
Figure 17	*Diffraction of the Entanglement for Beta, Collaborative Test Three*	p. 138
Figure 18	*Diffraction of the Entanglement for Beta, Collaborative Test Mind Map*	p. 140
Figure 19	*Smoke Encroachment at Western Sydney School*	p. 149
Figure 20	*Diffraction of the Entanglement for Zeta, Collaborative Test One*	p. 157
Figure 21	*Zeta's Classroom at Ceres Independent School.*	p. 158
Figure 22	*Diffraction of the Entanglement for Zeta, Collaborative Test Two*	p. 160
Figure 23	*Diffraction of the Entanglement for Zeta, Collaborative Test Three*	p. 162
Figure 24	*Diffraction of the Entanglement for Zeta, Collaborative Test Mind Map*	p. 163

Chapter 1: Introduction

The essence of this study revolves around the premise: if we continue with what we have always done, we will only achieve the same outcomes. My journey as a high school science teacher in both Australia and the United States exposed me to the pressures and limitations of high-stakes testing environments, sparking my interest in exploring alternative assessment methods. Initially, my assessment methods strictly followed the prescribed curriculum and the practices endorsed by more experienced colleagues, which primarily consisted of testing every topic. Fortunately, working with a head science teacher who modelled collaborative work in senior subjects broadened my perspective. As I gained confidence, I began integrating more collaborative activities into my teaching.

As the Higher School Certificate (HSC)[1] and Advanced Placement (AP)[2] exams approached, I shifted focus to exam preparation, ensuring I covered all topics and having students work through past exam papers. This approach increased anxiety levels in some students as the exam dates neared. Reflecting on these experiences, I began to question the impact of high-stakes tests on my students. While I recognised that I could not change the structure of significant exams like midterms, end-of-year exams, or external tests, I wondered whether there could be alternative methods beyond traditional individual testing to support student learning and development, especially in Australia, where school-based assessment plays a significant role in the final grade. I further observed a recurring pattern: I undertook collaborative activities within my classes; however, assessments remained individual, prompting me to think there has to be another way to assess students.

During my Chemistry Education Masters research, I came across a paper outlining collaborative testing. My investigation revealed that while many STEM (Science, Technology, Engineering, Mathematics) undergraduate courses describe benefits, such as improved understanding and engagement, few papers undertook collaborative testing in high schools (Eastwood et al., 2020; Reiger & Heiner, 2014). My discovery of collaborative testing in an undergraduate science class, which incorporated a blend of individual and collaborative tests (Miller et al., 2022; Newton et al., 2019; Reiger & Heiner, 2014), offered a new perspective. Inspired, I launched a pilot project of collaborative testing in my classes—in which positive feedback was gathered from an online anonymous survey,

[1] The Higher School Certificate (HSC) in New South Wales (NSW) is a qualification awarded to students who successfully complete senior high school level studies (Years 11 and 12), assessing them on a combination of coursework and final examinations (NESA, 2022).

[2] Advanced Placement (AP) are college level course that students take in high school, overseen by the College Board (CollegeBoard.org).

where students indicated improved understanding and reduced anxiety. I then undertook an action research study in AP environmental science classes over one semester, where the students undertook three purely collaborative tests. Consistent with the research from undergraduate courses, they mainly reported enhanced understanding of concepts and better communication and collaborative skills (Bremert et al., 2020). Consequently, this experience sparked a deeper exploration into collaborative testing, leading to the development of this thesis, which addresses the lack of research into what may promise to be an assessment method that improves knowledge and promotes the development of essential 21st-century skills.

Introduction to the Research Problem

High-stakes testing, a standard benchmark for student achievement, often imposes undue stress on students and distorts educational priorities (Baird et al., 2017; Wilson et al., 2021). Globally, education systems are increasingly shaped by policies emphasising accountability and performance metrics, driven by international comparisons and the quest for global competitiveness (Appel, 2020). Influential initiatives such as the "A Nation at Risk" report in the USA and the UK's Education Reform Act of 1988 have spurred extensive educational reforms aimed at standardisation and measurable outcomes (Lingard & Sellar, 2016; Mahon et al., 2019). This trend, often encapsulated by the Global Education Reform Movement (GERM) (Sahlberg, 2016, p. 132), has positioned high-stakes testing at the core of student and school evaluations, a trend observed in Australia and globally (Au, 2022; Ball, 2017; Fuller & Stevenson, 2019; Masters, 2013).

Critics of Australia's senior secondary assessment system, such as Andrade & Brookhart (2020) and Gonski et al. (2018), argue that there is a misalignment with modern educational and workplace demands. The focus on discipline-based knowledge overshadows essential 21st-century skills like critical thinking, collaboration, and communication (Masters, 2020; Wilson et al., 2021). These critiques underscore the need for a balanced approach that integrates both discipline-specific knowledge and general capabilities through flexible and varied assessments, including teacher assessments, online assessments, and collaborative testing, to measure students' abilities more accurately (Cumming, 2020; Masters, 2020; Wilson et al., 2021). Within this context, collaborative testing emerges as a viable supplementary method that promotes the often neglected 21st-century learning skills in high-stakes environments. This shift necessitates adopting a robust theoretical perspective; thus, agential realism and an understanding of materialism offer a framework that aligns with my study's proposed methodologies.

Research Aims and Significance

My thesis, '*Engaging Diffractive Ethnography to Explore Student and Teacher Perceptions of Collaborative Testing*,' outlines the focus of my study, which is exploring student and teacher perceptions of collaborative testing within high school science classrooms. I needed a methodology encompassing humans, matter, and nonmatter to understand all the elements that shape perceptions (Barad, 2007). A diffractive analysis encourages a researcher to be in the research, to immerse themselves within the setting, and to fully consider how various factors can impact students as they undertake collaborative testing (Barad, 2007). The research questions guiding this study are:

1. How do students and teachers perceive collaborative testing?
2. How does collaborative testing impact student learning?
    a. Which collaborative testing methods do students believe enhance their understanding?
    b. Why do they believe this?
3. How does implementing collaborative testing unfold in a real classroom?

My research aims to contribute to the broader discourse on educational assessment that is driven by globalisation and the managerialism of education, promoting outcome based policies impacting student learning and equity (Zajda, 2021). This is achieved by investigating student and teacher perceptions of collaborative testing, including its implementation and impact in a high school setting. This study was conducted in three science classes in two demographically different high schools in Australia. I describe the contexts and testing environments for the classes, discuss teachers' and students' perceptions of collaborative testing, and analyse how students and teachers perceive the efficacy of collaborative testing. This study also illustrates an innovative application of diffractive ethnography to explore complex educational phenomena, underscoring the importance of the material environment in shaping learning.

Theoretical Foundations: Bridging Agential Realism and Materialism with Educational Practice

The theoretical perspectives of agential realism and materialism are central to my thesis, offering insights into the complexities of learning environments and assessment. Developed by Karen Barad (2003, 2007), agential realism explores the intricate intra-actions between human and non-

human, material and discursive elements, framing these relationships as foundational to my understanding of how reality is shaped (Barad, 2007). This framework is crucial for comprehending how all the elements of learning environments mutually influence and reshape each other, impacting student learning and engagement (Lenz Taguchi, 2010; Taylor, 2013). Materialism complements this by underscoring the intra-connectedness of all components in the intra-action encompassing both temporal and spatial dimensions, as discussed by Plauborg (2018). Barad's (2007) concept of learning as a process of 'intra-actions' (Barad, 2007. P. 33) within an 'entangled world, where knowledge is not obtained in isolation but as part of the world becoming' (Barad, 2007, p. 185) enriches this method. This theory elucidates how physical layouts and spatial arrangements within schools can affect the teaching and learning processes, guiding the intra-actions within these spaces (Ahmed, 2006; Klykken, 2023).

Understanding the dynamic nature of 'spacetimematter' (Barad, 2007,p. 181, emphasis in the original) is pivotal when considering the implementation of collaborative testing. Research by Murris and Bozalek (2019) and Theil (2018) underscores the importance of considering material and environmental conditions in educational settings, revealing how these intra-actions shape perceptions and learning processes. This understanding is crucial for effectively implementing collaborative testing and other innovative learning strategies that are responsive to the needs of 21st-century education. The integration of agential realism and materialism shapes the methodological approaches used in my study. By adopting diffractive ethnography, this thesis aims to capture the complex interactions within classrooms, particularly how various educational practices and material conditions influence collaborative testing.

In this chapter on the theoretical foundations of this study, I sought to understand student and teacher perceptions of collaborative testing. For this purpose, a methodology that encompasses all elements of the testing environment while remaining adaptable was needed to illustrate the ever-changing nature of classroom environments. Diffractive ethnography, as advocated by Gullion (2018), serves as an ideal approach for this context. It is dynamic, focusing on emergent differences that arise from contextualised interactions, allowing each teacher's and their students' narrative to unfold naturally during collaborative testing. This approach facilitates understanding how knowledge is collaboratively produced by all involved elements continuously "becoming" (Gullion, 2018, p.125). Acknowledging every component of these interactions can enable a deeper exploration of student and teacher perceptions. Thus, this thesis aims to capture the complex interaction within classrooms, highlighting how various educational practices and material conditions influence the outcomes of collaborative testing.

Methodological Approach

Utilising agential realism and materialism to explore collaborative testing requires combining various data collection techniques to capture the nuanced perspectives of students and teachers. These techniques include teacher and student-led conversations, audio recordings of group testing, mind maps, detailed observations, and journal notes. I conducted this study in year 11 and 12 science classes, including physics, investigating, and agricultural science. The high-stakes nature of Australian senior subjects, which will be outlined in the Research Design chapter, informed my decision to undertake the research in senior (year 11 and 12) classes. As employed in this study, an iterative, flexible, and responsive diffractive analysis enriches the understanding of collaborative testing by viewing all the components as active participants in the educational process. The methodological intricacies outlined pave the way for a structured exploration, guiding the layout of each chapter to explore the perceptions and impacts of collaborative testing.

Thesis Outline

In this last section of the introduction, I explain how each chapter contributes to the broader development of my thesis.

In Chapter 2, I examine the literature by situating my thesis within established research and exploring recent gaps. First, I describe the various roles of different assessment types, focusing on norm-referenced, criterion-referenced, standards-referenced, formative and summative assessments. These tools are crucial in shaping educational strategies and influencing teaching effectiveness and student learning outcomes. I particularly highlight the challenges and limitations of high-stakes assessments for students and teachers. I then outline and critique the Global Education Reform Movement (GERM) for its focus on standardisation and accountability (Sahlberg, 2016). I argue that GERM's policies have instilled a managerialism-driven culture that impacts educators and students, often reducing teacher autonomy and limiting the breadth of students' educational experiences (Ball, 2016; Hendrix, 2021). The second part of this chapter explores collaborative learning and testing as a method that, while beneficial in reducing test anxiety and enhancing essential skills like teamwork and problem-solving, also presents challenges, such as effective group management and accountability, to lessen the risk of social loafing (Kennette & Myatt, 2018). Through this analysis, I address critical questions about optimising various assessment types to support diverse educational goals. I also discuss how these strategies can mitigate the adverse effects of high-stakes environments and managerial policies, advocating for a more balanced approach to student development.

In Chapter 3, I discuss the onto-epistemology framework that guides my study, specifically how diffractive ethnography and agential realism inform my methodological choices to explore the intra-connections within the classroom environment. This chapter describes agential realism and redefines the roles of phenomena, apparatus, and material-discursive elements. Emphasising the need for methodologies that capture the full complexity of educational entanglements, I engage with diffractive ethnography to understand the entangled nature of all elements (matter, nonmatter, human) involved in collaborative testing in school environments. Adopting agential realism and diffractive methodologies is an integral part of this research to explain how learning outcomes are (re)constituted through intra-actions within the classroom during collaborative testing (Barad, 2007; Murris, 2022). By actively participating with students, teachers, and the classroom environment, I aim to understand how various factors contribute to learning outcomes.

Chapter 4 depicts the research design I employ to understand each case in a 'natural' setting. I explain diffractive analysis, including the inclusions and exclusions that form the apparatus. I then outline and justify the various data collection techniques, including audio recordings, observational notes, and student and teacher-led conversations to construct a comprehensive picture of the collaborative testing landscape. Furthermore, I discuss the criteria for selecting schools and participants through purposeful sampling, which ensures the selection of information-rich cases that highlight the intricacies of collaborative testing, before discussing the study's research phases to illustrate the data-gathering techniques. Through this design, I seek to contribute to educational research by offering a more intra-connected and dynamic understanding of how collaborative testing unfolds in real-time and real-world contexts.

I have structured the data analysis in Chapters 5 through 7 to facilitate a detailed cross-case analysis in the discussion chapter. Each chapter begins with a descriptive vignette and diffractive diagrams illustrating the complex intra-actions among the teachers, students, and their classroom environments. To 'set the scene', I integrate comments from my journal notes in the present tense to provide depth to the vignettes. The analysis examines the complex intra-actions through a diffractive analysis, providing insights into the material and immaterial factors affecting the students and teachers when collaboratively testing. Following this, I describe collaborative testing's impact on student learning and skills, emphasising the unique observations from each case study and synthesising these insights to conclude the chapters. This analysis provides a comprehensive overview of the collaborative testing environment, highlighting its complexities and potential to transform classroom practices.

Chapter 8 provides a cross-case analysis of the three classes and synthesises the perceptions and impacts from Chapters 5 - 7 to understand how students and teachers perceive collaborative testing. The cross-class diffractive analysis is organised in three distinct parts. First, I analyse students' and teachers' perceptions of collaborative testing, identifying their commonalities and differences. Next, I synthesise the impacts of collaborative testing on student learning, integrating insights from the literature reviewed in Chapter 2. Lastly, I explore the practical implications of implementing collaborative testing in high school settings, offering actionable recommendations for teachers and schools.

Chapter 9 concludes this thesis by distilling the findings from the diffractive analysis of three high school science classes, focusing on the perceptions of collaborative testing among students and teachers. I delineate the findings and discuss my contribution to the literature on collaborative testing and high school assessment alongside the methodological innovation of diffractive ethnography, particularly its role in revealing how material factors impacted students' and teachers' perceptions of collaborative testing. This chapter is organised into four sections. The first section revisits the research questions that guided this study, summarising the key findings. The following two sections discuss the findings and their contribution to the literature on collaborative testing, as well as how high-stakes testing shapes the implementation of collaborative testing. The final section explains the novel application of diffractive ethnography and its contribution to the literature before discussing the broader implications and possibilities for future research. I then conclude the thesis.

Conclusion

In concluding this chapter, I set the stage for the following chapters to demonstrate how we might transform assessment practices in high schools to enhance student understanding and 21st-century skills. Across these chapters, this study weaves through the application of diffractive ethnography and the practical application of collaborative testing to transform teacher practice to reveal that there can be another way.

Chapter 2: Literature Review

Education systems globally are shaped by a complex interplay of social, economic, and political forces, with assessments emerging as a pivotal tool for steering teaching strategies, influencing student outcomes, and shaping broader educational policies. As mechanisms for measuring achievement, informing instructional practices, and driving accountability, assessments reflect both the aspirations and constraints of modern education. This chapter begins by situating assessments within the evolving educational landscape, where traditional priorities of knowledge dissemination and skill-building increasingly contend with performance driven paradigms fuelled by managerialism and global comparisons. I explore the implications of diverse assessment types-including formative, summative, norm-referenced, criterion-referenced, standards-referenced, and high-stakes testing-on teaching, learning, and policy. Against this backdrop, the chapter investigates collaborative testing as a transformative alternative, capable of challenging entrenched practices and fostering innovative, student-centred approaches. By unpacking the dynamics of assessment, this chapter seeks to highlight pathways for aligning educational practices with the diverse and changing demands of contemporary schooling.

This chapter is structured in two parts that explore the existing context of educational assessment practices and the literature around collaborative testing as an alternative approach. In this first part, I discuss high-stakes testing and the influence of managerialism on education policies, schools, teachers, and students. The second part of this chapter investigates how collaborative learning and testing can transform traditional educational paradigms. Through a detailed exploration and analysis, this chapter aims to understand how different assessment types can be enhanced to support diverse educational goals and effectively respond to the challenges of the ever-changing world.

Part 1: Assessments and Their Role in Education

Assessment is fundamental to education, serving as a multilayered tool to identify knowledge and understanding. Different educational contexts use various formal and informal assessments to measure different educational outcomes, informing stakeholders, teachers, and students on the depth of their knowledge and skills (Andrade & Brookhart, 2020; Klenowski, 2014). Through intra-actions with teachers, peers, and classroom environments, assessments identify learning gaps and evaluate the effectiveness of teaching practices (Masters, 2020). According to Binkley et al. (2012, p. 18), assessments should develop students' ways of thinking (creativity, problem-solving), ways of working

(collaboration, communication), and skills (personnel and social responsibility).' However, Baird et al. (2017) and Cumming (2020) stress that in many educational systems, assessments primarily focus on evaluating students' conceptual understanding, which may be used to influence classroom practices, both positively and negatively. Positive assessments can serve as communicative devices that convey expectations, feedback, and the knowledge and skills required by students (Au, 2011; Masters, 2013). However, Chen et al. (2017) and Cumming (2020) posit that in some cases, assessments may also become formulaic under pressure to meet standards, which can limit opportunities to enhance students' abilities and understanding. The prevailing perspective that assessment and grading strategies in secondary education are about determining student results is ingrained in policy and practices and accepted by teachers, students, and parents (Masters, 2020), thus preventing innovative assessments that may improve student ability, understanding, and skill development. Chen et al. (2017) mirror this sentiment:

> An authentic assessment of the student's ability to use their knowledge should be their ability to apply what they have learned to real-life situations and in collaborative groups. Hence, the ideal assessment should be creative, integrative, practical, and collaborative. This is rare (p. 13)

While traditional assessments have predominately focused on curriculum-specific content and skills, recent reforms, as highlighted by Polesel et al. (2021), advocate incorporating assessments more aligned with learning theories, such as computer-based tasks, problem-solving activities, or collaborative tests to better align with pedagogical practices. However, external standards still largely influence these to evaluate students' conceptual understanding and competence, as Polesel et al. (2021) and Masters (2013) attest. Understanding the balance between different assessment types is crucial for developing effective educational policies and practices that support all students. In the framework of this study, I refer to assessment as a broad range of activities that evaluate a student's skills, understanding, and abilities across a spectrum of learning domains. In contrast, 'testing' involves objective measurements to stratify students according to their achievement levels (Dixson & Worrell, 2016; Högberg & Horn, 2022). I will now move on to explore the managerialism of education and its impact on assessment and pedagogy in school environments.

Managerialism of Education

The concept of managerialism has influenced modern educational practices, integrating corporate management principles into school and university administration (Hendrix, 2021). This shift raises essential questions about balancing educational quality with performance metrics. This section explores the philosophical underpinnings of education, the transition towards a managerial-driven approach, global pressures impacting educational systems, and the resulting consequences. I conclude by examining alternatives to individual high-stakes testing and propose a model that might better align with foundational educational values. I will discuss managerialism's conception in educational policy to understand managerialism's impact fully.

Educational practices have evolved, reflecting shifts in societal values and the increasing dominance of economic and political objectives. A fundamental function of education is disseminating knowledge and skills, which Biesta (2015b, p. 77) termed 'qualification.' This role is essential for providing students with the competencies governments deem necessary. Biesta (2020) argues that the qualification process extends beyond mere knowledge transmission to include presenting a country's cultural and societal views, inevitably leading to the socialisation of students into cultural norms (p.11). As Biesta (2020) articulates, education should engage with the question of what it seeks to achieve and for whom. Reid (2019), drawing on an Aristotelian perspective, suggests education should 'enable individuals to develop their abilities to the fullest so that they can reach their potential, live enriching lives, pursue opportunities and further their life's ambitions' (p. 6). However, Biesta (2015b) defines education as an intentional act of teaching learners for specific purposes, typically determined by external entities such as governments. This process is recursive and 'evolves' through interactions between teachers and students. As Kemmis et al. (2014) contend, education's role is to 'foster the students' self-development, self-determination and self-expression' (p. 27).

The purpose of education has been subject to evolving discourse and policy, mirroring changes in societal needs and values (Ball, 2021). Biesta (2020) and Au (2008) contend that education not only imparts knowledge and skills but also fulfils social expectations mandated by political institutions. This shift has increasingly emphasised the qualification function of education, seen from economic and political standpoints (Chubb, 2020), transforming education into what is often termed 'schooling.' According to Kemmis et al. (2014), schooling, which they describe as a 'technical tool' (p. 25), is used to deliver and monitor information as outcomes, contrasting with more encompassing notions of education. Biesta (2009) criticises schooling for narrowing educational outcomes and relying on

performance measurements. Ball (2017) and Smith (2011) also illustrate this shift in educational discourse towards efficiency, competition and productivity, focusing on targets, comparisons, competencies and benchmarks. As Biesta (2015b) and Kemmis et al. (2014) argue, schooling has become individualistic and performance-driven, resembling other organised institutional processes. This shift has sparked concerns about adopting top-down, authoritarian educational practices that use corporate management principles to improve education outcomes. As managerial principles took hold, the nature of education began to shift.

Managerialism in education, applying corporate management principles to the administration of educational institutions, has shifted traditional goals towards a focus on efficiency and outcomes (Hendrikx, 2021). Hendrikx (2021) further explains that managerial practises aim to introduce efficiency and accountability to meet specific performance standards, often involving centralised decision-making. This numerical approach allows for comparisons, competitive markets, and an input–output view of education, shaping current policy and reform (Au, 2022). Despite this criticism, proponents of managerialism argue that its emphasis on standardisation and quantifiable outcomes has brought rigour and comparability to educational systems (Sahlberg, 2016). However, this focus can reduce individuals to mere numbers, undermining holistic educational objectives (Au, 2022). Researchers like Appel (2020) and Biesta (2015, 2020) have criticised this approach, highlighting its detrimental impact on the fundamental aspects of education, which is to educate the person academically and socially. Biesta (2015) points out that stringent controls and fears of falling behind other countries could stifle student autonomy and independence, reducing education to a mechanical, process-oriented system.

This trend is fuelled by governments enforcing standards and inspections without engaging in meaningful discussions about the essence of quality education. According to Appel (2020), many countries have adopted a business model in education reforms to promote competition among schools and improve teacher performance as a path to economic competitiveness. This strategy prioritises measurable outcomes like high-stakes testing and formal education standards (Appel, 2020). Hickey and Riddle (2023) assert that this culture of managerialism subjects educational institutions to constant surveillance, monitoring and supervision, diluting direct interactions and accountability within the academic sphere. Implementing managerial strategies in education is part of a global trend to change policy and practice.

Global economic pressures have compelled educational systems worldwide to adopt similar standards, exemplified by initiatives like the US "A Nation at Risk" and the UK's Education Reform Act of 1988 (Lingard & Sellar, 2016). These reforms, prevalent in many Anglo-Saxon countries, prepare students academically and as economic participants within a global marketplace (Ball, 2017; Lingard & Sellar, 2016). The Global Education Reform Movement (GERM), as defined by Sahlberg (2016), is driven by global comparisons and directs school reform through four principles: 'competition, the autonomy of schools, parental choice and a national curriculum to measure achievement, teaching and learning' (p. 130). Internationally, these reforms reflect a reconvergence towards standardised testing and curriculum reforms to align educational practices with economic principles (Ball, 2017; Sahlberg, 2016). This shift treats students as consumers and education as a service, focusing heavily on performance metrics from international tests like the Program for International Student Assessment (PISA) and its significant impact on national educational policies and practises (Fuller & Stevenson, 2019). In Australia, for example, PISA results have led to adjustments in local and national tests and educational assessments, aligning more closely with PISA benchmarks (Masters, 2013). While standardisation may streamline processes, it also significantly impacts the academic landscape.

Implementing educational managerialism has introduced numerous challenges, including diminished teacher autonomy and a narrow focus on quantifiable outcomes. Critics like Mahon et al. (2017, 2019) argue that GERM and globalisation have intensified external pressures on schools, promoting test-driven ideologies, over-regulation, and increased workloads on teachers, all aimed at enhancing outcomes. Such measures often disproportionally affect lower-performing schools, leading in some cases to funding cuts, school closures or staff reductions (Au, 2007; Mutereko, 2018). Furthermore, Klenowski (2012) argues that GERM-based policies reduce cognitive engagement by rendering education akin to an industrial production process, a concept Mahon et al. (2019) describe as 'cognitive capitalism' (p. 476). Recently, Polesel et al. (2021) expressed similar concerns about the industrial model of Australian education to ensure students have the knowledge and the skills required to engage productively in the world. The policies of the Global Education Reform Movement, adopted in various forms worldwide, offer several benefits alongside criticisms that have emerged.

Despite these challenges, Sahlberg (2016) and Smith (2011) recognise some benefits to this reform movement, such as improving teaching and student expectations by setting high standards, a focus on performance policies encouraging schools to use and distribute resources more effectively, and a competitive marketplace that can drive performance, thereby improving teaching practises. Hardy and Boyle (2011) assert that high-stakes tests are not the issue; instead, it is how stakeholders

use the results. As Ravitch (2010) points out, 'Tests can be designed and used well or badly. The problem was the misuse of testing for high-stakes purposes' (p. 151). Alternative models to managerialism, which focus on the holistic nature of education, offer promising pathways that prioritise educational quality over efficiency. O'Neill (2013) illustrates that assessment should not be attached to accountability measures; it should be informed by people who know and understand students and are sufficiently independent to report judicially to students, parents, schools, and governments. O'Neill (2013) terms this 'intelligent accountability' (p. 15). This section has explored the influence of the managerial system on educational practice. I now focus on assessment in Australian secondary schools, before discussing the impact of high-stakes testing, a prominent feature of managerial ideologies.

Assessment in Australian Secondary Schools

This study was conducted in New South Wales (NSW) and Queensland (QLD), highlighting the distinctive features of Australia's education system. At the national level, the Department of Education and Training (DET) sets policies and programs that span early childcare, K–12, higher education and vocational training. Although education is primarily a state and territory responsibility, the federal government wields significant influence through funding, the development of the national curriculum, and the establishment of a national assessment program, which includes overseeing public reporting of the results and participation in international tests such as the Program for International Student Assessment (PISA) and national tests such as the National Assessment Program Literacy and Numeracy (NAPLAN) test (Gurr, 2020). While individual states and territories primarily manage education in Australia, federal influence remains significant due to funding and assessment standards that aim to ensure consistency across the country.

Each state or territory in Australia manages its education departments, curricula, assessments and certification requirements. There are notable differences at the senior school level in subject choice, compulsory courses and assessment methods, ranging from high-stakes external tests to 'moderated school-based assessments in the ACT' (Tran & O'Connor, 2024, p. 8). In NSW, the NSW Education Standards Authority (NESA) governs the education standards and curricula through two key stages: primary (K–6) and secondary (7–12) (NESA, 2023). NESA provides a specific curriculum as a guideline for schools to tailor their teaching programs to their unique context. It offers flexibility in determining the frequency and type of assessments that meet their student's needs within the state guidelines. (NESA, 2023). The Higher School Certificate (HSC) is a pivotal aspect of the NSW education

system, representing the culmination of thirteen years of schooling and serving as a reliable basis for further education and training, with 76,839 students enrolled in one or more HSC subjects in 2023 (NESA, 2018). Similarly, Queensland offers an evolving framework that reflects changes in assessment approaches, highlighting differences from NSW.

Queensland has recently transitioned from a system focused on 'school-derived continuous assessment and teacher judgement of student work' (Cumming, 2020, p. 161) to four structured school-based assessments. These new assessments are now 'curriculum and criteria-based, that use psychometric tools to 'grade' students, and an external exam written by Queensland Certificate of Education (QCE)' (Cumming, 2020, 173). One of the in-class assessments is designed to be standardised across all schools, conducted simultaneously under uniform conditions, and marked according to a common scheme established by the Queensland Curriculum and Assessment Authority (QCAA) (Cumming, 2020). The new Queensland QCE includes a broad range of student learning, with options to 'combine applied subjects, Vocational Education and Training (VET) qualifications, and general subjects' (Shergold et al., 2020, p. 44). These end-of-school exams in Queensland, NSW, and other states aim to assess whether students are prepared for future endeavours. However, the effectiveness and relevance of these assessments have been increasingly questioned.

Recent reports, such as the Gonski et al. report (2018) and the Masters' Review of NSW Curriculum (2020), advocate for reforms to the senior examinations and curriculum to align assessment frameworks with the demands of modern education. These reviews highlight a problem facing Australia. While the reports acknowledge the concerns surrounding high-stakes external exams, there is pressure to continue using these examinations. As Gonski et al. (2018) assert:

> In senior secondary schooling, attention shifts to discipline knowledge, with less consideration given to developing an integrated and interconnected set of knowledge, skills, behaviours, and dispositions. This lack of consideration of capabilities such as personal and social capability and critical and creative thinking is accompanied by a strong focus on 'rating' or 'ranking' students to facilitate competition around university admissions (p. 50)

Of interest to this study are the concerns raised about the HSC in senior schooling, particularly its emphasis on content and exam skills, which tend to discourage the development of essential skills such as critical thinking, collaboration, communication, and problem-solving. To address these

concerns, Gonski et al. (2018) advocate for a more balanced assessment system that values discipline-focused knowledge and general capabilities. They recommend integrating teacher assessments and external examinations that move beyond traditional written tests to include online assessments and real-time student performance evaluations to assess students' abilities more accurately. Furthermore, NESA in NSW (2023) criticises this focus on exam preparation, calling for assessments to focus on critical skills, such as collaboration, problem-solving and critical thinking (Tran & O'Connor, 2024; Polesel et al., 2021). This recognition that students lack activities that utilise 21st-century skills in their senior years represents a gap in the literature that my study can address. Through collaborative testing and peer discussions, students can develop the teamwork, problem-solving, and argumentation capabilities needed for the changing workplace. Addressing these gaps through collaborative testing could help better prepare students for the demands of modern workplaces, emphasising the importance of skills beyond academic content. However, a focus on practical skills is often overshadowed by the pressure of achieving high ATAR scores in Australia's education system.

In Australia, the senior secondary certificate, such as the HSC, plays a crucial role in university selection through the ATAR. Polesel et al. (2021) note that, unlike other systems, 'ATAR is the sole mechanism through which most high school students gain entrance into university' (p. 567). Rather than a simple score, this rank has significant implications for schooling and teaching methods, often serving as a performance measure for students and schools relative to others (Shergold et al., 2020). Masters (2020) contends that pursuing a high ATAR overshadows other educational goals, becoming the primary objective for students and their families and exacerbating the high-stakes nature of the HSC. Cumming (2020) notes this prevailing emphasis students, teachers, and schools have on the HSC and ATAR, which influence students' subject choice in the senior years, with the focus on 'traditional' subjects over Vocational and Educational Training (VET) courses (Tran & O'Connor, 2024). However, recent reforms have aimed to include a broader range of learning assessments and subjects that reflect a variety of students' strengths and interests (Cumming, 2020; Gonski et al., 2018). O'Neill (2013) discusses how high-stakes testing data from the United Kingdom's 'A' levels and the General Certificate of Secondary Education (GCSE) produce ranking tables, holding schools accountable, with significant implications for teachers and students. This comparison underlines the need for a transformative vision for Australian senior secondary assessment to evolve from a narrowly focused, individualised, and competitive model to one that enhances students' outcomes and ensures students are equipped with the necessary skills to achieve post-schooling success. I now move on from a broad understanding of assessment to distinguish between different types of assessments to grasp their varied implications within educational settings.

Norm-Referenced and Criterion-Referenced Assessment

Norm-referenced and criterion-referenced assessments provide unique insights into student performance, each with varying focuses and implications for educational achievement. Norm-referenced tests involve students undertaking a bell-curved formal examination, providing a comparative ranking that identifies their relative performance on content, skills and understanding against their peers (Tognolini & Davidson, 2013). This assessment method can inform stakeholders and students regarding academic progress, knowledge gaps, and curriculum adjustments to meet diverse student needs (Burkett, 2018). Despite their widespread use in multiple areas to determine readiness for school, essential reading and math skills, and end-of-schooling examinations, norm-referenced tests have been criticised for being competitive, individualistic, and used to measure students' knowledge at a point in time (Burkett, 2018). According to Boghossian (2006), norm-referenced testing promotes rote memorisation, which could disengage students from the learning process and cause a change in memory instead of a deep conceptual understanding. Biesta (2015, 2015b) further contends that these assessments compare students' performance against different schools, negatively impacting students from lower socio-economic areas, minorities, and learners of English as an Additional Language or Dialect (EALD). Evaluating in this manner impacts how students see themselves, education, and society. These assessment methods are not designed to support the learning process and may, in fact, not demonstrate students' overall achievement (Burkett, 2018; Polesel et al., 2014). Recognising the limitations of norm-referenced tests, particularly their comparative performance, I turn to criterion-referenced assessment, which measures students' achievements against specified criteria.

Criterion-referenced assessment (CRA) can overcome the limitations of norm-referenced assessment and positively impact student learning. Biggs (1999) and more recent reviews claim enhanced learning and engagement with CRA if they are relevant, authentic, contain a range of higher-order thinking verbs, and have explicit, understandable criteria (Mason et al., 2018; Shay, 2008). Chen et al.'s (2017) current study on arts in K–12 schools supports Shay's assertion that criterion-referenced assessment can benefit student learning through explicit expectations and feedback from teachers and peers. Biggs (1999) further contends that effective CRA requires 'constructive alignment,' where clear objectives define teaching practises and assessment methods to develop student understanding (p.64) while cautioning against misalignment where objectives and assessment grading diverge. In New South Wales (NSW), the Board of Studies Teaching and Educational Standards (BOSTES, 2016a), now known as the NSW Education Standards Authority (NESA), and the Education Council (2020) attest that

this misalignment occurs with students in years 11 and 12 who undertake CRA. The objectives are written to promote depth of learning; however, the task is written similar to summative tests, promoting surface learning (BOSTES, 2016a, p. 5), poor retention (BOSTES, 2016a, p. 15) and increased anxiety (Education Council, 2020, p. 44). The literature describes the negative aspects of criterion-referenced assessment: a lack of explicit criteria, the different interpretations of the term criteria, lack of consistency when grading, and grading from a performative aspect (Burton et al., 2006; Lok et al., 2016). Another concern raised relates to the stated criteria. If the criteria are excessively detailed or too vague, students and teachers may not be able to interpret the assessment correctly, leading to grading anomalies (Tognolini & Stanley, 2007). The negative impacts of misaligned CRA are similar to those found with norm-referenced assessment. Standards-referenced assessments were developed to overcome the shortcomings of norm and criterion-referenced assessments based on a broader, standardised framework that may ensure consistency and comparability across different contexts.

Standards-Referenced Assessment

Standards-referenced assessment was introduced to provide descriptive measures of student performance based on a prescribed curriculum or syllabus that articulates the subject matter, essential goals, and learning outcomes that students work towards New South Wales Education Standards Authority (NESA), 2020; Sadler, 2005; Tognolini & Stanley, 2007). Masters (2020) observes that the widespread adoption of these standards-based assessments stems from their emphasis on individual progress through the attainment of benchmarks, which enhances student outcomes. In their discussion on standards-referenced assessment, Sadler (1987) states, 'The primary function of a set of educational standards is to enable statements about a student's quality of performance or degree of achievement to be made without reference to the achievements of other students, which conceivably could be either all poor or all excellent' (p. 196). Therefore, unlike norm-referenced tests, student performance is based on standards that measure what and how well students understand concepts (Bennett et al., 2012).

Similar to norm and criterion-reference assessment, there are concerns about using standards to measure student achievement. Masters (2020) and Tognolini and Stanley (2007) contend that if the standards are not aligned with the required assessments and class-based instruction, this may lead to inaccurate measurements, impacting student and teacher outcomes. Assessments are a necessary component of education, with each norm-referenced, criterion-referenced, and standards-referenced having a place due to their unique benefits and limitations. The balanced implementation of these assessments is crucial to promoting student achievement and learning and providing stakeholders,

schools, teachers, students and parents with an accurate assessment of teaching and assessment strategies and student outcomes. After discussing the role of these three assessment methods, it is essential to consider how formative and summative assessments intra-twine within a comprehensive educational assessment system.

Formative and Summative Assessment

Formative and summative assessments are essential components of education, each having different but compatible purposes in the learning environment. Deluca et al. (2019) describe formative assessments as a method that tracks student progress and provides teacher feedback. There are many formative assessments, including pre and post-learning quizzes, essays, computer tasks, class projects, and collaborative projects. Smith et al. (2022) contend that student understanding can be increased if teachers regularly use formative assessments across various contexts. However, summative testing may impact formative activities by focusing on test performance over learning processes (Black & William, 2018). These findings support adopting formative assessment as a continuous feedback mechanism that aids personalised learning.

Conversely, summative tests measure students' knowledge at the end of a learning sequence. Baird et al. (2017) state that these tests can be final exams, end-of-term projects, or standardised tests, evaluating student learning against prescribed standards to determine overall achievement. These tests are considered high-stakes due to the rewards or sanctions imposed on the outcomes, which can impact schools, teachers and students (Au, 2011; Baird et al., 2017). Moreover, in her study on assessments, Klenowski (2011) notes that the predominant focus on summative assessments, especially without immediate feedback, can affect student performance and anxiety levels due to the inability to identify and correct conceptual errors. Kulasegaram and Rangachari (2018) contend that current assessment practises may be detrimental to student learning as 'they promote unreal expectations of certainty, provide little or no practical preparation for future learning, and discourage the embrace of ambiguity and creativity' (p. 6). Cumming (2020) also points out that summative tests can reduce the number and style of in-class assessments, especially in secondary education, where the focus may tend towards tests that are structured similarly to external, high-stakes tests. The literature notes that testing in this manner causes adverse effects on intrinsic motivation, students' self-efficacy and metacognition, and how they view learning and schooling (Andrade & Brookhart, 2020; Au, 2008; Kulasegaram & Rangachari, 2018; William, 2011). As Zacharis (2010) describes, assessment can impact student learning and depends on students' motivation and the quality of feedback received from previous tasks. William (2011) supports this view; adding testing can constrain or enhance learning depending on the type of task and whether the learner can 'see' how the task will

enhance learning. He further asserts that effective assessment requires students to be active and involved with their peers.

Andrade and Brookhart (2020) contend that evaluations can positively impact learning if performance targets are explicit, the task engages students, and prompt feedback is given. Resnick and Schantz (2017) further add that examinations should include 21st-century skills such as critical thinking, creativity and problem-solving to encourage a deep understanding. While assessments are indispensable in education for measuring and enhancing learning, they require thoughtful consideration to ensure students meet learning and skill outcomes. In examining curriculum and assessment practises, Masters (2020) and Binkley et al. (2012) advocate that assessments should reflect and promote the conditions required for optimal learning and teaching, emphasising deep understanding, application, and mastery of the content by applying core knowledge across various contexts and demonstrating proficiency in critical thinking, problem-solving, communication, collaboration and information analysis. Such approaches allow a more multilayered view of students' understanding and skills.

Defining High-Stakes Testing

High-stakes testing (HST) has become a global central component of educational policy, influencing educational practises (Gonzalez et al., 2017; Wilson et al., 2021). Despite its widespread use, the implementation and impact of high-stakes testing are subjects of debate. Government policies and societal values shape high-stakes testing, mirroring broader educational objectives related to knowledge and skill acquisition (Appel, 2020; Klenowski & Wyatt-Smith, 2012). Bernstein (1990) and Sadovnik (1991) discuss how the values ingrained in an education system influence its structure, encompassing the sequencing, pacing, instruction, and assessment practises, all collectively defining what and how students are expected to learn. As Hendrikx (2021) and Sahlberg (2016) contend, focusing on managerial characteristics like accountability and performance outcomes has gained prominence, altering teaching and assessment methods to produce a numerical score.

As noted by Mutereko (2018) and Shepherd (2017), managerialism is devised by business models to quantify schools, students and teachers, determining which school is 'better' and who needs to improve their standing. Focusing on managerial behaviours through high-stakes tests reduces students and teaching to numerical values for comparison, creating competitive markets and an input–output view, which makes students more like objects (Au, 2022). As Mutereko (2018, p. 569) states, the managerialism of education has focused on the 'attention to performativity and the quantification

of education outputs through examinations and how such tests are used to reduce the learning of students over 12 years to mere numbers through test scores.'

In the USA and the UK, high-stakes testing is entrenched and commonly used by politicians and the public as a measure of accountability (Jerrim, 2021; Lewis & Biesta, 2015; Ravitch, 2010). In the USA, for instance, individual states develop these tests as part of accountability measures that have significant consequences for schools, teachers and students (Gonzalez et al., 2017). Acosta et al. (2020) describe how, in these tests, 'students' achievement scores on large-scale, state-wide tests serve as a criterion for making decisions about student academic growth, teacher instructional effectiveness, and school performance' (p. 329). Similarly, in Australia, Appel (2020) observes that high-stakes tests serve a comparable role, holding students, teachers, and administrators accountable to communities and governments. Furthermore, Sahlberg (2016) and Verger et al. (2019) attest that accountability systems that link learning achievements to incentives, sanctions, or reputational outcomes are not designed to improve teaching and learning at the school or classroom levels.

Research on high-stakes testing presents a mixed picture, reflecting its varied impact on students' achievement, learning and well-being. Macqueen et al. (2019) and Polesel et al. (2014) argue that data derived from these tests can inform educators and schools in identifying areas needing improvement, such as literacy or benchmark attainment, thereby improving educational quality and standards (Acosta et al., 2020). Demir and Özlem (2021) observe that such testing can motivate students to study harder and teachers to deliver more effective mathematics instruction. However, the emphasis on test performance also brings significant drawbacks, adversely affecting both students and teachers, as highlighted by Baker et al. (2016), Farvis and Hay (2020) and Wilson et al. (2021). Specifically, high-stakes testing has been linked to increased student stress and anxiety, a narrowed curriculum that limits broader educational opportunities, and a pressured educational environment that can reduce teacher job satisfaction (Au, 2022; Farvis & Hay, 2020). I will now discuss the impact of high-stakes tests on students and teachers.

Impact of High-Stakes Testing on Student Learning

High-stakes testing shapes learning outcomes and influences student achievement and experiences. These assessments, which have significant consequences for students, teachers, and schools, can have positive and negative results. The extent of their impact on student learning varies depending on the pedagogical techniques employed by teachers. Farvis and Hay (2020) highlight that while some teachers use inquiry-based learning, problem-solving activities, and collaborative projects

that encourage deeper engagement and understanding, others may prefer a more direct instructional method that prioritises individual performance, often in response to pressures from high-stakes testing. In their study, Morgan (2020) noted how teachers, particularly in disadvantaged schools, would start drilling students on the test structure and content before the test, while Rice et al. (2015) also noted that due to high-stake tests, teachers would undertake fewer collaborative activities while increasing individual tasks.

Andrade and Brookhart's (2020) research reveals that shifting from a constructivist teaching method to a teacher-centric approach can hinder the development of high-order cognitive skills essential for deep learning. This instructional narrowing is driven by the focus on test content delivery, which often prioritises memorisation over critical thinking, leading to a fragmented and superficial grasp of knowledge (Lobascher, 2011; Resnick & Schantz, 2017). This strategy undermines the complexity of learning and 'reduces cognitive stimulation' (Klenowski, 2012, p. 181) as students are less exposed to other experiences and practical activities that enhance real-world applicability, as Shergold et al. (2020) noted. Despite these challenges, some studies suggest that high-stakes testing can have motivational benefits, as the pressure to perform may encourage students to engage more deeply with the content, although these benefits are often limited to the scope of the test (Blazer, 2011; Perkins & Wellman, 2008). Furthermore, the shift towards a passive transmission of knowledge in high-stakes environments hinders the implementation of collaborative or problem-solving methods. This change can influence students' motivation and engagement, diminishing their depth of understanding and the development of high-order thinking (Appel, 2020).

Using high-stakes testing, where all students sit the same test simultaneously in the 'same' conditions, negates the differences in environmental factors that influence student test-taking ability. However, as Au (2022) explains, the results of the tests may not indicate a fair test due to differences in school environmental factors such as air-conditioning or lack thereof, overcrowding, and lack of computer access, in addition to socio-economic factors such as poverty. Hence, as the students sit their examinations, the factors around them influence their ability to engage with the test, impacting their performance (Au, 2022). Through the managerialism and quantification of education as students sit these exams, and through 'decontextualising' (Au, 2022, p. 70), the students are reduced to a number, which tells the schools, governments, parents how good or bad they are compared to their peers. As Au (2022) states, 'High-stakes testing not only pressures individuals by enforcing specific norms about who does or does not fit within education, defining who is or is not a good student, teacher or school but also by reducing students to numbers, it creates the capacity to view them as mere things, as

quantities stripped of their individual qualities.' (p. 70). This section has examined the impact of high-stakes testing on students' learning, highlighting how it shapes teachers' practises and the broader implication of reducing students to numbers. The following section discusses how high-stakes testing can influence surface or deep learning.

Surface and Deep Learning in the Context of High-Stakes Testing

Understanding the distinctions between surface and deep learning in high-stakes testing is critical for fostering educational environments that enhance meaningful learning. A significant amount of research (Cantwell et al., 2016; Gilley & Clarkston, 2014; Reiger & Heiner, 2014) has outlined the learning benefits of collaborative testing in undergraduate students, including the progression from a surface to a deeper understanding (Mahoney & Harris-Reeves, 2019). Understanding the intricacies of surface and deep learning becomes even more critical when considering the high-pressure environment of high-stakes testing. This understanding helps design educational practices that prepare students for tests and foster long-term retention and application of knowledge. In high-stakes scenarios, surface learning often prevails, as students resort to rote memorisation and lower-order cognitive skills due to the pressures and demands of the testing environment (Masters, 2020). Surface understanding is a passive approach to learning, employing lower-order cognitive skills to memorise facts (Asikainen & Gijbels, 2017). Students may adopt this method due to pressure, external life factors, high anxiety or performance-avoidance goals (Baeten et al., 2010; Howie & Bagnall, 2013). Teachers impact students' use of surface learning by teaching disjointed content, emphasising memorisation, and providing insufficient time to accomplish tasks (Dolmans et al., 2013; Howie & Bagnall, 2013).

Conversely, deep understanding requires students to engage with the material, actively developing higher-order thinking skills that lead to meaningful understanding and the ability to apply knowledge across different contexts actively, seek meaning, and extend their ideas (Hattie & Donoghue, 2016; Wang et al., 2013), developing a deeper understanding and progressing through various learning stages (Biggs, 1999). Initially, students undertake lower-order thinking skills such as memorisation, summaries, outlines, and descriptions to build and understand new information before developing a more integrated knowledge of the content to recognise and apply the information (Biggs, 1999). Eventually, students generate a deep understanding through metacognitive processes, applying their knowledge to different scenarios and utilising higher-order thinking verbs (Biggs, 1999; Dolmans et al., 2016; Morgan, 2013). In their study of deep learning, Hattie and Donoghue (2013) and Wang et al. (2013) note students could employ surface and deep learning at varying times depending on the subject, the new concept's difficulty, the student's viewpoint on the assessment's worthiness, and

their preferred learning styles. Teachers can facilitate students' progression to deep understanding by activating prior knowledge, setting precise goals, emphasising the depth of knowledge, modelling higher-order thinking verbs, and fostering a positive collaborative classroom (Mahoney & Harris-Reeves, 2019; Meseke et al., 2010). Moreover, Asikainen and Gijbels (2017) and Baeten et al. (2010) have found that the type of assessment influences a student's progression to deep learning. If the assessment process is not aligned with promoting a depth of understanding, students may resort to surface learning strategies, resulting in disengagement.

Collaborative assessments promote in-depth understanding through positive peer entanglements, where students listen, argue and problem-solve (Meseke et al., 2010; Siegel et al., 2015). This argumentation process allows students to compare and contrast their knowledge with their peers, gaining different perspectives. Positive peer intra-actions empower students to engage with their learning (Mahoney & Harris-Reeves, 2019). Groups with strong interdependence further enhance student understanding by increasing confidence and self-efficacy, enabling effective group discussions (Baeten et al., 2010). Nafziger et al. (2011, p. 11) assert that 'students potential for cognitive growth is limited by what they can do independently, but collaboration allows students to go beyond individual limitations.' There is a significant amount of research recognising that collaborative learning positively impacts students' understanding and metacognitive and social skills (Black & Allen, 2018; Mercer & Howe, 2012; Tolmie et al., 2010; Tran, 2013). However, teachers typically assess students in a high-stakes, competitive, individualistic manner, focusing on what they know rather than applying their knowledge and skills despite frequent participation in collaborative tasks (Giuliodori et al., 2008; Polesel et al., 2014). In their study on collaborative learning, Kulasegaram and Rangachari (2018) attest that this disconnect between how students learn and their assessment impacts students' ability to understand concepts and apply and transfer new knowledge. In recognition of this disconnect, some stakeholders, such as the New South Wales Education Standards Authority (NESA), have implemented questions designed for students to apply their knowledge (Masters, 2020). Masters (2020) notes that the HSC does not adequately assess the full range of learning valued in subjects, resulting in overemphasising the content and exam skills. As one teacher commented in the report:

> The stronger inquiry and investigative approach in the new science syllabuses are 'great, 'but we're still stuck with the HSC external exam. We can't get changes in senior subjects with the external exam still sitting there. Kids are reporting that they don't see the value in it (Masters, 2020, p. 34)

An emphasis on high-stakes testing can impact deep learning and long-term knowledge retention, potentially shifting the focus from broader educational goals to test performance. However, it can also drive focused curriculum improvements in some cases. Educational practice must balance high-stakes assessments' requirements with methods that foster critical thinking and genuine understanding. Additionally, it is essential to consider the broader impacts of high-stakes testing on students' overall well-being and the stress associated with these tests, which can significantly affect students' cognitive and emotional states. These will be discussed in the following section.

Impact of High-Stakes Testing on Students' Anxiety and Well-being

Student well-being encompasses physical, emotional and mental health, influencing their academic performance and life satisfaction. High-stakes testing environments can severely impair students' cognitive functions, affecting their ability to retrieve, organise and retain information, thus hampering their overall academic performance, as discussed by Soares and Woods (2020). Numerous studies have consistently highlighted the negative impacts on students, particularly focusing on the stress and anxiety associated with the need to perform. Soares and Woods's (2020) literature review on test anxiety asserts that students need the right amount of pressure to perform at their best, which Arroba and Bell (2001) term' zone of optimum pressure' (p. 313 in Soares and Woods 2020), noting that too little pressure will cause boredom and too much will overload their cognitive abilities. Stress from high-stakes testing affects students in several ways. It can impair their cognitive ability to retrieve, organise and retain information, which in turn hampers their capacity to access prior knowledge and effectively answer questions. Research by Appel (2020), Breedlove et al. (2004), Gilley and Clarkston (2014), Mahoney and Harris-Reeves (2019) and others have well documented this relationship between stress and cognitive function.

The stress and anxiety associated with high-stakes testing can affect students' academic performance and may lead to significant physiological and mental health challenges, impacting their confidence and self-esteem. However, some students may thrive in structured testing environments (Krispenz & Dickhäuser 2018). According to Breedlove et al. (2004), such testing environments increase feelings of helplessness and insecurity, diminishing students' motivation and adversely affecting their test performance. Similarly, research on the National Assessment Program-Literacy and Numeracy (NAPLAN) in Australia by Biesta (2015) and Cumming et al. (2018) and studies by Yates and Johnson (2018) in New Zealand reveal that inauthentic and unrelatable learning materials and test questions

further exacerbate these issues, particularly among disadvantaged students. These students become more disengaged and anxious, struggling to meet benchmarks, significantly impacting their overall academic achievement and experience at school. The publication of high-stakes test results adds another layer of stress. Cumming et al. (2018) elucidate how fear of failure and the possibility of peer labelling can diminish students' self-esteem and intensify their anxiety. Appel's (2020) literature review on high-stakes tests in Australia reflects this aspect, indicating an increase in reported depression among students due to performance pressures. Meanwhile, Högberg and Horn (2022), in their study on high-stakes testing and stress in Europe, emphasise that being labelled as a high-achieving student significantly impacts students' identity, social status and mental health. High-stakes tests and heavy workloads, including test preparation, exacerbate stress levels as they make academic achievement more prominent and linked to formal credentials.

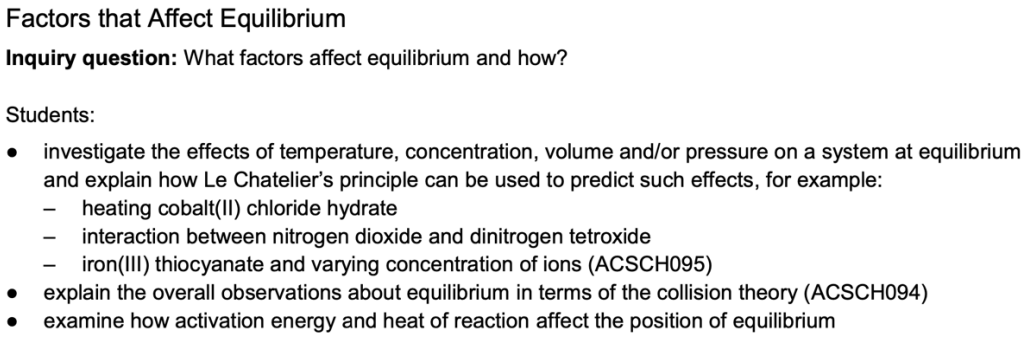
In addition to student well-being, high-stakes testing also shapes teachers' perceptions of their students and influences the educational dynamics within the classroom. Klenowski (2014) and Klenowski and Wyatt-Smith (2012) argue that competitive, individualistic testing environments cause teachers to adopt a more competitive approach to teaching. They may prioritise students on the cusp of moving to the next achievement level, potentially neglecting those who perform at the higher or lower ends of the spectrum, which can harm students' self-esteem and self-perceptions as learners, which has implications for their progression through school (Marklund & Taylor, 2016). High-stakes testing influences students' views of learning, their engagement with schooling, and how they and others perceive themselves. Moreover, Shergold et al. (2020) contend that the education system and the public's view of the importance of attaining high grades at the end-of-schooling exams demonstrate to students what society values most, impacting the importance of achievements in other areas. Resnick and Schantz (2017) summarise the implications for students of high-stakes testing. They state, 'We cannot ethically continue to use methods of assessment that do not match our learning goals for students, and that force us to make judgements that affect students' whole lives based on what is easiest to measure' (p. 431). High-stakes testing induces stress and anxiety, negatively affecting students' academic performance, confidence, and self-esteem. Educational systems should address these impacts by creating more supportive and holistic assessment strategies prioritising student well-being. In the next section, I will discuss the effects of high-stakes testing on the curriculum and how these changes influence student learning.

Impact of High-Stakes Testing on Curriculum

The terms curriculum and syllabus are often used interchangeably in the educational literature. The curriculum broadly encompasses the national or state frameworks that describe the

learning outcomes, goals, content, teaching methods, and assessment strategies used to achieve those outcomes (Au, 2007; Kelly, 2004; Luke et al., 2021). For instance, the Australian Curriculum and Assessment Reporting Authority (ACARA) (2020) refers to 'curriculum' as describing the comprehensive national programs from kindergarten to year 10. In the United States, this term aligns with the American Common Core State Standards (thecorestandards.org), while in the United Kingdom, it relates to the subjects taught and the standards to be achieved (Gov.UK). On the other hand, Bernstein (2000) views the syllabus as the detailed 'official' knowledge imparted to students, containing specific topics, themes, units to be covered and assignments, and can serve as a teacher guide (Chubb, 2020; Kelly, 2004). In an Australian context, the NSW National Education Standards Authority (NESA) defines 'syllabus' as a document detailing content, teaching and assessment methods. An example is shown in Figure 1.

Figure 1 *HSC Syllabus Point from Module 5: Equilibrium and Acid Reactions*



Factors that Affect Equilibrium

Inquiry question: What factors affect equilibrium and how?

Students:

- investigate the effects of temperature, concentration, volume and/or pressure on a system at equilibrium and explain how Le Chatelier's principle can be used to predict such effects, for example:
  - heating cobalt(II) chloride hydrate
  - interaction between nitrogen dioxide and dinitrogen tetroxide
  - iron(III) thiocyanate and varying concentration of ions (ACSCH095)
- explain the overall observations about equilibrium in terms of the collision theory (ACSCH094)
- examine how activation energy and heat of reaction affect the position of equilibrium

Note. The question was taken from the 2017 NSW Education Standards Authority (NESA). https://www.educationstandards.nsw.edu.au/wps/wcm/connect/ff0f1e84-3e7c-45bd-9ed4-b1972546e166/chemistry-stage6-syllabus-pdf.pdf?MOD=AJPERES&CVID=

For this study, the term 'curriculum' will denote a combination of the broader educational framework (curriculum) and the specific content and instructional guidelines (syllabus). This definition encompasses the overall educational objectives, goals, and outcomes, as well as the detailed content, teaching methods, and assessment strategies used to achieve these objectives. Teachers are the primary agents who transform the official curriculum, governed by national bodies such as the Australian Curriculum and Assessment Reporting Authority (ACARA) and state education departments such as NESA in NSW, into planned instructional activities (Tran & O'Connor, 2024). According to ACARA (2020), the curriculum details what is taught and the outcomes expected by the students. Still, it notes that schools can decide 'how best to deliver the curriculum and determine pedagogical and other delivery considerations' (p. 10) to suit their students' needs best (Polesel et al., 2021). Research undertaken by Tran & O'Connor (2024) indicates that the effectiveness of curriculum implementation

is dependent on how teachers engage with the documents, whether the curriculum is too over-detailed, and the degree of an individual teacher's ability to adapt the school-based curriculum to their individual student needs. Teachers' beliefs and experiences also shape how they enact the curriculum. The curriculum does not state how teachers deliver the content: depending on teachers' orientation, they may follow the curriculum explicitly, or they may consider it a guide (Tran & O'Connor, 2024). Additionally, Priestley et al. (2016) contend that the teachers' knowledge and beliefs about teaching and learning can impact how they teach the curriculum and student outcomes. Teachers' transformation from policy to practice highlights their crucial role in shaping the curriculum, directly affecting student learning.

The curriculum is shaped by political ideologies and societal values, reflecting a society's educational and ideological beliefs (Bernstein, 1999; Clark, 2005; Pluim et al., 2020). Placier et al. (2002) note that there has been political opposition to constructivist-based curriculums in the USA, with some politicians favouring managerial-driven policies. Westbury (2008) explains that many countries develop curriculum documents to respond to criticisms of their education policies and to establish specific criteria for evaluating students and teachers. Additionally, Wood et al. (2010) explore the instability of curricula, highlighting their influence on government policies, public concerns, media coverage and school performance in national and international high-stakes testing. According to Masters (2020), governments design curricula to dictate the content and methods of learning and assessment. In countries such as Australia, the UK and other Anglo-Saxon countries, managerial policies that emphasise performativity and high-stakes testing substantially impact the nature of the curriculum.

High-stakes testing and a performance-driven culture have reshaped the curriculum and teaching practices across countries such as Australia, England and the USA (Tran & O'Connor, 2024). Research indicates that materials not included in assessments are often neglected, limiting students' learning experiences (Appel, 2020; Ellery, 2017; Gonzales et al., 2017). Smith (2011) observes that high-stakes testing has modified the curriculum to incorporate performative measures, such as outcomes, benchmarks and competencies. In the US, Belcastro (2015) finds that performativity language shapes the curriculum around essential knowledge defined by stakeholders. Some researchers argue that aligning the curriculum with high-stakes tests can enhance performance (Gonzalez et al., 2017; Plank & Condliffe, 2013), but others warn that it narrows and standardises education (Au, 2007, 2008). Ellery (2017) notes that a detailed, scripted syllabus based on high-stakes tests dictates what constitutes 'legitimate knowledge and learning' (p. 910). Polesel et al. (2014)

emphasise the adverse effects of prioritising a test-based syllabus, leading to the undervaluation of non-tested areas. Wilson et al. (2021) agree that high-stakes tests, such as the NAPLAN, have constrained teachers' ability to implement the curriculum, causing them to spend less time on non-tested subjects, such as art and music, and more time spent on English and mathematics, during the junior years of school (Au, 2007, 2008; Klenowski, 2011, 2012; Polesel et al., 2014). Thompson and Harbaugh (2012) reveal that this narrowing affects students' ability to integrate knowledge, impacting their long-term learning and understanding. Moreover, Belcastro (2015) and Macqueen et al. (2019) posit this focus on high-stakes external testing changes teachers' pedagogical practises to ensure that tested material is well-covered.

Nonetheless, there is recognition of the issues with the current curriculum approaches. When teachers have a well-designed curriculum that does not prescribe the methods of teaching but enhances teachers' 'capacity for curriculum work, including pedagogy and assessment' (Tran & O'Connor, 2024, p. 6) along with sufficient support, they can positively impact students by providing more explicit learning goals and enabling a focus on student-centred learning (Zepeda, 2014). Appel (2020) notes that an effective curriculum enhances transparency and accountability, helping establish clear baselines and practises for teachers and students. In Australia, Masters (2020), in his NSW Curriculum Review, highlights that an overloaded and scripted curriculum negatively impacts student learning. The NSW government has introduced a 'new syllabus for mandated subjects' (p. v), allowing for more flexible teaching and learning schedules in the early to middle school years. The new curriculum emphasises explicit teaching of essential knowledge, requiring students to demonstrate competence based on set standards. However, balancing achieving these outcomes with creating a rich, engaging, supportive learning environment may still be challenging (Masters, 2020). High-stakes testing can lead to a narrowed curriculum that undervalues non-tested areas, even though policymakers aim to ensure a more balanced curriculum that promotes a broad range of skills and knowledge. In the next section, I discuss how high-stakes testing can change teachers' practice and its effect on students.

Impact of High-Stakes Testing on Teachers' Pedagogy

High-stakes testing influences teachers' pedagogical practices, often requiring a shift towards more restrictive teaching methods. Au (2007) describes these tests as 'evaluative rules' (p. 644) that measure standard attainment, transforming teachers into 'technicians' who focus primarily on preparing students for external examinations, a phenomenon echoed by Mahon et al. (2019) and Polesel et al. (2021). This technical approach restricts both the content taught and the delivery method

(Sahlberg, 2016). Moreover, in accountability-driven environments, teachers often shift from constructivist methods to more competitive and individualistic strategies, as leadership evaluations based on test results can pressure teachers to teach to the test, contradicting their educational beliefs (Appel, 2020; Klenowski & Wyatt-Smith, 2012; Wright & Froelich, 2012). While these tests aim to ensure quality and uniformity in education, they often lead to unintended consequences that affect core teaching practices.

High-stakes testing can curtail teachers' delivery, where they may prioritise test-orientated teaching over creative educational methods (Sahlberg, 2016). Such tests influence teachers' practice, pushing them to mimic the content and format of the tests to enhance student performance; this results in decreased creativity and homogenisation of content delivery (Binkley et al., 2012). Martin (2015) illustrates that 'teaching to the test' confines teachers to an unimaginative, didactic approach, resulting in student disengagement. Berliner (2011) further explains that teaching this way diverts from engaging, project-based or problem-solving lessons that could captivate student interest. The focus on test content increases rote learning, diminishes broader educational skills and encourages competitive tasks over collaborative learning opportunities (Au, 2007; Lingard & Sellar, 2016). Alexander (2010) supports this view and observes that this focus on accountability often restricts teaching to lower-order skills that are directly tested, further narrowing the educational experience. Furthermore, Au (2022) notes that beyond the immediate constraints on teaching content, high-stakes testing also impacts the deeper aspects of pedagogy, such as teacher creativity and authenticity.

Conversely, despite prevalent criticisms, a technical approach influenced by high-stakes testing data can also be a valuable tool for improving teaching methods. For instance, Appel (2020) and Martin (2014) suggest that data from these tests can help refine and personalise teaching methods, aiding student progression. Despite the predominance of negative impacts, some studies show that high-stakes testing can also promote a focus on student-centred and inquiry-based methods, fostering higher-order thinking skills (Lobascher, 2011; Polesel & Rice, 2014). Nonetheless, the overarching impact of high-stakes testing reshapes teachers' pedagogy towards transmitting limited content, often in a less engaging and demotivating style for both students and teachers (Au, 2007; Hannaway & Hamilton, 2008; Lakes & Carter, 2011). Au (2011) articulates this well by asserting that 'such testing is promoting the standardisation of teaching that both disempowers and disengages teachers' (p. 30). Acosta et al. (2020) note that this focus on performance can impact teachers' agency in what they can do in their classrooms for their students. However, it is not all detrimental; aspects of high-stakes testing can potentially foster improvements in teachers' practice. Recognising the challenges and

opportunities high-stakes testing presents, teachers have employed various strategies to navigate and leverage this educational landscape for better academic outcomes.

Teachers often develop adaptive strategies to cope with the pressure of high-stakes testing environments, seeking professional development opportunities to improve student performance and, for example, participating in exam assessor training to better understand the marking process, enhancing their students' performance on these tests (Comber, 2012). Klenowski (2012) notes that high-stakes environments cultivate a 'test wiseness' (p. 179) where teachers instruct students on the intricacies of test-taking strategies. However, this focus can inadvertently affect the teacher-student relationship, limiting opportunities for individualised learning, particularly for underperforming or ELAD students (Masters, 2013). Reid (2019), in their book on the Australian Curriculum, notes that teachers are trying to provide their students with contextualised and quality learning amid a focus on high-stakes testing and accountability. In Hargreaves's (2020) study on assessments' effects on teachers' practice, they discuss how the use of accountability, particularly the use of test results to reward or punish schools, is one of the four 'wrong drivers' (Fullan, 2011, p. 8, in Hargreaves, 2020) that assumes teachers will simply react to the pressures by making the necessary improvements. Instead, leading with accountability fails to promote the instructional achievements required for teachers to implement contextualised teaching to aid their students (Hargreaves, 2020). Finally, while high-stakes tests are designed to assess knowledge at a specific time, their broader impacts on pedagogy and teacher development highlight the complex nature of high-stakes assessments in educational systems. Educational strategies should balance high-stakes testing data with methods that support innovative and student-centred teaching methods. In the next section, I discuss how these tests affect teachers' well-being.

High-Stakes Testing and Teachers' Wellbeing

High-stakes testing significantly influences teachers' well-being and their views on the profession. Gonzalez et al. (2017) explored the effects of a performative culture on teachers' self-efficacy and morale, finding that while high student performance can boost teachers' confidence, accountability pressures are particularly intense for those teaching low-performing students. Black et al. (2010) echoed this sentiment, noting that pressure from parents, students and the school to perform undermines teachers' confidence in their professional practice. As discussed by Daliri-Ngametua and Hardy (2022), Australian media coverage of NAPLAN results has eroded public trust in teachers and schools, further diminishing teachers' job satisfaction and potentially driving them away from the profession (Ball, 2021). In the USA, Farvis and Hay (2020) describe how high-stakes testing

has created a competitive, individualistic culture that increases workloads and stress due to reduced collaboration among teachers. This stress is linked to how teachers' performance is viewed by their schools, peers and students (Gonzalez et al., 2017). In their study on educational assessment in the USA, Smith & Holloway (2020) discuss the impact of Value-Added (VA) measures on teachers' self-efficacy. They found that high-stakes testing, which uses VA to assess teacher performance based on student outcomes, increases teacher stress, pressure and anxiety. In Australia, Emslander et al. (2022) argue VA scores are used to evaluate school effectiveness, which may obscure the achievements of teachers in lower-performing environments, thus not fully recognising their efforts to enhance student outcomes.

Despite these challenges, Jerrim et al. (2024) and Nichols and Harris (2016) recognise some benefits of high-stakes testing, such as motivating teachers to improve their students' performance and providing a clear measure of teaching effectiveness, which can help to reduce instructional ambiguity and associated anxiety (Acosta et al., 2020). These benefits can further teachers' professional growth and satisfaction when the tests are managed effectively. Nonetheless, the overall impact of high-stakes testing can increase stress, erode confidence, foster a competitive culture, and undermine public trust, shaping teachers' perceptions of their profession.

Summary

Assessments are crucial in shaping educational outcomes, offering critical insights into student achievement and informing teaching practises. However, the prevalence of high-stakes testing in many educational systems poses significant challenges to student learning, undermining deep learning and student well-being. The current education system's managerial practises, which prioritise efficiency and accountability, further compound these challenges by narrowly focusing on quantifiable outcomes at the expense of holistic educational goals. This tension between measurable achievements and students' comprehensive development highlights the need for more nuanced policies. By addressing these complexities, educators and policymakers can align assessments with education's broader and more meaningful objectives.

Part 2: Collaboration and Collaborative Testing

Collaboration

Collaboration in educational settings promotes enriched learning experiences and deepens understanding through active peer engagement. This method encourages an environment where 'inner thinking' (Mercer & Howe, 2012, p. 15) and the collective solving of complex problems occur, thereby acquiring a deeper understanding of the concepts as noted by various studies (Hämäläinen & Vähäsantanen, 2011; Johnson and Johnson, 2005). McKay and Sridharan (2024), in their research on collaborative learning, suggest that unlike the structured specialisation in cooperative learning or the loose formation of group work, collaborative learning promotes an active, authentic learning environment, facilitating peer interactions, argumentation, reasoning, reflection, and external dialogue (Slavin, 2014). Although the advantages of collaborative learning are significant, it is equally important to address its challenges and limitations, as identified by recent studies. This balanced approach ensures a comprehensive understanding of the methods' impact in educational settings.

Collaborative learning effectiveness has received mixed reviews, particularly regarding varied group dynamics and the complexities of implementing collaborative tasks within a culture of high-stakes tests (Hammond, 2017; Loyens & Gijbels, 2008). Studies indicate that perceptions of collaborative learning's effectiveness can vary significantly, influenced by past experiences or group dynamics. Premo et al. (2018) discuss how unsupportive group dynamics and past negative experiences can lead to oppositional entanglements where students compete rather than cooperate, thus negating the potential benefits of collaboration. These dynamics can adversely affect academic performance, emotional well-being and social skills. As suggested by Törmänen et al. (2021) in their collaborative learning study, addressing these challenges requires a supportive environment that maximises collaborative learning benefits. A supportive classroom environment, as noted by van Leeuwen and Janssen (2019), enhances student engagement and achievement by fostering a sense of belonging and motivation, reducing anxiety and aiding in the development of positive peer relationships (Gillies, 2014; Juceviciene & Vizgirdaite, 2012). In their collaborative testing study, Eastwood et al. (2020) describe a supportive environment that enhances academic performance and significantly improves students' overall satisfaction and engagement with learning.

Equally important is promoting positive goal interdependence and promotive interactions. Johnson & Johnson (2005) attest that positive goal interdependence occurs when students perceive their success as linked to the success of the group as a whole, encouraging them to support each other.

This concept ensures that students work together to achieve the objective. Promotive interactions involve students actively encouraging, sharing knowledge, providing feedback, challenging each other's ideas and collectively solving the problem, thereby attaining a greater understanding of the concept (Johnson & Johnson, 2005). Encouraging a positive environment and explicitly teaching collaborative skills fosters academic achievement and social and interpersonal skills (Juceviciene & Vizgirdaite, 2012). Conversely, negative interdependence occurs when group members compete to attain goals, leading to oppositional interactions. In this dynamic, students are unwilling to share information or help their group members and may use coercion to achieve goals, affecting their ability to collaborate effectively (Johnson & Johnson, 2009; Tran, 2013). The distinction between positive and negative interdependence illustrates the delicate balance required to ensure collaboration leads to effective learning outcomes. While encouraging a supportive environment is essential, effective implementation hinges on adept teacher management. The following section delves into strategies that can be employed to ensure effective collaboration.

Teachers play a critical role in creating an active, non-competitive classroom. Giving students enough time to finish their work and forming small, diverse groups highlights the complexities of enacting collaborative learning and tasks. Miller et al. (2022) and Johnson and Johnson (2005) assert that for effective collaboration, teachers need to consider students' abilities, gender, and personality to encourage diverse perspectives, improve students' motivation, and enhance students' metacognitive, interpersonal and social skills. Failure to implement heterogenous groups may affect group cohesion, adversely affecting performance and self-esteem (Miller et al., 2020; Van Mechelen et al., 2015). These advantages and challenges are illustrated in the diverse perceptions of students and teachers. Some studies highlight a strong appreciation for the intra-active learning style, while others cite frustrations with group dynamics and implementation difficulties (Le et al., 2018; McKay & Sridharan, 2023). Successful collaborative learning can be enhanced by integrating inquiry-based practices.

Educational frameworks in countries like Australia, the USA, Canada and Singapore support inquiry-based practice, which promotes collaborative learning by ensuring authentic and relevant instruction (Bhagat, 2017). Specifically, the Australian Curriculum (ACARA, 2020) and Bhagat (2017) advocate for inquiry-based practices across disciplines, particularly in science, to deepen understanding of scientific processes through questioning, investigating, analysing data, reflecting on the process and communicating findings. Research undertaken by Black and Allen (2018) and Tran (2013) underscores the importance of explicitly teaching social and interpersonal skills necessary for

effective collaboration. These include metacognitive skills like self-reflection and questioning, which activate prior knowledge and help construct new knowledge. Ensuring engaging content, designing attainable lessons, ensuring accountability and providing enough time for group processing are essential strategies for promoting an inclusive and productive collaborative environment (Hmelo-Silver et al., 2013; Loyens & Gijbels, 2008; Slavin, 1996). While collaborative learning offers numerous benefits, including enhanced metacognitive abilities, increased intrinsic motivation and improved social skills, its success depends significantly on creating a supportive classroom environment and effectively implementing strategic teaching practices. Exploring collaborative dynamics leads me to consider collaborative testing, specifically, a method that combines the benefits of collaboration and testing.

Collaborative Testing

Collaborative Testing Overview

Supported by positive group dynamics, collaborative testing can enhance understanding and address the challenges associated with high-stakes testing. Collaborative testing is a strategy whereby students in a group work together to answer questions on a quiz or test (Siegel et al., 2015). This method often includes formats like two-stage exams comprised of an individual section followed by a collaborative section (Siegel et al., 2015), or the collaborative section may precede the individual test (Eastridge & Benson, 2020). Studies undertaken by Eastwood et al. (2020), Leight et al. (2012), and others describe the benefits of collaborative testing in science-based undergraduate degrees, such as gaining a deeper understanding of the concepts and improved 21st-century skills (Efu, 2019; Gilley & Clarkston, 2014; Mahoney & Harris-Reeves, 2019). To grasp the full potential of collaborative testing, I need to delve into the concept of active learning, which lays the groundwork for effective collaborative testing.

Active Learning Environment

Active learning environments characterised by student-centred, interactive teaching methods represent a shift from lecture-based approaches, offering a foundation for collaborative testing. The effectiveness of this model is well-supported by extensive research, including studies by Freeman et al. (2014), Ting et al. (2019) and Doolittle et al. (2023), demonstrating enhanced student outcomes in exam performance and critical thinking skills. Nguyen et al. (2021) further emphasise that these dynamic settings encourage students to actively engage in knowledge construction through diverse classroom activities, such as problem-solving, collaborative activities, and peer-to-peer discussions,

promoting an immersive educational experience. However, these studies often focus on specific disciplines and contexts, which may not universally translate to all educational settings, suggesting a need for cautious implementation.

While the benefits of active learning environments are well documented, discussing the limitations and practicalities of implementing such pedagogical strategies is essential. Critically, Bernstein (2018) reviews active learning with a degree of scepticism towards its unchecked adoption, arguing that its success is contingent on specific classroom contexts, levels of student engagement and the teachers' pedagogical methods. Instead, he suggests a multifaceted stance that incorporates elements of direct instruction and active learning to optimise learning by balancing 'learning by viewing' and 'learning by doing' (p. 299). Bernstein (2018) notes that while active learning is popular with students, its effectiveness, as measured by course examinations, 'shows little or no correlation with academic performance' (p. 297). Questions about the reliance on traditional assessments that may not fully capture the benefits of active learning strategies designed to foster a broader set of cognitive and interpersonal skills are raised. Expanding on this critique, Biesta (2015, 2015b) argues that traditional examinations often fail to measure the comprehensive range of skills and outcomes that active learning aims to develop. Supporting this perspective, Andrade and Brookhart (2020) advocate for assessment methods that align with instructional strategies, promoting assessments that engage students in cognitive processes similar to those experienced during active learning tasks.

Collaborative testing emerges as an innovative application of these theories, where students actively engage in discussions, leveraging prior knowledge and argumentation to reach conclusions. Supported by research from Lombardi et al. (2021), Levine et al. (2018), Mahoney and Harris-Reeves (2019), Reiger and Heiner (2014) and Siegel et al. (2015), this method not only enhances critical thinking and argumentative skills but also deepens students' understanding of concepts. Plauborg (2018) extends the concept of active learning, suggesting that learning is not confined to a specific time, place or individual but is dynamically (re)configured through human and nonhuman interactions. This methodological shift has proven to narrow achievement gaps (Theobald et al., 2020) and improve self-regulated learning and metacognition (Lombardi et al., 2021), illustrating how collaborative testing actively utilises and reinforces these skills, thereby cultivating a richer, more effective educational experience for many students.

Enhancement of Critical Thinking

Students undertaking collaborative testing may gain a deeper understanding when the test is designed with questions that promote critical thinking and peer discussions (Mahoney & Harris-Reeves, 2019). In many undergraduate studies on collaborative testing, critical thinking enhanced understanding by students analysing and reasoning through the questions, defending their opinions, and listening effectively to other viewpoints (Levine et al., 2018; Meaders & Vega, 2022; Meseke et al., 2010). Despite these positive effects, some studies indicate that high-performing students do not benefit as much as middle- and lower-performing students (Cooke et al., 2019; Giuliodori et al., 2008). To offset this problem, Eastridge and Benson (2020) suggest adding higher-order thinking and open-ended questions to the statistics collaborative test to improve critical thinking, enabling students to use more in-depth learning strategies that benefit all students. While critical thinking is undeniably crucial, the success of collaborative testing also hinges on group formation dynamics. Understanding how strategic group formation influences outcomes can provide valuable insights into optimising collaborative testing methods.

Group Formation Dynamics

Effective collaborative testing relies on the strategic formation of the student groups, which can influence both the perceptions and outcomes of collaborative testing (Cao & Porter, 2017; Miller et al., 2022). Groups can be randomly assigned through computer programs or teacher-directed seating arrangements (Nafziger et al., 2011; Siegel et al., 2015), or students can form their groups (Gilley & Clarkston, 2014; Giuliodori et al., 2008; Leight et al., 2012). Mixing students by ability and personality helps maintain group cohesiveness, increasing intrinsic motivation, lessoning disagreements, and enabling students to feel more independent (Gilley & Clarkston, 2014; Nafziger et al., 2011). However, when considering gender dynamics in collaborative testing groups, the composition and management of these groups have significant implications for learning outcomes, especially for women in science courses (Miller et al., 2022). Miller et al. (2022) highlight that while diverse groups generally enhance learning, placing women in predominately male groups can adversely affect their experiences, as women thrive in cooperative, supportive environments. Notably, Miller et al. (2022) state that the ability of female students to choose their group members is more influential in shaping their educational experience than the actual gender composition of the group. This suggests that, while careful group formation by teachers is crucial, empowering female students to select their collaborators could enhance their collaborative testing experience. Cohesiveness in collaborative testing groups can break down due to dominant personalities who may override

discussions and pressure individuals, preventing students from articulating ideas and impacting their experience (Cantwell et al., 2016; Efu, 2019; Slavin, 1996). Prior experience in structured collaborative settings can enhance group dynamics and outcomes. Effective group composition sets the stage for productive peer interactions, which are vital for conceptual change during collaborative testing by fostering an environment where students feel supported and engaged.

Peer Interactions and Conceptual Change

Constructive peer interactions during high-stakes collaborative testing can enhance student understanding, as constructivist theories attest that peer interactions drive conceptual change through the active exchange of ideas (Mahoney & Harris-Reeves, 2019; Reiger & Heiner, 2014). There have been mixed reviews in higher education as to whether collaborative testing deepens understanding, with Fournier et al. (2017) and Sandahl (2010) asserting no improvements in retention, while others, like Green et al. (2018) and Vogler and Robinson (2016) report positive benefits in retention for their students. The benefits to student understanding can arise as students actively discuss questions, critique each other's answers, and construct new solutions to reach a consensus. In effect, this method enables students, through peer interactions, to intellectually cooperate, leading to cognitive growth (Levine et al., 2018; Mahoney & Harris-Reeves, 2019; Reiger & Heiner, 2014; Siegel et al., 2015). Moreover, as Giuliodori et al. (2008) assert, peer-to-peer instruction in collaborative testing resembles Vygotsky's (1978) zone of proximal development, where students peer teach their group members to gain the necessary understanding, benefiting both high and low-performing students (Giuliodori et al., 2008; Juceviciene & Vizgirdaite, 2012). Additionally, students' opinions indicate that discussions with their peers led to increased confidence and communication skills, increased learning through explanations and arguments, and the ability to 'see' problems differently (Cooke et al., 2019; Eastridge & Benson, 2020; Leight et al., 2012; Mahoney & Harris-Reeves, 2019).

However, negative peer interactions can impact the effectiveness of collaborative testing (Nafziger et al., 2011). Eastwood et al. (2020) attest that if the groups are not correctly formed to ensure positive interdependence, competition may occur, causing group dynamics to weaken and leading to social loafing. Social loafing occurs when students do not participate in the group and depend on others' work (Nafziger et al., 2011). Competitive groups can hinder understanding, as a student may threaten or coerce others in the group, preventing high-quality discussions and the exchange of ideas (Levy et al., 2018; Reiger & Reiger, 2020). To mitigate these effects, Cantwell et al. (2016) and Siegel et al. (2015) posit that students must have experience working collaboratively in low-

stakes assessments and should be placed into small-assigned groups. Peer entanglements facilitate conceptual understanding and enhance metacognitive awareness and intrinsic motivation.

Metacognitive Awareness and Intrinsic Motivation

In collaborative testing, student interactions and discussions can increase metacognitive awareness, positively impacting achievement (de Carvalho Filo, 2010). As Hattie and Donoghue (2016) note, these interactions allow students to observe how their peers use their testing strategies to answer questions effectively. Moreover, Levy et al. (2018) note in their study on collaborative testing that such interactions help the student to focus on the task, enabling deeper processing of the test questions and leading to greater understanding. Furthermore, self-esteem and self-efficacy may increase as students observe their peers listening and responding positively to their ideas while answering test questions. Improvements in self-efficacy benefit student learning by increasing intrinsic motivation and persistence to problem-solve (Kapitanoff, 2009; Parkinson, 2014). With growing social skills and self-efficacy, students' sense of empowerment increases, leading to improved learning (Kapitanoff, 2009, Meaders & Vega, 2022). However, failure to meet these conditions may prevent some students from collaborating with their peers. Due to feeling overly dependent on others, the group may be unproductive, or conflicts may arise (Parkinson, 2014; Saunders et al., 2019). In addition to metacognitive and social skills, intrinsic motivation can be enhanced in a collaborative testing environment (Parkinson, 2014). When testing groups are formed to promote positive interactions, students are more motivated to participate, which increases students' self-efficacy and self-confidence, leading to higher engagement, which in turn enhances their competence and cognitive understanding (Cantwell et al., 2014; Gilley & Clarkston, 2014; Hargreaves, 2007). As I consider the cognitive and motivational benefits of collaborative testing, it is also essential to address how this method can alleviate the anxiety typically associated with high-stakes testing, thereby improving student performance.

Collaborative testing can reduce test-related anxiety in students (Krispenz & Dickhäuser, 2018). Stress associated with high-stakes testing affects student understanding in various cognitive and physiological ways. Cognitively, students may not be able to retrieve, organise or retain information, which impacts the student's ability to access prior knowledge, affecting their ability to answer questions (Breedlove et al., 2004; Krispenz & Dickhäuser, 2018; Levine et al., 2018). Additionally, physiological or mental health reactions impact the student's confidence and self-esteem, increasing feelings of helplessness and insecurity and lowering motivation, further impacting the student's test

performance (Breedlove et al., 2004; Krispenz & Dickhäuser, 2018). To mitigate negative impacts on students, Premo et al. (2018) assert that peer-to-peer interactions in collaborative testing may reduce testing anxiety by sharing information, discussing questions, and receiving academic and social support from their peers. Multiple studies on collaborative testing have demonstrated that this assessment method reduces testing anxiety in students, enabling them to focus more on activating their prior knowledge and reflecting on their performance (Breedlove et al., 2004; Gilley & Clarkston, 2014; Krispenz & Dickhäuser, 2018; Levine et al., 2018). Having discussed the effects of collaborative testing on students and teachers, I now move on to explore how collaborative testing can address students' 21st-century skills.

Collaborative Testing and 21st Century Skills

The concept of 21st-century skills encompasses a broad range of competencies essential for success in today's dynamic and interconnected world. Geisinger (2016) emphasises that cognitive skills, including critical thinking, problem-solving and creativity, are at the forefront. Meanwhile, Kennedy and Sundberg (2020) assert that critical thinking is essential for navigating complex problems, as it involves analysing situations and making reasoned judgements. They further explain that problem-solving requires identifying issues and developing effective strategies to address them. At the same time, creativity encourages innovative, out-of-the-box thinking, which is invaluable in an ever-changing landscape.

Interpersonal skills are indispensable for effective collaboration, including communication, teamwork, cultural sensitivity, and the ability to work effectively in diverse groups (Geisinger, 2016). According to Thornhill-Miller et al. (2023), verbal and nonverbal communication skills are essential for conveying ideas clearly and effectively. Geisinger (2016) discusses teamwork as a collaborative effort to achieve common goals while maintaining respect for the diverse backgrounds within groups. These skills are crucial for building and maintaining productive relationships. Furthermore, intra-personal skills, which involve self-management, adaptability, self-regulation, and metacognitive skills, enable students to manage their learning by setting goals, planning, and organising their efforts to achieve them (Lombardi et al., 2021). Adaptability and self-regulation are crucial for responding to changing situations. At the same time, metacognitive skills facilitate planning, monitoring and evaluating one's learning and performance, as Thornhill-Miller et al. (2023) highlight. These intra-personal skills are increasingly vital in settings that demand innovative thinking and global awareness.

As a pedagogical method, collaborative testing can enhance 21st-century skills within educational settings. This assessment method not only evaluates student knowledge but also promotes the development of critical thinking, problem-solving and interpersonal skills through peer interactions and teamwork (Dolmans et al., 2016; Mahoney & Harris-Reeves, 2019; Siegel et al., 2015). Significant research in higher education illustrates that collaborative testing promotes critical thinking and problem-solving, encouraging students to defend their viewpoints, reason through problems collaboratively, and integrate diverse knowledge and perspectives, as noted by various studies (Levine et al., 2018; Rempel et al., 2023). Efu (2019) notes that as students articulate their thoughts and respond to peers, they refine their understanding and recognise the strengths and weaknesses of their reasoning. Effective collaboration during these tests also hinges on clear and efficient communication. Students must convey their thoughts, listen actively, and negotiate solutions, enhancing verbal and nonverbal communication skills (Levine et al., 2018; Morreale et al., 2017; Siegel et al., 2015). In a positive environment, collaborative testing fosters a sense of shared responsibility and interdependence among the groups as they learn to appreciate different perspectives, manage conflicts and build consensus, as noted by Headrick et al. (2021) and Johnson and Johnson (2008). These experiences are crucial for building teamwork skills and the ability to work with diverse people, equipping students with the necessary competencies to succeed. Collaborative testing assesses knowledge and is a powerful tool for developing a broad spectrum of 21st-century skills, thereby supporting the holistic development of students for future challenges and opportunities.

Despite the extensive research on collaboration and collaborative testing in higher education, there is a noticeable gap in the literature regarding its application in a high school setting, particularly in science classrooms. My study aims to fill this gap by exploring students' and teachers' perceptions of collaborative testing in three high school science classrooms. By examining all elements involved in collaborative testing, I seek to illustrate that learning is not merely a static acquisition of knowledge but a dynamic process. Understanding the complexities of active learning environments at the higher education level sets the stage for exploring how collaborative testing impacts student understanding and the development of collaborative skills in a context that has been underexplored in the existing research.

Summary

In Part 2, I explored collaboration and collaborative testing, determining how they impact student learning through enhancing engagement, fostering deeper understanding, and improving critical thinking and problem-solving skills. Through peer discussions, collaborative learning

encourages students to intra-twine and share knowledge to develop a robust understanding of concepts. This approach promotes an active learning environment where students engage in meaningful dialogue and reflect on their ideas. Collaborative testing extends these benefits by utilising group dynamics to deepen understanding as students clarify concepts through discussions. Collaborative testing contributes to an educational experience that can enhance academic outcomes and prepare students for future challenges.

Conclusion

This chapter's exploration of diverse assessment types has underscored their critical roles in reinforcing and challenging existing educational paradigms. My review demonstrates that assessments shape educational practices and policies, influencing everything from individual student trajectories to national educational standards. While traditional methods continue to dominate, there is a growing recognition of the need for more innovative and integrative assessment methods as an integral part of the teacher-designed curriculum that responds to the needs of the learners. I have advocated for a balanced assessment approach in this literature review, emphasising strategies that measure learning outcomes while actively contributing to the educational process. The insights underline the potential for assessments to enhance learning environments and teaching efficacy, calling for ongoing innovation to better align with educational goals. In looking forward, this review sets the stage for exploring the philosophical foundations of diffractive ethnography, aiming to transform educational practices through innovative research methodologies that challenge traditional boundaries and encourage a deeper understanding of the entangled nature of educational phenomena.

Chapter 3: Methodology

In the evolving landscape of educational research, the call for methodologies that observe and engage with the complexities of learning environments has become increasingly pressing. This chapter introduces diffractive ethnography inspired by Karen Barad's foundational theories of agential realism (Barad, 2007). By adopting this innovative perspective, I seek to move beyond conventional observational methods and engage with the material-discursive elements that constitute educational settings. Diffractive ethnography positions me as an active participant in the collaborative testing phenomenon. This approach enables an exploration of complex settings where all elements play critical roles in shaping outcomes.

I begin the chapter by exploring the philosophical underpinnings of diffractive ethnography (Barad, 2007; Haraway, 1991), drawing on concepts from quantum physics to demonstrate how educational research can benefit from understanding the entangled nature of reality (Barad, 2007). I then delve into the practical implications of this approach in educational settings, highlighting how it redefines the roles of data, phenomena, and myself to produce meaningful insights. Through this exploration, the chapter aims to demonstrate how diffractive ethnography enhances our understanding of educational processes and promotes a transformative space where new forms of knowledge emerge.

From Observer to Participant: Evolving Research Paradigms

My perspective as a researcher has been shaped by my experiences as a science teacher in Australia and the United States, where I initially held a social constructionist view of 'knowledge-making' (Taylor, 2021, p. 29). I undertook a mix of explicit teaching and collaborative work in my science classes to promote student understanding. However, I could see the effects of the classroom environment on my students: the heat or cold, the overcrowding of the classroom, and the impacts of high-stakes tests, which affected their concentration and anxiety levels. However, when I came to start this thesis, I perceived these as separations between humans (students, teachers and researcher) and the objects in the classroom (such as desks, chairs, light, and ambient temperature)(Talbot, 2023). My early thoughts on methodology involved observing classroom dynamics as an outsider to gather the 'facts' about collaborative testing outcomes objectively. This approach assumed knowledge could be obtained through detached observation and quantifiable measures. As I delved deeper into the

literature, I encountered a critical problem: this detached approach seemed inadequate for understanding the nuanced perceptions of students and teachers regarding collaborative testing. This realisation prompted me to re-evaluate the research questions guiding this study, shifting my focus towards understanding the perceptions of teachers and students regarding collaborative testing and its influence on their understanding and skills. Through reading and reflexivity, I recognised the importance of moving away from the conventional bracketing method, where researchers distance themselves from their own experiences and perceptions to remain objective observers (Tufford & Newman, 2012).

In contrast to a distanced or outsider approach, I embraced the classroom environment, recognising its inseparability of human influences and material aspects. At this point, I engaged with Karen Barad's (2007) onto-epistemology, a framework that emphasises the intra-connectedness of all elements within the research setting. Barad (2007) captures this concept, asserting that 'all real living is meeting' and 'each meeting matters' (p. 535). Their framework emphasises that meaningful research results from intra-actions (Barad, 2007), allowing me to recognise matter and other than matter's role within the environment. Intra-action presents a dynamic alternative to the notion of interaction (Haraway, 1991). Barad (2007) explains that the conventional understanding of interaction assumes the existence of entities in isolation before they engage. In contrast, intra-action posits that entities do not pre-exist; instead, they actively emerge and materialise through their mutual entanglement (Murris & Bozalek, 2019). This active process of intra-action, as further explored by Juelskjaer et al. (2021), not only forms new entities but also transforms existing ones, engaging the audience in a continuous reconfiguration of reality.

I introduce the term *intra-active materiality* (Barad, 2007, p. 170) to describe the complex, dynamically entangled entities within the classroom, encompassing everything from students, smells, temperature, and sounds, each exerting agency within dynamic entanglements (Barad, 2007; Murris, 2022). In Bennett's (2010) book on *Vibrant Matter*, she emphasises the dynamically entangled nature of matter and nonmatter, highlighting the "vibrant, vital, quivering, effervescent and energetic materiality" (p. 112) that exists both alongside and within humans. Gullion (2018) further illustrates that 'things' are 'lively' (p. 129), engaging in constant intra-action and evolution, and having an equal effect on outcomes (Bryant, 2016). For example, the ambient temperature of the classroom might affect students' concentration as much as the physical presence of a teacher moving around the room. This perspective underscores that all components, regardless of their conventional classification, contribute to the perceptions of collaborative testing. By integrating the participants' direct

experiences and insights, I could explore the intra-actions within the classroom, offering a more holistic understanding of the impact of collaborative testing on teaching and learning.

Introduction to Onto-epistemology and Educational Implications

In a dynamic universe, I am not merely an observer but an active participant, intricately connected with the world and its myriad components. I strive against representational perspectives that often establish dichotomies between subjects and objects and between humans and nature (Barad, 2007). From my personal teaching experiences, I have observed how the principles of Karen Barad's (2010) performative philosophy play out in the dynamic intra-actions between students, teachers, and the school environment, impacting the learning process. Barad's (2007) philosophy reimagines our role in the universe as active participants, entangled with each other and our environments. This theoretical perspective, further explored by Murris (2022), suggests that our understanding of our environment evolves through continuous entanglements. Supporting this perspective, Lenz Taguchi (2010) argues for a holistic view of learning environments, stating, 'The learner and the world cannot be separated but are of the world in a co-dependency' (p. 47). Acknowledging a dynamic world underscores the need to move beyond a human-centred perspective to fully grasp our place within Earth's broader ecosystem, a sentiment echoed by Blom (2022). This recognition of the interdependence of all elements allowed me to explore how factors like temperature and classroom layout are active in the learning process. Extending beyond the classroom, Spyrou (2019) critiques traditional knowledge practises for their narrow focus, which often marginalises the 'other' (p. 38–39), nonmaterial entities, and 'other than human man' (Murris, 2017, p. 46) that are crucial components of what Murris (2017) refers to as the 'fold of knowers' (p. 46). In practice, this means acknowledging all elements of the classroom environment to broaden the scope of what is considered valuable knowledge.

Moving beyond a focus on human-centred research, I am committed to (re)positioning[3] and developing a broader entanglement of all elements in this study, as proposed by Orlikowski and Scott (2015). This method is essential to understanding teachers' and students' perceptions of their environment, where knowledge and reality are not static but are dynamically intra-connected,[4] as both

[3] The brackets symbolise the dynamic nature of the world, which is continually shaped and reshaped in the environment (Gullion, 2018).
[4] Intra-connectedness involves all parts of the phenomena, as they impact and are impacted by each other in that time and space (Barad, 2007).

Braidotti (2019) and Lenz Taguchi (2010) highlight. Barad (2007, p. 379) articulates this concept powerfully: 'Knowing is a direct material engagement and process of intra-acting with and part of the world in its dynamic material configurations and ongoing articulations.' This inseparability of meaning deepens our understanding of knowledge and reality, facilitating knowledge construction (Taylor, 2021). The entanglement of 'what is real and how we know that reality' (Gullion, 2018, p. 125) forms the foundations of Barad's (2007, p. 44) 'onto-epistemology,' a framework that encompasses all material and discursive elements of the environment, which is essential to grasp the nature of reality and knowledge. The following section builds on these foundational ideas to delve into agential realism, exploring how it deepens our understanding of the intricate relationships within educational environments.

Agential Realism

Building on insights from Niels Bohr, Judith Butler, Donna Haraway, and others, Karen Barad (2003, 2007) developed the concept of agential realism. This philosophical framework helps us to understand how phenomena with specific temporal and spatial characteristics come into existence. Agential realism extends beyond individualistic perspectives, by focusing on the 'relational nature' of space, time, and entities as described by Lenz Taguchi (2010, p. 166) and further explored by Murris (2022). This approach immerses us in the intra-connections and intra-dependencies that shape our reality, offering a nuanced understanding of our position within the world. According to Barad (2007), agential realism is:

> An epistemological-ontological-ethical framework that provides an understanding of the role of human and nonhuman, material and discursive, and natural and cultural factors in scientific and other social-material practises, thereby moving such considerations beyond the well-worn debates that pit constructivism against realism, agency against structure, and idealism against materialism (p. 26)

In this framework, relational practices emerge not from pre-existing interactions but through intra-actions, processes through which entities and their relationships are continually formed and reformed. Agential realism challenges views that see representations as some form of pre-existing reality, arguing instead that representations are active participants in their existence (Barla, 2023). For instance, a flower in a painting is not just a passive image; it is affected by and affects the conditions of its

depiction (Barla, 2023; Murris, 2022). In the context of education, agential realism illustrates how spatial arrangements in the classroom can affect learning outcomes or how the temporality of the school day impacts student engagement (Orlikowski & Scott, 2015). An agential approach recognises the intricate intra-dependencies between learning, lived experiences, and environmental factors (Barla, 2023). This viewpoint sees the classroom as a physical place and active in the educational process, which is crucial for enhancing educational outcomes (Plauborg, 2018). By adopting agential realism, I acknowledge the intricate web of dependencies that shape educational experiences, recognising that phenomena are not fixed but are continuously (re)shaped in the intra-actions (Barad, 2007; Brown et al., 2020). The transition from theory to practice is critical, as I discuss in the next section, where I apply agential realism to concrete examples of phenomena and apparatus.

Phenomena and Apparatus in Research Design

Phenomena

Karen Barad's (2007) introduction of agential realism marked a significant theoretical advance, describing phenomena as 'continuously emerging through the co-action with the world' (p. 333). They applied Niels Bohr's ideas from quantum physics to explore how matter and meaning become entangled. This outlook challenged traditional representationalism, which distinctly physically and conceptually separates objects with fixed properties from the subject (Barad, 20014; Hultman & Lenz Taguchi, 2010). Expanding on Bohr, Barad (2007) states that phenomena include:

> Ontological units at the core of reality, not predetermined by measurements or the inseparability of observer and object, existing as part of a continuous intra-action, a dynamic intra-weaving that determines and redefines boundaries, properties and causality within a constantly shifting network of relations (p. 33)

In educational settings, this translates to a complex intra-action where teachers, students and phenomena entangle to shape learning experiences. Educational phenomena arise through the iterative (re)forming (dis)connecting within these intra-actions, shaped by the specific circumstances in that spacetime (Barad, 2003; Bergstedt, 2021). The concept of 'agential cuts' (Barad, 2007, p. 148) is crucial in determining what aspects of the environmental process are included or obscured. These cuts are not mere separations but are processes that dynamically enact and redefine boundaries within the learning environment (Klykken, 2023). According to Bergstedt (2021) and Thiel (2018), intra-

actions play a crucial role in shaping educational phenomena by dynamically enacting these boundaries rather than predetermining them. Murris and Bozalek (2019) further explain that these agential cuts are not separations but processes of 'cutting together-apart' (Barad, 2014, p. 168), revealing the intra-connectedness within environments. With a refined understanding of phenomena conceptualised by agential realism, my analysis examines the apparatus: the tool through which we observe and influence these educational phenomena.

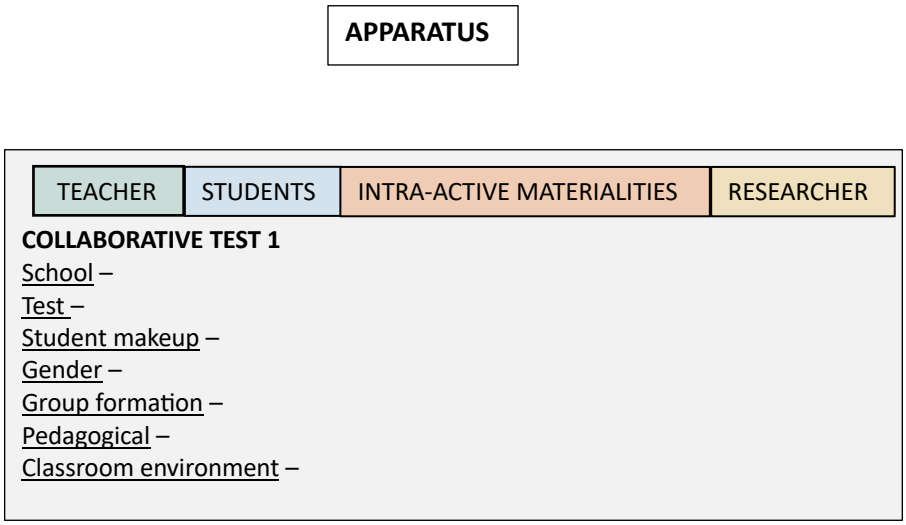
Apparatus

In Karen Barad's (2003) agential realism, the term' apparatus' is not merely a collection of sayings, tools, or instruments but a dynamic set of material-discursive intra-actions (re)configured by the phenomena under study (p. 816). The researchers' apparatus, therefore, is not static but contextual and dynamic, intra-acting with phenomena, policies, and histories to make meaning of the concepts under study (Barad, 2003, 2007; Charteris et al., 2019; Holford, 2020). Bozalek and Fullagar (2021) describe this intra-action as a complex intra-relationship of apparatuses for perceiving reality's ever-evolving nature. Moreover, Murris (2022) contends that apparatuses have flexible arrangements with human, matter and non-matter subjectivities, which help the researcher determine what to include or exclude based on the contextual situation. These apparatuses subtly shift, leading to multiple outcomes based on their components and relationships with other apparatuses (Barad, 2007). As Barad (2007) describes, 'apparatuses are the material conditions of possibility and impossibility of mattering; they enact what matters and what is excluded from mattering' (p. 148). As such, apparatuses set specific *boundary-drawing practises* (Barad, 2007, p. 206, emphasis in the original) in the course of the intra-action, defining what is intelligible and what remains excluded in an ongoing, open-ended practice (Hayes et al., 2020). Apparatuses then enact agential cuts that materialise the phenomena at a given spacetime.

My research journey into the study of collaborative testing within educational settings has been guided by an iterative process framed by Karen Barad's (2007) concept of the apparatus. This journey began with a broad overview, considering the roles of students, teachers, the classroom environment and broader societal influences such as the marketisation of education. However, as my understanding deepened through further research and contemplation on factors influencing students during tests, I realised that the most critical influences shaping perceptions were those occurring within the classroom. These include the complex entanglements of teachers, students, and phenomena during collaborative tests (Talbot, 2023). This realisation led me to narrow the apparatus to examine the nuances of these classroom intra-actions more closely.

Further exploration introduced an ethical dimension to constructing the apparatus for studying collaborative testing. In shaping my research framework, I meticulously considered which variables to include or exclude as part of the intra-action (Barad, 2007). The variables I included define the factors that transform perceptions of collaborative testing into discernible knowledge. Meanwhile, those excluded remain outside the current scope of inquiry but hold the potential to be clarified through new intra-actions (Barad, 2007). Chapter 4 details these methodological decisions, including the inclusions and exclusions specific to this study. They reflect a deliberate and ethical engagement with the apparatus, acknowledging that each decision impacts how collaborative testing is constructed and understood. Figure 2 illustrates the dynamic interplay of various elements within the apparatus used in this study, providing a visual representation of the complex intra-actions that shape collaborative testing environments.

Figure 2. *Diagram of the Apparatus for each Collaborative Test*



Material-discursive practices

Grounded in the theoretical framework of Foucault and Bohr, material-discursive practices offer a transformative perspective on the 'dualisms of discourse and materiality' (Barad & Gandorfer, 2021, p. 29). Discursive practices are traditionally viewed as distinct from the material components of the world, such as physical objects, bodies, and environments, and involve the symbolic and linguistic aspects of creating meaning (Lenz Taguchi, 2010). However, within the framework of 'agential realism, these methods are seen as dynamic, ongoing boundary-making processes' (Barad & Gandorfer, 2021, p. 29) that are continuously enacted, thus shaping the world (Plauborg, 2021). These practises are not

merely reflective but are causally intra-connected, marking and being marked by other components in a complex web of intra-actions (Holford, 2020). This reconceptualisation leads to a rethinking of materiality: phenomena is not static or passive but is a dynamic process of (re)stabilisation within the entanglements (Barad, 2007; Juelskjær et al., 2020).

The relationship between discursive practices and phenomena is characterised by mutual intra-dependence (Plauborg, 2018). According to this perspective, humans, other than humans, and meaning cannot be articulated in isolation (Holford, 2020); they are co-constitutive and mutually implicated in phenomena' entanglement (Barad, 2003). This concept challenges conventional views that treat matter and discursive elements as separate, emphasising their intra-twined nature, where each continually influences and shapes the other in an ongoing process of intra-action and boundary formation (Juelskjær et al., 2020). While discourse in education often focuses on the content delivered by the teacher and absorbed by the students, this framework recognises the significant yet frequently overlooked role of matter in the learning process. Here, discourse encompasses content, teaching methods, and the physical and social contexts (Evans, 2016). Lenz Taguchi (2010) conceptualises learning as a dynamic process that emerges from intricate intra-actions within the classroom, including the entanglement among students, the dynamics of the learning environment, classroom materials, and the specific temporal and spatial conditions particular to that classroom. According to this view, learning is a collective activity involving various elements co-existing in the educational space, enabling the unknown to become known (Gullion, 2018; Murris, 2020). Exploring material-discursive practice redefines our understanding of the relationship between discourse and phenomena, highlighting the potential for these concepts to innovate and transform educational strategies, fostering an environment where learning is seen as a dynamic, co-creative process.

Collaboration as a Material-Discursive Practice

Collaboration in educational practices involves intra-actions across multiple dimensions between the teacher, class and classroom. These intra-actions, as Barad (2007) suggests, are not limited to interactions between pre-existing entities but also include the entanglement of material-discursive elements. For example, the effectiveness of student and teacher intra-actions is influenced not only by interpersonal dynamics but also by the material environment—classroom layout, available technology, and curriculum design, all of which are entangled with teachers' perceptions of collaboration as a pedagogical strategy (Akiba & Liang, 2016). By understanding collaboration as material-discursive, I recognise that it emerges from the intra-connectedness of discursive practices

(such as pedagogical approaches and policies) and material conditions (such as classroom space and environmental conditions).

Muckenthaler et al. (2020) identify several levels of teacher collaboration, each involving distinct material-discursive entanglements. At the basic level, the exchange of ideas, information or materials allows teachers to maintain individual autonomy, while fostering shared influence. These exchanges are material-discursive in nature, as they are shaped by the tangible (e.g. lesson plans or teaching materials) and intangible (e.g. professional trust and cultural norms) elements of the educational environment (Muckenthaler et al., 2020). A higher level of collaboration necessitates more involvement from the teachers, who must share information and implement common goals and tasks. For instance, two physics teachers might simultaneously plan and implement a unit of work yet remain autonomous. The most complex form of collaboration entails co-constructing standards, goals, syllabi, implementation and assessment—a process deeply influenced by material-discursive intra-actions. Beninghof and Leensvaart (2016) argue that collaborative lesson planning introduces multiple perspectives, enhancing pedagogical practises. This deeper form of collaboration engages material-discursive dimensions by requiring teachers to integrate their shared knowledge with institutional resources and constraints to design innovative lessons that improve student engagement and learning outcomes. Moreover, Akiba and Liang (2016) attest that effective collaboration on implementation strategies can help teachers focus more on content delivery, thereby increasing student learning and motivation in the classroom. In a supportive school environment, teacher collaboration may reduce stress and boost confidence in adopting innovative teaching methods that enhance learning outcomes (Akiba & Liang, 2016).

Despite these benefits, its practical application in schools remains limited. Vangrieken et al. (2017) observed that teachers often avoid deeper, more meaningful peer collaboration. This reluctance can stem from various reasons, including a desire to remain autonomous and maintain control over their teaching approaches. These barriers highlight the entangled material-discursive constraints on collaboration. For instance, workload concerns are not only material (increased planning time) but are also shaped by discursive pressures such as the perception that collaboration may dilute individual accountability, as Johnson et al. (2014) highlighted. Reeves et al. (2017) further highlight the critical role of the school environment—a site of material-discursive entanglement—in fostering or hindering collaboration. A non-competitive atmosphere and supportive leadership can enhance teachers' openness to collaboration, leading to more positive experiences and increased classroom collaboration use (Reeves et al., 2016). Conversely, performance-driven environments emphasising individual performances create discursive tensions and material barriers, such as rigid

evaluation framework, that fragment collaborative efforts (Reeves et al., 2016). According to their studies on teacher collaboration, Muckenthaler et al. (2020) and Vangrieken et al. (2017) attest that forced collaboration by school leadership may exacerbate these issues, disrupting the balance of intra-actions by imposing external expectations without addressing the material and discursive conditions necessary for effective collaboration. By reframing collaboration through material-discursive approach, it becomes evident that effective collaboration requires more than shared goals or strategies. It depends on creating intra-actions that align discursive practices (polices, pedagogical approaches) with material realities (resources, scheduling, and classroom environments) (Murris, 2023; Talbot,2023). Without addressing these entanglements, efforts at collaboration may remain fragmented or superficial, limiting their impact on both teacher practices and student learning outcomes.

Entanglement in Educational Practice

With the theoretical foundation of agential realism laid out, I progress to a focused analysis of the nature of entanglement in diffractive ethnography. I will examine the evolution of ethnography to diffractive ethnography and discuss how educational practices can transcend conventional methods, fostering a collaborative and transformative space that integrates the material-discursive.[5] This section will underscore their consequential role in knowledge construction, particularly through their application in educational research, moving from abstract theory to concrete methodology. By understanding and applying the concepts of entanglements, phenomena and apparatuses, the classroom environment can be seen as a vibrant space where knowledge is actively (re)shaped through continuous intra-actions.

Ethnography and its Evolution Towards Diffractive Ethnography

'All research is anchored to basic beliefs about how the world exists' (Pascale, 2010, p. 3).

Pascale's (2010) quote emphasises the fundamental role of beliefs in shaping research and underscores the importance of clarifying our perceptions of the world. It posits that our connections with our surroundings profoundly influence our ability to generate new knowledge. Drawing on Gullion (2018), I argue that viewing the world through the perspective of material-discursive practices fosters a more nuanced exploration of knowledge creation. This perspective facilitates a unique approach to

[5] Material-discursive practices configure reality by intertwining discourse and the material world, as Barad (2007) explains. They note that neither discourse nor material reality can exist independently; they define and influence each other, shaping our understanding and interactions with the world.

understanding phenomena and emphasises the importance of acknowledging the intra-connectedness of all elements involved in the learning process. Ethnography has been a cornerstone in exploring human behaviours and interactions in educational settings. However, as Hammersley (2018) and Reeves et al. (2008) illustrate, the focus has predominantly revolved around human interactions, with the dominant culture taking precedence, potentially overshadowing and neglecting the importance of matter. The advent of approaches such as institutional ethnography (Talbot, 2020) marks a shift towards recognising participants as embodied beings actively existing within temporal and spatial contexts. I take this further by embracing an entangled view of ethnography, influenced by Donna Haraway (1997) and Karen Barad (2007), where all elements are entangled and mutually influence one another to (re)configure our understanding of collaborative testing. As Murris and Fullagar (2022) argue, entanglement is more than an intertwining; it decentres the human to equalise all phenomena in material-discursive practises. Such a perspective is crucial for reimagining the scope of ethnographic research in education, allowing for a more inclusive and comprehensive exploration of how different elements contribute to the perceptions of students and teachers in collaborative testing. By integrating material and discursive elements, diffractive ethnography recognises the contributions of all components within the educational landscape.

A new approach to understanding matter and meaning challenges the separation of humans from matter, which is often seen in research and education. This perspective emphasises that intra-actions and knowledge are context-dependent, emerging through active entanglements with social and physical environments. Entanglements are pivotal in Karen Barad's framework of agential realism, emphasising how the material elements gain significance. In their seminal work, Barad (2007) clarifies this concept:

> To be entangled is not simply to be intertwined with another as in the joining of separate entities, but to lack an independent, self-contained existence. Existence is not an individual affair… rather individuals emerge through and as part of their intra-relating (p. ix)

Our existence is not separate; I do not exist alone; I entangle with phenomena and events to co-construct our realities (Barad, 2007; Taylor, 2021). When applied to educational contexts, this perspective provides a nuanced position to explore how students and teachers perceive and are entangled within the learning environment. For example, tangible objects like books and desks and intangible elements like curriculum and classroom dynamics participate in the educational process.

Here, academic concepts and pedagogical strategies are not merely tools but active participants in the learning experience (Barad, 2007; Gullion, 2018). In this framework, Murris (2022) and Spyrou (2019) note that the entangled relations between human and the material-nonmaterial continuum give rise to a dynamic and enhanced awareness of the constraints and possibilities that shape practices. The concept of entanglements in collaborative testing underlines the complex intra-actions and relationships that shape perceptions and practises within education settings (Barad, 2007). These entanglements involve the dynamic connections among students, teachers, and the physical and abstract dimensions of the classroom. Understanding these entanglements is crucial, particularly in the context of high schools, where collaborative testing is a novel practice. Various elements intra-connect, creating a dynamic ecosystem where participants and their environment influence classroom dynamics (Murris, 2022). Adopting a diffractive method acknowledges the complexity and equal significance of all elements involved in collaborative testing.

My conceptualisation of entanglements differs slightly from Murris and Fullagar's (2021) approach, which advocates decentring the human to equalise all elements within material-discursive practices. While they emphasise an equalisation between all human and materialities and immateriality's, my approach centres on the experiences of students and teachers, acknowledging their primary role in education settings. However, I also recognise that matter can significantly influence these entanglements (Hultman & Lenz Taguchi, 2020). This thesis entangles human elements (teacher, student, researcher) with materialities and immateriality's (time-space, temperature, noise) to examine the teachers' and students' perceptions of collaborative testing. This performative approach, considering physical space, social facets, cultural influences, power dynamics and environmental factors, facilitates a holistic understanding of the implementation and impacts of collaborative testing. In this framework, meaning emerges through phenomena informed by these entangled and performative strategies.

In the preceding section, I explored the implications of onto-epistemology for educational practice, challenging prevalent hierarchal structures and advocating for an intra-connected learning environment that equally values all participants and components. Based on Karen Barad's (2007) agential realism, this approach underscores the relational nature of knowledge and reality, where learning is seen as a dynamic process shaped by the entanglement of all components of the phenomena. Building upon these philosophical foundations, the next section will delve into diffractive ethnography. By fostering a transformative space and integrating the material-discursive, I will examine

how this methodological shift can help transcend conventional educational practices and reimagine the construction of knowledge.

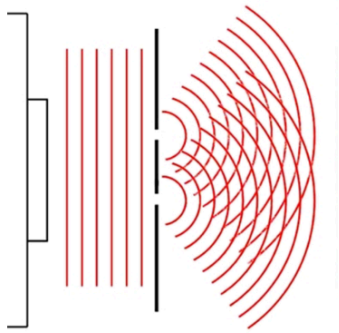
Diffractive Ethnography

The evolution from ethnography to a post-modern approach marks a significant transformation in social research methodologies. Initially, ethnography sought to identify universal laws through objective observations, aiming to generalise behaviours across various contexts (Howell, 2012). As Reeves et al. (2013) argue, this approach tended to elevate the researcher's perspective above that of the participants, assuming it to be superior and more valid. However, by recognising the limitations of this stance, contemporary ethnography has evolved to acknowledge that realities are not singular but multiple, constructed through social interactions (Hammersley, 2018). As Howell (2012) notes, this shift renders the established stance of researcher detachment impractical and ineffective for grasping the nuances of cultural experiences. Current ethnographic practices emphasise the importance of researchers immersing themselves within the cultural settings they study, directly engaging with participants to learn their unique perspectives (Reeves et al., 2013). This immersive approach fosters a more empathetic and inclusive approach, with research findings deeply informed by the lived experiences of the community involved (Hammersley, 2018). However, this human-centric focus sometimes overlooks environmental factors. To address this gap, Gullion (2018) advocates for a modern ethnographic approach that embraces the researchers' dynamic and evolving experiences as part of the intra-actions within temporal and spatial contexts. This perspective has inspired new methodologies such as 'Transqualitative' research (Blom, 2022, p. 113), which captures the fluidity of research by encompassing both material and nonmaterial elements, providing a more comprehensive view of how individuals and their surroundings co-influence each other.

Diffractive ethnography introduces a novel research approach that deeply engages with all elements of intra-actions. Unlike methods that often prioritise linguistic representation and reflexivity, which can oversimplify real-world complexities by maintaining a binary view of various dualities (Murris, 2021), diffractive ethnography considers our dynamic world where 'knowledge represents a stable and observation-independent reality' (Barla, 2023, p. 5). Inspired by the physics concept of diffraction, diffractive ethnography highlights wave-particle duality, where light waves give rise to intricate patterns (Barad, 2007). Figure 3 shows the diffraction patterns produced when light passes through two slits, describing how diffractive ethnography reveals new insights through the intra-action of various elements. However, these patterns are not merely reflections; they reveal new insights as the experimental conditions change (Peralta, 2020). Donna Haraway (1997) uses the term diffraction

in her work *Modest Witness Second Millennium: FemaleMan Meets OncoMouse: Feminism and Technoscience* to advocate for a method that immerses the researcher in the educational environment. This approach emphasises the entanglements of words, objects, individuals, their intra-actions and their centrality to knowledge production.

Figure 3. *Illustration of the Interference Pattern Resulting from the Diffraction of Light through Two Slits.*



Note. This figure illustrates the interference pattern observed when light passes through a double slit illustrating the wave-like properties of light as varying experimental conditions result in distinct patterns. This phenomenon underscores the dual wave-particle nature of light, as demonstrated by the formation of interference patterns rather than simple reflections.

The diffractive diagram was taken from John D. Norton's *Diffraction Diagram*, published in 'The Quantum Theory of Waves and Particles,' Pittsburgh 2001. © John D. Norton 2018.

This method's non-linear and unpredictable nature allows knowledge to emerge from the coming together of objects, ideas, and apparatus, all becoming determinate in that specific *spacetimemattering* (Barad, 2007, p. 181, emphasis in the original), which demands active researcher engagement. As Barad (2007, p. 91) notes, 'the researcher does not stand outside the world but is part of the very material entanglements which reconfigure the world at every moment, making it possible to produce knowledge differently.' This responsibility means that the researcher must carefully consider the material-discursive elements they engage in while being mindful of how these choices impact the phenomena being studied and the broader educational landscape (Barla, 2023). For instance, my role as a science educator or a white woman could influence where the cuts[6] are made and what is brought forth. Thus, I bear responsibility for both the human, matter and the meanings enacted and excluded within the apparatus, constantly discerning 'what comes to matter, how it comes

[6] Cuts are dynamic, context-dependent splits in the practises of observing and understanding phenomena. They create connections and separations without fixed boundaries, enabling us to untangle intra-actions to better perceive phenomena (Barad, 2007).

to matter and for whom it matters' (Barad, 2007, p. 90). Further exploring this concept, Hayes et al. (2020) reveal that a diffractive methodology not only uncovers but also emphasises emerging differences, and importantly, it delineates what is omitted. Inspired by both Donna Haraway (1997) and Karen Barad (2007), I embrace an entangled view of ethnography that considers all elements as mutually influential to (re)configuring our understanding of phenomena like collaborative testing. By embracing an active role in the collaborative testing environment, I reveal how various educational practices, matter, and intra-personal relations shape learning outcomes. I recognise the intricate entanglements[7] between humans (teachers, students, researcher) with material elements (desks, teacher and students' views on collaboration, heat, overcrowding, collaborative tests and the timing of tests) to generate student and teacher perceptions.

This exploration of Karen Barad's diffractive methodology highlights how intra-actions between various elements, apparatus, participants, and phenomena create new perspectives. This method challenges binary distinctions and promotes the inclusion of all entities in research activities, irrespective of their physical nature as noted by Barad (2014) and Murris (2022). By developing this methodological framework, I illustrate how intra-actions significantly affect our perceptions of collaborative testing. I establish that every participant, human and material-nonmaterial actively contributes to co-constituting the learning environment and influencing outcomes. As Murris (2022) and Plauborg (2018) describe, a diffractive analysis is keenly aware of these differences, their creation, and their impacts on responsible knowledge creation, (re)configuring assessment practices to open new avenues for understanding and improving pedagogical practises.

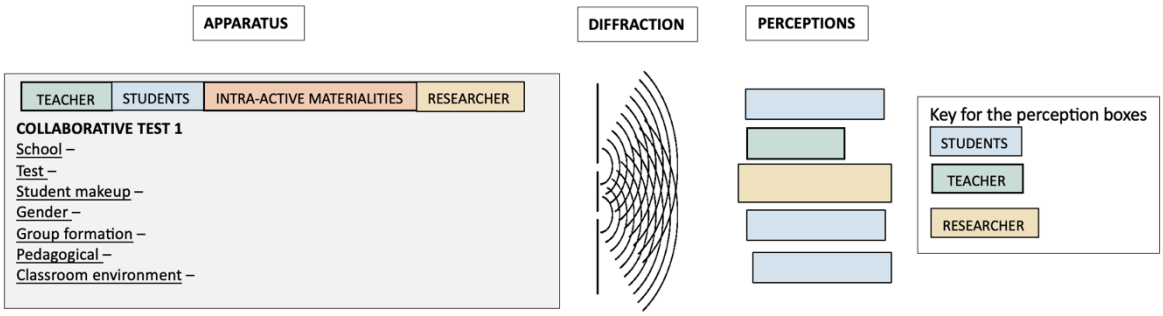
Components of the Diffraction

This study employs a diffractive methodology to analyse the dynamic intra-actions among human and material elements in the environment, shaping educational outcomes through complex relationships, as illustrated in Figure 4. To describe the rich settings of the students and teachers, I will use a case-study approach, where each class, including the teachers, students, and material and environmental conditions in that classroom, represents a separate case study. The apparatus used in this research comprises a combination of human and material elements within classroom settings and assessments, shaped by the data encountered and distinctions made in the apparatus during the research process (Barla, 2023). Each apparatus represents a particular reality: one class, one

[7] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, teachers, researcher and the material elements of the classroom environment (Barad, 2007).

collaborative test, one mind map, and myself as the researcher. While other realities are excluded from any specific apparatus, they can exist in other apparatuses (Barad, 2007). Thus, this study does not cover all possible realities but is detailed within its scope, incorporating twelve distinct apparatuses. The apparatuses share categories: subject, classroom, test and student composition, but there are variations in test content and environmental conditions such as ambient temperature, classroom setup, noise levels, and test timing. Chapters 5, 6, and 7 will describe these apparatuses in detail for each class, while Chapter 4 extensively details each apparatus's inclusions and exclusions. Each teacher's involvement is elaborated upon in their respective chapters: Halo, Beta, and Zeta, which detail the composition of the diffractive apparatus used for each test.

Figure 4. *Diffraction of the Intra-Action for ….. Collaborative Test*



Note. Diffractive diagram based on Barad (2007), Gullion (2018), Talbot (2024), and Hayes et al. (2020). The diffractive diagram was taken from John D. Norton's *Diffraction Diagram*, published in '*The Quantum Theory of Waves and Particles*,' Pittsburgh 2001. © John D. Norton 2018.

Data was collected via audio recordings of students' perceptions during each test, followed by student-led discussions, and transcribed for analysis. Teachers' insights on collaborative testing were gathered through pre- and post-test conversations and transcribed. The techniques section of Chapter 4 provides detailed descriptions of the tests, teacher conversations, student-led conversations and audio recordings. Transcriptions were processed using Otter-ai software (Chia et al., 2020), an artificial intelligence speech-to-text program chosen for its efficiency and accuracy in managing extensive data from over fifty recordings. Iterative reflexive readings encourage new questions, the discovery of connections, and the formation of intricate interpretations (Srivastava & Hopwood, 2009), enabling a layered understanding of how collaborative testing is perceived by students, teachers and the researcher. The diffractive diagrams visually differentiate perceptions using coloured text boxes: dark grey is used for students, white corresponds to teacher perceptions and light grey is used for my insights. This visual differentiation distinguishes the nuanced insights gained, providing a more comprehensive understanding of collaborative testing.

Transformative Nature of Data in Diffractive Methodology

I explore data's transformative nature in this discussion through a diffractive methodology. This approach considers data as deeply intra-twined with the researcher in the process of knowledge creation, emphasising its active and agentive role within the constantly evolving landscape of educational practice (Koro-Ljungberg et al., 2018). By employing a diffractive methodology, I highlight the entanglement among myself, students, teachers and material elements, fundamentally shaping what emerges from the collaborative testing intra-action. Drawing on the insights of Barad (2007) and Taylor (2013), I challenge the conventional perception of data as inert, suggesting instead that a diffractive methodology animates data by commingling it with the apparatus[8], participants and other elements to understand what becomes significant (Barad, 2007). This perspective, supported by Hayes et al. (2020) and Kylkken (2023), describes diffractive methodologies as dynamic and emergent, contrasting with methodologies that unintentionally reduce complex phenomena into simple relationships, even going so far as to assume cause and effect.

Data are not static but part of a fluid state of knowledge-making (Taylor, 2021). Building on transformative theories that decentre the human and challenge binary distinctions like here/now and this/then, Koro-Ljungberg et al. (2018) suggest viewing data as fluid, merging in various directions yet remaining latent until brought forth to the present. We can also constrain data within time-space captured in settings like schools, notes, experiences, or recordings until liberated. Koro-Ljungberg et al. (2018) advocate that data are integral to dynamic knowledge, exist beyond binary distinctions, and remain dormant until made relevant. This perspective frees data from constraints, underscoring its complexity and active role in shaping real-world outcomes in collaborative testing.

Reflecting on these concepts, I recognise data's potential to transform, morphing into something novel and insightful. By actively engaging with data, I give voice to students' and teachers' perceptions of collaborative testing. This perspective is echoed by Tsing (2015) and Taylor (2019b), who contend that the researcher's role involves deep engagement with the data, fostering emergent knowledge through the 'art of noticing' (Tsing, 2015, p 17). Embracing this dynamic, inclusive and intra-connected view of data acknowledges the complex entanglement between myself and data, challenging established approaches to knowledge creation. Moreover, this practice encourages a

[8] The researchers' apparatus is not static but contextual and dynamic, intra-acting with and being shaped by the researcher, participants, and phenomena to make meaning of their practise under study (Barad, 2007).

deeper understanding of collaborative testing, marking a shift from viewing data as static and inert to recognising it as dynamic and intra-twined with the research process.

Diffraction Methodology Applied

This section serves as a bridge to Chapter 4, which will detail the research design, focusing on the application of a diffractive methodology. This methodology acknowledges and integrates material and environmental conditions as essential factors within educational processes. By doing so, I examine how various elements within the classroom setting, including students, teachers, researchers, and environmental conditions like classroom lighting, odours, and temperature, intra-act and influence one another. Understanding this dynamic entanglement is crucial for comprehending the complex realities of collaborative testing.

To explore these intricate intra-actions, during the initial phase of my PhD research, I sought to investigate how students' and teachers' interactions, experiences, and environments shaped their experiences with collaborative testing. Inspired by Dewey's philosophical ideas of a socially constructed, practical, active, and reflective process, I initially planned a mixed-methods study (Calvet-Minor, 2014; Morgan, 2013). The goal was to evaluate perceptions of this testing method, incorporating survey data alongside qualitative methods such as interviews, focus groups and mind maps. However, after a deeper examination of the literature and discussions with my supervisors, I determined that surveys could not adequately capture the nuanced perceptions of students' and teachers' during each test. These reflections, coupled with Karan Barad's (2007) insights on the entanglement of knowledge production in *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning,* influenced me to adopt an exclusively qualitative approach. This shift allowed for a more detailed exploration of the dynamic intra-actions between students, teachers, learning, and their classroom environments.

This methodological pivot also revealed a critical tension in my research—the challenge of working within the posthuman space of diffraction, while initially being inspired by the humanist process of reflection (Morgan, 2013). Reflection, as conceptualised by Dewey, assumes a bounded, intentional subject actively constructing meaning through introspection and dialogue, as noted by Ultanir (2012). In contrast, Barad's (2007) diffractive approach destabilises this framework, emphasising the entangled and relational nature of subjects, objects, and environments in producing knowledge (Blom, 2022). This tension revealed the limitations of reflection as a human-centered process, leading to its reconceptualisation in my research. By shifting towards diffraction, I began to

view reflection not as a discrete activity but as something co-produced and entangled within the phenomena it seeks to understand.

Adopting a multi-method research design facilitated a thorough exploration of teacher and student perceptions regarding collaborative testing. This approach emphasised the transformative and intra-twined nature of practice, context, and material elements in each class environment (Kemmis et al., 2014; Lenz Taguchi, 2010). Partnering with teachers was essential to exploring the relatively new practice of collaborative testing in high schools. Grounded in a cyclic, dynamic 'intra-vention' described by Barad (2007, p. 177), this exploration contextualised the research within the classroom to improve educational outcomes through a diffractive analysis of various qualitative methods (Bozalek & Zembylas, 2017; Mik-Meyer, 2020). By collecting data through multiple methods, this study reveals the diverse perceptions and differences that emerged when undertaking collaborative testing. This process of navigating the tensions between diffraction and reflection not only shaped the study's methodology but also invited a deeper engagement with how knowledge is co-produced in educational environments. These insights underscore the potential of diffractive analysis to challenge traditional concepts like reflection, emphasising entangled and relational ways of knowing that are particularly suited to understanding complex intra-actions inherent in collaborative testing.

Conclusion

In conclusion, diffractive ethnography, influenced by the rich concepts of entanglements, agential realism, phenomena, and apparatus, offers a transformative and robust methodological approach to exploring the complexities of collaborative testing environments. A diffractive approach goes beyond prevalent methods by actively engaging with the dynamic and co-constitutive entanglements between myself, students, teachers, and the material and environmental conditions within the educational setting. By focusing on how the elements intra-act rather than interact, diffractive ethnography unveils the emergent and ongoing production of knowledge and reality through an immersive experience, which challenges the separation of the observer and observes, placing the researcher in the intra-action. I consider agential realism to play a crucial role in this process, highlighting the dynamic nature of educational phenomena, which are not static but shape and are shaped by their intra-actions, offering a deeper understanding of students' and teachers' experiences. Furthermore, the concepts of phenomena and apparatus in diffractive ethnography help to see how specific educational practices and environments come to matter. I then discussed the importance of stating the agential cuts made by the apparatus to bring the phenomena into being at that spacetime. By integrating these concepts, diffractive ethnography illustrates the complexity of educational settings and seeks to understand the influences that shape the collaborative testing experience. Thus, diffractive ethnography contributes to the ongoing (re)configuration of what it means to teach and learn, acknowledging and embracing the intra-connectedness of all elements within the educational ecosystem.

Chapter 4: Research Design

This chapter outlines the study's research design, explaining the techniques[9] for understanding student and teacher perceptions of collaborative testing. The primary aim is to articulate the rationale behind the chosen research techniques and diffractive methodologies and justify the selection of case studies as the most suitable approach to this research. By integrating theoretical concepts with practical applications, this chapter provides a comprehensive framework for exploring the complex dynamics of collaborative testing in high school settings.

Approach to Research

Having established the theoretical underpinnings of agential realism and diffractive ethnography in the previous chapter, I now turn to how these concepts can be utilised to understand student and teacher perceptions of collaborative testing. My past experiences as a science teacher significantly shaped this study. Entangling[10] with the students and classroom environment revealed that repeated individual testing often induced anxiety, prompting an exploration of alternative assessment methods to supplement individual testing and enhance conceptual understanding. This exploration led to collaborative testing, a technique previously attempted mainly in undergraduate courses, which informed my research questions. Research questions guide the focus and processes used in this study (Hayes et al., 2020).

1. How do students and teachers perceive collaborative testing?
2. How does collaborative testing impact student learning?
    a. Which collaborative testing methods do students believe enhance their understanding?
    b. Why do they believe this?
3. How does implementing collaborative testing unfold in a real classroom?

A qualitative study was employed to understand these perceptions. Qualitative studies enable researchers to explore problems and develop a detailed understanding of concepts. Creswell and

[9] Techniques are flexible and iterative approaches to research that focus on the organic and evolving process of knowledge co-creation, as opposed to methods which often imply a fixed and predefined strategy (Fox & Aldred, 2018; Springgay, 2015; Taylor & Hughes, 2016).

[10] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, teachers, myself and the materiality of the classroom environment (Barad, 2007).

Guetterman (2019) note that qualitative studies are best suited when understanding the phenomena from the participants' perspectives, particularly when little prior research has informed the concept. Qualitative methods gather participant data, generally texts from audio recordings, observational notes, interviews, or videos, to develop a rich, complex picture (Creswell & Gutterman, 2019). In their discussion on qualitative methods in medicine, Crowe et al. (2011) argue that conducting qualitative research provides a comprehensive, nuanced understanding of real-life situations in a natural setting. Hence, this study uses qualitative methods to explore perceptions of collaborative testing. These methods support diffractive ethnography, which considers all elements to comprehend the phenomena fully. In choosing diffractive ethnography over more conventional qualitative methods, I am tapping into a richer, more nuanced way to explore the complex dynamics of collaborative testing. This methodology isn't just about observing; it is about engaging deeply with the material-discursive intra-actions that traditional methods overlook. By engaging in diffractive ethnography, I can better capture how all elements, human and material, contribute to the educational experience. This approach is particularly effective in understanding the layered realities of collaborative testing, revealing how students and teachers intra-act with each other and their environment in real-time, real-world educational settings.

Embracing Diffraction over Reflexivity in Qualitative Research

A diffractive methodology allows for a richer, more contextually grounded exploration by moving beyond the limitations of reflexivity, offering a framework that deeply engages with the entangled and co-constitutive nature of research environments (Barad, 2007). Reflexivity, as traditionally applied in qualitative studies, emphasises the researchers external reflection on their biases to ensure rigour (Bozalek & Zembylas, 2017). However, critics such as Haraway (1991) and Barad (2007) argue that reflexivity inadequately acknowledges the researchers' intrinsic impact on the phenomena under study, as it conceptualises the researcher as external to the observed entanglements. Gullion (2018) and Lumsden et al. (2019) highlight the limitations of reflexivity in dealing with the entangled nature of social research, where objectivity is not only unattainable but also an inappropriate goal. This limitation is particularly problematic in complex educational settings, where intra-actions between students, teachers, materials and environments shape the research outcomes. As Gullion (2018) asserts, 'Reflexivity is incompatible with social scientific research in the ontological turn because there is no external objectivity and because we cannot represent outside of our entanglements' (p. 90).

In contrast, Barad's (2007) diffractive methodology shifts the focus from external reflection to the intra-active process through which phenomena materialise. Diffraction, as a method, emphasises the relational and dynamic nature of knowledge production, allowing researchers to examine how specific outcomes emerge through entangled intra-actions (Plauberg, 2021). For example, while reflexivity might centre the researchers positionality, diffraction recognises that the researcher, participants, and material conditions are all co-constructive elements of the research process. This approach aligns with Bozalek and Zembaylas (2017) assertion that focusing on differences is critical for understanding diversity and complexity, as it acknowledges the multiplicity of perspectives, experiences, and contexts that shape educational phenomena.

In the context of collaborative testing, a diffractive method enables a detailed examination of how intra-actions among students, teachers, and material conditions (such as, lighting, seating arrangements, and classroom dynamics) influence the perceptions and outcomes of testing practices (Barla, 2023). Barad's (2007) terms these emergent effects "agential cuts,"(p. 337) which include both inclusions and exclusions that define the phenomenon under study. For example, by focusing on the differences and specificities of each classroom environment, diffraction can reveal nuanced insights into how collaborative testing unfolds uniquely within each setting (Barad & Gandorfer, 2021). Studies by Blom (2022) and Talbot (2023), exemplify how diffraction captures the entangled nature of educational environments. These studies highlight how factors like the physical environment and student entanglements during class intra-act with social and pedagogical dynamics to shape both teacher and student experiences. By adopting a diffractive methodology my study recognises these intra-actions as integral to understanding collaborative testing, providing a richer exploration of the emergent phenomena. Thus, diffraction not only grounds the research contextually but also offers deeper insights into the diversity of perceptions and experiences within school environments. It shifts the focus from merely cataloguing differences to understanding how these differences materialise, evolve, and influence the outcomes of collaborative testing (Barla, 2023). This approach produces more nuanced and impactful findings, challenging traditional notions of representation and objectivity, while embracing the entangled complexity of educational processes.

Methodological Approach

The diffractive approach is combined with a qualitative case study methodology to analyse the intricate phenomenon of collaborative testing within classrooms. In the context of this study, each class, encompassing the teacher, students, classroom environment, and material factors involved in collaborative testing, constitutes a case study. As Cresswell and Guetterman (2019) suggest, case

studies provide rich insights into complex phenomena within their natural contexts, making them particularly suitable for the intricacies of educational environments. Crowe et al. (2011) further argue that case studies permit an authentic exploration of educational settings without manipulating the environment, which is essential for capturing the true dynamics of collaborative testing. The methodological choices, deeply embedded in qualitative traditions and enriched by firsthand experiences, facilitated a contextually rich exploration of the three case studies, emphasising the importance of human-matter-nonmatter[11] to capture the diversity of educational experiences. As Gullion (2018) asserts in her ontological turn, a focus on all the elements of the intra-action does not decentre the human participants but instead 'acknowledges the creative complexities of all sorts of agencies that may or may not assemble with the human' (p. 20). In this study, teachers and students are bound in the spacetimematter (Barad, 2007) of the classroom and test, all intra-acting in what Taylor (2021) describes as 'the messy multiplicities of co-creating and co-constituting knowledges' (p. 30).

Criteria for Selecting and Excluding Classes

This section outlines the criteria for selecting and excluding classes, focusing on factors such as the students' year levels, science subjects, and the content covered in the tests. These selection criteria are deliberate choices that actively shape the scope of the study. By implementing agential cuts[12] and intra-actions, this study explores which elements are considered significant and which are excluded. The primary research questions aim to understand student and teacher perceptions of collaborative testing and its influence on student learning. The necessity for agential cuts involves deliberate decisions regarding inclusions and exclusions, emphasising the critical aspects related to the phenomena (Barla, 2023). Although cultural and socio-economic factors, technological integration, and teacher professional development are acknowledged as potentially influential on collaborative testing, they were excluded from this study to maintain a manageable scope.

In this study, purposeful sampling was employed to select the classes. Purposeful sampling is a strategic method in qualitative research that targets the most informative cases for in-depth analysis (Staller, 2021). Staller (2021) highlights that the strength of this approach lies in selecting rich cases

[11] Dashes symbolise the crucial intra-action between all elements, material or otherwise, highlighting the equal importance of both in shaping perspectives on collaborative testing (Gullion, 2018). Other elements contribute to this intra-action but are too numerous to detail individually.

[12] Agential cut refers to the idea that boundaries and properties of objects are not inherent or pre-determined but enacted through specific intra-actions involving phenomena, thus making these boundaries consequential and meaningful within the context of those intra-actions (Barad, 2007).

that provide detailed information to gain deep insights and enhance understanding of critical research issues. Similarly, Guetterman (2015) emphasises that the primary value of qualitative research is its ability to capture the richness of diverse perspectives within a specific setting. Abrams (2010) describes purposeful sampling as a process where researchers select the participants likely to offer the most valuable perspective on the studied phenomena. This targeted approach aligns with the researcher's deep exploratory objectives.

The recruitment process began after securing ethics approval from the University of Sydney and the NSW Department of Education through the State Education Research Application Process (SERAP). Co-educational schools were included to comprehensively address the research questions regarding collaborative testing from both students' and teachers' perspectives. Examining only single-sex schools would limit the diversity of perceptions and entanglements in mixed-gender groups (Miller et al., 2022). An agential cut was made to focus on co-educational public schools, which represent a significant segment of the NSW educational landscape. This decision was made to comprehensively address the research questions regarding collaborative testing from both students' and teachers' perspectives. According to the Centre for Education Statistics and Evaluation (2023), there are four hundred and twenty-four co-educational high schools compared to forty-five single-sex schools in New South Wales (NSW) as of 2022, broadening the potential participant pool. Eligible teachers must be employed and possess the qualifications to teach general science and their specific senior science subjects. This focus leverages my prior experience as a science teacher, aiding in test development and allowing for direct engagement and support during the testing phases. Moreover, as Harrison et al. (2017) and Wan (2019) described familiarity with the research site can further provide a more nuanced understanding of classroom dynamics and build trust with participants.

After defining the inclusion criteria for the schools (public, co-educational), the first step in selecting classes involved compiling a comprehensive list of schools in metropolitan Sydney. A shortlist of fifty high schools within a 10 Km radius of the city centre was created, considering factors like transportation time, co-educational status, and year 7–12 student composition. This selection included a mix of selective[13], selective and community, and only community, ensuring a diverse mix of student populations to provide deeper insights. Guetterman (2015) emphasises that a diverse mix of classes encompassing varying student abilities, socio-economic backgrounds, and cultural identities

[13] In New South Wales (NSW), Australia, the Department of Education and Communities administers an exam to identify high-ability students, who are then placed in advanced classes in certain schools. Meanwhile, 'community students'—those living within a school's local catchment area—are automatically eligible to attend without taking any entrance exams.

enhances the breadth of perspectives obtained from diffractive analysis. This heterogeneity ensures that the research captures a broad spectrum of experiences and viewpoints, leading to more nuanced insights into collaborative testing phenomena and more robust conclusions.

A study summary was attached to the first fifteen emails sent to principals. These schools were chosen for their varied environments (a mix of selected, mixed, and community schools) and proximity. However, the initial emails yielded no responses, prompting me to send emails to an additional thirty schools around metropolitan Sydney. In this outreach, I received two responses indicating interest, but both declined to participate due to increased demands on their teachers caused by the pandemic. Post-COVID-19 restrictions resulted in increased demands, stress, and exhaustion for teachers, as well as an increase in their workloads (Fray et al., 2023). The effects of the pandemic contributed to a limited response, leading to the postponement of outreach until the third term of 2022. The inclusion criteria were expanded to include co-educational independent (private) schools to enhance participation rates. Additionally, information about the research was posted on the Association of Independent Schools of NSW website and the social media site X (Twitter) to reach interested teachers. Although five school principals responded and facilitated contact with their science department heads, none expressed interest in participating.

Due to the impacts of COVID-19 on schools in NSW, which resulted in a lack of participant response, I had to rethink my approach to attracting classes to the study. Therefore, I spoke to my university colleagues and supervisors, asking them to reach out to their contacts, which led to the participation of a school from metropolitan Sydney. Polaris[14] High School, a co-educational public school with both selective and community students, provided a diverse range of participants of varying abilities, which was crucial for gaining a broad understanding of student and teacher perceptions of collaborative testing. After presenting the research objectives to the science department during an in-person meeting, three additional teachers requested information and consent forms. To ensure positive engagement, I organised a follow-up meeting at Polaris to discuss the practical aspects of implementing collaborative testing. During this meeting, two additional teachers joined the study. To address any concerns about coercion, I held an informal conversation in the science staffroom, without the head teacher present, to confirm that the teachers' participation was entirely voluntary. Simultaneously, I expanded recruitment efforts to Ceres[15] Independent School in Southern Queensland, tapping into a colleague's connection. As a non-selective boarding school in a rural

[14] A pseudonym, for the school. Details will be outlined in the Table 1.
[15] A pseudonym, for the school.

setting catering to students from kindergarten to year 12, Ceres offered a contrasting demographic that enriched the study's diversity. An initial meeting at Ceres led to the participation of another teacher.

The strategic selection of eight diverse classes enhances the depth and applicability of this study. Table 1 introduces the two schools and the teachers, with pseudonyms used to protect their identities. The table also describes whether the students were selective, community or mixed ability. These eight teachers represent the classes where I gathered and analysed data; ultimately, three classes were included in this study, with the reasons for including them explained later in this section.

Table 1. *School Pseudonym, Teacher Pseudonym, Selective, Community or Mixed Students, and Subject Taught.*

School (pseudonym)	Teachers (pseudonym)	Number and year level of students. Selective, Community or Mixed Students	Subject taught
Polaris	Halo	Year 11 – 6 Females – 1 Selective, 5 – Community Year 11 – 6 – Males – Community Year 12 – 8 Females – All selective Year 12 – 4 Males – All selective	Investigating Science
Polaris	Beta	Year 11 – 4 Females – 3 Selective, 1 – Community Year 11 – 14 Males – 12 Selective – 2 Community	Physics
Polaris	Altair	Year 11 – 3 Females – 2 Selective, 1 – Community Year 11 – 15 Males – 12 Selective, 3 – Community	Physics
Polaris	Castor	Year 11 – 10 Females – 5 Selective, 5 – Community Year 11 – 8 Males – All Selective	Chemistry
Polaris	Ajax	Year 11 – 9 Females – 6 Selective, 3 – Community Year 11 – 13 Males – 10 Selective, 3 – Community	Chemistry
Polaris	Gamma	Year 11 – 3 Females – All Community Year 11 – 14 Males – 7 Selective, 7 – Community	Physics
Ceres	Zeta	Year 11 – 9 Females – 7 Males – Mixed Ability	Agricultural Science
Ceres	Luna	Year 10 – 13 Females – 11 Males – Mixed Ability	General Science

Note. Polaris High School is in metropolitan Sydney, and Ceres Independent School is in rural Southern Queensland.

This group of eight classes provides a broad spectrum of instruction across various science subjects. Such diversity is crucial for examining the differences and similarities across these distinct disciplines. The inclusion of classes from both the same and different subjects enriches the dataset through the unique material-discursive elements present in each classroom. Priya (2021) supports this approach, highlighting that having multiple participants from varied backgrounds ensures a comprehensive analysis and captures a broad spectrum of experiences. This diversity enables a deeper, diffractive

analysis of student and teacher perceptions, leading to more robust and nuanced findings from collaborative testing.

Furthermore, this study focused on the final two years of high school, specifically targeting students in years 11 and 12, a critical period characterised by the unique pressures of preparing for highly competitive external tests such as the HSC and the Queensland Certificate of Education (QCE). The selection of these grades is justified by the significant impact that repeated testing has on students' educational experiences, as documented in the literature review (Chapter 2). However, due to the scheduling of the final HSC[16] and QCE examinations, the timing of the study was adjusted to the first term of 2023. I utilised the video conferencing platform Zoom to coordinate the next steps with all eight participating teachers. This technology bridged geographical distances and provided the flexibility needed to accommodate the diverse schedules of the teachers involved, as Gray et al. (2020) noted. This adaption was essential in maintaining consistent communication and ensuring all participants were well prepared for the study, emphasising the role of technology in facilitating modern educational research.

I had to narrow the number of cases from the initial eight to three to ensure a thorough and focused analysis of my case studies. This decision was guided by a diffractive analytical approach, which, according to Koro-Ljungberg et al. (2018), prompts me to consider the practical implications of my data ('of what data do,' p. 462). The overwhelming volume of data necessitated a pragmatic approach, recommended by Holford (2020), who advocates for assessing the usefulness of data in contributing to knowledge, saying, 'how is this any use to us' (p. 122). Consequently, after all the data was gathered and diffracted, I selected three of the eight classes for detailed analysis in this thesis.

I included Halo's class because it was the only one undertaking a collaborative summative test featuring a mix of year 11 and 12 students. The year 12 students were undertaking their HSC, making this collaborative test high-stakes for those students. This class provides insight into how students and Halo perceived collaborative testing in both graded and ungraded contexts. Additionally, this inclusion allows for exploring different group dynamics and student behaviours when undertaking graded versus ungraded tests.

[16] The Higher School Certificate (HSC) in New South Wales (NSW) and Queensland Certificate of Education (QCE) in Queensland, Australia, are qualifications awarded to students who successfully complete senior high school level studies (Years 11 and 12), assessing them on a combination of coursework and final examinations.

Next, I selected Beta's class due to the significant gender imbalance within the class, with only four female students and fourteen males. This selection aims to explore potential differences in student perceptions influenced by these gender disparities. During the study, Beta expressed that there was an excessive amount of collaboration in classrooms; thus, understanding their and their students' views on collaborative testing would add another layer to the insights gained.

The third class I chose was Zeta's class because of its distinct demographics and rural location, offering a contrast to metropolitan Polaris High School. This class was also included due to timing constraints brought about by fixed assessment schedules, which impacted students cognitively and, consequently, their perceptions of collaborative testing. These three classes provide distinct differences for cross-case analysis and a broad perspective on collaborative testing.

The content of the tests was developed directly from the subject curriculum and aligned with each teacher's respective positions in their teaching sequence. Consequently, any content not covered by the teacher at this point was excluded from the study. Table 3 provides information on the specific content for each test, reflecting the tailored approach to each subject matter, ranging from basic principles to complex applications in science.

Table 2. *Collaborative Test Content for each Collaborative Test*

Class	Test 1 subject matter	Test 2 subject matter	Test 3 subject matter
Halo	Cause and Effect, Describing variables, Impact of technology, Isotope decay	Scientific modelling, Technologies, Impacts related to natural disasters	The impact of technology on science, Genetics and cloning
Beta	Kinematics – Motion in a straight line	Kinematics – Acceleration in a straight line, Vectors in one dimension	Kinematics – Motion on a plane, Vectors in two dimensions
Zeta	Agricultural enterprise, Animal and Plant Biology	Agricultural enterprise, Genetics and breeding	Plant production

Note. This table includes a breakdown of the subject matter covered in each class test.

Diverse Educational Contexts: Polaris High School and Ceres Independent High School

The learner and the world are inseparably intra-twined; they constitute the learning environment through the intra-actions among students, teachers, researcher, and collaborative tests (Murris, 2020). These intra-actions encompass both the school and classroom environment and the land on which the classes are situated, playing a crucial role in student engagement in learning.
The discussion now separately examines two distinct educational settings: Polaris High School in Sydney and Ceres Independent High School in Southern Queensland. I explore their demographics, which are pivotal in shaping students' and teachers' perceptions of and engagement with collaborative testing.

Polaris High School, a public, co-educational school governed by the NSW Department of Education, serves students in grades 7 – 12. According to data from ACARA (2023), the school features a diverse student body, with 2% identifying as Indigenous and 85% from culturally and linguistically diverse backgrounds. The gender distribution is nearly balanced, with 51% males and 49% females. The school offers two academic streams: a selective stream for students admitted through the NSW Department of Education and Training examination and a comprehensive stream for students from the local catchment area, referred to as the community stream.

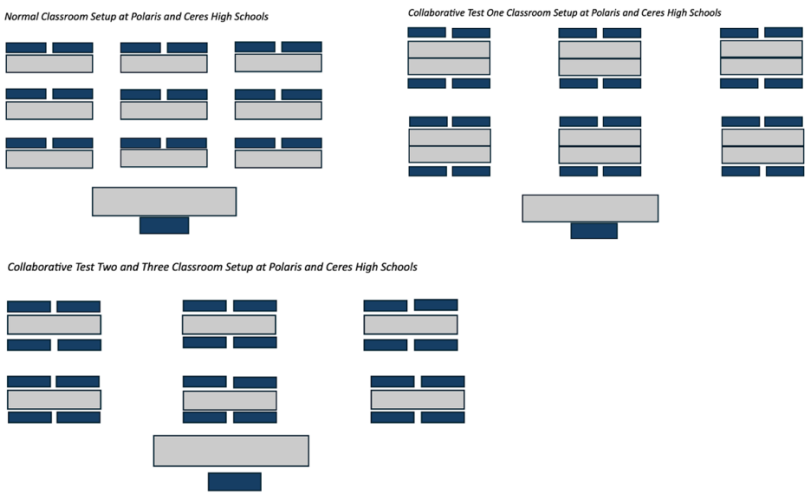
In contrast, Ceres Independent School is a co-educational Catholic institution that serves students from kindergarten to year 12, imposing compulsory yearly fees. As of 2022, this school exhibits less cultural diversity, with approximately 8% of the student body identifying as Indigenous and another 6% coming from culturally diverse backgrounds. This school features a similar gender distribution to Polaris (ACARA, 2022).

Despite similar gender distributions, the schools exhibit significant differences in cultural composition and educational systems. Polaris High School has significant cultural diversity, contrasting with the more homogenous cultural makeup of Cere's Independent School. These variations are substantial because they could influence the dynamics and effectiveness of collaborative testing. Polaris High School provides a multicultural public education through its selective and community-based streams. At the same time, Ceres Independent School operates within the Independent Catholic system and features a less culturally diverse student body. These contrasting demographic and educational frameworks are critical for understanding their potentially different impacts on the effectiveness of collaborative testing.

Engaging with the Material-Nonmaterial Elements

In this diffractive ethnography study, matter and the environment is prominently recognised, playing a crucial role in shaping student and teacher perceptions of collaborative testing. For instance, an audio recorder was centrally placed on the desks during tests to capture student group discourse. This recorder becomes entangled with the testing environment as students turn it on and off and adjust it throughout the testing period. In the Techniques section of this chapter, the choice of using an audio recorder will be explored in greater detail. The classroom setup also plays a significant role in the dynamics of collaborative testing. All classes across the two schools featured a similar arrangement of desks in rows with spaces facing the front, as depicted in Figure 5. According to Morris et al. (2023), this arrangement tends to support more content-focused teaching and less emphasis on collaborative skills due to the physical separation of students. However, recognising the need for a more conducive environment for collaboration, adjustments were made to the desk arrangements during the tests. For the first collaborative test, desks were rearranged by placing two tables together with chairs on either side, allowing students to face each other, as shown in Figure 5. While facilitating face-to-face intra-action, this setup proved suboptimal for audio clarity as the recorder failed to capture the voices of the quieter students. Consequently, the arrangement was modified for subsequent tests by grouping four chairs around a table, enhancing auditory capture and student intra-actions. These modifications and their impacts on student engagement and entanglement are critical to assessing the effectiveness of collaborative testing environments.

Figure 5. *Classroom Setup for the Three Collaborative Tests*



I now discuss the following data collection techniques designed to capture the rich, qualitative data necessary for analysing students' and teachers' intra-actions during collaborative testing.

Evolving Research Methodologies: Integrating Fixed Methods with Dynamic Techniques

As I embarked on my doctoral journey, my foundational beliefs about the nature of knowledge, its development, and the structured research approaches were influenced by my scientific background and experience as a high school science teacher. Initially, I viewed data collection as a linear process akin to following a set recipe. This perspective was challenged by Taylor and Hughes (2016, p. 17), who posed the question, 'How can one access, know about, and represent the experience of another reality?' This question initially led me to explore humanistic methods, which conflicted with my belief in the researcher's intra-connectedness with the data, participants, and environmental conditions. Fox and Aldred (2018) advocate unlearning these methods in favour of embracing a practice that breaks down and reconstructs our processes. This unlearning meant abandoning established ways of doing things and embracing a 'methodology-to-come' (Lather, 2013 p. 635), where my development mirrors the evolution of the research itself. I became an emergent researcher by opening myself to discovery and continually redefining my experiences and understanding. According to Snaza (2013, p. 49), this is akin to engaging in 'bewildering education,' where the path unfolds unpredictably.

Given the dynamic nature of this journey, the term methods seems inadequate, as it implies a fixed and predefined research strategy, presupposing an understanding of the outcome. In adopting a diffractive methodology, I acknowledge that we cannot separate ourselves from the instruments used to construct understanding (Thiel, 2020). Pickering (2017) describes humans and instruments as 'intertwined, positioning and emerging themselves as they work together, islands of stability in the flux of becoming' (p. 140). This perspective led me to reconsider traditional terminology. Springgay (2015) proposes using 'techniques' (p.77), a term that better captures the organic and iterative co-creation of knowledge. Taylor and Hughes (2016, p. 19) further clarify this by stating, 'Techniques activate modalities of thought, rhythms, affects from inside the act, techniques activate a practice from within.' As I navigate these techniques, the outcomes remain uncertain but promise a richer, more intra-connected understanding of the phenomena under study. To capture the nuanced perspectives of students and teachers on collaborative testing, I integrated a variety of techniques, including conversations, audio recordings, student-led conversations, mind maps, observations and journal notes. These diverse techniques are woven together to explore the research questions and explain the practice of collaborative testing.

Collaborative Tests

Collaborative tests employ various methods to evaluate student learning effectively. Research by Green et al. (2018), Meseke et al. (2010), and Vogler and Robinson (2016) demonstrates the effectiveness of multiple-choice questions in collaborative testing. These studies highlight that well-designed multiple-choice options enhance in-depth student discussions, similar enough to provoke discussion but distinct enough to require careful consideration. In contrast, Newton et al. (2019) and Jang et al. (2017) incorporated both multiple-choice and short-answer questions to probe deeper understanding. Siegel et al. (2015) examined the advantages of open-ended questions in the collaborative section, supplemented by a mix of multiple-choice and open-ended questions in the individual test. Cooke et al. (2019) and Eastridge and Benson (2020) focused on short-answer items, with the latter study also incorporating matching and statistical problem-based questions to challenge students. In this study, I adopted a mixed approach that includes short and long-response questions along with calculation-based items, a decision informed by the work of Mahoney and Harris-Reeves (2019) and Reiger and Reiger (2020). These researchers advocate for the use of Bloom's higher-order thinking verbs, such as 'analyse,' 'evaluate,' and 'create', which they found to enhance peer discussions and foster deeper understanding in collaborative testing environments.

I meticulously tailored the test questions to align with the specific content requirements of each science subject. Initially, teachers provided the content they wanted covered for each test. I then wrote the tests and sent them back for review. The teachers ensured the questions matched the intended depth and difficulty appropriate for their student's abilities. After making any necessary adjustments based on their feedback, I finalised and printed them. Examples of test questions are in Appendix D. The initial tests for investigating science primarily included short-response questions. Subsequent tests incorporated a blend of short-response and two extended questions using analytical verbs like 'evaluate', 'assess' and 'predict,' as suggested by the literature on collaborative testing (Mahoney & Harris-Reeves, 2019). In Physics, the progression involves moving from short-response items with minimal calculations to more complex problem-solving scenarios, reflecting the mathematical nature of the subject. In Agricultural Science, each test consisted of both short and at least one long-response question, ensuring a comprehensive evaluation of student understanding. Following teacher approval, the tests were finalised and printed. This collaborative review process illustrates the tailored and rigorous approach to test design, ensuring the assessments are both challenging and pedagogically sound.

Conversations

This study adopts the term *conversation* instead of *interview* or *focus group* to capture the diversity of intra-active dialogues integral to this research. The use of the term *conversation* in qualitative studies has a long history. For example, Webb and Webb (1932) in the United Kingdom and Palmer (1928) in the United States incorporated conversations as a central method in their social inquires. In early ethnography research, scholars such as Margaret Mead (1928) and researchers associated with the Chicago School employed conversations as a means of data production (Swain & King, 2022). More recently, Swain and King (2022) have advocated for the use of conversations in qualitative research, emphasising that "both informal and formal conversations are a co-construction between researcher and participant" (p. 8). Similarly, Johansson (2016) employed conversations in their post-qualitative study, *Post-qualitative line of flight and the confabulative conversation: a methodological ethnography*, highlighting the evolving methodological significance of this approach.

Both interviews and focus group discussions typically involve a question-and-answer exchange between the interviewer and interviewees, where participants and researchers engage in discussions centred around open-ended questions to discuss perspectives and experiences related to the topic (Nathan et al., 2019; Roulsten and Choi, 2018). Denzin (2001) and, more recently, Roulsten and Choi (2018) characterise interviewing and focus group methods as human-centric. This approach is reflected in transcripts that emphasise a discursive and cultural understanding using the 'apparatus of the interview machine' (Denzin, 2001, p. 30) while possibly influencing the data gathered through participant and researcher interactions, including dialogue, empathy and nonverbal cues (Brinkmann, 2011; Johansson, 2016). However, the demeanour of interviewers or focus group moderators may also lead participants to respond in accordance with the researcher's expectations rather than expressing their true beliefs. Moreover, focus group participants may influence one another, fostering conformity to the prevailing opinion, a phenomenon referred to as the 'halo effect' by Nyumba et al. (2018, p. 30). Despite these limitations, interviews and focus groups have proven invaluable in numerous qualitative studies, where insights and detailed data can be systematically gathered and analysed (Busetto et al., 2020; Thelwall& Nevill, 2021).

In contrast, Barad (2007) emphasises that discourse is not solely about what is said but also encompasses the constraints and possibilities of what can be said. This perspective underpins diffractive ethnography practice, which embraces human storytelling through intra-actions among participants, responsive listening, and the way in which they move in specific contexts (Kuntz &

Presnall, 2012). Such conversations extend beyond verbal exchanges to encompass material-discursive practices, facilitate the emergence of relational knowledge and construct meaning in collaborative testing (Taylor & Hughes, 2016). By engaging teachers, students and myself in these enriched dialogues, we co-create understanding, moving beyond simple question-and-answer dynamics to explore our collective entanglement in this innovative assessment method.

The initial conversation with teachers and myself occurred in early December 2022, as the school year concluded. Due to the pandemic, the NSW Department of Education restricted outsiders from entering public schools, affecting my face-to-face conversations with the teachers. Consequently, the first conversation with teachers took place over Zoom. The NSW teachers were located in their science building, while one teacher from Queensland participated from the school, and the other joined remotely from home. The second conversation occurred shortly after the conclusion of the third test in term 1, April 2023. I conducted this conversation in person with the six NSW teachers and over Zoom with the two teachers in Queensland as their school term had ended. These conversations were audio-recorded and transcribed. Tables 3 and 4 describe the environment for the two conversations with teachers. During these conversations, open-ended questions were posed to elicit their thoughts on collaboration and collaborative testing. In the second conversation, teachers were also asked to share their insights into the impact of collaborative testing on their students and their practice. The conversations had no set time limit and concluded when the teachers had fully articulated their thoughts.

Table 3. *The Setting for Pre-Collaborative Testing Conversations with Teachers*.

Teacher	Pre-test conversation situated
Halo	The conversation unfolded in Halo's office, a space struggling against the heat. Despite the air conditioner and fans whirring in full force, they barely made a dent in the room's stubborn warmth. Outside, the day was hot and dry, contributing to the office's dark ambience, an attempt to ward off the sun. During the Zoom call, Halo chose to blur their background, focusing the conversation solely on the intra-action. Positioned behind their desk, Halo engaged with keen eye contact, their laughter punctuating the discussion, and they occasionally swung leisurely in their chair while delving into points.
Beta	The dialogue occurred in the head teacher's office, which doubled as an enclosed storeroom bathed in the harsh glow of fluorescent lights. The space had a fan and a single window that offered a view over the science block's hallway. Despite the active air conditioner and the fan's loud hum, the room maintained a sense of confined academic rigour. Beta was seated at the desk, animatedly gesticulating and nodding to emphasise their points. Technical difficulties plagued this Zoom call, necessitating the video to be turned off and adding a layer of challenge to our communication.
Zeta	This was the only face-to-face discussion in 2022, set against the unique backdrop of an empty classroom adjacent to the gymnasium at Ceres High School. The rendezvous unfolded during the lunch hour amidst the cacophony of student chatter permeating the space. Inside, the desks stood in orderly rows; we repositioned two, creating a more conducive setup for our conversation. My nerves were palpable; I stumbled over my words and struggled to keep pace with Zeta's points, and my speech hastened before I prematurely turned off the recorder. Then, a teacher entered, engaging Zeta in a brief exchange. Paradoxically, this interruption served to ease our nerves, allowing us to adopt a more relaxed demeanour as the conversation progressed.

Table 4. *The Setting for Post-Collaborative Testing Conversations with Teachers.*

Teacher	Post-test conversation situated
Halo	Unplanned and impromptu, our conversation unfolded outdoors in the school's quadrangle, a necessity born of the absence of available classrooms. A gentle drizzle hung in the air, and our conversation occurred under a large green awning, where droplets found their way through, sprinkling us with water. The wind was erratic, swirling around us and scattering my papers in a playful yet inconvenient dance. Throughout our discussion on the nuances of collaborative testing, Halo was animated—gesturing for emphasis, leaning forward with interest, and occasionally tilting their head back in contemplation, embodying the thoughtful process of pondering the questions posed.
Beta	This exchange occurred in a prep room within the science block, directly adjacent to the classroom, where the students were absorbed in their workbooks. During our conversation, Beta demonstrated engagement with the questions through animated hand gestures, nods and gesticulating. Seated forward, they embodied attentiveness and consideration, pausing to reflect on each question thoughtfully before articulating their perceptions on collaborative testing.
Zeta	This dialogue unfolded virtually over Zoom, a concession to the onset of school holidays. The setting was Zeta's home office, where they had opted to blur the background, focusing attention solely on the exchange at hand. Engaged and responsive, Zeta navigated the conversation with hand gestures that complemented their verbal responses. They nodded in agreement and occasionally cast their gaze upwards, a reflective pause, illustrating a thoughtful process behind each articulated perception.

In addition to the two teachers' conversations, students participated in conversations immediately after each test. Following the first test, students discussed and answered two questions about what they liked and disliked about collaborative testing, with each group recording their responses. After the second test, students answered three questions regarding what they appreciated and did not appreciate about collaborative testing and whether they believed it impacted their learning. Upon completing the second test, I observed that the students were less engaged with the discussion questions. Six questions were presented to the students to foster more thoughtful introspection and discussion after the third test, as displayed in Figure 6. Before recording their individual opinions, the students deliberated on them. These more structured questions prompted the students to consider how collaborative testing influenced their learning and collaboration skills. I then transcribed the recordings of the students' conservations.

Figure 6. *Student-Led Discussion Questions*

1. What did you like about working collaboratively?
2. What did you not like about working collaboratively?
3. Do you think your collaborative skills have improved? Why or why not?
4. How do you think your group worked together?
5. What would you change if we did the collaborative test again?
6. What would you think about collaborative testing if these tests counted towards your grades? Why?

Note. After the third collaborative test, students received the questions in advance.

Audio Recording

Recording discourse is a vital aspect of ethnographic research and is considered integral by many researchers (Rutakumwa, 2020). Audio recordings encapsulate individuals' viewpoints, thoughts, emotions and reactions at a specific moment in timespace, making them active participants in the knowledge-creation process. The audio recorder interacts with students, researchers, teachers and nonmatter in this 'unstable stage' (Cobussen, 2022, p. 15). Its materiality does not restrict its role in the entanglement; it is active, can be manipulated, turned on and off, and wields power as voices are or are not recorded. It can amplify silence or noise and narrate the story the participants wish to convey. Various methods for recording participants exist, including phones, small hand-held audio recorders, video recordings, and platforms like Zoom and FaceTime. In this study, Olympus hand-held digital voice recorders were employed. These compact mobile devices were placed at the centre of each collaborative desk during the three tests for each subject. Students gave verbal consent before each test and provided written permission at the study's outset. The recorders were positioned between the teacher and myself for in-person teacher conversations. The Zoom platform was used for remote conversations, with teachers granting consent before conversations commenced. I used Zoom's transcription feature to transcribe audio files of teachers' conversations and student test group recordings, integrating my observations during transcription.

Mind Maps

Wheeldon and Ahlberg (2019) argue that an individual's account of their experiences and perceptions privileges their constructed realities (p. 1114). In line with this perspective, my goal was to capture individual students' perceptions of collaborative testing. Initially, I considered using a survey with a five-point Likert scale and some open-ended questions to probe why students responded in a particular way. However, I soon realised that a survey would not allow me to gather perceptions that were not influenced by the questions posed. Instead, I chose mind maps as a more spontaneous and personal technique that enables students to freely express their perceptions based on their experiences (Wheeldon & Ahlberg, 2019).

Mind maps are creative graphical techniques that facilitate the expression of feelings and perceptions. Unlike cognitive maps, which outline individuals' ideas, beliefs, values, attitudes and relationships, or concept maps, which are structured and hierarchical, mind maps encourage introspection around a central concept, in this case, collaborative testing (Burgess-Allen & Owen-Smith, 2010; Fearnley, 2022; Rowell et al., 2021). This approach allowed students' unique perspectives to emerge, providing a subjective exploration of their experiences and collaborative testing. Given the

subjectivity of mind maps, it is important to contextualise them to fully understand the perceptions of the students who created them. This context is discussed in detail in the participation section. Based on Wheeldon and Ahlberg (2019), the mind map is displayed in Figure 7. The central concept is collaborative testing, with four probing questions emerging from this central point to gauge students' perceptions. Each consenting student in each subject received the mind map upon completing the third test. After collection, I subjected the mind maps to diffraction, with each mind map's diffraction analysis detailed in the chapters on Beta, Halo and Zeta.

Figure 7. *Individual Student Mind Map Questions.*

Note. The individual mind maps were handed to students at the completion of the third collaborative test. The mind map was based on Wheeldon and Ahlberg (2019).

Observation and Journal Notes

Standard qualitative studies rely on field notes, a crucial method in rigorous qualitative research. Ethnography, a qualitative research approach, uses field notes to capture researchers' experiences by documenting observations and conversations in specific environments (Phillippi & Lauderdale, 2017). Field notes as a data-gathering method emerged in the last century, originating from anthropological ethnographic studies before extending into sociology and other fields (Phillippi & Lauderdale, 2017). They serve as a sense-making tool, categorising descriptions of the environment, people, or culture before being compiled into a narrative that contextualises a particular place and time (Hayes et al., 2020; Phillippi & Lauderdale, 2017). Thompson and Burkholder (2020) attest that writing traditional field notes should be reconsidered as they carry positionality and power, which could affect the participants' agency. In contrast, diffractive methodologies depict all intra-action

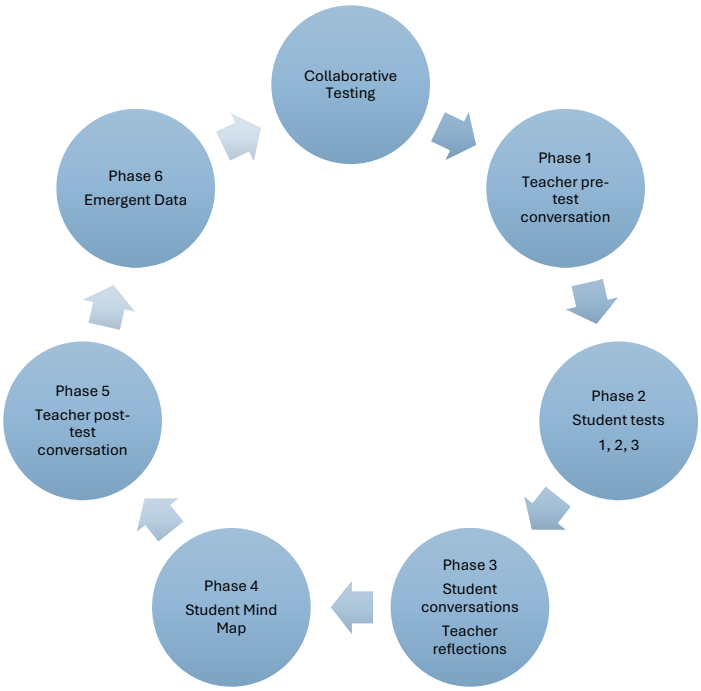
components. As Rust and Altman (2020) state, 'Entanglements insist on tracing the ways in which the observer and the observed are intertwined with various technologies, materials, emotions, histories, and timespaces evoked in any given moment' (p.221).

I am part of the encounter, engaged in an ongoing entanglement to understand the meaning of the phenomena, collaborative testing. Unlike fieldwork, where the 'field' is seen as separate and the researcher observes and then leaves (Hayes et al., 2020), entanglements are continuous. I propose the term 'observation notes' to describe all the interactions and components, addressing the static separateness implied by the term field. These notes encompass all details and derive meaning through an iterative equalising method. I selected this method to document not only the intra-actions between students and teachers but also, crucially, the materiality of the classroom. In practice, I messily wrote my observation notes in a small book while wandering the classroom and observing all the components of the testing, writing how students moved, talked, intra-acted, whether the classroom was hot or cold, fans on or off, open or closed windows, and the odours. Alongside other techniques, observation notes facilitated a complex understanding of the phenomena.

Research Phases

As I was equipped with diverse data collection techniques, my study is structured into six distinct phases, each designed to explore different aspects of the collaborative testing process. Figure 8 illustrates the phases, which will be discussed in detail below.

Figure 8. *Illustration of the Phases and Techniques for Collaborative Testing*



1. Pre-test Conversation: Teachers and I discussed their perceptions of collaboration, collaborative work in their classroom, and the test structure and content over Zoom.
2. Implementation: To observe and document how collaborative testing is conducted in real-time settings.
3. Feedback Collection: Gathering immediate reactions and participant feedback after tests to gauge initial perceptions.
4. Individual Student Perceptions: Students were given a mind map for individual comments.
5. Analysis Using Diffractive Methodology: Applying a diffractive methodology to analyse the data, aiming to understand how different perceptions and experiences intra-twine influence the efficacy of collaborative testing.
6. Synthesis and Evaluation: Evaluating the overall findings to address the research questions and assess the implementation and impact of collaborative testing on learning outcomes.

Each phase is guided by specific research questions that seek to explore the perceptions of students and teachers regarding collaborative testing, which are outlined in Table 5. The use of diffractive methodology throughout the study aims to illustrate the ways in which various data collection methods can clarify and enhance our understanding of these perceptions, as well as the overall effectiveness of collaborative testing.

Table 5. *Phases, Methodological Approaches, Techniques, and Emergent Data for Collaborative Testing*.

Phase	Focus	Research question	Methodological approach	Techniques	Data documentation	Emergent Data
Teacher Pre-test	Teacher perceptions, implementation and views on the effect on learning	How do teachers perceive and implement the intervention, and what are their views on its effect on learning?	Diffractive ethnography	Conversations	Audio-recording Researcher notes	Agential realism, Collaboration, Material-discursive practices
Collaborative Tests 1, 2 and 3	Student perceptions and views on the impact on their learning	How do students perceive collaborative testing? Which collaborative testing methods do students believe enhanced their understanding?	Diffractive ethnography	Student-led conversations Tests Researcher notes Teacher reflections	Audio recording Tests Researcher notes	Agential realism, Collaboration, Material-discursive practices
Student Mind map	Student perceptions and views on the impact on their learning	How do students perceive collaborative testing? Which collaborative testing methods do students believe enhanced their understanding?	Diffractive ethnography	Mind map	Mind map Researcher notes	Agential realism, Collaboration, Material-discursive practices
Teacher Post-test	Teacher perceptions, implementation and views on effect on learning	How do teachers perceive and implement the intervention, and what are their views on its effect on learning?	Diffractive ethnography	Conversations Researcher notes	Audio recordings Researcher notes	Agential realism, Collaboration, Material-discursive practices
Emergent Data		What emergent data and unexpected themes arise from the study?	Diffractive ethnography	Apparatus Agential cuts		Agential realism, Collaboration, Material-discursive practices

Phases of Data Collection

Phase One (Early December 2022)

Teacher pre-test conversations occurred at the end of term four over the Zoom platform for all but one Queensland teacher, which occurred in person. These conversations focused on teachers' perceptions of collaboration, collaborative classroom activities, individual testing, and their understanding of collaborative testing. After this phase, we discussed the content and structure of the tests. The conversations were generally twenty minutes long and audio-recorded before being transcribed. During and immediately after the discussions, I took notes on material elements, nonverbal cues, and body language, which I added during the transcription phase.

Phase Two (Late February – April 2023)

The tests started three weeks into the new school year. Three tests were administered every two weeks and were completed by early April. Students were placed in groups for some tests, and some chose their groups, generally composed of three or four students. Their conversations were audio-recorded to capture their discussions. I also intra-twined with the classroom, noting or drawing nonverbal cues and the material environment. After the tests, students discussed their perceptions of collaborative testing. Recordings were transcribed and diffracted through the apparatus[17] to bring forth commonalities and differences.

Phase Three (Immediately after Test Three)

Students were given mind maps just before the end of test three, which they completed after the test to illustrate their perceptions of collaborative testing. I collected the mind maps and undertook a diffractive analysis, enabling differences to emerge.

Phase Four (Ending of Testing Period)

Teachers from NSW and I were entangled at Polaris High School, in the classroom, outside on benches, and in prep rooms, while the Queensland teachers' conversations were over the Zoom platform due to the commencement of school holidays. These conversations centred on teachers' perceptions of collaboration, the differences between individual and collaborative testing, and the impact of collaborative testing on their students. Discussions included considerations of matter and nonmatter intra-actions. I transcribed the teachers' conversations and conducted an analysis to identify similarities and differences.

[17] Apparatuses actively reshape the world through specific practises, intra-actions, and performances that establish distinct boundaries of inclusion and exclusion (Barad, 2007).

Discussion of Data and Methodology

Diffractive methods for data are inherently agential and dynamic. They acknowledge that data does not exist in isolation but materialises through complex intra-actions (Murris, 2021). According to Koro-Ljunberg et al. (2018), data in a diffractive methodology is mysterious, unknown, tentative, and uncertain until it manifests. This goes against the idea that data is fixed or predetermined. As Taylor (2013) states, a diffractive approach to data makes it 'lively' (p. 692) as it actively participates in collaborative testing; data materialises and gains meaning through the entanglement of human and material agents. Barad (2007) emphasises that new meanings are 'brought into being through the co-constitutive enactments of human/nonhuman opportunities' (p. 701). This conceptualisation suggests that understanding data requires a dynamic and intra-active methodology, as discussed below.

The research began with an iterative review of various data forms, including pre- and post-test teacher conversations, recordings and transcriptions from each test and class session, and mind map responses. This stage was crucial for initial insights and laid the groundwork for deeper engagement. The second stage delved into how the components of the research apparatus, school, student body, teachers' perspectives of collaboration and assessment, and classroom environments intra-twine with the data. This involved a detailed analysis of how these components influence teacher and student perceptions of the testing process. I analysed their perceptions against material-discursive practises and the influence on collaboration and emergent understandings in this diffractive process, illustrating the concept of agential realism. The final stage involves critically re-assessing the data in light of the relevant assessment literature and theoretical frameworks. This stage aims to answer the research questions and provide insights into the perceptions of collaborative testing among students and teachers.

This methodology underscores the need for a flexible, responsive approach to data analysis that adapts to new insights and understandings that emerge from intra-active processes (Hayes et al., 2020; Taylor, 2013). Recognising that data emerges through these entanglements; I seek to elucidate the complex, dynamic nature of educational assessments and the multiple factors that shape their effectiveness and the perceptions of students and teacher.

Research Design, Strength and Limitations

Adopting a case study method is essential for examining the complexities of collaborative testing in educational settings. This approach, endorsed by Cresswell and Guetterman (2019) and Crowe et al. (2011), provides detailed insights and an authentic exploration of classroom dynamics. By immersing myself in each classroom environment, I aimed to capture collaborative testing dynamics comprehensively. As Barad (2007, p. 88) suggests, the researcher 'understands the world from within and as part of it.' According to Tight (2022), the strength of case studies over other qualitative methods is their ability to holistically explore multiple factors, in contrast to other qualitative methods that may focus on a limited range of variables. According to Priya (2021), each case study presents a unique, ever-changing ecosystem, providing 'thick' descriptions that reveal the complexities, benefits and challenges of implementing collaborative testing. The advantage of using case studies is their capacity to 'bind' the sites, creating determinate boundaries through agential cuts (Murris & Bozalek, 2019). The vignettes describe each site in detail, including timespace, environmental conditions, tests and participants.

Case study methods have limitations. Their most significant drawback is the lack of generalisability compared to other methodologies (Crowe et al., 2011). However, Hyett et al. (2014) note that case study research inherently focuses not on comparisons but on delivering insights into particular phenomena. Moreover, as discussed by Barad (2007, p. 181), an agential realism perspective acknowledges the 'uniqueness of each entanglement in spacetimematter disrupting generalisations' (Barad, 2007). Each classroom environment is unique and dependent on the students, teacher, day, time, and activity, and each entanglement changes slightly at the point of intra-action. This study does not aim to universally represent the efficacy of collaborative testing but rather to explore it within the specific contexts of each class. It is the 'production' (Bozalek & Zembylas 2017, p. 117) of differences that emerge through the diffractive reading that is insightful. The perceptions of students and teachers towards collaborative testing are unique to their classroom and environment.

Three case studies and three different tests generated extensive data providing a detailed description of the three classrooms. This ensures 'quality' (Priya, 2021, p. 102) conclusions, as defined by Lincoln and Guba (1985) in qualitative research, a term they prefer over reliability or validity. The quality of the conclusions in this study is strengthened by the varied techniques used to gather students' and teachers' perceptions. The entanglement of three tests in Halo, Beta, and Zeta's classes and myself led to more authentic intra-actions and perceptions as students and teachers became

familiar with the technique. Through the transcripts, observations, and my journal notes, a rich, inclusive understanding of collaborative testing emerged. As Thomas (2011, p. 33) states, to seek '*generalisable knowledge*, in whatever form—everyday or special—is to miss the point about what may be offered by certain kinds of inquiry, which is *exemplary knowledge*' (emphasis in original).

The case study method used in this research on collaborative testing offers comprehensive insights into classroom dynamics and student and teacher intra-actions by allowing an in-depth exploration and 'thick' descriptions of each unique educational setting. The primary limitation of the case study method is its lack of generalisability because each site is bound to its specific context; the findings cannot be broadly generalised across different educational contexts (Merriam & Tisdell, 2015). The agential realism framework acknowledges this limitation, highlighting the uniqueness of each situational intra-action in spacetimematter (Barad, 2007). The case studies followed similar implementations to ensure the 'quality' of the findings. I grouped the students, gave one test to each group, set a timer, and the students undertook the test. After testing, the students and teachers expressed their thoughts on the collaborative test. The only deviation from this procedure occurred in Halo's class during the third test, where an individual test preceded the collaborative portion. During data analysis, I constructed each test apparatus to accommodate emerging elements. The transcriptions, along with my journal and observation notes, provided clear descriptions of the apparatus and enacted agential cuts, including descriptions of spacetimematter (Barad, 2007).

Ethical Considerations

Prior to conducting this study, ethics approval was received from the University of Sydney Human Research Ethics Committee (Appendix A) and the NSW Department of Education State Education Research Application Process (SERAP) (Appendix B). During the research, I made amendments to include independent schools and placed a recruiting advertisement on social media, X (Twitter) and the Association of Independent Schools of NSW website. Additionally, a short information video was generated to be attached to the social media posts and websites; these amendments were approved.

I provided participant information forms (Appendix C) to ensure participants' willingness to participate. Participants were informed about the study's purpose, procedures, and benefits. Students and teachers were given opportunities to withdraw from the study at any time, as verbal consent was requested before each test. An essential part of maintaining ethical research is ensuring the protection of participants by safeguarding their privacy and confidentiality (Priya, 2021). Morse and Coulehan

(2015) note that qualitative researchers often give alternate names to participants. This study changed the teachers' names to gender-neutral names (Halo, Beta and Zeta). Additionally, the schools' names and specific locations were not included. The particular length of teaching and gender were omitted as they might identify the teachers and were not crucial to the study. In order to protect their future identification, teachers were referred to by pseudonyms or gender-neutral pronouns (they, their, them). Within a group, the students were identified as female or male or F1 or M2, and no group-identifying features were disclosed.

To maintain regulatory requirements, the researcher adhered to the university's data management protocols to protect research materials and data during the study. The hard copies of the mind maps are stored in a secured filing cabinet in the School of Education & Social Work, University of Sydney. Audio recordings and student information are held in the university's One Drive (Office 365) in a password-protected folder only accessible to the researchers. The State Records Authority of NSW requires that I keep the data in a secured location for a minimum of 20 years or until subjects are 25 years old.

Chapter Structure and Methodological Approach

The main focus of this study is to understand student and teacher perceptions of the innovative assessment method of collaborative testing through a diffractive methodology that incorporates all human, matter and nonmatter in the intra-actions. To capture these perceptions effectively, I will focus on their expressed thoughts and actions while also considering the impact of the material aspects. This requires a comprehensive narrative method that encompasses all relevant elements. Skilling and Stylianides (2020) define vignettes as 'written, visual, or oral stimuli designed to reflect realistic settings and provoke responses such as perceptions, emotions, and decision-making, in alignment with research paradigms and methodologies' (p. 542–543). Vignettes offer a rich portrayal of the complexities of collaborative testing (Simon & Tierney, 2011). I crafted the chapters detailing each class, Chapter 5, Halo, Chapter 6, Beta, and Chapter 7, Zeta, to situate the reader in the classroom environment, enabling the reader to 'see' and 'feel' the intra-actions among students, teachers, and the classroom environment. I adopted a specific formatting approach to clearly highlight the students' and teachers' comments within the data. I bolded the students' voices and italicised the teachers' voices to make distinguishing between the two groups easier. This choice was made to ensure that the insights from each participant group were emphasised and easily identifiable. Additionally, I bolded the teachers' quotes when both students and teachers expressed the same or similar comments. This was done to draw attention to the commonalities between the perspectives of students and teachers,

further reinforcing the shared experience and insights across the groups. Notably, the quotes have not been altered for grammar or spelling, as I aimed to preserve the authenticity of the participants' voices. This method ensures their direct, unfiltered expressions remain intact, providing genuine illustrations of their thoughts and contributions.

These narratives are inherently subjective, as I actively participate in the intra-actions. Restricting the narrative to an objective perspective separates the researcher, the participants, and the materiality of the classroom. Mitchell and Clark (2018) point out that, while writing is messy, exploring alternate methods to encapsulate the voices can capture the depth of these intra-actions. Hultman and Lenz Taguchi (2010) further articulate this, suggesting the goal is to communicate the complex intra-relationships and mutual transformations within these encounters rather "than trying to reveal and incarnate a specific phenomenon or quality of 'being-in-the-world' (p. 537). The chapters are organised into four distinct parts.

Part 1 introduces the collaborative testing environments through vignettes that highlight the intra-action among the teachers, students, and classroom environment. These vignettes, accompanied by diffractive diagrams, visualise the classroom's complexity. The three case studies illustrate the nuances of collaborative testing by directly quoting participants, allowing them to articulate their perspectives authentically. To acknowledge their perspectives, I bolded student quotations and italicised teacher responses. Additionally, they meticulously describe the materiality of the classrooms, demonstrating how these factors intra-twine with the testing process. My journal entries are inserted in the first person within coloured text boxes, enriching the narrative with personal comments. This part culminates with reflections from the teachers on the nature of collaboration.

Part 2 transitions to examining the logistical setup of the tests and how these arrangements influence the perceptions and experiences of teachers and students. This section connects the theoretical framework to practical experiences, illustrating how each element contributes to the effectiveness of collaborative testing.

Part 3 delves into the influence of the material environment on collaborative testing. By analysing these environmental factors, I underline how physical space and material conditions affect student engagement and perceptions of collaborative tests.

The final part, **Part 4**, synthesises observations from the collaborative tests, discussing broader implementations of educational methods, particularly focusing on the impact of accountability measures and teacher practises. This part connects specific classroom experiences to broader educational theories, providing insight into how these dynamics influence and are influenced by broader pedagogical and institutional contexts. Each part is consistently structured across the three sites, facilitating an integrated analysis that enhances understanding of the complex intra-actions between human and the material-nonmaterial continuum in shaping educational outcomes.

Conclusion

This chapter outlines the research design I crafted to explore teachers' and students' perceptions of collaborative testing. I aim to capture the intricate intra-actions in educational environments through detailed case studies by combining qualitative methodologies with diffractive ethnography and agential realism. This approach aimed to describe the techniques used to examine the intra-actions. I have also sought to illustrate an innovative approach that presents a more intra-connected and dynamic view of educational phenomena, thereby enriching the subsequent analysis. The forthcoming chapters will present the data gathered for each case study teacher, illustrating the practical application of these methodologies to understand collaborative testing's impact in high school settings.

Chapter 5: Halo's Entanglement

This chapter explores the complex dynamics of collaborative testing in Halo's classroom, structured into four comprehensive parts. Each part builds upon the previous, weaving a coherent narrative that explores the various elements influencing collaborative testing environments.

Part 1 begins with a vignette that sets the scene for the collaborative tests and the mind map. Each vignette leads with diffractive diagrams, illustrating the dynamic intra-action[18] between Halo,[19] students, the classroom, and the environment. The apparatus components and agential cuts concerning what has been included and excluded were detailed in the techniques section of the Research Design Chapter. *Italic text* is used to accentuate the teachers' comments, while **bold texts** are used for students,' and the quotes have not been amended in any way for grammar. Regarding each collaborative test and the mind map, the complex entanglements[20] between Halo, students, and the classroom environment are detailed. This analysis not only considers the perspectives of Halo and the students but also acknowledges the tangible influence of the classroom's environment on student experiences. The diffractive patterns emerging from these intra-actions are illustrated through excerpts from my reflexive journal. These excerpts are presented in shaded boxes using the present tense, weaving these documented experiences into the narrative of the ongoing collaborative environment. To conclude Part One, I describe Halo's pre and post-test conversations regarding collaboration and collaborative testing. Setting the stage for the deeper analysis in subsequent sections.

Part 2 examines Halo and students' perceptions of collaborative testing's effects on student understanding, metacognition, and 21st-century skills.

Part 3 explores Halo's classroom environment and the material elements, particularly focusing on how desk setup and crowding influence students' perceptions of collaborative testing.

Part 4 discusses unique observations from Halo's classroom, analysing the impact of high-stakes testing, accountability, and teacher practice, which affect both students' and Halo's perceptions of collaborative testing. By investigating these elements, I aim to understand how different factors

[18] Intra-actions refer to the intra-connectedness between students, teacher, myself, and the material classroom environment (Barad, 2007).
[19] Halo is a pseudonym
[20] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, teachers, myself and the environment in the classroom (Barad, 2007).

contribute to the efficacy of collaborative testing in high school settings. The chapter concludes with a summary of the analyses provided.

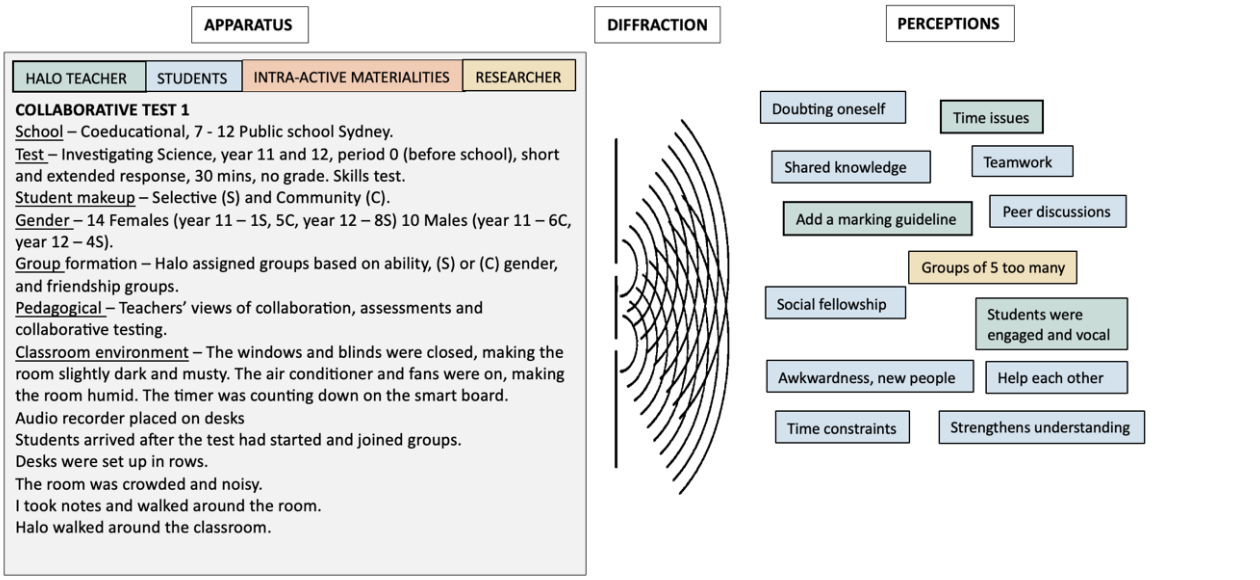
Part 1: Setting the Stage with Collaborative Tests and Mind Map

Collaborative Test One

In this exploration of collaborative test one, I aim to examine the complex, intricate entanglements among all elements that shape perceptions of collaborative testing (Barad, 2007). This method emphasises the significance of human participants and the material environment in shaping the perceptions of both students and Halo towards collaborative testing. To underscore the equal significance of these factors, I use dashes in my descriptions (e.g., students-Halo-environment), symbolising that no single element outweighs another in contributing to the test's dynamics. The discussion will include environmental factors such as noise, light levels, temperature, desk setup, and the degree of crowding. By undertaking a diffractive approach, a more nuanced understanding of collaborative testing can be achieved.

The diffraction[21] of collaborative test one, Figure 9, is structured into three components.

Figure 9. *Diffraction of the Entanglement for Halo, Collaborative Test One.*



Note. This figure illustrates the analysis of collaborative test one unfolds in three stages. It begins with examining the test environment, which includes physical and discursive elements such as classroom conditions and teacher opinions. It then moves to a dynamic analysis of varied perceptions from student and teacher reflections. It culminates in presenting these insights through a colour-coded scheme to highlight differences between students' and teachers' viewpoints based on the results of the diffraction analysis.

[21] A diffractive analysis considers the intra-action between students, teacher, phenomena, and the researcher. It is iterative and evolving, revealing the emergence of phenomena and differences to understand the possibilities of collaborative testing (Barad, 2007).

Apparatus description: The first section describes the apparatus, which consists of the material-discursive elements that can impact perceptions of collaborative testing. Initially, this section details the school, subject (Investigating Science), year level (11 and 12), period (0), test structure (short and long-response) and timing of the test (30 minutes). It then presents the class composition, detailing the mix of selective[22] and community students and their gender distribution. Following this, the apparatus outlines factors that could influence students' and Halo's perceptions of collaborative testing. These factors encompass the teachers' opinions on collaboration and assessments and the physical classroom environment, including noise and light levels, temperature, air quality, and the arrangement of the desks.

Diffractive analysis: The second component is the diffraction, which focuses on analysing the varied perceptions of students, Halo, and the researcher regarding collaborative testing by examining diffractive patterns. This dynamic analysis unfolds as I examine the transcripts from six student groups alongside Halo's reflections to discern similarities and differences in their perspectives, aiming to reveal the factors that influence their perceptions of collaborative testing.

Emergent perceptions: The final component elaborates on the perceptions that emerge from the diffractive analysis. The collaborative test entanglement highlights the colour-coded perceptions within the study: blue represents student perceptions, green corresponds to teacher perceptions, and orange is used for the researchers' insights.

The first collaborative test occurred in mid-February during Sydney, Australia's summer. I arrived at Polaris High School before 8 am on a warm and humid day. Around me, numerous eucalyptus trees housed squawking magpies and cockatoos, filling the early morning with their sounds. Walking to the science block, I wondered how collaborative testing could (re)configure science testing practices and enhance student understanding. I stepped into the science block and immediately noticed a dimly lit hallway adorned with old wooden boxes and Bloom's taxonomy verbs lining the walls.

The noticeable difference between the bright outdoors and the dimly lit hallways led me to consider how this contrast affects students moving between these environments. Morales-Bravo and

[22] In New South Wales (NSW), Australia, the Department of Education and Communities administers an exam to identify high-ability students, who are then placed in advanced classes in certain schools. Meanwhile, "community students" — those living within a school's local catchment area — are automatically eligible to attend without taking any entrance exams.

Navarrete-Hernandez's (2022) research indicates that natural light[23] can positively influence mood, energy levels, and emotions. Thus, when students move from the bright outdoors to dimly lit hallways and classrooms, their emotional state may shift, leading to decreased energy, potentially affecting their engagement and participation in classroom activities and collaborative testing (Morales-Bravo & Navarrete-Hernandez, 2022). As I walked down the hallway, I noticed that the low ceilings further accentuated the aged character of the science block. Passing by the classrooms, [24]

there are rows of old wooden desks seating two students with blue plastic chairs facing the front of the classroom and the teachers' desk facing the students

I reflected on traditional classroom layouts and instructional methods centred on the teacher at the front, overseeing both behaviour and learning. Despite significant shifts in society, technology, and education methods, these traditional classroom arrangements remain prevalent. Recently, there has been an increased focus on results and outcomes, coinciding with the emergence of managerial policies that emphasise outcomes, competition, and individualism. This context, as analysed by Smith and Benavot (2019), might explain why standard classroom arrangements persist. Entering Halo's classroom, Figure 10, illuminated by bright overhead florescent lights, presented a materiality strikingly different from the hallway's dimness and the corridor's antiquity.

overhead fluorescent lights shone bright white, subduing the little natural light flowing into the classroom

Figure 10. *Photo of Halo's Classroom at Polaris High School, taken at the End of the School Day.*

[23] Light was chosen as an essential aspect affecting learning environments and students' emotional and cognitive behaviour, which may impact their perceptions of the effectiveness of collaborative testing (Manca et al., 2022).

[24] The shaded box contains excerpts from my reflective journal, presented in the present tense to emphasise my role in the intra-connectedness of collaborative testing. These excerpts enhance the description of the collaborative test by illustrating my personal engagement and observations during the test.

Upon entering the classroom, I encountered dim natural lighting due to the drawn blinds, sharply contrasting with the bright artificial overhead lighting. As I entered the classroom, I noticed the classroom furniture was noticeably different from the other classrooms, with tall modern benches with light grey plastic legs and dark grey tops, green high stools, and laboratory benches extending from the room's perimeter. Despite the modern furnishings, the benches were arranged in rows, reflecting the conventional layout prevalent in Polaris High School's science classrooms.

Walking into the classroom, I greeted Halo at their desk. Halo sat at their desk facing the front of the classroom, orchestrating the first collaborative testing setup on the interactive whiteboard. The collaborative test group dynamics were organised around six student groups, considering gender, year level, academic aptitude and social affiliations and displayed on the whiteboard. As Halo prepared the student groups, I rearranged the classroom by pairing desks and adding stools, creating a configuration designed to encourage student collaboration.

as I move the desks, I think where will they all go? is there enough space?

However, this setup presented challenges; the proximity of the desks hindered effective collaboration due to noise from adjacent groups, making it difficult for some students to hear each other. I opted to utilise the laboratory benches around the classroom's perimeter to address this issue. However, this also posed challenges as students could not pull their chairs close enough, remaining distanced from each other. However, I separated the desks as best I could to give students as much distance between each group as possible.

To capture the dynamics of student intra-actions, I placed a hand-held digital audio recorder at the centre of each group. These recorders were marked with a distinctly coloured paper star and were intended to document the levels of reasoning and how students intra-acted with their peers, the test and the group. The recorders were active. They can be adjusted, turned on and off, demonstrating their influence by amplifying silence or noise and narrating the comments and reflections the students wish to share. Aware of the presence of the recorder, some students were self-conscious, which may have influenced their natural responses. Each group table was also labelled with a Post-it note detailing the student names and the star's colour to identify which recorder was placed on what table and to maintain organisation.

As students entered the classroom, they became active participants in the test environment intra-action.[25] This involved their dynamic engagement with various elements: the classroom setup, the arrangement of the benches and stools, and their peers. The presence of the teacher, Halo, and myself played significant roles in guiding this process. This moment underscored the adaptable nature of the classroom, where human and material components intra-twine in the evolving collaborative testing entanglement.

I observed

> the classroom buzzing with a sense of curiosity and anticipation

The students found their places at the benches marked with their names amidst soft murmurings of '**what's going on**'[26] and '**who's she**.' Halo stood facing the students, establishing a sense of control and watching. After a brief introduction of myself, I explained the collaborative testing procedure. Students would work in groups, collaboratively combining their knowledge to answer each question. After ensuring no lingering questions, Halo started a thirty-minute timer on the whiteboard, signalling the start of the test. The arrival of latecomers necessitated a swift reconfiguration of seating arrangements, which can alter the groups' dynamics.

From the front of the classroom, Halo monitored the proceedings carefully; their attention focused on the students as they displayed curiosity and apprehension while embarking on their first collaborative test, so different from their typical individual tests. '**Everyone's a little hesitant**,' one student observed out loud, voicing the group's collective nerves. I moved among the groups, offering guidance on how to approach the collaborative test. 'Discuss the questions together', I explained, and then 'agree on an answer to write down; it's the teamwork that counts here.' While most students appeared nervous, one said, '**this is going to be fun**.'

I navigated through the room entangling with the student-test, observing and listening to student responses as they began the test.

> students leaned in, making eye contact, moving their chairs, gesticulating with their hands playing with the test and animatedly discussing the questions

The shift in the room was palatable as they began to warm to the idea, turning a tentative start into an opportunity to collaborate with their peers. Halo noted the overall positive engagement of the

[25] Intra-actions refer to the intra-connectedness between students, Halo, the researcher, and all the components of the classroom environment (Barad, 2007).

[26] Bold texts illustrate student comments and bring student voices to the forefront above the sea of text.

students, saying the students '*appear to be having fun*.'[27] As the test progressed, I noticed students moving their heads forwards to listen, tapping their shoes as they answered questions, smiling, encouraging each other.

However, amidst the focused discussions, the impact of the late arrivals became evident, altering the dynamics within one group and leading to varied levels of participation. Two students appeared

> disengaged, their attention wandering, yawning, not talking and then talking loudly over off-topic subjects

The issue of students not contributing to the group can arise from various factors. According to Kanevsky et al. (2022), these include personality differences, gender dynamics and weak social ties within the group. Additionally, an underappreciation of collaborative work and a lack of individual accountability contributed to the students' social loafing. In this phenomenon, some group members exert less effort, relying on others to complete the task, as noted by Boren and Morales (2018). This issue was particularly evident in larger groups, where students felt less connected and motivated, as observed in the five-member group. The lack of individual accountability further reduced incentives for students to participate actively, leading to increased distrust, disagreements, and disengagement (Kanevsky et al., 2022). Halo, who monitored the class from the front, noticed the two disengaged students, prompting them to consider further adjustments in group formations to accommodate varying personalities better and optimise group sizes to enhance participation and reduce social loafing.

As the test progressed, I saw the gradual emergence of quieter students as their initial reticence gave way to active participation in the collaborative test entanglement.

> quieter students are active, rearranging their chair closer to the group, more vocal in the discussion, leaning in, gesticulating

It is pertinent to highlight that certain reserved or quieter students, whose participation may not be overtly evident in my observation notes, were actively engaged in the discussions, as revealed upon reviewing the transcripts. Previously reserved students became more engaged in group discussions, underscoring the inclusive nature of collaborative testing revealing the potential within each student to contribute meaningfully when given the right environment.

Some students expressed concern over the difficulty of the questions, highlighting a gap in their learning or the challenge of revisiting content after a lengthy period. As a group of students described,

[27] Italics illustrate Halo's comments and bring their voice to the forefront above the sea of text.

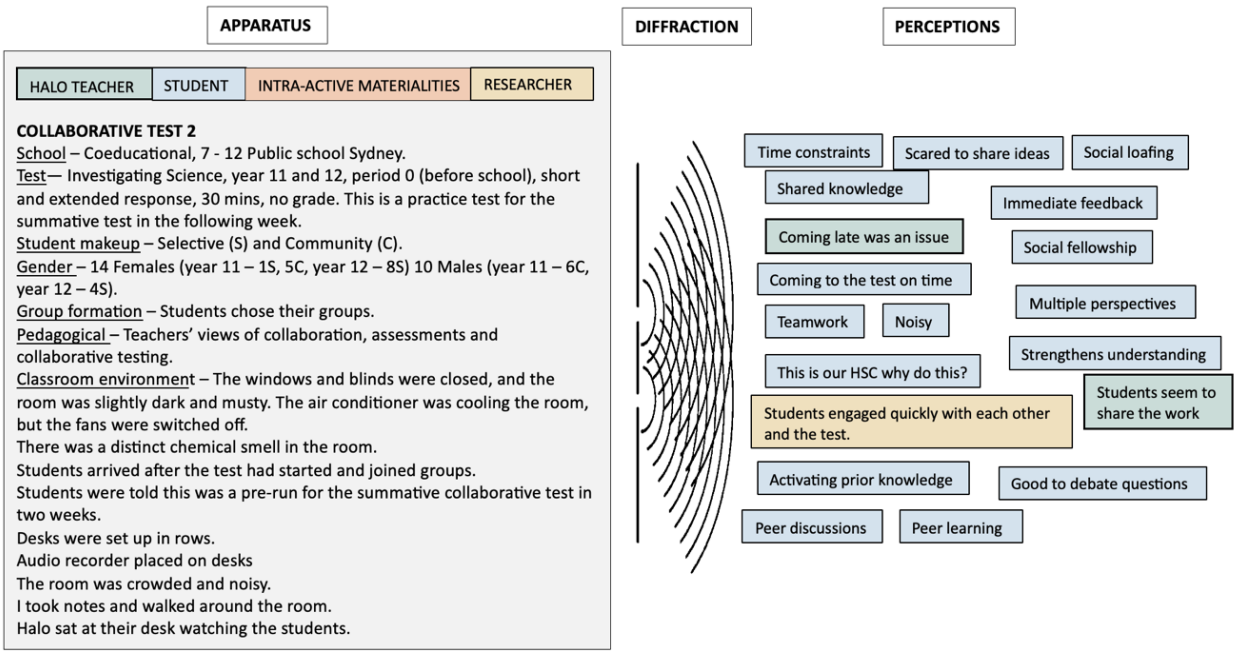
'**Oh, we haven't learned this,**' '**we haven't learned anything on this**,' '**remember modules five and six**,' '**it's like been two years**,' and '**the questions are really hard**.' This mix of engagement and challenge illustrates the collaborative testing entanglement in action.

As I approached Halo's desk, I overheard students expressing concerns about the insufficient time allocated to the test, with comments like, '**there's too little time**.' Halo acknowledged these challenges, stating there were '*too many questions, and they would reduce the number of questions for the second test to give them more time to collaborate*.' Time to complete the test had a significant impact on students' perceptions of collaborative testing. Providing enough time for constructive discussions promotes critical thinking skills, group cohesiveness and lessens anxiety, which could impact student reflections on collaborative testing. This adjustment highlights how test timing affects the perception and efficacy of collaborative testing. Research by Martin (2018) supports this observation, noting that time constraints can exert cognitive and emotional pressures, impairing students' ability to complete the tests efficiently. The timed nature of the test exacerbated anxiety, particularly among those unable to finish, as they felt rushed through discussions. This is problematic, as restrictive time allocations can undermine the very purpose of collaborative testing, which Vogler and Robinson (2016) and Meaders and Vega (2022) argue should encourage consensus-building discussions. One student's feedback, '**it takes much longer because you have to discuss everything, and then we get rushed**,' captures the tension between the necessity for thorough discussion and time constraints. Additionally, the lack of adequate time hampered students' ability to engage in key activities such as listening, defending viewpoints, reasoning, and expanding on their responses. The resulting anxiety and hasty decision-making underscored the need to align time allocations with the goals of collaborative testing, ensuring sufficient discussion time to support critical thinking and meaningful participation.

As the test approached its end, I wrote three reflective questions on the board, encouraging the students to share their thoughts on what they liked, disliked, and would change about the collaborative test. This was their opportunity to express their opinions on the first collaborative test. When the timer signalled the conclusion, the students recorded their answers, marking the end of the collaborative test. I collected the papers and recorders as they streamed out of the classroom, leaving me to ponder their views on this novel testing approach.

Collaborative Test Two

Figure 11. *Diffraction of the Entanglement for Halo, Collaborative Test Two.*



The second collaborative test occurred in early March, three weeks after the initial test. I arrived at Polaris High School before 8 am on a cool and sunny morning. Walking towards Halo's classroom, I wondered whether this practice test would change how students would entangle with the test and their peers. This test is structured just like next week's high-stakes test. Halo suggested that having the students participate in a collaborative, non-graded test would increase their comfort levels. Navigating through the slightly dim hallway to Halo's classroom, I took note of the few students scattered about, engaged in casual conversations. I entered the classroom, becoming part of the collaborative test intra-action, noticing

a distinct chemical and musty scent

a lingering reminder of the room's frequent use for chemistry classes. I observed that the chemical odour was more noticeable in the morning when the classroom's windows and doors remained closed after school hours, trapping the smells inside. As I wandered into the classroom, I noticed the new fans were switched off, but the air conditioner was on.

Halo was seated at their desk facing the front of the classroom, typing the test directions on the smartboard. Before this test, Halo and I discussed a revised format aimed to accommodate 30 minutes better: four questions, blending short and extended responses. The physical setup of the classroom was (re)arranged to facilitate collaboration, opting for single benches accompanied by

flanked stools instead of the paired desks used in the first test, a change driven by the need to capture quieter student contributions better. I placed the recorders with their stars in the middle of the benches. Students entered the room (re)configuring the entanglement[28]as they became part of the second collaborative test. Halo then told the students they *'could form their own groups of three or four students*.'

Noisily, students began to form their groups and moved to the benches, writing down their names on the test and the colour of the recorder. The composition of the groups chosen by the students was intriguing, with groups forming and blending genders and year levels, each bringing a unique dynamic to the intra-action. From their position at the front of the classroom, Halo explained to the students, '*There are no points for this test, just a practice for their summative assessment next week*', and the '*questions would be similar to what they would find in their HSC*.' With the commencement of the timer, the test began. Students' lateness was again an issue as two female students were ten and twenty minutes late, respectively, and had to join a pair of female students. Halo acknowledged this concern, noting, '*students coming late was an issue*.'

As I wandered among the groups, one notable difference from the first collaborative test was the observation that students swiftly adapted to the second test compared to the initial test. I observed them

> starting the test straight away, they leaned in, made eye contact, moved in their chairs, touched their faces, listened, and were enthusiastically saying their responses.

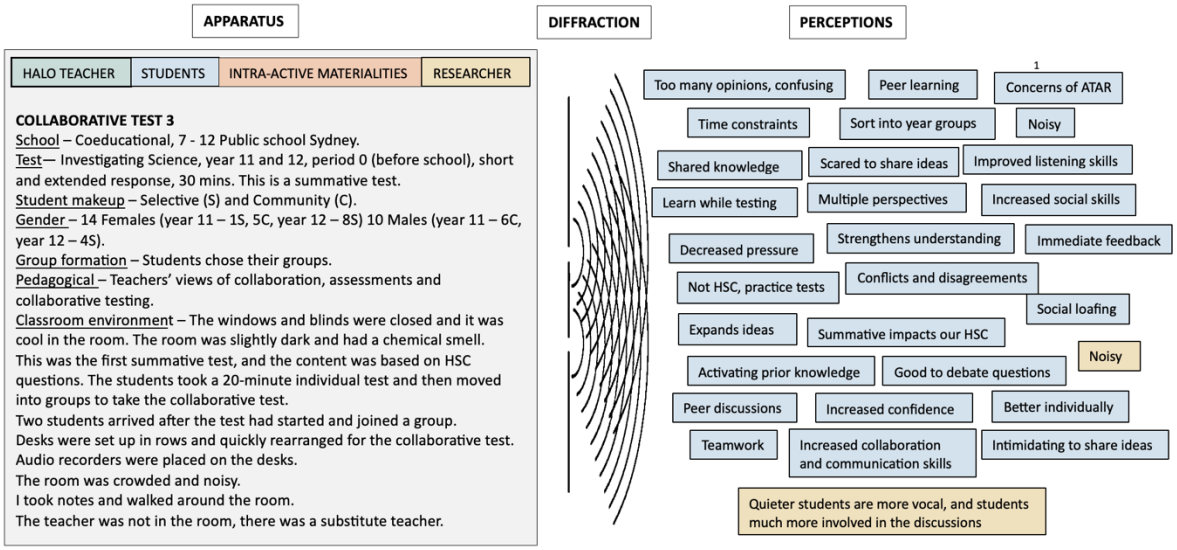
As students engaged in animated discussions about the test questions, the classroom noise levels inevitably rose. This increase in volume, a sign of active engagement, poses challenges in maintaining a conducive learning environment. One student remarked, 'I**t got kinda loud with all the groups collaborating**,' underscoring the difficulty of balancing group engagement with manageable noise levels. The active exchange of ideas led to louder discussions, which created significant obstacles for some students. According to Manca et al. (2020), excessive noise can hinder students' ability to listen attentively, concentrate on the task, and fully engage in the collaborative process. Additionally, the situation was likely exacerbated by the overcrowded classroom, which tended to amplify noise and can lead to irritation among students, detracting from the effectiveness of collaborative testing. These environmental factors, such as noise and overcrowding, play crucial roles in shaping students' experiences of collaborative testing.

[28] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, Halo, myself and environment of the classroom (Barad, 2007).

Halo and I circulated the classroom during the test, observing the students. After the timer sounded, marking the end of the test, Halo collected the completed tests. The students then began to reflect on their experiences, focusing primarily on the newly added third question about the test's impact on their learning. As the students departed, I gathered the recorders, thanked Halo, and exited the classroom.

Collaborative Test Three

Figure 12. *Diffraction of the Entanglement for Halo, Collaborative Test Three.*



[1] The Australian Tertiary Assessment Rank (ATAR) is used to gain entrance into university by comparing students undertaking multiple subjects across Australia. ATAR scores are not a measure of a student's individual performance (Shergold et al., 2020).

The third collaborative test unfolded on a cool morning, a week following the preliminary practice test. This high-stakes test was uniquely significant, as it was the only summative test in this study, featuring an individual test accounting for 80% of the total grade and the remaining 20% stemming from the collaborative portion of the test. As I navigated through the slightly dim hallway towards Halo's classroom, and entered the classroom with its customary rowed layout, Halo informed me of a change of plans due to the illness of a science teacher, which necessitated their absence from this class. As Halo explained the procedure, I noticed a strong chemical odour, indicative of the classroom's multi-use.

Students began assembling outside and, as they entered, became part of the collaborative test intra-action. I greeted them and handed out the tests, advising them to find a balance between

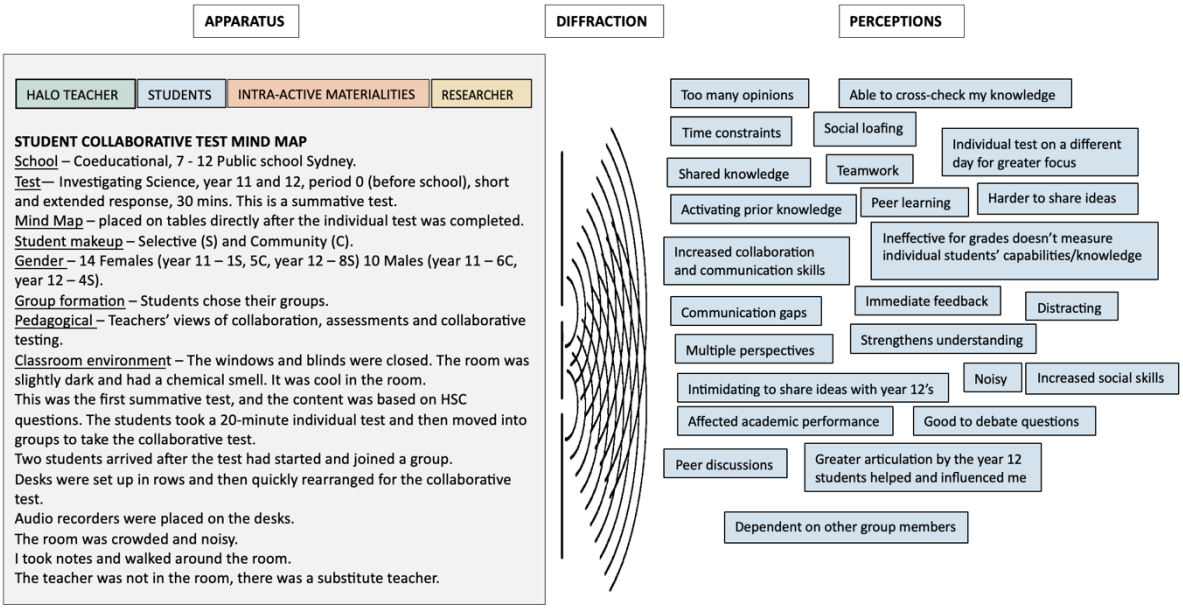
spreading out to maximise the space available and sitting close enough to their intended collaborators for the group portion of the test. I explained that the test would start with a twenty-minute individual section followed by a thirty-minute collaborative segment. With the timer set for twenty minutes, the students began the individual component of the test. At the conclusion of the individual test, the students transitioned to the collaborative phase. The realignment of the classroom heightened the noise levels as students moved into their groups, and I collected the tests. However, the students required the individual tests for the collaborative test, and their redistribution added a layer of challenge, momentarily heightening the room's tension. Despite the initial stress, engagement levels soared, particularly among those typically more reserved. I circulated among the groups, observing. As the collaborative dynamics unfolded, I noticed both the verbal exchanges and some groups' quiet, intense focus. I noticed one particular student in another group who had been very quiet in previous tests, waving their hands, leaning forward and stressing their point with an intense look.

Post-test, this student confided to me '**that they feel more comfortable now and knew they had to participate so they could maximise marks**.' This was the third and final collaborative, which enabled the students to adapt to the established format and engage immediately and quickly. As the test neared completion, each group was provided with a set of questions designed to provoke discussion on the group's perceptions of their experiences. Additionally, individual mind maps designed to further reflect on the collaborative testing process were handed to students. As I collected the tests, the students were

noisy, lots of laughing, frowns on faces, heads down, chairs scraping, desks being moved

Mind Map

Figure 13. *Diffraction of the Entanglement for Halo, Collaborative Test Mind Map.*



This technique separated individual reflections from the collective insights that emerged during the group discussions. As the students engaged with the discussion questions, they simultaneously addressed the mind map prompts on an individual basis. This dual approach facilitated personal reflection on their experiences in collaborative testing and promoted students' metacognitive skills. According to Rivas et al. (2022), as students engage in structured reflection that considers their personal experiences, they better understand their own cognitive processes, learning preferences and collaborative behaviours. This enhanced awareness helps students regulate their learning more effectively and refine their strategies for improved outcomes in both individual and collaborative settings. Despite the logistical challenges of setting up the collaborative test and redistributing completed tests, the students enthusiastically tackled the mind map questions. The lesson concluded with the bell signalling the end of classes; however, students remained seated, completing their mind maps. They handed me their mind maps and tests as they walked out of the classroom, marking the conclusion of this class's entanglement with collaborative testing.

Halo-Researcher-Conversation-Collaboration-Environment[29]

The pre-test conversation occurred in early December 2022; I connected with Halo via Zoom from my location in Sydney and their office at Polaris High School. Our conversation initially focused on Halo's thoughts on student collaboration before shifting to the structure of the collaborative tests and the number of teachers participating. Halo expressed their enthusiasm, remarking, 'I'm *quite excited about this idea [collaborative testing], and it's just taking it to the next level, which is really good*.' A proposal to use a rubric for evaluating students' collaborative skills was developed and later approved by Halo. Following this discussion, the Zoom meeting concluded.

The post-test conversation occurred on a windy and drizzly day in April 2023. Halo and I met in an outdoor learning space at Polaris High School, sheltering under a canopy at a wooden table where I set up the recorder. Halo considered collaboration to be '*students working with their peers and teacher to accomplish tasks in an environment of trust within the classroom*.' They highlighted a significant issue: '*there is a huge disconnect between primary and high school collaborative work. In high schools, collaboration peters out. It's all about me, the individual … we've got to tap those skills that worked in primary schools*.' They further explained that due to this disconnect, '*we don't teach soft cognitive, academic and scientific skills like collaboration really well … not just in this school but across the entire schooling system … you can't have an individualistic outlook as that's not how it's done these days*.' Emphasising the need for genuine collaborative efforts, Halo critiqued superficial group activities, stating:

> *Authentic, not tokenistic … if we were going to do a collaborative exercise on an experiment or something … But are they actually collaborating? Or are they just dividing the work? And I think that's probably from my experience; probably 80% of the so-called collaborative work isn't collaborative. It's a delineation of work tasks. It's cooperative. Yeah, it's not collaboration it's cooperation*

[29] Dashes symbolise the crucial intra-action between all elements, material or otherwise, highlighting the equal importance of both in shaping perspectives on collaborative testing (Gullion, 2018). Other elements contribute to this intra-action but are too numerous to detail individually.

During our conversation, Halo pointed out the reluctance among high school teachers to adopt collaborative activities driven by fears of deviating from standard instructional methods. Halo cites the pressure to adhere to explicit criteria and deliverables as hindrances to collaborative methods.

> *I think there is a bit of fear with some teachers as well that if I'm not teaching them something every single lesson … and displaying explicit criteria on the walls, then the kids don't know what they're learning … so the idea of collaborative learning, I think, makes people a little bit reticent to take it on board. If suddenly your head teacher, DP, or someone walks past your class, that might be a little bit of organised chaos because they're discussing with each other. Well, what are you doing? Where's your explicit instructions? We're in this generation of teaching bureaucracies. You've got to have deliverables, you got to have data to show this, that and the other.*

Furthering this discussion, Halo emphasised the value of collaborative work in the classroom before noting that they had little time to engage in effective collaborative work. The '*time required for deep discussions in collaborative testing is a limitation; we just don't have that class time*.' Collaborative testing does take longer than individual tests as the focus is on students critically thinking through the questions to reach a consensus about the answer, involving arguing, debating, and resolving conflicts as they co-construct the answer.

In our post-test conversation, Halo underscored the importance of encouraging senior students to engage in collaborative work, particularly in the competitive environment at Polaris High School. Halo noted, '*selective students in this school are competitive, so they need to learn how to collaborate to lift everyone's grades up*.' I asked how this was accomplished, and Halo responded that they undertake '*lots of cognitive work like revisions and self-assessing to encourage a culture shift from individualistic, me I need to be the best, to a working together … if we all do well then, we all do better*.' As our conversation drew to a close, Halo revisited the essential concepts of collaboration versus cooperation. Their insights highlight the need for a shift in educational strategies at Polaris High School towards more authentic and effective collaborative activities. As Halo returned to their office, they left behind a compelling narrative of the challenges and opportunities that collaborative testing presents, echoing their initial enthusiasm for a deeper, more authentic engagement in collaborative educational practices.

Part 2: Analysis of the Collaborative Testing Entanglement

Now that I have outlined the logistical aspects of setting up the tests, Part Two shifts focus to the perceptions and experiences of Halo and their students. This section examines how collaborative testing intra-actions influence students' understanding and engagement. By connecting the intra-actions between students' and Halo's experiences, I bridge the gap between theory and practice, showing how each element contributes to the efficacy of collaborative testing.

Student Understanding

This analysis delves into the intricate dynamics of collaborative testing within educational settings, focusing on the intra-action between human and material elements. Through active peer discussions, students enhance their conceptual understanding. For instance, students shared diverse viewpoints during discussions, noting the benefits of second opinions and confirming their thoughts through peer engagement. One student effectively summarised this, stating, '**it allowed us to get different inputs and ideas and allowed us to build on the question with knowledge that we wouldn't have known prior to it. So like, it helped us learn as well while we were answering the questions**.' These entanglements enrich conceptual understanding by integrating new information and fostering a deeper comprehension of the concepts. The transformative impact of collaborative testing is supported by Mahoney and Harris-Reeves (2019), who observed that this dynamic between students-test-discussions leads to the construction of new knowledge as students defended, reasoned, and elaborated on their opinions. Plauborg (2018) also illustrates the importance of the entanglements in evolving students' understanding through the intra-action of spacetimematter (Barad, 2007)[30] amid collaborative testing. However, as Premo et al. (2018) attest, not all students engaged during collaborative testing, leading to resentment and disengagement among some students.

As I consider the outcomes of collaborative testing, it becomes clear that active participation is crucial for deep cognitive engagement. These active entanglements often led to peer-peer teaching scenarios, wherein group members benefit from explanations presented in accessible language. A student's reflection encapsulates this dynamic: '**we have like, a wide sort of knowledge to answer all the questions… so we notice someone might not know something, and then we can explain it like we are a collective**.' Another student noted, '**it expanded my approach to questions**.' This environment not only broadens understanding across all student levels but also cultivates a spirit of

[30] Spacetimematter per Karen Barad (2007), challenges the separation of space, time, and matter, presenting them as inseparable and co-constitutive. This concept is exemplified in the intra-action of collaborative testing, where these elements are integrally and dynamically involved.

collective knowledge. Research by Bowman-Perrott et al. (2013) links peer teaching and discussions to improved academic performance, highlighting the benefits of leveraging diverse subject strengths, backgrounds, and life experiences.

Furthermore, the benefits of peer-to-peer teaching are reiterated by Slavin (1996) and, more recently, Mahoney and Harris-Reeves (2019), who note that higher-performing students reinforce their knowledge through explanations while lower-performing students gain from receiving explanations that are more relatable to them. Additionally, all students benefit from engaging in the critical analysis of the questions and each other's ideas. Meseke et al. (2010) also affirm that student understanding is significantly enhanced by discussing questions, explaining their reasoning and reflecting on each other's responses. Levine et al. (2018) further elaborate that active dialogue among students initiates various metacognitive processes, enhancing their metacognitive awareness and test-taking skills. According to Rivas et al. (2022), metacognition encompasses the awareness and comprehension of one's thinking and learning processes during problem-solving. Hattie and Donoghue (2016) further highlight that engaging in active discussions allows students to (re)think their test-taking strategies by integrating approaches from their peers. They caution, however, that the classroom environment must be characterised by positivity and a lack of competitiveness to maximise benefits.

The analysis of student reflections on collaborative testing offers insightful perspectives on the impact of group dynamics on learning outcomes. Students report that collaborative testing in positive environments enabled them '**to gain different inputs**' and '**working and communicating with other people**' helped to '**build my knowledge base**' while in a '**less stressful environment**.' In these settings, students reported that collaborative testing gave immediate feedback on their '**level of understanding**' while revealing '**gaps in our knowledge**.' This enhancement comes from effectively identifying knowledge gaps and reducing test anxiety, as Siegel et al. (2015) supported. The shared experiences and positive intra-dependence lessened test anxiety and improved confidence by fostering a sense of camaraderie and removing a sense of competition. Highlighting the benefits of this approach, students remarked on the relief felt when knowledge is shared among group members, contrasting it with individual testing. One student stated, '**It takes the pressure off when you're working as a group than when you're by yourself**.' Increasing confidence in students through collaborative testing occurred as students expressed their ideas and engaged in dialogue. The importance of a supportive environment is underscored by Cao and Porter (2017) and Johnson and Johnson (2009), who assert that classroom settings play a significant role in the success of collaborative testing. Positive group dynamics facilitated knowledge exchange and effective collaboration in a less stressful atmosphere. In their studies on

collaborative testing, Nafziger et al. (2011) and Levine et al. (2018) have noted that positive group dynamics, marked by empathy, shared responsibility, and collective effort, create an environment conducive to learning.

In contrast, the dominance of some group members and the passive involvement of others can breed resentment and reduce the effectiveness of collaborative testing. Despite their enthusiasm for participating in authentic, collaborative tasks that promote the adoption of multiple perspectives, as noted in Talbot's (2023) research, challenges were also reported. Students described, '**there were a lot of people and a lot of opinions. So it was hard to prioritise or like, you know, understand who was right or wrong**,' '**some people gave more information than others**,' and lastly, some students felt '**there were people doing the work and people not doing the work**.' As noted in the reflections, some students encountered conflicting opinions, peer intimidation, and uneven contribution levels, which complicated the consensus process. Mahoney and Harris-Reeves (2019) posit how negative dynamics, characterised by competition, conflict, and unequal participation, can affect the collaborative testing experience.

Negative experiences can erode students' confidence and diminish their engagement, ultimately impacting their learning efficacy. A concern raised in this class related to mixing student year groups in the first test. Some students felt the groups '**should be sorted by year levels**' as '**it can be intimidating to share ideas with the year 12's**.' However, a student felt '**greater articulation by the year 12 students helped and influenced me**.' Negative group dynamics, especially in mixed-year groups, impacted students' comfort levels and willingness to contribute, as disparities in knowledge and confidence levels prevented open communication and mutual understanding. Negative group dynamics can increase test anxiety as students are unable to recall prior information, leading some students to lose confidence in their knowledge. These concerns underscore the importance of careful group structuring and fostering positive intra-dependence, as advocated by Meseke et al. (2010) and Siegel et al. (2015), which lessen test anxiety and enhance student confidence and self-efficacy. To improve the collaborative testing experience, Mercer and Howe (2012) advocate for the deliberate structuring and facilitation of collaborative work. Such an approach aims to counteract competitive instincts and promote an inclusive, supportive environment essential for effective collaboration and deep learning. Collaborative testing can enrich student understanding through collective knowledge creation but depends on the entanglement between human and material elements in the classroom environment.

Halo's Perceptions of Students' Understanding

Consistent with the students' perceptions of collaborative testing, Halo underscored the positive impact of such testing methods on student understanding through multiple perspectives. Halo elaborated on the significance of engaging with diverse viewpoints, highlighting its advantages for students in expanding their knowledge of the concepts beyond their initial interpretations. This approach aided in broadening their perspectives and significantly contributed to mitigating inherent biases. Dacey (2020) defines cognitive bias as the tendency of individuals to form conclusions about knowledge and individuals in a manner that is not logically sound, often leading to flawed interpretations and a misrepresented perception of their peers' thinking and actions. Halo expounded on the benefits to student understanding, stating:

> *I think getting **multiple viewpoints**[31] on something. You know, the idea that students have this one idea of what a concept is, but if they're talking to someone else, they've got a slightly different idea … is also a bigger chance of um overcoming biases that kids have as well. Because I think there's this cognitive bias in itself of kids saying, well, you're just trying to teach me whatever's there. But if one of their peers is giving them a **different opinion**, in terms of their cognitive bias, it probably has a greater bearing on changing thought*

Halo pointed out the inclination of students to question educational content based on cognitive biases. This suggests that the sharing of alternate opinions among peers might more effectively influence their thought processes and attitudes towards learning (Siegel et al., 2015).

Throughout the reflection on collaborative testing, Halo observed improvements in student dynamics and understanding. Students were encouraged to ask clarifying questions, promoting deeper inquiry and collaborative response formulation. As Halo observed, students asked, '*what does this mean here? What do you think? And I think there seemed to be a lot more conversation around what the question meant and we'll write this little bit down there, and then we'll go there*.' Halo described collaborative testing as a strategy that deepened students' understanding through questioning and '*peer discussions where students are bringing in different bits of knowledge and views … while enabling them to stop for a second and think about what they are writing*.' These intra-actions introduced varied knowledge and perspectives, encouraging students to pause and reflect on their written responses,

[31] The bolded and italicised text illustrates that both students and Halo made this comment.

concluding that students '*come up with better responses without attacking the integrity of the answer.*' Engaging in the entanglement of the Halo-student-test underscored the effectiveness of collaborative testing in improving educational outcomes. Collaborative testing can successfully broaden student understanding, reduce cognitive biases, and promote a more inclusive and reflective learning process by fostering an environment where students are encouraged to share and discuss diverse perspectives.

Collaboration – 21st Century Skills

Collaborative testing can promote students' 21st-century skills, particularly enhancing social capabilities and conceptual understanding. Active discussions necessitate clear and effective communication, a critical skill when navigating complex concepts (Tolmie et al., 2010). One student said, '**I think my collaborative skills have definitely improved since the first day. Because it's a lot easier to … like, explain my opinions, and I feel a little bit more confident**.' Other students reported increased effective listening skills and an enhanced ability to articulate and refine their concepts through peer engagement. Further reflections reveal growth in collaborative skills and confidence. As one student articulated, '**I think definitely my collaborative skills have improved because we're working with people, we like have never talked to before … from the first group task I didn't talk much but now I introduce ideas I've had even though they could be wrong**.' These experiences highlight how collaborative testing enhanced communication skills and cultivated constructive engagement with diverse viewpoints, which is essential for effective communication. The test-student-group environment also promoted conflict resolution and consensus building. A student appreciated the teamwork aspect, noting they '**liked the teamwork aspect of this task … because we all have to think collaboratively and talk on different topics**.' Students learned to present ideas convincingly and persuasively by navigating disagreements and understanding peers' perspectives, thus enhancing empathy and teamwork, as Johnson et al. (2000) discuss. These intra-actions strengthened persuasive abilities and deepened their understanding and group cohesion.

Active participation in these collaborative testing intra-actions improves social skills by encouraging cooperation, support, and mutual understanding among peers (Johnson & Johnson, 2009). Students reported improvements in social intra-actions: '**it helped me improve my social skills as it introduced me to people I wouldn't usually interact with,' 'it lets me talk to people that I wouldn't have usually talked to.'** Another noted, **'I find it easier to work in groups now**,' indicating a positive shift in group dynamics. However, social conflicts and discomfort with unfamiliar peers can detract from the learning experience, shifting focus from collaborative knowledge construction to navigating group tensions (Lee et al., 2015). Halo assigned groups in the first test based on criteria such

as gender, personality, year and friendship groups, transitioning to self-selected groups for the remaining tests. Students expressed mixed feelings about this: '**preferred to do [work] with people I know it's easier to talk to and like you can open up but with random people I don't feel confident enough to say anything**' and '**like working with people you don't know makes you less likely to actually voice your opinion, especially if it goes against what someone else is saying**.' This analysis demonstrates that the effectiveness of collaborative testing in skill enhancement was linked to the careful management of group dynamics and the resolution of social conflicts or discomfort.

Halo's Perceptions of Students' 21st-Century Skills

Halo's reflection on the development of students' 21st-century skills initially noted that the class exhibited '*minimal collaborative skills as they hadn't had experience in working in that setting*.' This observation points out a significant gap in current educational methods, which often fail to cultivate essential 21st-century skills. Polesel et al. (2014) critique the prevalent focus on high-stakes testing and outcomes-based education, arguing that such practices promote competitive and individualistic behaviours over collaborative skill development, crucial to preparing students for future challenges. Supporting this, Kapitanoff (2009) and, more recently, Jang et al. (2017) suggest that collaborative testing, which requires effective idea exchange and discussion, can significantly improve these skills. Despite initial challenges, such as students talking over each other and ineffective communication, improvements were evident. Halo notes, '*I think some of our kids have this tendency to talk over each other and not necessarily have that talk, think, talk process. And I think that that seems to have improved a little bit*.' This shift highlighted the positive impact of collaborative testing. Halo further observed:

> *It's been really good. It was really interesting seeing the evolution of that first task to the last one … it's been really positive in terms of also seeing the* ***dynamics of the students cooperating with each other*** *… particularly in this school where traditionally kids have been quite competitive with each other over marks … When you say, well, to get marks, you've got to cooperate with each other. I think that's just changing the interest and the dynamic within the class a little bit for the positive for the most part*

Halo's observations provide a narrative on the evolving nature of collaborative testing in the classroom. Alternatively, the comment also highlights the significance of the HSC in influencing

teachers' pedagogies. Halo noted that if students collaborated, they could collectively raise their grades, benefiting the schools' performance in high-stakes external exams. Although the environment[32] was open and engaging, students also viewed each other as rivals due to the ranking system of the HSC, influencing some students' approach to assessment. The introduction of collaborative testing encouraged a more positive and collaborative atmosphere for many students, suggesting broader educational benefits.

Through collaborative testing, Halo initially saw a lack of effective communication, where students often talked over each other without engaging in reflective listening. Through the process of undertaking the three collaborative tests, students improved their collaboration skills, engaging in more thoughtful discussions and demonstrating a willingness to incorporate diverse viewpoints. This shift challenged students to work together towards common academic goals and facilitated the development of essential communication skills. It encouraged active listening, clear articulation of thoughts, and negotiation of understanding, fostering a learning environment where diverse ideas can be explored. Juceviciene and Vizgirdaite (2012) support this perspective, positing that a collaborative and inclusive classroom environment promotes positive interdependence, reduces competition, and alleviates the fear of failure, thereby improving students' cognitive abilities. This approach, as observed by Halo, represents a valuable pedagogical tool in preparing students for the complexities of a 21st-century world where collaboration and effective communication are as important as academic competencies.

Part 3: Exploring the Material of the Classroom Environment

Building on the insights from Halo's and students' perceptions, Part 3 explores the less tangible aspects of the classroom environment, such as timing and spatial considerations, and their impact on collaborative testing. This section reconnects with initial descriptions of the classroom setup to deepen understanding of how environmental factors shape the testing experience. By examining the material aspects, I highlight their critical role in student engagement and learning.

[32] The classroom environment, as described, includes not only the students and Halo but also the material elements in that particular classroom. Furthermore, it incorporates Halos' teaching methods, highlighting their significance in student learning and engagement.

Timing of the Tests

The early scheduling of the Investigating Science class, at 8 am, an hour before the official start of the school day, poses significant challenges to students' alertness, engagement, and overall performance (Ingham-Hill et al., 2024). Adolescents are particularly impacted by early morning start times due to shifts in their sleep-wake cycles during puberty. Research highlights that adolescence is marked by a preference for later sleep and wake times, resulting in cognitive functioning peaking in the afternoon rather than the early morning (Crowley et al., 2018; Saxvig et al., 2021). Minges and Redeker (2016) note that the consequences of early start times include memory difficulties, reduced focus and poorer grades. Comments from students indicating that it was '**too early**' and observations of yawning and struggling to focus align with findings that early morning performance is hindered by circadian misalignment and fatigue (Dikker et al., 2020). Reduced alertness during this time impacts students' ability to contribute effectively to group discussions, retrieve prior knowledge, and problem-solve collaboratively. These limitations compromise the benefits of collaborative testing, which relies on active participation and mutual engagement to enhance learning outcomes. Research supports this link, as Ingham-Hill et al. (2024) attest early school start times misalign with this circadian rhythm, exacerbating issues such as insufficient sleep, inattention, and reduced cognitive performance.

A growing body of evidence supports the benefits of delayed school start times, with meta-analyses showing increased sleep duration, improved attendance, and reduced tardiness (Bowers & Moyer, 2017; Minges & Redeker, 2016). Furthermore, delayed start times have been linked to reductions in daytime sleepiness, depression, and even motor vehicle crashes, emphasising broad behavioural and health benefits (Wheaton et al., 2016). Despite logistical challenges in implementing later start times, these findings underscore their potential to improve academic outcomes and align with adolescents' natural circadian rhythms. Morning classes often show diminished retention compared to mid-morning classes, reflecting a "U-shaped" relationship between performance and time of day. This is influenced by both circadian misalignment and fatigue as the day progresses (Dikker et al., 2020). Thus, addressing the mismatch between adolescent chronotypes and traditional school schedules remains critical for enhancing learning outcomes and overall well-being.

The early scheduling also exacerbated issues of tardiness among students. The impact of late arrivals disrupted not just the individuals' learning but also altered groups' dynamics, which is crucial for collaborative testing. Late arrivals forced groups to reconfigure, hindering their cohesion. This disruption was compounded by the frustration of punctual students, with one commenting, '**this is bad. Everyone must come in on time**.' Gottfried (2014) notes that tardiness leads to disruptions in

group discussions and necessitates time spent updating latecomers, reducing the time to engage with test questions. Tardiness undermines the cohesion and efficiency of collaborative work and negatively impacts students' perceptions of the collaborative testing experience, potentially demotivating them from actively participating. Given these factors, it is evident that the early timing of the Investigating Science class has a tangible impact on students' attention, engagement, and group cohesion, limiting the effectiveness and perceived value of collaborative testing. Addressing the alignment of class schedules with adolescents' circadian rhythms could mitigate these issues, fostering an environment more conducive to the success of collaborative testing initiatives.

Spatial Arrangements

The physical environment of a classroom, including spatial arrangements, plays a critical role in shaping learning outcomes and collaborative dynamics. According to Hernández (2023), elements such as movement, speech, and the physical environment significantly influence student learning. Llorens-Gámez et al. (2022), in a systematic literature review, emphasize that the design of learning spaces impacts memory and attention through factors like spatial geometry, colour, texture, height, width, light, sound, and temperature. While Talbert and Mor-Avi (2019) found that active learning spaces enhance learning. Conversely, Demir-Yildiz and Tatik (2019) observed no significant differences in outcomes between flexible and traditional classrooms, highlighting the nuanced effects of spatial design.

In Halo's classroom, the bench arrangement created a sense of overcrowding, hindering movement and interaction within the tightly packed rows of three benches without gaps. This setup limited Halo's ability to reach students in central seats, potentially affecting their understanding of the subject (Brooks, 2012). The overcrowded arrangement favoured certain pedagogical approaches, restricting student-to-student interactions and promoting passive learning, as noted by Ozkan Bekiroglu et al. (2022). Halo's positioning at the front of the room, combined with the focus on individual tasks, reflected a preference for individual learning that undermined the potential benefits of collaborative approaches, which are known to foster diverse skills and competencies (Ozkan Bekiroglu et al., 2022). Additionally, the close proximity of desks limited effective collaboration, while overcrowding increased noise levels during testing, aligning with Meader's (2023) findings on how noise impacts perceptions of collaborative testing. These spatial constraints disrupted the intended collaborative learning environment, underscoring the challenges of aligning physical classroom arrangements with the needs of collaborative pedagogies.

This section examined how timing and spatial arrangements influence student experiences in collaborative testing environments. I highlighted the challenges posed by an early start, which disrupts the student's circadian rhythms, reducing alertness and exacerbating tardiness, thereby affecting group dynamics and test performance. Additionally, cramped seating arrangements limited intra-action and movement, hindering effective collaboration. Overall, these factors impacted the efficacy of collaborative testing, underscoring the importance of optimising environmental conditions to enhance educational outcomes.

Part 4: Unique Observations and Broader Implications

This final part of this chapter synthesises observations from the preceding sections to discuss broader implications for educational practice. Here, I analyse how specific factors like high-stakes testing, accountability, and Halo's practice influence both the process and outcomes of collaborative testing. This section concludes my detailed examination and connects to the chapter's overarching argument, reinforcing the importance of considering the material-nonmaterial continuum in collaborative testing settings.

Perceptions of Halo and Students on the Summative Collaborative Test

High-stakes testing, exemplified by the HSC, has prompted a pedagogical shift towards prioritising exam preparation over innovative learning strategies. This shift is reinforced by the intra-actions of students, parents, teachers, the media and government policies, prompting rote learning and exam skills above collaborative learning, problem-solving, and critical thinking skills (NESA, 2020). Critics like Resnick and Schantz (2017) argue that this emphasis on narrow exam-centric teaching methods undermines active learning environments and distorts students' perceptions of assessment methods. In Halo's Investigating Science class, two instances illustrate the impact of this educational culture. Upon exiting the classroom, a student lamented, '**This is our HSC. why do this? I think I'll cry; this is terrible**.'

A group of year 12 students debated the fairness of the assessment's contribution to their final scores, saying,

> Male: '**Yeah, really don't think it should go towards our full marks**,' Female: '**I don't think this should be used for my HSC assessments**,' Male: '**Yeah, exactly. Especially for the HSC, I understand practise tasks**,' and Female: '**Totally and maybe even do it for preliminary [exams], not the HSC'**

These anecdotes underscore a broader educational trend prioritising quantifiable success, such as HSC outcomes and ATAR[33] scores over the learning process (Lingard, 2010). This culture places immense pressure on students, as reflected in their anxiety about losing ATAR points due to collaborative tests. This discussion highlights the significant weight students and stakeholders place on ATAR scores as definitive indicators of a student's ability over thirteen years of school, intensifying the pressure on students (Shergold et al., 2020). After leaving the classroom, two male students were overheard voicing their fears about losing ATAR points, remarking, '**I probably just lost half a mark of my ATAR; my ATAR has been decreased by 30 points; maybe by like two points. Just because of this, you can't tell**.' Such attitudes diminish the value of diverse educational experiences and hinder the adoption of collaborative testing methods, revealing a broader issue of an educational focus that prioritises grades over learning.

Despite the benefits of collaborative testing for enhancing understanding and teamwork, there is a deep-seated preference for individual high-stakes assessments. An illustration emerged with students articulating concerns regarding the test, describing it as '**ineffective for grades as it does not measure a student's capabilities and knowledge**.' This preference, supported by education's focus on individual achievement, persists despite the potential of alternate assessments like collaborative testing to foster critical thinking and collaboration. This reluctance is amplified by stakeholders who perpetuate an educational paradigm prioritising grades over diverse learning experiences.

Collaborative Testing and Accountability

In observing a student group discussion during the test, a notable intra-action caught my attention. A female student's appeal for group attention and collaboration highlighted the accountability issue. She expressed frustration over her peers' lack of engagement: '**we need to talk about this 'cause we haven't talked about that. You guys aren't listening**.' The other female student replied, '**of course, they're not**,' before a male student in the group said to the other, '**try to focus mate**.' This situation led to a discussion with Halo, who identified a lack of accountability as the underlying problem. Nevertheless, many students in the class actively collaborated with their peers during the first two non-graded tests. In their collaborative testing study, Hattie and Donoghue (2016) note that students who have greater intrinsic motivation and self-confidence, in addition to a positive

[33] The Australian Tertiary Assessment Rank (ATAR) is used to gain entrance into university by comparing students undertaking multiple subjects across Australia. ATAR scores are not a measure of a students' individual performance (Shergold et al., 2020)

group, enhance their testing experience, leading to greater motivation and active involvement in the group.

However, it was observed that more students were not engaged in their group during the first two tests than the summative third test. This frustration with their peers' disengagement highlighted the accountability issue, pointing to the need for collective responsibility and individual contribution, essential for fostering active participation (Gillies, 2016; Johnson & Johnson, 2008). The observation that students actively engaged with their group to secure higher marks lends substantial support to the idea that the integration of grades in collaborative testing significantly influences student intra-action dynamics. This observation underscores the relationship between assessment methods and student behaviour. When grades are at stake, students are more likely to invest effort in undertaking the collaborative test by articulating their thoughts and listening to their peers. This heightened engagement can lead to a deeper comprehension of the subject matter, as students are not merely passive recipients of information but actively engage in their learning.

Prior to the study, Halo and I had developed a grading scheme for these tests that mirrored those of individual topic tests alongside a rubric evaluating students' 21st-century skills. However, a critical decision was made before the commencement of testing: grades for collaborative tests would not be officially recorded, nor would the rubric be utilised to assess student skills. This decision effectively nullified the structured accountability framework I had envisioned. The decision to abandon the accountability framework is attributable to the high-pressure context of senior school education in NSW, where students in years 11 and 12 frequently undertake assessments that emulate the HSC exam in format and complexity. The high-stakes associated with these examinations and the public release of HSC results place significant scrutiny on schools and teachers, as highlighted by Masters (2020). This scrutiny often discourages the adoption of varied assessment methods, leading to a preference for conventional approaches closely aligned with HSC expectations, thereby influencing pedagogical strategies.

Collaborative Testing and Halo's Practise

Reflecting on the collaborative testing intra-action, Halo plans to incorporate more collaborative work in regular lessons due to the observed progress of the students over the tests, noting, '*I think the one thing we might look at changing is just embedding some more of those group skills into some of that general classwork*.' This approach aligns with Hayes et al. (2020), who view educational practises as dynamically entangled through the coalescence of time, bodies, spaces and

objects, (re)constituting each other during the different tests (p. 367). Halo further proposed incorporating more collaborative work over all the science grades to engage both mainstream and selective students, describing,

> *We just need to start embedding in there. But particularly, like our mainstream kids, we tend not to do that. Because a lot of people say, no, you just need to sit down and do your work. And the selective kids have the exact same problem because they just want to sit down and do their work. They don't want to cooperate. So you've got these two extremes*

Collaborative testing offers a way to address the lack of 21st-century skills in schooling but faces substantial obstacles, from rigid curriculum mandates and entrenched assessment methods to inertia in stakeholders to embrace the value of trying collaborative methods (Cairns, 2021; Klenowski, 2011). Halo further proposed incorporating collaborative testing assessments, '*Particularly with the year 12 class, as we get closer to trials, we'll probably do more of that collaborative sort of formative assessment*,' recognising the benefits amidst the pressure from high-stakes testing in the HSC. This strategy reflects critiques by Klenowski and Wyatt-Smith (2012) and a 2020 evaluation by the New South Wales Education Standards Authority (NESA) of the HSC framework, advocating for greater pedagogical flexibility. Despite these intentions, Halo acknowledged the challenges of integrating collaborative testing and discussions amid the pressures of content-heavy curricula, articulating, '*the only limitation is that it's finding time to give them this task where they could have deep discussions about it just as hard to do with the time*.' This concern aligns with Lingard and Sellar (2016), who contend that high-stakes testing often limits the teaching of broader skills and competencies.

Nevertheless, Halo sees the value in aligning summative tests to real-world expectations,

> *In terms of collaborative summative stuff, it's at university so that's an easy justification of saying, well, okay, you're going to go to university. Well, this is how they're assessing in university; oh, you're not going to university; what does it matter? So when you get to the workplace, you're not necessarily judged on your own work; you're judged on your team's work. I think it's an important skill; you've got to learn these skills*

Halo's exploration into the potential of collaborative assessments for summative evaluation underscores the importance of aligning educational assessment with real-world expectations, addressing the critical need for teamwork skills and challenging the prevalent focus on solo performance. Halo acknowledged student concerns about collaborative tests with the imperative of (re)examining teaching and assessment strategies to cultivate the development of 21st-century skills within the educational framework.

Conclusion

This chapter provided a comprehensive analysis of collaborative testing in Halo's classroom, structured into four detailed parts that collectively delve into the intra-actions between various elements that shape the testing environment. Students and Halo articulated the benefits, such as an increased enhanced conceptual understanding due to peer discussions and multiple perspectives and the development of collaborative and communication skills. Concerns regarding student lateness, class timing, length of tests and group structure were commented on. Notably, the lack of accountability in the first two tests was highlighted as affecting group effectiveness. The only summative test in the study was explored, with some students remarking on the impact of this collaborative test on their Higher School Certificate and ATAR scores. Halo's views of collaborative testing and its influence on their practice were discussed. They noted they would increase the number of collaborative tasks in their classes while commenting on the difficulty of undertaking collaborative work and testing in their classroom due to syllabus constraints. The intra-action between collaborative tests and the other influencing elements created new insights into how Halo and their students perceived this novel assessment method.

Chapter 6: Beta's Entanglement

This chapter shares a similar structure with the previous chapter on Halo but focuses on a different class with new participants. Here, I describe the experiences of Beta' and their students.' The diffraction[34] of the collaborative tests, illustrated in Figure 14, is organised into the apparatus, diffraction, and perceptions, which are unique to Beta, their students, and their classroom. The apparatus[35] details the school, subject (Physics), year level (11), period (2), test structure and timing of the test (40 minutes). It also describes the class composition, including the mix of selective and community students,[36] gender distribution, and the physical classroom environment.

This chapter is structured into four comprehensive parts, each examining a distinct aspect of collaborative testing unfolding in Beta's classroom.

Part 1 introduces the four diffractive diagrams that depict the intricate entanglements among all the elements influencing perceptions of collaborative testing. These entanglements illustrate the intra-connectedness between students, Beta, myself, and the material aspects of the classroom environment. Exclusions from these collaborative tests, detailed in the techniques section of the Research Design Chapter, address the selection of only year 11 students and the tests' subjects, content, structure, and timing. This part concludes with a description of Beta's post-test conversation regarding collaboration and collaborative testing. A deeper analysis is deferred to subsequent sections.

Part 2: This part examines the similarities and differences in perceptions of collaborative testing's effects on student understanding, metacognition, and 21st-century skills among Beta and the students.

Part 3: Here, I explore the material elements of Beta's classroom environment, particularly how heat, exam stress, and a predominant test format centred on calculation and problem-solving questions influence students' perceptions of collaborative testing.

Part 4: This section offers unique observations from Beta's classroom, analysing how group formation, gender imbalances, and Beta's emphasis on individual performance affect perceptions of collaborative testing. Each part aims to provide a thorough understanding of the different factors contributing to the

[34] A diffractive analysis considers the entanglement between all matter. It is iterative and evolving to reveal the emergence of phenomena and differences to understand the possibilities of collaborative testing (Barad, 2007).
[35] The researchers' apparatus is a contextual and dynamic entity, intra-acting with the researcher, participants, material-nonmaterial continuum, to interpret their practice and the phenomena under study (Plauborg, 2018).
[36] In New South Wales (NSW), Australia, the Department of Education and Communities administers an exam to identify high-ability students, who are then placed in advanced classes in certain schools. Meanwhile, "community students" — those living within a school's local catchment area — are automatically eligible to attend without taking any entrance exams.

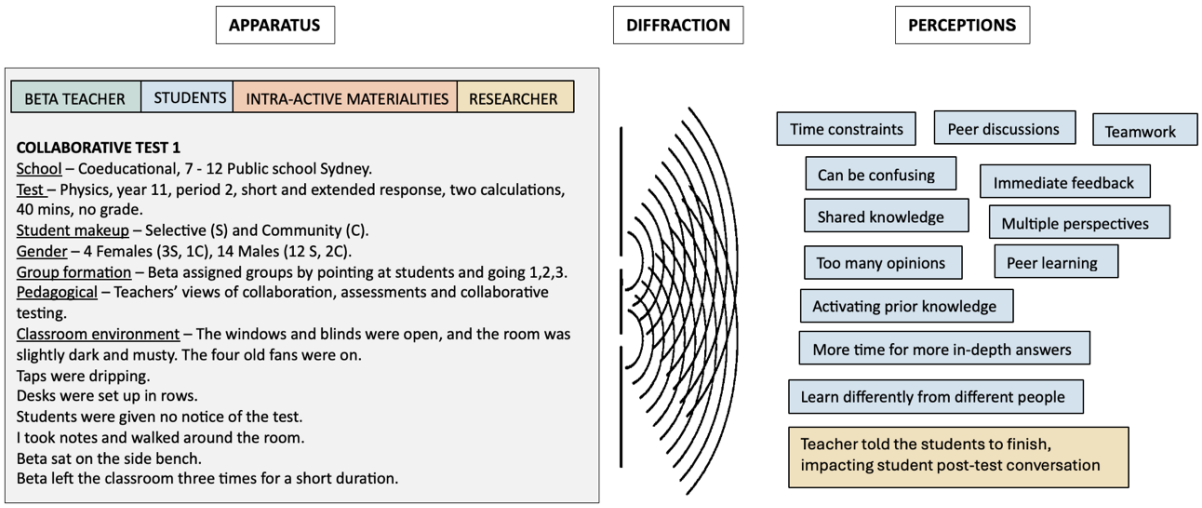
efficacy of collaborative testing in Beta's classroom. The chapter concludes with a summary that synthesises these insights.

Part 1: Setting the Stage with Collaborative Tests and Mind Map

The chapter begins by introducing the four diffractive diagrams that illustrate the complex intra-connections among various elements influencing perceptions of collaborative testing. This setup is crucial for understanding the entangled dynamics within Beta's classroom and details the components of the apparatuses. This part concludes with an overview of Beta's post-test reflections, setting the stage for deeper analysis in subsequent parts.

Collaborative Test One

Figure 14. *Diffraction of the Entanglement for Beta, Collaborative Test One.*



Note. This figure illustrates the analysis of collaborative test one unfolds in three stages. It begins with examining the test environment, which includes physical and discursive elements such as classroom conditions and teacher opinions. It then moves to a dynamic analysis of varied perceptions from student and teacher reflections. It culminates in presenting these insights through a colour-coded scheme to highlight differences between students' and teachers' viewpoints based on the results of the diffraction analysis.

I walked into Polaris High School's[37] science staff room after finishing the first collaborative test with Halo. As I was chatting to Beta about when they would like to undertake the first test, Beta stated, *'Why not in period two today?*' [38] further remarking that the '*students would be fine with it*.' As I pondered how to implement the collaborative test in Beta's classroom without information on students' personalities or abilities, a full evacuation drill occurred as period one started, with the students and teachers moving to a designated area away from the school grounds.

[37] This is a pseudonym for the school.
[38] Italics illustrate Beta's comments and bring their voice to the forefront above the sea of text.

I entered the warm classroom for period two and noticed a stark contrast in ambience and layout compared to Halo's room. Beta's classroom featured antiquated wooden desks and plastic chairs, surrounded by wooden cupboards housing minimal laboratory equipment and minimal displays of student work on the walls, creating an austere feel. This setup is documented in Figure 15.

Figure 15. *Beta's Classroom Layout, Polaris High School.*

The contrast between Halo and Beta's classroom is striking. Halo's classroom is a modern, well-lit, air-conditioned space equipped with new desks, stools, and lab benches, which, according to Manca et al. (2020), may boost student engagement by creating a comfortable learning environment. In stark contrast, Beta's classroom is dimly lit and sparsely furnished with outdated amenities, creating a less inviting atmosphere. This discrepancy supports findings by Cheryan et al. (2014), who argue that the physical aspects of the classroom, from its furnishings to its overall ambience, play a role in influencing students' perceptions of their learning environment. However, as Zapatero et al. (2012) note, the impact of the classroom environment is often overshadowed by the more dominant influence of the teacher's pedagogical methods.

Beta and I entered the classroom before the start of period 2 to discuss the collaborative test. Despite the diverse personalities, gender imbalance and abilities of their students,' Beta declined to assign groups beforehand, suggesting they would 'be fine.' Similar to Halo's and the other science teachers' classrooms, the desks were arranged in rows, which I quickly rearranged into five groups to facilitate collaboration. The desk layout was the same as for Halo, with two desks together, chairs on either side and the audio recorder placed in the middle of the desks to capture intra-actions.

As the bell rang, students promptly filed in, lining up on the sides of the classroom. Beta smiled at the students and numerically assigned groups based on '*1, 2, 3, you go there*' as they moved through the room. I am left wondering[39]

how will the groups work, will this affect the test and their perceptions

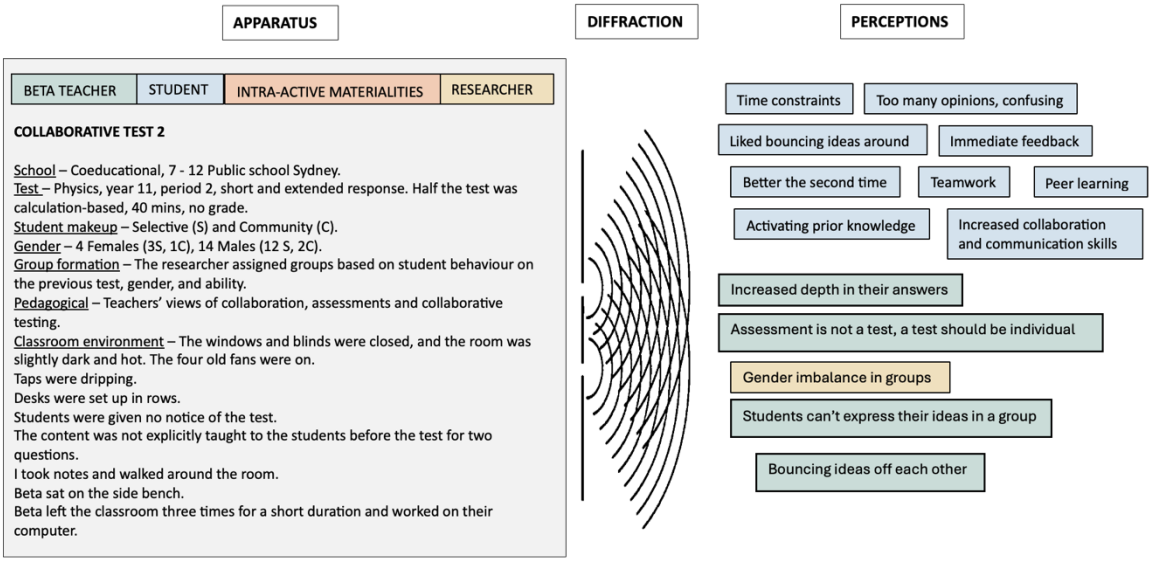
This improvisation led to various group dynamics: one consisted of three male students and one female student, all selective.[40] Two other groups were each made up of three male students, mixing selective and community students. Additionally, one group included three female students, and two were sisters. Once the collaborative test commenced, I observed varying levels of engagement and intra-active processes across these groups. For instance, the three female students demonstrated high enthusiasm and collaboration, as evidenced by their positive verbal expressions, remarking between them, '**this is really fun; I like this. It's fun to collaborate**,' before delving back into their discussions—conversely, the mixed-gender group comprising three extroverted males and one reserved female faced notable challenges. The males' off-topic behaviour and subsequent exclusion of the female member highlight critical aspects of group cohesion and participation. The physical and interactive withdrawal of the female from the group underscores the potential barriers to inclusive engagement within spontaneously formed groups.

While circulating the classroom, I noticed Beta stepping out several times and often focusing on their computer rather than observing the students. As the students neared the end of the test, I wrote three reflective questions on the board for them to discuss the post-test. The timer sounded, and Beta resumed a more active role as the students engaged with the questions. After collecting the tests, I exited the room with insights into the critical role that both classroom environment and teacher involvement play in shaping the efficacy of collaborative testing strategies. Beta declined to reflect on the first collaborative test.

[39] The shaded box contains excerpts from my reflective journal, presented in the present tense to emphasise my role in the intra-connectedness of collaborative testing. These excerpts enhance the description of the collaborative test by illustrating my personal engagement and observations during the test.
[40] Group composition is an important physical consideration for the efficacy of collaborative testing because it influences the diversity of perspectives, skills, and knowledge, enhancing student understanding and 21st-century skills.

Collaborative Test Two

Figure 16. *Diffraction of the Entanglement for Beta, Collaborative Test Two.*



Collaborative test two occurred in early March, two weeks after the initial test. The day of this test was sweltering and humid, so the school decided to restrict outdoor activities, including sports, during recess and lunch. This test was aligned with the physics syllabus and approved by another physics teacher before the week of testing. Compared to the previous test, this collaborative test incorporated more calculation-intensive questions. Beta chose to review the content on the morning of the test, at which point they expressed surprise at some of the questions, exclaiming that there was content not covered in their class. Nevertheless, Beta was curious about how students would answer unfamiliar questions in a collaborative testing environment. As I made my way to the classroom, I contemplated what would unfold as students tackled these challenges.

When entering Beta's classroom, I noticed

> It's pretty hot and humid, and the fans were doing little to alleviate the heat

The physical setup of the classroom was (re)arranged to facilitate collaboration. Instead of the paired desks used in the first test, I opted for single desks accompanied by flanked chairs, which I had instigated for Halo's second test. This change was driven by the need to capture quieter student contributions better. In preparation for the students' arrival, an audio recorder and individualised Post-it notes bearing the group names were placed on the tables. I organised the student groups based on ability, gender, personality, and how they entangled with the test-students in the previous test. For this second test, I placed the four female students in four different groups with two male students. Students entered the room and (re)configured the entanglement as they became part of the second

collaborative test. The students were not told the test would occur today, as in the previous test, and their lack of enthusiasm for the collaborative test was palatable,

students groaning, mutterings of why

when they observed me in the classroom. I informed the students that they were in specific groups and asked them to find their group. The students moved to their groups, and the test began after I reminded them how to collaborate for the test. As the students undertook the test, Beta sat on a bench, working on their computer and left the classroom several times. Students undertaking test questions that they have not been taught can increase anxiety, affecting students' performance and perceptions of collaborative testing. As one male stated, '**what's this**?' before a female answered, '**oh, I don't know**' and the other male in the group said, '**we did not do this**.' Testing students on subjects they have not been taught can increase their anxiety, impacting their confidence and self-esteem and lowering their motivation to persist in the test (Breedlove et al., 2004; Krispenz & Dickhäuser, 2018). However, peer-to-peer interactions may reduce students' anxiety by sharing information, discussing questions and answers, and having academic and social support from their peers.

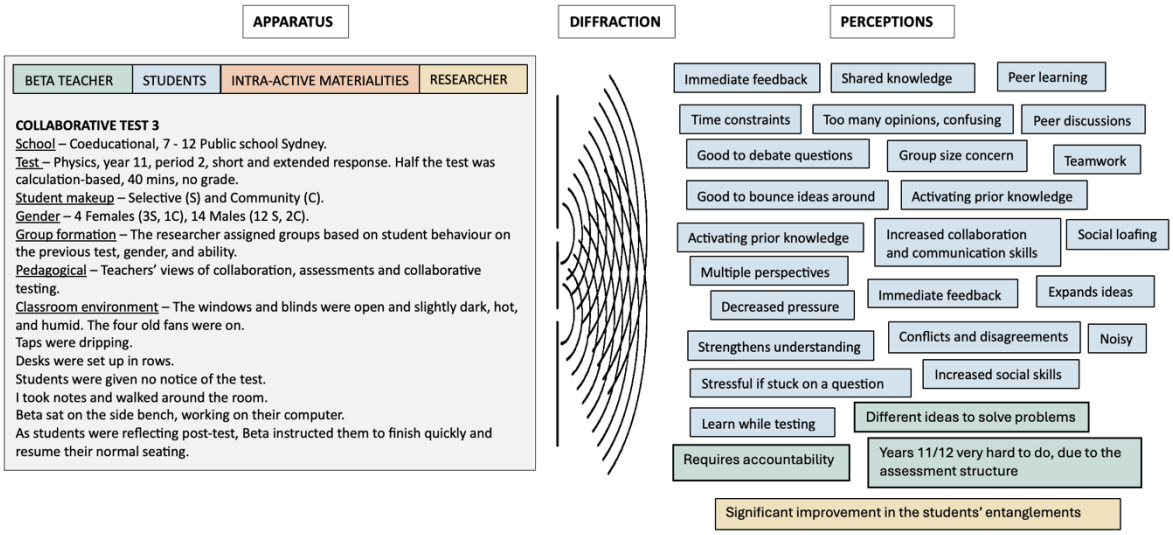
During my observations, I noticed that similar to Halo's class, most students quickly engaged with their group members as they embarked on the second test. However, one student remained disengaged throughout the test. I asked why they were not participating, and the male student stated, '**I don't know, it's hot**.' As the test progressed, students grew increasingly fatigued and uncomfortable in the stifling classroom. I noticed the students'

red faces, sweat, yawning

At the completion of the forty-minute test, I wrote the discussion questions on the board, and the students answered them. However, Beta's urgency to conclude the test and resume regular instruction was evident. After collecting the tests, I left the classroom. Expressing gratitude to Beta, I left the staffroom, reflecting on the day's insights and challenges.

Collaborative Test Three

Figure 17. *Diffraction of the Entanglement for Beta, Collaborative Test Three.*



I arrived at Polaris High School on a warm, balmy day two weeks after the second collaborative test. Prior discussions with the physics teachers emphasised a shift towards more calculation-based questions, resulting in a format that included two short-response questions and six extensive calculation-focused problems. Before the test, Beta and I convened in the science staff room to review the test content and discuss the student group discussions and individual mind map questions. During our conversation, Beta voiced concerns about the curriculum's emphasis on group work at the expense of individual problem-solving, particularly in senior classes. Stepping into Beta's classroom just before period two, I observed,

> It's hot even though the windows are open, musty, and dim

I promptly reorganised the tables from their usual rows into spaced-out clusters in the corners of the classroom, aiming to maximise the distance between the groups to facilitate a new collaborative test setting. This group arrangement, which was based on ability, gender, personality and how they entangled with each other, mirrored the setup in the previous test. In preparation for the students' arrival, an audio recorder and individualised Post-it notes bearing the group names were placed on the tables.

The bell signalled the start of period two, and students streamed into the classroom. Upon greeting Beta, the students noticed the changed classroom layout, eliciting groans. I instructed the students to find their groups, emphasising the collaborative nature of the impending test. Once settled, the test commenced. Beta observed the students on the perimeter laboratory bench while I circulated among the groups. The group dynamics appeared well matched; students engaged more

readily with both the test material and their peers compared to previous tests. Similar to the prior test, students worked collaboratively while I observed them undertaking the calculation-based problems. However, one group divided the questions as they thought they '**might not finish in time**.' I intervened, encouraging them to collaborate to overcome this limitation, which they ultimately did. Some groups experienced vocal disagreements, momentarily talking over each other before resolving the issue and resuming their discussions. As I navigated between the groups, I observed Beta maintaining a watchful eye on the students from the room's perimeter. The challenge of time, a common issue for Halo's students, became increasingly apparent as the test progressed. Many groups struggled to complete all the questions within the allocated time. Recognising this, I advised the students to skip the final question, allowing more time for the other questions. Reflecting on these time constraints, one group of students shared their frustrations:

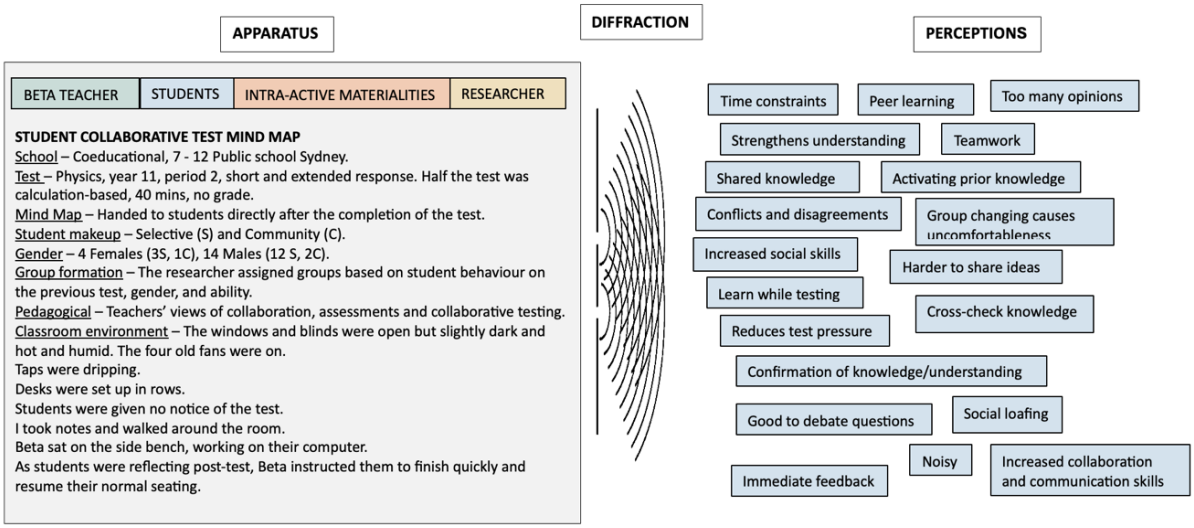
> **I didn't like time restraints because we should have more time.**' '**But one thing I didn't like is that we only had 30 minutes to do it. I wish we had more time to do it**,' '**one thing that I disliked is the amount of time we'll need more time to like to understand**

These comments underscore the need to re-evaluate the time allotted for calculation-based problems, which involve complex problems requiring multiple equations, as detailed by Tanujaya (2016). More time should have been allocated for this test to allow students adequate time to comprehend the questions, select the appropriate formulas, and solve the problems to find the answers.

As the test period ended, I prepared the classroom for the next phase: a student-led discussion. Each group's table was supplied with the questions and individual mind maps. Once the test officially ended, I took a moment to outline the discussion process, detailing the six questions designed to capture their insights on the test, including their likes and dislikes, how the test influenced their understanding and skills, and their thoughts on collaborative testing, especially if the test counted for grades. Despite my intentions for a smooth transition, I observed that several groups had not yet completed their test questions and continued working on them. I navigated through the classroom, encouraging groups to shift their attention to the questions. As discussions began and I collected the completed tests, I noticed Beta's presence at the front of the classroom, standing behind their desk.

Mind Map

Figure 18. *Diffraction of the Entanglement for Beta, Collaborative Test Mind Map.*



Upon concluding the group discussion, I asked the students to individually reflect on their collaborative experience using the mind maps, which featured four prompts: what did they like and dislike about collaborative testing, how do you think this method affected your understanding or learning, and what are some problems with collaborative testing? However, before they could fully engage with this reflective task, Beta announced, '*Okay, that's it. Finish this and go on to this*,' causing the students to rush through the reflective process. This rush made it hard for the students to think through the prompts, leaving them just to write a few words, affecting my assessment of their perspectives on collaborative testing. After collecting the mind maps and the recorders, with the students hurriedly rearranging the desks back to their regular rows, I departed the classroom, contemplating the day's insights and challenges.

Beta-Researcher-Conversation-Collaboration-Environment[41]

I had arrived at Polaris High School before recess, as Beta and I had scheduled a meeting to discuss their perspectives on collaborative tests. On that morning, I found Beta and the other science teachers discussing concerns about senior science assessments in small groups. Noticing Beta's particularly strained expression, I asked if we should reschedule our conversation; when they suggested period two, I was puzzled, given Beta's teaching commitments. As I entered Beta's classroom, the reason behind the timing of the meeting became clear.

[41] Dashes symbolise the crucial intra-action between all elements, material or otherwise, highlighting the equal importance of both in shaping perspectives on collaborative testing (Gullion, 2018). Other elements contribute to this intra-action but are too numerous to detail individually.

Beta stood at the front of the classroom, sternly addressing the students about their inadequate grades on the last individual test.

> Students are quiet, no sound, they are looking at the floor, agitation in their hands as they twiddle their fingers

After admonishing the students, Beta instructed them to work individually on their booklets, creating silence that filled the room. Beta and I relocated to the adjacent prep room with the class settled, commencing our conversation. Initiating our dialogue, I queried Beta's stance on student collaboration, which they defined as the ability to exchange ideas and broaden perspectives through cooperative task allocation. Beta described collaboration as the '*ability to bounce ideas off each other … to get diverse ideas … widening your vision rather than tunnelling it*.' Further, Beta characterised collaboration as a cooperative task where '*students allocate work. Okay, you're doing this bit, you're doing this bit, I'm doing this bit. Then let's come together*.'

Beta pointed out a discrepancy between students' in-class participation and their performance on solo tests, underscoring the potential benefits of collaboration testing to bridge this gap. Beta differentiated between collaborative learning and testing as our conversation deepened, expressing, '*I'm all for collaborative learning. But testing is another thing altogether*.' Further conversations revolved around the potential of collaborative testing to enhance students' conceptual understanding, with Beta noting the potential for students to share ideas, addressing gaps in individual knowledge through '*students may not know how to answer this question, but another kid does … you know, some of the ideas that are bouncing off each other … So, in that sense, I hope that's what they gathered*.' Beta further mentions in our conversation that collaborative testing '*may incorporate learning with assessment*.' However, Beta acknowledged the absence of data on improving students' collaborative and communication skills, describing, '*we never actually collected any data on that to see what it was like*.' Before commenting, '*I don't have all the data. Yes, because they were answering the types of questions, and we saw that they were discussing and talking with each other. So yeah, now I reckon it does, it does*.'

Acknowledging issues with student engagement in collaborative testing, Beta attributed it to social loafing and other factors, such as students who are '*not very talkative. Now, that doesn't necessarily mean that that student doesn't know the word; it could be that they very well know it. But sometimes it could be shyness or the fear I might be wrong*.' When asked about the impact of collaborative testing on Beta's teaching strategies, they admitted the absence of collected data

prevented them from knowing whether it affected them before expressing the intent to incorporate collaborative assessment when undertaking practice exam questions. As our conversation concluded, Beta returned to the classroom, leaving behind thoughts on how collaborative strategies could be integrated more effectively into assessment methods at Polaris High School. While Beta's classroom provides a microcosm of collaborative testing's challenges and benefits, it is essential to understand these insights to infer their broader implications for educational practice.

Part 2: Analysis of the Collaborative Testing Entanglement

Having described the setup and initial observations in Part One, I now turn to Part Two, where I delve into the perceptions of collaborative testing from the perspective of Beta and their students. This part examines the impact of specific test formats and classroom dynamics on group intra-actions and individual performances. Deliberating on Beta's classroom environment provides valuable insights into the practical challenges and nuances of implementing collaborative testing in real-world educational settings. As I transition to a broader analysis, it is crucial to consider how the observed intra-actions and educational strategies underscore the complex dynamics of collaborative testing while highlighting the entanglement between phenomena.

Echoing experiences from Halo's class, numerous students recognised the benefits of this dynamic intra-action, noting that it confirmed and expanded their knowledge through exposure to diverse perspectives. For example, students commented, **'learn while testing**,' '**I was able to, like get ideas that I would not have thought about**,' and '**I liked hearing other people's ideas and confirming my own ideas. And this allowed me to see another perspective on these types of questions**.' These reflections underscore the significance of multiple perspectives in deepening students' understanding, a perspective that aligns closely with Halo's observations on the advantages of collaborative testing. Beta and Halo's students appreciated the opportunity to 'bounce' ideas off one another, facilitating a richer exploration of the concepts. Beta also commented positively on the process, noting:

> *I was pleased with how the students worked **together**[42], especially when they were **bouncing off ideas**. And coming up with different types of ideas and trying to solve a problem, this is great. They come up with responses that are a bit more angled than just being focused on one particular aspect*

[42] Bolded and Italicised comments represent the same remarks made by both students and Beta.

Collaborative testing enhances problem-solving and stimulates decision-making by encouraging the active retrieval of prior knowledge and insights from other group members. Critical thinking within a collaborative testing context yields more sophisticated responses than individual testing. This observation is supported by studies from Leight et al. (2012) and Mahoney and Harris-Reeves (2019), demonstrating that active collaborative discussions can deepen students' conceptual understanding. Deliberating on this, Beta noted: '*One person doesn't need to know everything. A key essence of collaboration is that if I know something, someone else knows something, so let's synthesise it*.' This viewpoint is further supported by research from Siegel et al. (2015), highlighting the cognitive growth of students, which arises from effective group intra-actions.

Despite these advantages, challenges remained. Students in Halo and Beta's classes reported that negative group dynamics often lead to conflicts, preventing some group members from articulating their thoughts and leading to disengagement. According to Reiger and Heiner (2014), students may disengage from the group or merely conform to the group's consensus instead of developing and voicing their thoughts. As one student explained, '**some people kind of take over and then some people don't get much work to do, so it's kind of leaving them out, which is an issue with working collaboratively**.' These negative group entanglements, as described by Meseke et al. (2010), undermine the effective synthesis of knowledge to clarify understanding. Such dynamics are not merely incidental but reflect deeper issues of group compatibility and communication styles, echoing group formation theories that emphasise the importance of effectively managing diverse abilities and personalities to foster productive collaboration (Johnson & Johnson, 2008; Miller et al., 2020). Group dynamics will be explored further in Part Three.

In Beta's physics class, the strategy involved deliberately changing the composition of groups for the second and third tests by mixing students of different abilities, selective or community, personalities, and genders. This approach contrasted with Halo's class, where such mixed groups occurred only in the first test. Varying levels of familiarity among peers within these diverse groupings occasionally hindered effective participation, as students were not comfortable intra-acting with each other. The complexity of physics exacerbated the students' reluctance to share ideas, particularly when unsure of their correctness. For example, one student admitted, '**sometimes I didn't contribute; I was scared to share my ideas**.' Another noted that some group members '**lacked comprehensive physics skills, affecting their performance**.' Fear of being incorrect may discourage active participation, limiting the ability to engage in critical thinking by defending, reasoning and elaborating on responses, thus constraining their potential to improve their understanding, as Mahoney and Harris-Reeves

(2019) noted. Beta pointed out the detrimental effects of a non-supportive environment, where students may feel their ideas are undervalued, leading to reduced engagement, stating:

> *Students are **not able to express their own views**. … when somebody does put in their ideas, [they may] feel like their ideas are not good. And that can also be negative on students because it'll hold other students from saying what they have on their mind*

Moreover, Beta commented, '*some students might be very shy, and because of fear of getting it wrong, they may probably not be as forthcoming with the group*.' Levine et al. (2018) suggest that the fear of being incorrect is a significant barrier in collaborative settings, particularly in academically mixed groups, such as those in Polaris High School's year 11 cohort. These groups require careful implementation of collaborative testing strategies to mitigate academic disparities and encourage an inclusive and effective learning environment.

Halo regarded collaborative activities and testing as necessary to instruction and student understanding while noting that they do not engage in 'real' collaborative work due to time constraints. Conversely, Beta acknowledges the benefits of student understanding and skills but also stresses the importance of individual expression, remarking, '*there was too much group work occurring these days*' and '*students need to express themselves individually as well, especially when you are preparing them for HSC … there should definitely be individual assessments … so students put forward their individual ideas on paper.*' The balance between collaborative and individual work presents a significant educational challenge, especially when collaborative testing is implemented within a predominately teacher-centred instructional framework (Meseke et al., 2010). Such an approach prioritises content delivery over diverse learning activities, limiting students' opportunities[43] to apply knowledge in varied contexts.

Furthermore, Breedlove et al. (2004) and Meseke et al. (2010) suggest that students accustomed to passive learning may struggle with collaborative tests and entangling with their peers due to a lack of experience and skills in communication and collaboration, which can hinder effective argumentation, leading to conflicts among peers. However, as students gain more experience with collaborative methods, their communication and collaboration skills improve, boosting confidence in

[43] Students' prior experiences in collaborative learning are an intangible factor affecting their self-efficacy and confidence in collaborative testing and impacting their perception of collaborative testing.

their abilities to discuss and solve problems with peers, as evidenced by comments like, '**my skills did improve because I got to discuss more with other people. When usually in class I don't normally get to do that**' and '**I got used to working collaboratively, with more practise**.' After participating in the three collaborative tests, Beta remarked on the increase in skills and abilities of their students:

> *This group work, particularly if they are senior students, works well as they practise more. So, the more group tasks you can give them, the more they get used to the idea. So they know what the expectation is and can get on with it. They got more* ***used to the idea of what collaboration is***

As Meseke et al.'s (2010) study affirms, increased exposure to this assessment method increases students' cognitive processing as they engage in deeper critical thinking, reasoning, and self-reflection.

Time management was a common issue for both Beta and Halo students. As the tests progressed, many groups struggled to complete the questions within the allocated time. Students' feedback underscored their frustration with the time constraints. As expressed by students, '**time limit was stressful.**' These comments on timing highlight the need to re-evaluate the time allotted for tests involving open-ended questions and complex calculations. In the physics test, the necessity to comprehend the questions, select the appropriate formulas, and solve the problems highlights why more time is essential.

Similarly to Halo's class, the limited time affected students' ability to thoroughly engage with and understand the test material. This analysis focused on the similarities and differences between Halo, Beta and their students' entanglement with collaborative testing. Both classes appreciated the active exchange of ideas, which promotes cognitive development by enhancing critical thinking and problem-solving skills. However, challenges such as group formation and time constraints also impact student's ability to engage fully with collaborative testing. As I transition to the next part, it is essential to explore the intra-actions between all components of the collaborative testing environment.

Part 3: Exploring the Material of the Classroom Environment

After examining the entanglements of collaborative testing in Part Two, Part Three shifts the focus to material and environmental conditions. Here, I explore how factors like heat and the test format shape the students' experiences and perceptions, continuing my comprehensive analysis of the collaborative testing environment.

Enhancing Critical Thinking Through Test Design

Collaborative tests two and three took place two weeks following the previous test. These tests differ from the first collaborative test as they include more calculation-intensive questions requiring students' critical thinking and problem-solving skills. Such tests are designed to enhance higher-order thinking, encouraging students to engage with their peers as they discuss and formulate responses (Levine et al., 2018). This design drew on findings by Tanujaya et al. (2017), who attest that engaging in calculation-intensive questions in mathematics education requires applying critical thinking skills, which are equally crucial in physics. The inclusion of these questions was deliberately aimed at stimulating cognitive processes where discussion acts as a catalyst for generating solutions as described by Hattie and Donoghue, (2016) in their work on higher order thinking skills. Engaging in long-response questions, whether focused on calculation or problem-solving, encourages students to utilise higher-order thinking skills, as noted by Newton et al. (2019). This method assesses a broader range of knowledge and promotes reflection through the active exchange of ideas. Mahoney and Harris-Reeves (2019) also observed that collaborative testing could enhance conceptual understanding, attributing improvements in their study to higher-order thinking questions.

This collaborative testing model can enhance problem-solving skills, a perspective supported by Reiger and Heiner (2014). In their study of an introductory undergraduate physics class, they posit that collaborative testing improves students' problem-solving abilities from exposure to extended questions and the diverse perspectives of group members. A student commented, '**I was able to, like, get ideas that I would not have thought about**,' and '**it allowed me to see another perspective on these types of questions**.' These comments underscore the value of multiple perspectives in enhancing students' problem-solving strategies, in line with research undertaken by Loyens et al. (2016) and Siegel et al. (2015). Nonetheless, there must be positive interdependence and accountability for collaborative testing to be effective.

This section conveyed the critical role of test question structure in a physics classroom, emphasising the use of calculation-intensive questions to promote critical thinking and group collaboration. This testing method enhanced problem-solving and higher-order thinking skills and exposed students to diverse perspectives through extended problem-solving questions. Student reflections further underscore the benefits of these style of test questions, with many noting how diverse viewpoints both enriched and confirmed their understanding. Beyond the design and execution of test formats, environmental factors such as classroom conditions also play a crucial role. Temperature, as demonstrated during the second collaborative test, is a particularly impactful element affecting student performance and engagement.

Temperature Effects on Student Understanding

On the second collaborative test day in Beta's class, high temperatures prompted the school to restrict outdoor activities. The heat was especially problematic in Beta's older classroom, which lacked air conditioning, creating an environment that was physically and mentally taxing. [44] One student, frustrated by the environmental conditions, disengaged entirely, while others showed signs of fatigue and distraction as the test progressed. Such observations align with research highlighting the adverse effects of heat on cognitive performance (Park et al., 2021; Roach & Whitney, 2022). Excessive heat has been shown to increase sleepiness, reduce attention, and impair memory and verbal learning, all of which critical for effective collaboration during assessments (Cho, 2017: Wargocki et al., 2019). Heat stress not only affects cognitive abilities but also shapes students' emotional responses to the testing environment, likely influencing their perceptions of collaborative testing as a negative experience, diminishing their willingness to engage fully. Roach and Whitney (2022) emphasise the role of extreme temperatures in creating barriers to learning, with studies showing measurable declines in student performance on standardised tests during periods of heat. In collaborative testing contexts, where intra-action and sustained focus are essential, such conditions can significantly hinder group dynamics and outcomes. This highlights the importance of addressing environmental factors when designing equitable leaning environments. Verlie and Blom's (2022) concept of "atmos-fears" (p. 619) provides a framework for understanding how nonhuman elements like heat become active participants in the learning environment. In Beta's classroom, the heat not only disrupted cognitive processes but also generated a collective emotional atmosphere of discomfort

[44] Excessive heat in the classroom is an essential material factor affecting students' concentration and tiredness during collaborative testing. These effects can alter students' perceptions and, ultimately, their performance on the test.

and unease, further complicating collaborative efforts. This dynamic underscores the importance of considering both human and nonhuman factors in designing effective learning environments.

The physical discomfort in Beta's classroom also highlights broader issues of resource disparity. While one classroom in the school's science block were air-conditioned, Beta's was not, reflecting inequalities in resource allocation. These inequities exacerbate the challenges posed by climate change, disproportionately affecting students in under-resourced settings (Park et al., 2021). Without climate-adaptive infrastructure, the potential benefits of collaborative testing may be undermined, as the material conditions of the classroom become a barrier to equitable learning. While Beta's classroom exemplifies the immediate effects of heat on collaborative testing, these challenges are part of a broader pattern observed globally, as climate change increasingly disrupts educational systems.

Beyond Beta's classroom, climate change is increasingly disrupting education systems worldwide. Extreme weather events, including heatwaves, wildfires, floods, and cold snaps, are causing widespread school closures and long-term academic setbacks (Marin et al., 2024). Students today are projected to experience significantly more extreme weather events over their lifetimes than previous generations, intensifying the challenges faced by educators and learners alike.

Heat remains one of the most significant threats to student learning. Research demonstrates that a 1°F increase in average maximum temperatures corresponds to a measurable decline in standardized test performance, with particularly severe effects observed during days exceeding 37.8°C (Roach & Whitney, 2022). Schools without adequate climate adaptation measures, such as air conditioning, are disproportionately impacted, further widening educational inequities (Groppo & Kraehnert, 2017; Park et al., 2021). In collaborative testing, where sustained focus and group interaction are critical, extreme heat exacerbates cognitive fatigue, disrupts group dynamics, and shapes negative perceptions of the assessment method (Wargocki et al., 2019; Roach & Whitney, 2022).

Cold weather, though less frequently discussed, also poses significant challenges. Cold snaps and winter storms can damage infrastructure, cause power outages, and force school closures, disrupting learning continuity (Hyndman & Button, 2023). Johnston et al. (2021) found that unexpected cold days in typically warm regions, such as Australia, resulted in substantial declines in test scores, highlighting the regional variability in climate impacts. Both extreme heat and cold reveal

the intricate entanglement of environmental factors with educational practices, emphasising the need for climate-conscious policies to support innovative assessment methods like collaborative testing.

Verlie and Blom's (2022) analysis of Australia's "Black Summer" bushfires provides a striking example of how the human and nonhuman elements of climate change interact to disrupt education. Thick smoke infiltrated classrooms, creating atmospheres of discomfort and fear that profoundly shaped children's experiences and learning processes. As shown in Figure 19, the smoke enveloping schools during Australia's Black Summer illustrates the tangible and affective disruptions caused by climate change. This physical intrusion into the classroom exemplifies how nonhuman forces shape learning environments and affect student experiences.

Figure 19. *Smoke Encroachment at a Western Sydney School*

Note. This photo illustrates the smoke effects on one school in Western Sydney in 2019, where the temperature reached 41 degrees Celsius. From *Education in a Changing Climate: Reconceptualising School and Classroom Climate through the Fiery Atmos-Fears of Australia's Black Summer* by B. Verlie & A. Blom, 2022 (p. 626). Reprinted with permission from Dr. B. Verlie.

These "atmos-fears" highlight how climate change is not just an external threat but as an active force reshaping educational spaces. In collaborative testing, such disruptions complicate the already delicate balance of group dynamics and learning objectives, emphasising the need for adaptive strategies that consider the entanglement of human and nonhuman forces.

Addressing these challenges requires systemic investment in climate-resilient educational infrastructure. Wargocki et al. (2019) demonstrate that reducing classroom temperatures significantly improves cognitive performance, while Park et al. (2021) advocate for equitable resource distribution to ensure all students benefit from climate adaptation measures. By acknowledging and addressing

the impacts of climate change on both individual classrooms and broader educational systems, educators can create more supportive and equitable environments for innovative practices like collaborative testing.

In Part 3 of the analysis, I examined how the structure of tests and classroom temperature influence student performance in collaborative tests. I discussed the strategic design of the test questions to enhance critical thinking and problem-solving skills. Additionally, I addressed the impact of temperature on students' cognitive abilities, underscoring the need for climate-controlled classrooms to optimise learning outcomes. Moving to Part 4, I explore additional factors unique to Beta's classroom, such as gender imbalance and group formation, and how these intersect to influence the dynamics and outcomes of collaborative testing.

Part 4: Unique Observations and Broader Implications

In the final part of the chapter, I delve into the unique observations from Beta's classroom, focusing on the effects of group formation and gender imbalances on collaborative testing. I critically evaluate the impact of pedagogical strategies and classroom dynamics on both the process and efficacy of collaborative testing. Furthermore, I discuss the broader educational implications of these findings, especially concerning curriculum design and instructional practices.

Strategic Group Formation in Collaborative Testing

Enrolment in Physics for the HSC in NSW, and its equivalent in other states, has been declining, notably among female students in mixed-gendered schools (Lyons & Quinn, 2010). Lyons and Quinn (2010) highlight several factors contributing to this trend, including the perceived difficulty of the subject and a lack of engaging content. Additionally, they point out that a competitive and individualistic approach to content delivery further discourages female participation. This gender disparity was evident in Beta's class, which had a ratio of four female students to fourteen male students, underscoring the issue of gender inequality.

During Beta's initial test, groups were formed randomly, resulting in diverse groups and various levels of engagement and intra-actions within the groups, particularly affecting one female student. Miller et al. 's (2022) study researching gender grouping in two-stage collaborative exams asserts that while diverse groups can enhance the effects on collaboration, significant issues may arise regarding gender groupings. One group of three female students, entangled from the beginning of the test, articulated a positive experience of the testing environment. Conversely, one female was isolated in a

group with three extroverted males faced challenges in collaborating. In their study, Miller et al. (2022) posit that isolating female students in all-male groups or a male in an all-female group can adversely affect their collaborative testing experience due to discomfort and competitive group dynamics, which was evident where male students excluded the female student from the intra-action.

These classroom observations reinforce the critical role of thoughtful group composition. Research by Curşeu and Pluut (2013) and Miller et al. (2022) advocate for intentional group assignments based on gender, race, ability and personality to enhance educational outcomes. According to Cao and Porter (2017) and Nafziger et al. (2011), in their collaborative testing studies, deliberate group assignments can enhance performance levels when there is low to medium heterogeneity in the group. Conversely, significant differences in ability can diminish motivation, leading to social loafing and detracting from the effectiveness of collaborative efforts. Aligning with the observed classroom phenomena, random group assignments that did not consider gender, personality and extroversion levels often failed to create optimal learning environments. Previous research into the effectiveness of group dynamics in collaborative tasks, such as the study by Curşeu and Sari (2015, p. 426), suggests that mixed-gendered groups enhance the 'group cognitive complexity.' This complexity enables students to effectively blend their knowledge, drawing on their cognitive abilities and the social dynamics within the group.

In response to these initial challenges, Beta's subsequent tests used strategic groupings based on ability, personality, and how they intra-twined in the previous test, resulting in noticeable improvements in engagement, particularly among female students. Four female students were placed with two male students, and this change significantly enhanced the participation of a previously marginalised female student in her new mixed-gender group. This highlights the importance of carefully managing group compositions to tap into students' diverse task-related knowledge structures, relational attitudes, and expectations, factors critical to the success of collaborative testing. Curşeu et al. (2018) support this view, arguing that deliberate groupings, grounded in a comprehensive understanding of students' characteristics, can mitigate negative social dynamics and enrich the collaborative testing experience. However, Miller et al. (2022) contend that allowing female students to select their own group members, regardless of gender, creates a more comfortable and familiar group environment. However, male students may feel uncomfortable and less effective if isolated in groups of female students. Therefore, teachers must thoughtfully balance personalities and abilities in mixed groups to create more effective and inclusive learning environments.

A notable challenge emerged from the frequent group changes.[45] Students expressed discomfort, commenting, '**the group changed every time, which meant we didn't have time to get comfortable with the group**' and '**if you're collaborating with people you don't know, it may lead you not to voice your opinions**.' This lack of consistency decreases comfort and familiarity, essential for promoting effective communication and collaboration. Research by Meaders and Vega (2022) and Premo et al. (2022) suggest that a familiar and positive social environment within groups enhances student comfort levels, thereby improving engagement and skill development. Conversely, they note that negative groups can hinder cognitive processes, such as information retrieval, reduce motivation, and impair collaboration.

Furthermore, arbitrary and informal grouping can lead to competitive behaviours, relationship conflicts, and stereotyping, resulting in social loafing and a diminished view of the groups' value (Premo et al., 2022). As one student asserted, '**some people kind of take over … and some people don't get much work to do, and you get left out**.' To counter these issues, I suggest that groups be assigned based on an understanding of student's characteristics and the dynamics of previous group entanglements,[46] which can ensure effective collaborative participation. While both the literature and observations from Beta's classroom advocate for structured group assignments, considering gender, personality, and ability, implementing the ideal group assignment may present practical challenges. For instance, early in the school year, a lack of familiarity with students' personalities may hinder the formation of effective groups, as seen in Beta's class. These insights from Beta's classroom highlight the complexities of managing group dynamics in collaborative testing. Teachers must thoughtfully navigate these challenges to foster an environment that enhances learning and aligns with their students' evolving needs.

Messaging by Beta on Collaborative Testing

Beta's perspective on collaborative testing reveals a nuanced stance that intra-twines the benefits of group learning with adherence to conventional assessment methodologies. This duality not only influenced instructional choices but also shaped students' perceptions of the educational value of different learning and assessment modalities, potentially constraining the adoption of innovative educational practices within the constraints of curricular demands. Beta's approach during

[45] The composition and frequent changes of student groups impact students' communication and collaboration skills, which may impact their perceptions of collaborative testing.
[46] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, teachers, myself and environment in the classroom (Barad, 2007).

collaborative testing revealed an underlying tension. This tension aligns with Biesta's (2015b) observation that 'Students learn not only from what we say but from what we do' (p. 79). Beta frequently encouraged students to 'hurry up and finish so they can do some work' and remarked, 'okay, that's it. Finish this and go on to this,' which might have signalled a lesser emphasis on the educational value of collaborative testing. This approach contrasted with the potential of these tests to promote deep intra-actions crucial for (re)configuring and deepening students' conceptual understanding, as suggested by Meaders and Vega (2022). Ideally, a more active engagement in the collaborative process would demonstrate a clear valuation of collaboration as a critical component of learning and assessment.

During post-test discussions in the science staff room, Beta and I considered the advantages and potential of collaborative testing in creating a more integrated learning environment. Beta acknowledged collaborative testing notably supported students who struggle to articulate individual thoughts during assessments. Beta pointed out the discrepancy between students' in-class participation and their performance on solo tests, highlighting the potential benefits of collaboration testing to bridge this gap. However, Beta's behaviour during the testing sessions often displayed a scepticism towards the method's efficacy. Beta non-verbally communicates a disengagement from the collaborative testing process and a preference for traditional solitary assessments by physically distancing themselves, leaving the room or focusing on their computer. Such behaviour risks influencing students' perceptions of the legitimacy of collaborative tests, as evidenced by a student's comment that such testing 'takes time of classwork' (Hayes et al., 2020). Through this perspective, Beta's approach to the collaborative tests demonstrates the impact of teacher behaviour on student learning perceptions and a lack of willingness to engage with innovative assessment practises.

Our conversation deepened as Beta articulated a clear distinction between individual testing and what they term 'collaborative assessments.' They staunchly defended the traditional testing format: 'I'm all for collaborative learning. But testing is another thing altogether, testing is individual only.' Beta suggested the term collaborative testing might be a misnomer, advocating for collaborative assessments, especially considering their observation that 'there's too much group work and insufficient time spent on problem-solving, especially in the senior classes.' Beta further emphasised the importance of individual achievement in preparing for high-stakes tests like the HSC, asserting, 'I still believe that students need to express themselves individually, especially when you're preparing

them for HSC.[47] Klenowski and Wyatt-Smith (2012) and the 2020 evaluation by the New South Wales Education Standards Authority (NESA) of the HSC framework suggest that existing assessment methods constrain pedagogical flexibility and encourage a competitive individual performance ethos, often through external tests, thus creating significant barriers to the adoption of collaborative testing.

Beta further differentiated individual testing and collaborative assessment concepts by suggesting that while collaborative activities can benefit active discussion and peer-test-environment entanglement, formal testing remains an individual endeavour. This stance was reinforced when Beta stood at the front of the classroom, reprimanding the students for their poor test performance and instructing them to complete the problems in their workbooks individually, stating, 'there would be no collaboration between students today.' As the students dispersed to work silently on their assignment, the classroom was filled with a palpable silence. With the class settled, Beta and I relocated to the adjacent prep room to commence our conversation. Beta elaborated, 'I mean, right now, I have given them work. Because of the individual results at the moment, I'm not happy.'

Beta's dissatisfaction with the student's performance on an individual test illustrates a broader educational perspective that views assessment primarily as a measure of student success, a view deeply ingrained in academic practises and societal expectations (Cumming, 2020). The pressure from high-stakes exams compels teachers to adjust their instructional methods in response to additional pressure from parents and students to ensure optimal performance (Cheng, 2021). This pressure has led some parents to question the validity of in-class assessments, potentially undermining teachers' confidence in their evaluative methods and discouraging their willingness to employ innovative assessments, such as collaborative testing. Black et al. (2010) contend that such dynamics illustrate the challenges in integrating collaborative testing into a curriculum that predominately values individual performance metrics. Beta's approach, while acknowledging the benefits of group intra-actions during tests, reflects a preference for individual assessments. Their scepticism, evident in Beta's sayings and actions, underscores the inherent tension between innovative and traditional assessment methodologies in their educational philosophy. Beta's mixed approach to collaborative testing illustrates the broader challenges of integrating innovative educational practices within traditional assessment frameworks, highlighting the need for educational systems to evolve, accommodating more collaborative and integrative assessments that can truly reflect and enhance student learning.

[47] The Higher School Certificate (HSC) in New South Wales (NSW), Australia, is a qualification awarded to students who successfully complete senior high school level studies (Years 11 and 12), assessing them on a combination of coursework and final examinations.

Conclusion

This chapter has examined the multifaceted dynamics of collaborative testing within the context of Beta's classroom at Polaris High School. By combining entanglements and classroom observations, I have identified key factors that influence both the process and perception of collaborative testing. I discussed the physical environment and classroom setup, shaped by heat and stress, significantly influencing student engagement and performance and impacting their perceptions of tests. Furthermore, I detailed how group composition influenced by personality and gender significantly impacted group cohesion. The analysis also demonstrated that while collaborative testing offers substantial benefits in developing critical thinking and deeper understanding through peer interactions, it also presents some challenges. These include whether the teacher actively participates in and endorses collaborative methods and the structuring of effective collaborative groups when there are significant gender imbalances. These impacts highlight the complex intra-action between various factors in the implementation and effectiveness of collaborative testing.

Chapter 7: Zeta's Entanglement

While structurally similar to Halo's and Beta's cases, this chapter explores the unique setting of Zeta's classroom, involving new participants and stories. I will highlight the importance of spatial configuration, student comfort, and cognitive fatigue to perceptions of collaborative testing. The diffraction[48] of collaborative tests, Figure 19, is structured into the apparatus,[49] diffraction, and perceptions, which are uniquely observed in Zeta's classroom environment. The apparatus encompasses the material-discursive elements that influence perceptions of collaborative testing. It details the school setting, subject (Agricultural Science), year level (11), periods (2 and 4), test structure and timing of the test (30 minutes) and the physical classroom environment. The varied perceptions of students, Zeta, and the researcher emerge by analysing the diffractive pattern produced by examining student group transcripts and Zeta's reflections.

This chapter is structured into four parts:

Part 1 introduces the diffractive diagrams that depict the intra-connectedness between students, Zeta, and myself and the classroom environment. The apparatus components and agential cuts concerning what is included and excluded are detailed in the techniques section of the Research Design Chapter. To conclude Part 1, I will describe Zeta's post-test conversation regarding collaboration and collaborative testing.

Part 2 examines the effects of collaborative testing entanglements and their impact on student understanding.

Part 3 shifts focus to specific environmental factors in Zeta's case that affect students' perceptions, such as student comfort and the activation of prior knowledge

Part 4 offers unique insights into how cognitive fatigue and teacher scheduling affect students and Zeta's perceptions of collaborative testing. The chapter concludes with a summary of the analyses provided.

[48] A diffractive analysis considers the intra-action between students, Zeta, phenomena, and the researcher. It is iterative and evolving, revealing the emergence of phenomena and differences to understand the possibilities of collaborative testing (Barad, 2007).

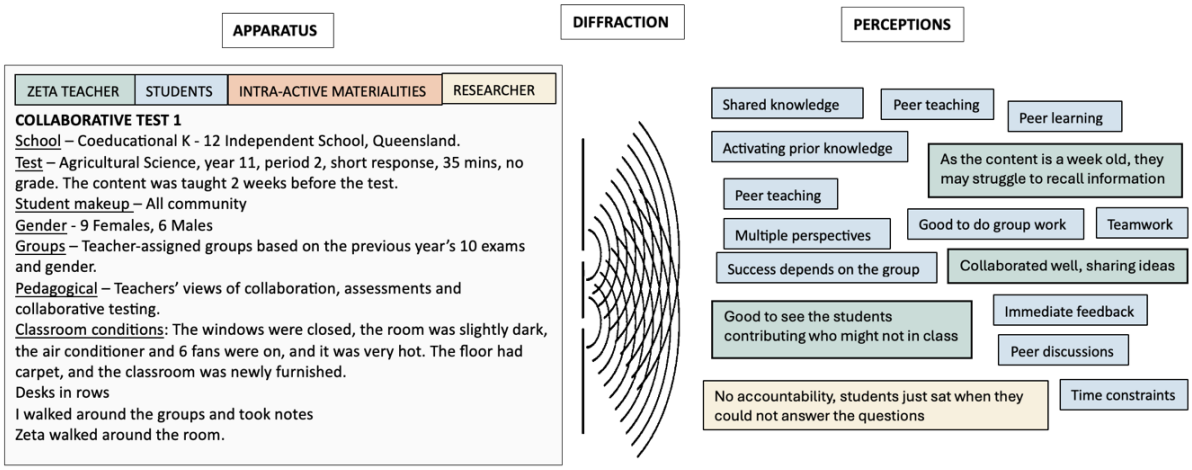
[49] The researchers' apparatus is a contextual and dynamic entity, intra-acting with the researcher, participants, phenomena to interpret their practice under study (Plauborg, 2018).

Part 1: Setting the Stage with Collaborative Tests and Mind Map

I begin Part 1 by introducing the foundational elements of collaborative testing in Zeta's classroom, detailing the diffractive diagrams that illustrate the intra-connectedness of all participants and the classroom environment. I now turn to the specific dynamics of the first collaborative test.

Collaborative Test One

Figure 20. *Diffraction of the Entanglement for Zeta, Collaborative Test One.*



Note. This figure illustrates the analysis of collaborative test one unfolds in three stages. It begins with examining the test environment, which includes physical and discursive elements such as classroom conditions and teacher opinions. It then moves to a dynamic analysis of varied perceptions from student and teacher reflections. It culminates in presenting these insights through a colour-coded scheme to highlight differences between students' and teachers' viewpoints based on the results of the diffraction analysis.

The first collaborative test occurred in mid-February, during Queensland's summer. I arrived at Ceres Independent School[50] before 9 am on a hot and cloudless day. I immediately noticed a difference between the two schools involved in the study as I walked into the school, large, well-maintained grounds, bright red brick buildings, a swimming pool, playing fields, and a new set of classrooms. Zeta greeted me enthusiastically at the office and shared their anticipation as we walked to the science block. '*I'm really looking forward to watching my students undertake the test and seeing what they would do*,' they expressed. This statement made me ponder the potential impact of collaborative testing on (re)configuring science testing practices and enhancing student understanding in Zeta's class. The science block awaited us at the back of the grounds.[51]

an older building, red brick, prep rooms in front of the classrooms, well-lit, large classrooms, new furniture

[50] This is a pseudonym for the school.

[51] The shaded box contains excerpts from my reflective journal, presented in the present tense to emphasise my role in the intra-connectedness of collaborative testing. These excerpts enhance the description of the collaborative test by illustrating my personal engagement and observations during the test.

Zeta teaches Agricultural Science during periods 2 and 4, scheduled before student breaks. Entering the classroom, I noticed its resemblance to Halo's, with new desks methodically arranged in rows, reflecting a traditional classroom setup at the centre of the room and lab benches lining the periphery. However, unlike Halo's classroom, Zeta's was more spacious, allowing students to navigate easily between the desks. Figure 21 illustrates the layout of Zeta's classroom.

It's hot outside waiting but when I walk into the room it's cool with the air conditioner on and fans silently moving around

Figure 21. *Zeta's Classroom at Ceres Independent School.*

Similar to the two other classrooms, the desks were rearranged into five groups of two tables pushed together and separated as far apart as possible. The tests and audio recorder were placed in the centre of the tables. Zeta had grouped the students for this test based on their year ten end-of-year exam results and gender. As the students filed into the classroom, Zeta greeted them enthusiastically, telling them they would do the collaborative test today. After introducing me to the class, Zeta explained the collaborative test process to the students before mentioning that this test had no grades.

I wonder about telling students before the test about no grades, will it make them off-topic, not engage, I'm concerned

The timer was set on the interactive whiteboard, and the test commenced. Mirroring the two other classes, the students were initially hesitant and uncertain about what to do. Zeta went around to each group, telling them to '*give input and to talk to each other*.' The groups consisted of two groups of two males and one female, two groups of two females and one male, and one group of three

females. In one group of two females and one male, one female student did not engage with the test or their peers and sat apart from the table. I asked Zeta about the student, and they said the student was having some '*settling-in issues*.' The effectiveness of collaborative testing significantly depends on the student's comfort[52] in their groups.

As the test proceeded, Zeta and I walked around the classroom, with Zeta commenting that it was '*good to see different students contributing where they might not have before*' before noting that some positive groups could '*share their ideas and discuss their thoughts and clarify them*.' However, one group encountered difficulties. Unable to recall the content[53] taught two weeks prior, a delay by the school's scheduling conflicts between the start of year activities and clashes with timetabling of the other teachers' testing times, they quickly became frustrated. A male in the group expressed, '**they can't think anymore**.' This incident illustrates the negative impact of the time gap between when the content was taught and tested, affecting students' ability to access their prior knowledge and engage with the test intra-action. As different group members expressed their frustration, it became clear the delay had impacted their perceptions of the collaborative test. One student tried to engage the group, saying, '**Okay, let's think about it**,' while a male student replied, '**I don't even remember learning this**,' another female in the group said, '**oh that was like a week ago**.' Meanwhile, a female student in another group mentioned she had not '**looked at this stuff in a couple of weeks**.' As Zeta and I observed the students undertaking the test, it became clear that the gap between learning the material and taking the test affected some students' ability to engage successfully with their peers and test. The delay impacted students' ability to retrieve prior knowledge.

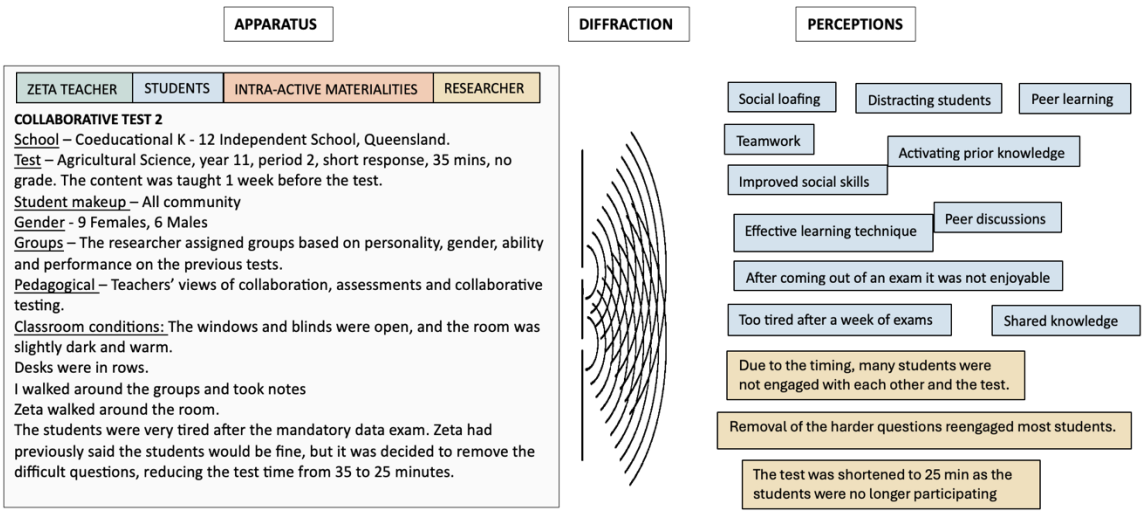
Nearing the end of the test, I wrote three reflective questions on the board, encouraging the students to share their thoughts on what they liked, disliked, and would change about the collaborative test. This was the students' opportunity to express their opinions on the first collaborative test. I collected the papers and recorders as they streamed out of the classroom, leaving me to ponder their views on this novel testing method. With the first test completed, the chapter progresses to describe subsequent tests.

[52] Students' capacity to engage effectively in collaborative testing hinges not merely on their academic prowess but equally on their ability to navigate the interpersonal landscape of group work.
[53] The extended time between the taught content and the collaborative test can impact students' perceptions of this method and increase test anxiety.

Collaborative Test Two

After completing the initial test, Zeta and I reviewed our schedules to plan the date for the subsequent two tests. Considering our time constraints and the approaching end of Term One, we decided that the students would sit the second test following their ninety-minute mandatory data test. This data test, as outlined by the Queensland Curriculum and Assessment Authority (QCAA) in 2021, constitutes 10% of the total grade. It requires students to analyse at least two data sets to discern trends and evaluate data limitations to draw informed conclusions. Furthermore, the timing of this second test coincided with the year 11 mid-semester exams, a period where students undertake multiple exams in a single day. Concerned about the potential overload, I asked Zeta if this scheduling could adversely affect the student's performance on the collaborative test. Zeta believed the students would manage well, saying, '*the students would be fine*.'

Figure 22. *Diffraction of the Entanglement for Zeta, Collaborative Test Two*



The second collaborative test occurred in early March, two weeks after the initial test. I arrived at Ceres Independent School on a cool and cloudy morning. As I walk to the classroom, I am apprehensive about the student's ability to sit the collaborative test straight after a ninety-minute data test and how this will impact their perceptions of this testing method. However, when I arrive, Zeta greets me, smiling and saying, '*I'm looking forward to what they will do*.' The classroom was warm and brightly lit, with open windows and fans whirring gently in the background. The desks were arranged in spread-out rows facing the front of the classroom for the data test. The students were told to take a five-minute break after finishing the test. The physical setup of the classroom was (re)arranged to facilitate collaboration, opting for single benches accompanied by flanked chairs instead of the paired desks used in the first test, a change driven by the need to capture quieter student contributions better.

I placed the recorders with their stars in the middle of the benches. Students re-entered the room (re)configuring the entanglement[54]as they became part of the second collaborative test. The students filed in and found their groups, which were rearranged from the first test by gender and personality and how they performed in the first test, which Zeta checked.

students groaning, I'm too tired, why

The thirty-five-minute test began with Zeta setting the timer on the interactive whiteboard. As I wandered around the groups, observing the students, I noted their initial enthusiasm for the collaborative test intra-action. However, around ten minutes into the test, some students disengaged, with one student saying, '**I'm too tired to do this**.' In contrast, two groups maintained their engagement, actively participating with each other to answer the questions, with two students in one group saying, '**it's good**' and '**I enjoy doing the group work**.'

why are some motivated

During the test, Zeta and I endeavoured to motivate the students, with Zeta actively encouraging them. Observing the students, I noticed that they were becoming more fatigued. At this stage, Zeta and I discussed the possibility of omitting the final two more challenging questions to reengage them. Ultimately, Zeta decided to shorten the test from a thirty-five-minute test to a twenty-minute test by removing the questions. While intended to alleviate stress in the students, this decision raised concerns about the student's perceptions of the value and purpose of collaborative testing.

Test length[55] and scheduling multiple tests in a single day can significantly impair students' cognitive functions, as two students expressed, '**I can't find the brain power for the exam**' and '**This was my third test today, so I didn't enjoy it**.' Moreover, facing multiple high-stakes tests can prompt students to disengage from collaborative testing due to feelings of fatigue. This disengagement was notably observed in one group, which stopped participating after thirteen minutes, turning off their recorder and discussing unrelated topics. This act of withdrawal underlines the tangible impact of cognitive fatigue on students' ability to perform collaborative tests. The timer went off, and I collected the tests as the students engaged in a student-led discussion, reflecting on their likes and dislikes of collaborative testing, particularly the new third question about the test's impact on their learning. As the students left the classroom, I collected the recorders, thanked Zeta and exited the classroom,

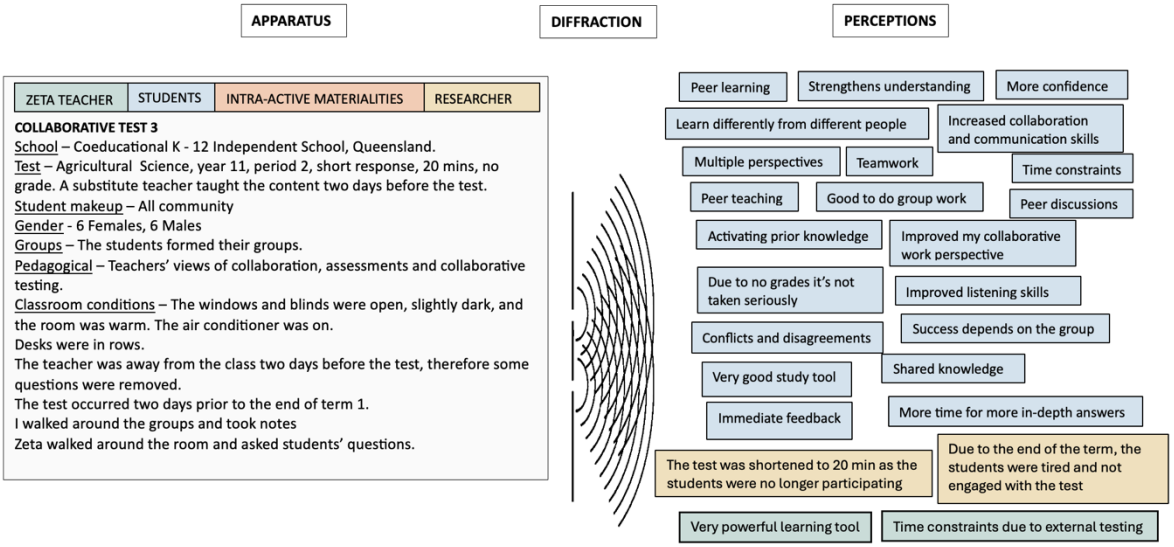
[54] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, teachers, myself and environment in the classroom (Barad, 2007).

[55] Scheduling multiple tests in one day can lead to cognitive fatigue and can alter students' perceptions and, ultimately, their test performance.

pondering the effects of test scheduling on students' ability to participate successfully in collaborative testing.

Collaborative Test Three

Figure 23. *Diffraction of the Entanglement for Zeta, Collaborative Test Three*



Two weeks after the second collaborative test, I conducted the third test. As I walked through the school grounds towards the classroom, Zeta informed me of their absence during the previous week; therefore, a substitute teacher had taught the students some of the content on the test. Consequently, we decided to exclude two questions from the test that the substitute teacher had not explicitly taught. The third test was scheduled two days before the end of the first term when three students had already departed for home. It is important to note that many students in this class are boarders[56] set to leave for their homes the following day. This context made me consider how the test timing might affect these students' overall engagement and performance and how the challenges posed by Zeta's class schedule impact students and Zeta's perceptions of collaborative testing.

As the students entered the classroom, greeting each other and Zeta, they noticed the grouped tables.

they look tired, not energetic

Zeta told the students they could form their groups today. The groups were then split along gender into four groups. At the start of the test, the students engaged with the test and each other, encouraging each other, vocalising their answers, and gesticulating when responding. However, as the test difficulty increased, signs of distraction emerged, particularly in one female and male group. Their conversations became increasingly distracted and off-topic. As the exchange between one male group

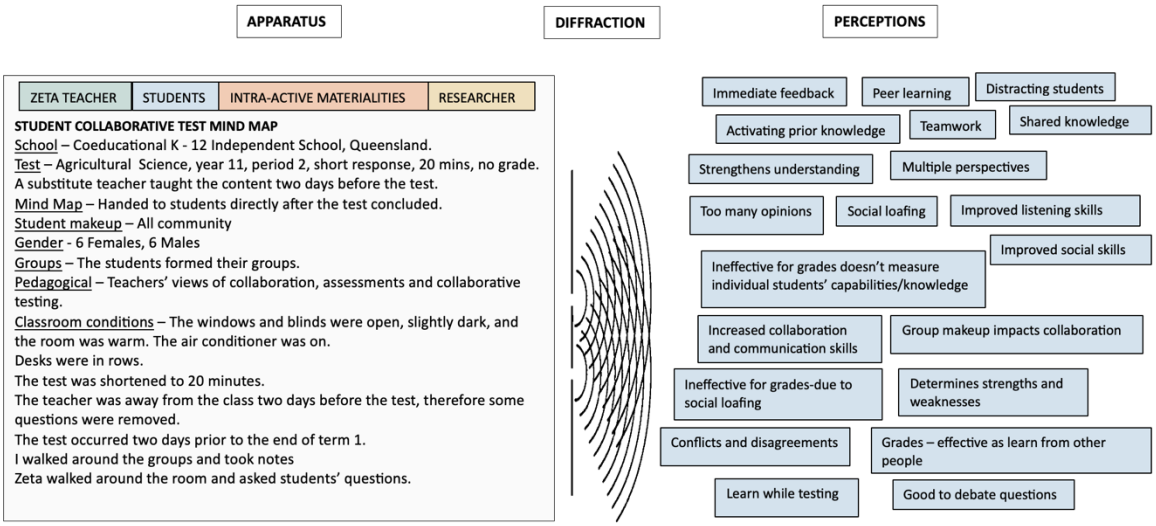
[56] Boarders at Ceres Independent School are students living on the campus during the school term.

illustrates, Male student one, '**we should be rapping that thing**?' Male two replies, '**all right start collaborating and listen**' (singing with M1). Male three sings, '**let me tell you what you already know: the world ain't sunshine and rainbows; it's mean and nasty**.'

Zeta attempted to refocus these students by circulating among the groups and emphasising the importance of the material, reminding them to '*take it seriously, especially because we've talked about what questions a lot*' The students' engagement fluctuated, with moments of focus, followed by remarks like, 'yeah, it's just the last week, miss.' At this stage, Zeta instructed the students to finish the test. As the test period ended, I prepared the classroom for the next phase: a student-led conversation. Each group's table was supplied with a set of questions and individual mind maps. Once the test officially ended, I took a moment to outline the discussion process, detailing the six questions designed to capture their insights on the test, including their likes and dislikes, how the test influenced their understanding and skills, and their thoughts on collaborative testing, especially if the test counted for grades.

Mind Map

Figure 24. *Diffraction of the Entanglement for Zeta, Student Collaborative Test Mind Map*



After the discussion concluded, students reflected on their collaborative testing experience using the mind maps.

> students are done; they are laughing and making silly comments in the recorder. I need them to be serious. What will they write on the mind maps

These mind maps addressed four critical areas: what they liked and disliked about collaborative testing, how they thought this method affected their understanding or learning, and what were some problems with collaborative testing. Students who were mainly distracted during the test were reminded to complete their reflections seriously. Once prompted, all students answered the

prompts. After collecting all the mind maps and tests, I thanked the students and left Ceres Independent School. Reflecting on this being the last test of the study, I'm eager to see what insights emerge from the diffractions.

Zeta-Researcher-Conversation-Collaboration-Environment[57]

In December 2022, Zeta and I had an in-person pre-test conversation at Ceres Independent School. Our discussion centred on their thoughts on implementing collaborative methods in their classroom. Zeta expressed enthusiasm for collaboration, stating, '*It's about bettering the whole group through the process … So I think collaboration is good. I wish I could do more of it. I feel that I probably don't do as much as I could*.' When asked why they didn't engage in collaborative activities more frequently, Zeta mentioned the pressure to cover content, saying:

> *I think, probably, it's the pressure to teach content that has been like the number one priority, um. But I still feel that sometimes I pressure myself to get content done and then look after the other stuff, which probably should be the other way around. It should be like noise in the background, student-centered first, and then the content. But yeah, I think it's probably that pressure and that fear of Oh, but what if it doesn't work*?

Our discussion then shifted to Zeta's perspective on collaborative testing. They remarked,

> *I feel like it's like, a peep into this world of like the future possibility of actually engaging kids, particularly, I think, senior kids … I think that you'd be building their skills … and you'd be enhancing their ability to have those 21st century skills while also engaging with a subject that they really enjoy. So, bringing the two together, as opposed to being like, well, it's science and boring. You know, rote learning, and you get an assignment or an exam, and that's it, versus, you know, actually bringing their contribution in terms of that collaboration space*

We concluded the conversation, and Zeta returned to class.

[57] Dashes symbolise the crucial intra-action between all elements, material or otherwise, highlighting the equal importance of both in shaping perspectives on collaborative testing (Gullion, 2018). Other elements contribute to this intra-action but are too numerous to detail individually.

The post-test conversation occurred over Zoom, as Ceres Independent School had closed for the school holidays. Zeta expanded on their thoughts about collaboration, saying, '*As a teacher*, *it's another option … it engages more kids … [while] also deepening awareness and understanding … collaboration fosters all the best parts of learning [as] kids share their knowledge and also improve it as well*.' Zeta acknowledged the challenges posed by the,

> *Dictates of the content … having to push content dictates how you can do it … I also think different student backgrounds and what they bring in, some are keen … and it's getting over the social hurdle particularly when students are new to the school or subject*

Zeta added that as they had the same students for two years, years 11 and 12, they think '*it's something the students would probably embrace as collaborative testing can build students' confidence*.' They also remarked upon the difficulty in fitting collaborative testing into the schedule as '*timing like the second test straight after an exam. The last test was like their last day of classes. So I think that those sort of external factors played into it*.' Zeta then commented there is '*So much content that revision and reflection period is often either short or not at all. So I think for them. [it] gives an opportunity for them to reflect and to refine and understand*.' As our Zoom conversation ended, I was left with rich insights into Zeta's perceptions of the challenges and opportunities of collaborative testing in school settings.

Part 2: Analysis of the Collaborative Testing Entanglement

In Part Two, I delve into the analysis of collaborative testing by exploring its impacts on student understanding and the dynamics within Zeta's classroom compared to Halo and Beta's cases. I emphasise the intra-action between phenomena and their influence on students' conceptual understanding. Through active dialogue and exchanging ideas, collaborative testing emerges as a dynamic method for enhancing conceptual understanding and developing students' 21st-century skills, such as social capabilities. Engaging in these discussions requires students to articulate their reasoning, which promotes clear and effective communication, a crucial skill for navigating complex concepts. This approach aligns with findings from Tolmie et al. (2010), who emphasise the importance of communication in learning processes. Moving beyond test experiences, I now focus on the broader educational implications and the impact of collaborative testing on student understanding.

Students in Zeta's class acknowledged that collaborative testing improved their understanding of the content. One student mentioned, '**I understood and learned a lot more with people around me, helps [my] understanding and learning**.' Additionally, another appreciated '**learning from others and hearing other opinions and interpretations of the questions**.' Another student reflected, '**adding their information to others increases our knowledge, and we get to acknowledge [our] strengths and weaknesses**.' In our post-test conversation over Zoom, Zeta reflected on the benefits of collaborative testing in enhancing student understanding. Zeta emphasised how this approach facilitated peer explanations, helping students articulate their responses and connect to past and new information. Zeta observed, '*If someone's talking about something … it also deepens awareness and understanding that way.* ***If you can hear the points of view or things in a different way, you can make those connections.***' Learning and the deepening of students' conceptual understanding are dynamic processes. Intra-acting[58] through cognitive elaboration, students adjust, clarify, and restructure their knowledge of the concept (Slavin, 1996). Zeta observed that when students were actively engaged in entanglements[59] they (re)configured their prior knowledge and made connections to new understandings (Plauborg, 2018). Zeta acknowledges that '*collaboration fosters all those best parts of learning, allowing kids to share their knowledge and improve it.*' Biggs (1996) emphasises that for students to grasp concepts truly, they

[58] Intra-actions refer to the intra-connectedness between students, Zeta, researcher, and the material classroom environment (Barad, 2007).
[59] Entanglements refer to all the elements that shape perceptions of collaborative testing. These entanglements represent the intra-connectedness between students, Zeta, myself and environment in the classroom (Barad, 2007).

must integrate action and thought, denoting that meaningful learning arises from the intra-action between doing and reflecting. However, challenges such as a content-heavy syllabus hinder teachers' ability to encourage reflection among their students, affecting students' ability to apply learned information across contexts and underscoring the need for strategies such as collaborative testing.

This section highlights how collaborative testing enhanced conceptual understanding and fostered 21st-century skills among students. Building on the cognitive benefits of collaborative testing, Zeta's students felt that collaborative testing enhanced their communication, collaborative and social skills. Two students voiced they '**learned how to work as a team**' and that '**my skills have improved because normally I would just shut up and sit at the back of the class and not contribute**.' Zeta observed that their students' collaborative skills were notably lacking, with Zeta remarking that they displayed '*pretty minimal skills, I'd say, all around. I don't think they were great collaborators at all'*. This observation points to a widespread issue in educational settings. Despite the recognised importance of collaboration as a critical 21st-century skill, Zeta observed that their students seldom engaged in meaningful collaborative work, remarking students had "little opportunity to collaborate" and that the students' unstructured group experiments were not '*a collaboration*.' The focus on high-stakes external examinations exacerbates this lack of structured collaborative practise. Binkley et al. (2012) argue that this focus on test content detracts from teaching and practising essential skills like collaboration, which are necessary for student development.

Zeta concurs, stating, '*it's not just only learning of the theory but also … a lot of opportunity for collaboration has been cut short*.' In response to this challenge, Zeta proposed that integrating collaborative elements into formative assessments could enhance communication and collaboration skills. As part of end-of-topic tests, for example, collaborative testing could motivate students to actively defend, reason and elaborate on their responses (Eastridge & Benson, 2020). This technique promotes a deeper understanding of the subject matter and enhances their communication skills, preparing them for future success. Zeta posits that collaborative testing enhances students' skills by '*improving their ability to actively speak to each other and work together as opposed to just maybe one person saying something and someone else is writing it down*.' This method is not passive. It encourages students to communicate and collaborate while working cohesively as a team. While noting the benefits, it is also crucial to address the challenges presented by collaborative testing, which are explored next.

However, adopting collaborative testing also presents challenges, particularly concerning individual accountability. As Meijer et al. (2020) assert, the lack of accountability can encourage free-riding, where some students rely too heavily on their peers due to a lack of motivation or belief that the group's output will not suffer from their minimal contribution. This sentiment is echoed in students' comments, '**If the group doesn't want to work well, there is no hope**,' '**some people are off track and distracted**,' and '**If the group doesn't work well together, you can't get the questions done**' potentially jeopardising the task at hand. Social loafing, where individuals exert less effort in a group than they would individually, further complicates group dynamics. In Zeta's class, as opposed to Beta and Halo, students commented on the lack of seriousness by group members as the reason for not getting the full benefits from collaborative tests, as exemplified by one student, '**If it was a more serious thing, everyone would contribute a lot more**.' According to Hattie and Donoghue (2016), such negative dynamics brought about by disagreements and students' off-task behaviour can impact students' understanding through a decline in cognitive functions. In light of these insights, it is evident that collaborative testing has the potential to enrich students' learning experiences by deepening their understanding and engagement. However, its effectiveness is intricately linked to the group's dynamics and the overall emotional and cognitive atmosphere in the classroom. To maximise the benefits of collaborative testing, it is important to address these challenges by fostering an environment that encourages active participation and constructive intra-actions among students. Additionally, students need to be held individually accountable during the test to offset concerns of social loafing.

Adopting collaborative testing requires careful integration within existing frameworks, particularly when considering assessment strategies. The lack of accountability in these collaborative tests in Zeta and Beta's classes impacted how students engage in collaborative testing. In practical applications, different techniques have been used to hold students accountable while undertaking collaborative testing on undergraduate students. Siegel et al. (2015) conducted a collaborative test followed by an individual test in their undergraduate biotechnology course. Students could write their own answers to offset group disagreements, and marks were given on the depth of the written responses. Mahoney and Harris-Reeves (2019) and Newton et al. (2019) implemented a split of 80% for the individual test and 20% for the collaborative test, providing an incentive for active participation. Leight et al. (2012) in their first-year Biology course, added percentage points to the individual score based on group performance: 5% for scores between 90–100%, 3% for 80–90%, and 1% for 70–79% Reiger and Reiger's (2020) analysis of studies utilising collaborative testing, note that for fairness, if high-performing individuals scored lower in the collaborative section, the group grade should not be

used, and only the individual grade will be assigned, notably to support students who are anxious about or are opposed to collaborative testing.

In evaluating the effectiveness of collaborative testing in Zeta's class, it is necessary to consider the changed assessment frameworks that are in place in Queensland and the impact on their QCE.[60] Previously, in Queensland, in year 12, the internal assessments were school-derived and census-moderated between schools, along with agreed-upon standards. However, in 2019, the Queensland government brought in a mix of school-based assessments and external tests, which resembled the Higher School Certificate in NSW, focusing on standardising grades (Cumming, 2020). Zeta feels the need to ensure the content is taught and how grading can impact the students' and schools' grades. In their class, there were no grades assigned, hence no accountability, due in part, as Zeta mentioned in a post-test reflection, that they felt they could not assign grades to these tests due to,

> *In Queensland, obviously can't speak for other states, but the value and emphasis placed on external exams … [now] they do an experiment and research investigation. In the old system, the school could do at least two, if not three [internal assessments], to go towards their final grade throughout year 12*

The hesitancy to embrace collaborative testing by assigning grades, despite Zeta and the students recognising improvements in their comprehension of the subject matter and their teamwork, communication, and social skills, can be traced back to the prevailing belief that individual testing is the sole indicator of academic achievement, as reflected in this comment:

> **I don't think test counting for grades is a very good idea because it really depends on who you have in your group, which can affect your grades. So, it won't really be fair marking; that's why the individual tests are fairer, as they test your knowledge and skills**

The concerns raised by this and other students in Zeta's class are reflected in the concerns of students in Halo and Beta's classes; the lack of contribution, group members being off task, and ineffective group dynamics all contributed to some students' negative perceptions of collaborative

[60] The QCE is similar to the Higher School Certificate in New South Wales, it provides evidence of senior schooling achievements generally at the end of year 12 (Queensland Curriculum and Assessment Authority, 2023).

testing, primarily when related to the assignment of individual grades. Conversely, another student remarked on the advantages of collaborative testing, noting that diverse knowledge within the group can lead to better overall results, as she stated:

> **I think it would have been a lot better. Like everybody knows different things … Some people might be really good at one part of the test and bad at the other, but if you all work together, then you get the best results. I think everyone's grades will be better**

Group composition is an essential factor influencing both perceptions and efficacy of collaborative testing. Research by Silm et al. (2020) suggests that the combination of group composition alongside students' emotional and cognitive states can impact collaborative testing perceptions. Careful design and management of collaborative work in the classroom can offset negative perceptions associated with collaborative testing. Teachers can provide essential guidance and support by implementing structured collaborative tasks featuring explicit criteria for skills development, such as rubrics or reflective practise.

In Zeta's class, thoughtful consideration was given to group compositions in the first two tests, with Zeta and I assigning groups based on specific criteria. However, for the third test, students were allowed to choose their groups, which introduced different dynamics and engagement levels. Generally, female students were more on task than their male counterparts. Curşeu et al. (2018) support the observation that gender can influence interpersonal relations and engagement in educational tasks. The observed lower engagement among some male students could be attributed to reduced motivation and the perception that the outcomes of these tests were inconsequential. Furthermore, it was noted in Zeta's class and the two other cases studied that female students valued the opportunity to engage with their peers and discuss the questions, positively influencing their perceptions of collaborative testing. This insight demonstrates the importance of carefully considering group formation and actively fostering motivation among all students to enhance the effectiveness and inclusive of collaborative testing as an assessment method.

This section examined the implementation and effects of collaborative testing in Zeta's classroom, distinguishing the positive impacts on student understanding and 21st-century skills, as seen in Chapters 5 and 6. This section also addressed the challenges of implementing collaborative testing, such as the lack of individual accountability, the potential for social loafing, and the hesitancy

to integrate collaborative testing into grading systems due to concerns over fairness. While collaborative testing can enhance learning and skill development, its success depends on effective group dynamics, proper structuring of the testing process, and addressing concerns regarding individual contributions and accountability. Having established how collaborative testing affects students' understanding and entanglements, Part 3 will shift focus to environmental factors that play a crucial role in shaping these perceptions.

Part 3: Exploring the Material of the Classroom Environment

In Part 3, I shift to the factors that influence the efficacy of collaborative testing, such as student comfort and the activation of prior knowledge.

Student Comfort

Student comfort is pivotal in educational settings, impacting individual and collaborative testing experiences. This section explores how a supportive social environment enhances student engagement and increases communication and collaborative skills while examining the negative consequences of poor group dynamics. Alongside Beta and Halo, Zeta's students felt that a positive social environment fosters a sense of connectedness crucial for effective collaboration. Students felt that collaborative testing improved their communication and teamwork skills, with one noting, '**I learned how to work as a team**.' This sentiment is supported by Meaders and Vega (2022), who argue that such environments increase comfort, thereby enhancing engagement and developing critical interpersonal skills. Furthermore, positive group entanglements also contribute to significant skill development, as students reported that working collaboratively allowed them to leverage their peer's strengths, facilitating learning and skill application in real-time. One student commented, '**You get to work with friends and share stuff that you might know**,' and another student found increased comfort from these three tests, '**have worked with just about everyone, now I feel comfortable**.' Conversely, negative group dynamics can severely hinder these benefits. Negative emotions or learning environments can lead to decreased engagement and motivation, as group members may withdraw from actively contributing to the discussions. This was evident when one student expressed reluctance to work with a particular group, '**I didn't like my group much**' before telling Zeta, '**I don't want to be with that group**.' Premo et al. (2022) support this, commenting that such barriers to communication and collaboration can hinder the development of essential skills.

The literature underscores the importance of interpersonal skills such as effective communication and negotiation in managing group dynamics and achieving successful collaborative outcomes. Meseke et al. (2010) and Slavin (1996) underline the critical role of these skills, while Siegel et al. (2015) and Tolmie et al. (2010) further point out that positive group experiences can bolster skills like elaboration and self-reflection. Black and Allen (2018) recommend that interpersonal skills be explicitly taught and modelled before such activities to optimise collaborative testing, ensuring all students can contribute positively. Expanding on this, Loyens and Gijbels (2008) and Tran (2013) emphasise that comfort in groups not only improves interpersonal intra-actions but also enhances the application of metacognitive strategies like questioning and argumentation, which are crucial for knowledge construction. A supportive group environment and explicitly teaching critical interpersonal skills are essential for effective collaborative testing, leading to improved academic performance and collaborative competencies. After discussing the influences of psychological and cognitive factors, I shift to examine specific incidents that underscore the importance of these elements.

Activating Prior Knowledge

Activating prior knowledge is crucial for effective collaborative testing, as evidenced by the challenges and successes observed in Zeta's classroom. Observations from both Zeta and myself during testing revealed that the gap between learning phases and subsequent tests reduces student engagement, making it harder for students to retrieve previously learned material. Rice et al. (2015) highlight this issue, noting that students often perceive questions as too difficult when overwhelmed. A typical student complaint, '**I didn't like some of the questions, couldn't answer them**,' reflects this sentiment, eroding confidence, inducing negative emotions, and increasing test anxiety, further affecting performance. Research by Breedlove et al. (2004), Gilley and Clarkston (2014), and Krispenz & Dickhäuser (2018) demonstrate that test-related stress impairs cognitive and physiological functions, causing panic, dizziness, and an inability to retrieve or organise information. This was particularly clear when students failed to recall concepts taught two weeks prior without interim reviews. A student's admission, '**I don't even know what this is**,' highlights how failure to activate prior knowledge complicates answering questions and may reduce motivation to proceed with the test. Wininger et al. (2019) argue that disengagement and lack of motivation can lead to encoding failures, where information is not adequately stored in memory. However, the 'forward testing effect,' as Yang et al. (2021) noted in their meta-analysis, suggests that acquiring new information during testing can enhance understanding through the activation, transfer, and retention of knowledge.

Additionally, collaborative testing encourages the activation of prior knowledge as students engage in peer-to-peer problem-solving and teaching, which enhances comprehension and recall. This is supported by findings from Levine et al. (2018), Meseke et al. (2010) and Nafziger et al. (2011). Many students across the three classes confirmed that collaborative testing improved their recall and deepened their learning, with one noting, '**I felt like I learned more in this test**.' Nonetheless, Yang et al. (2021) report inconsistent findings on the effectiveness of collaborative testing. Efu (2019) attributes this to variations in the amount of time spent engaged in constructive collaborative learning activities. These factors underscore the complex entanglement between testing formats, student engagement, and educational outcomes, suggesting that the timing of reviews and the testing environment significantly influence learning efficacy.

This section has examined how delays between learning and testing can impair students' ability to recall information, leading to increased test anxiety and diminished performance. The section also highlighted the potential for collaborative testing to enhance recall and understanding, although its effectiveness can vary depending on the nature of the collaborative activities involved. After considering the subtle yet significant effects of the classroom environment, in Part 4, I will synthesise observations related to cognitive fatigue and scheduling challenges specific to Zeta's classroom context.

Part 4: Unique Observations and Broader Implications

In Part 4, I synthesise observations unique to Zeta's case, focusing mainly on how cognitive fatigue and scheduling impact collaborative testing.

Cognitive Fatigue

Cognitive fatigue significantly impairs students' reasoning functions, negatively impacting their perceptions of collaborative testing. Cognitive Fatigue (CF) refers to the reduced processing ability caused by prolonged cognitive activity, affecting attention, problem-solving, and decision-making, thus decreasing academic performance (Earl et al., 2023). This section examines how repeated testing exacerbates cognitive fatigue, diminishing the benefits of collaborative testing. According to Earl et al. (2023), cognitive fatigue increases with heightened academic demands, social comparisons, and excessive emphasis on grades, which overloads students' working memory and impairs concentration and focus. Furthermore, students with lower class engagement are more prone to fatigue over the term. At the same time, those who are cognitively engaged exhibit less fatigue, stronger intra-

connectedness with their peers, better self-regulation, and are proactive in their learning (Earl et al., 2023; Linnenbrink-Garcia et al., 2016). The timing of tests can also impact students' ability to engage with the materials, their peers and the tests. Regarding Zeta's class, the timing of the tests affected their students' ability to deepen their understanding of the test content. Zeta noted, '*I think also with the timing like the second test was straight after an exam. The last test was their last day of classes. So I think that those sort of those external factors played into it*.' Tests scheduled immediately after another exam or on the last day of classes may increase student cognitive fatigue, negatively impacting group intra-actions.

This lack of participation and enjoyment in learning situations further impacts students' motivation to persist, reducing their academic achievement in collaborative testing. A student stated, '**they did not like anything about collaborative testing**.' The relationship between test timing and student motivation is complex. Factors like the time and day, the classroom environment, and the test format significantly influence student motivation (Linnenbrink-Garcia et al., 2016). High-stakes tests naturally elicit motivation due to their consequences, whereas low-stakes tests often fail to engage students unless perceived as directly beneficial (Simzar et al., 2015). This variance in motivation was observable in Zeta's class, mirroring broader educational research that suggests students approach tests with differing attitudes and motivations, influencing their behaviour and engagement levels during collaborative testing.

Furthermore, the testing environment and students' backgrounds are crucial in developing active learning through collaborative testing, where students (re)construct knowledge and apply it in various contexts. Black and Allen (2018) and Meseke et al. (2010) highlight how varying levels of motivation, attitudes and social intra-actions can lead to negative dynamics when the classroom environment is not conducive to learning. For instance, after a particularly data-intensive test, one group in Zeta's class chose not to continue, turning off the recorder and not attempting any further questions. However, three other groups persisted, reflecting Eklöf's (2010) suggestion that students' learning goals influence their engagement; some focus on mastery and understanding, while others are motivated by competence and comparison. To address differences in motivation, it is essential to implement interventions to mitigate cognitive fatigue, such as reducing the number of tests in one day, scheduling tests in the mornings, and incorporating breaks and less cognitively demanding activities. Additionally, developing explicit strategies to enhance students' metacognitive abilities and cognitive engagement can further improve their experience and outcomes in collaborative testing.

Modelling Collaborative Testing

Modelling collaborative testing techniques provides a practical framework for understanding and implementing these methods effectively in the classroom. Researchers can demonstrate the reasoning, benefits, and concerns through modelling before explaining the process. Specifically, as Moore and Bell (2019) detail, instructor modelling involves educators demonstrating pedagogical approaches, vocabulary, or practices directly in the classroom, facilitating contextualised learning for teachers. Tondeur et al. (2012) emphasise the importance of authentic and scaffolded experiences. Live modelling in real classroom situations allows teachers to observe collaborative testing within their situated classroom. This approach shifts the gap from a theoretical view of collaborative testing to a practice-based view. In Zeta's classroom, I demonstrated how to construct groups, manage the classroom, and conduct collaborative tests. The practical impact of this approach is exemplified by the experience of Zeta, who observed collaborative testing in their classroom through explicit modelling on how to structure groups, manage the classroom, and conduct the tests. Zeta remarked, '*I would do it for a whole host of reasons, but definitely because it was modelled, and I was able to see it. It wasn't that hard, and it had benefits, and I can see it working.*' This hands-on experience helped Zeta adapt their pedagogy to utilise collaborative testing as a tool for deeper content understanding and enhanced social skills among students. Zeta expressed their desire to implement this approach more broadly, commenting that:

> *I would love to see it go over, say over a whole unit, not necessarily just one topic or one kind of thing; I'd love to see that whole start to finish and be able to reflect on refining that process. But that's also really hard*

Collaborative testing modelling has been used as an agent of change, improving Zeta's pedagogical skills and practices. However, despite recognising the value of collaborative testing, Zeta and Halo also pointed out the practical challenges of implementing collaborative techniques in their classrooms. With Zeta noting:

> *It's not really valued as much as we should … I think it speaks to a larger issue in Australia … it's not only the learning of the theory but also the skill in collaboration, but I'm just not sure this system allows it … it's quite restrictive*

This reflection underscores the tension between the potential benefits of collaborative testing and the restrictive nature of Australia's educational system, where standardised curricula and accountability measures are overseen, limiting opportunities to develop collaborative skills (Ball, 2016, 2017; Hickey & Riddle, 2023). As a result, many teachers adhere to established methods rather than exploring innovative methods like collaborative testing.

This section discussed the benefits of modelling pedagogical techniques through live demonstrations to enhance teachers' adoption of new methods. Zeta's reflections reveal how practical modelling can bring about change and develop new knowledge and practices. However, the restrictive nature of educational systems continues to limit the broader application of collaborative testing. Having explored these unique elements, I conclude the chapter by summarising the broad implications of these findings and emphasising the need for tailored approaches in implementing collaborative testing.

Conclusion

In this chapter, I have explored the multifaceted nature of collaborative testing within Zeta's classroom and highlighted critical insights from this class. By examining the complex entanglement of the testing environment's elements, I have gained deeper insights into how they collectively influence students' and Zeta's perceptions and the dynamics of the testing environment. The discussion reveals that the physical layout of the classroom, the timing of the tests, and the emotional and cognitive state of students significantly affect their ability to participate and perform in collaborative testing. It also underlines the critical role of student comfort and the activation of prior knowledge in shaping the testing experience, pointing out that these factors can either enhance or hinder the learning process. I discussed Zeta's desire to implement more collaborative practices but felt constrained by the curriculum's content-heavy nature and the emphasis on individual testing. These insights underline the need for thoughtful implementation and support to maximise the benefits of collaborative testing in educational settings.

Chapter 8: The Intra-Actions of Collaborative Testing

Introduction

In this chapter, I present a cross-case analysis of collaborative testing across three diverse classes, where each class defines a bounded case (Meriam & Tisdell, 2015). The analysis highlights recurring patterns of intra-actions, revealing how students and teachers navigated the complexities of collaborative testing while underscoring its challenges and transformative potential.

A key finding is the quality of active peer discussions in building students' argumentation skills and understanding. Through a fine-grained analysis, I specify the elements that promote peer discussions during collaborative testing, including group dynamics, the impact of positive and negative intra-actions, and the influence of gender imbalance on student perceptions. Additionally, 21st-century skills, such as communication and collaboration, are examined, alongside the importance of ensuring prior skill development before implementing collaborative testing.

The intra-action of human and material elements plays a critical role in fostering an environment conducive to collaborative testing and peer discussions. This chapter highlights how high-stakes testing affects the implementation and perceptions of collaborative testing among students and teachers. Practical guidance is also provided to help teachers implement collaborative testing strategies effectively, ensuring they are equipped with the necessary tools for success. The chapter synthesises the key elements influencing student and teacher perceptions during collaborative testing.

Cross-Case Analysis

In the detailed cross-case analysis of Halo, Beta, and Zeta's classes, I observed recurring patterns of intra-actions that revealed comparability's between students and teachers across all three classes. I began by reviewing the diffractive diagrams for each class, Halo, Beta, and Zeta, as detailed in Chapters 5-7, paying close attention to the intra-actions between students and teachers during each collaborative test to discern 'what comes to matter' (Hayes et al., 2020, p.372). This examination was enriched by a variety of data sources, including student and teacher reflections, student discourse during collaborative tests, mind maps, and my own observations and journal notes, all of which provided the contextualised analysis of the specific conditions of each class, necessary for a more

thorough understanding. As stated in the Research Design, Chapter 4, bold text was used to highlight student voices and italics for teachers to ensure they are 'heard' and not hidden in a sea of text.

Moving through each case, I began identifying patterns, noting key commonalities and differences in the diffractive patterns. Peer discussions and argumentation emerged as central pattern, although each class displayed unique variations in how these dynamics played out. To bring structure to these findings, I organised them into categories, such as classroom environment, group composition, and test length, which helped trace how different practices unfolded in each class. This process was cumulative, with each test adding another layer of analysis, allowing the diffractive diagrams to evolve as the study progressed. Once all data was collected and the analysis completed, the final diffractive diagrams were created, visually representing the intra-actions that shaped each class. These diagrams, alongside the broader cross-case analysis, revealed key emergent patterns, illustrating the significance of collaborative testing in encouraging peer discussion and argumentation across diverse classes.

Peer Discussions and Argumentation in Collaborative Testing

Students initially experienced uncertainty, primarily due to the unfamiliarity of working within new group dynamics. Still, as they engaged in collaborative testing, it became evident that peer discussions were crucial in deepening their conceptual understanding. In test one, students from Halo and Beta's classes commented, '**We don't know each other well**' and '**I would prefer it to be with people I know**.' Despite this awkwardness, a common thread emerged: students and teachers noted that peer discussions enhanced their conceptual understanding. To encourage peer discussion, test questions need to be written with higher-order thinking verbs, as seen in Appendix D. A common assertion by the students across the three classes was: '**the combined effort allowed us to pool our ideas together and create the best response**,' '**brought insight into the stuff I didn't know I needed to know**.' In contrast**,** one student noted, '**I feel like I learnt more in this test**.' As Nguyen et al. (2020) discussed in their research, active peer discussions encourage knowledge construction as students debated answers, gaining different perspectives, which helped build their knowledge base. The gaining of different perspectives was acknowledged across all three classes as aiding in conceptual understanding, as students commented, '**improved my concepts, while listening to other people talk**,' and '**helped me to understand other people's ways of solving problems**.' The active nature of collaborative testing encouraged students to engage with each other and their environment, aligning with the principles of a diffractive analysis, where the intra-action of students, material and environmental conditions become central to knowledge construction (Murris, 2022).

As the collaborative testing sessions progressed, I observed a growing comfort and proficiency among the students, who reported increased communication and collaborative skills. The impact of this practice was particularly notable among students, who reported greater clarity in expressing their ideas and an improved ability to listen and respond constructively to others. Many students in Beta's, Zeta's and Halo's classes made comments like, '**I got used to it over the three tests**,' '**improved teamwork**,' '**I work better with people now**,' and '**my social skills have improved, I can talk to people now**.' In our post-test conversation, Zeta noted that collaborative testing '*socially and in terms of thinking about 21st-century skills, massively beneficial, they had to talk to their peers*.' Halo and Beta remarked upon this in the second test: '*the active listening between each of the people in the group had improved*' and '*pleased to see the way the students worked together*.' The benefits of improved communication and collaborative skill development in positive peer group intra-actions are reflected in the literature on collaborative testing. Collaborative testing studies with tertiary students conducted by Levine et al. (2018), Morreale et al. (2017), and Siegel et al. (2015) note improved communication skills as students effectively listened, negotiated solutions and communicated their thoughts. Teachers' observations across the three cases in this study reported improvements in student engagement and the quality of discussions. Halo observed that students appeared to '*share the work, and no one was dominating*,' highlighting a clear benefit of repeated exposure to collaborative testing. Beta, however, pointed out a challenge where some students may not be able to express their ideas fully or might feel their ideas were not valued. In line with findings from existing studies on the benefits of active learning (Biggs 1996, 1999; Hattie & Donoghue, 2016), repeated exposure to collaborative testing increased student comfort and participation and deepened engagement with the content.

However, there were a few persistent concerns within and across groups. Teachers and students noted challenges such as dominant personalities and social loafing. These issues often led to an incomplete representation of the group's knowledge, as discussed in Chapter 2, p.33. Many students in Zeta's class discussed the effect of poor group dynamics, '**they don't listen**,' '**some were too headstrong**,' and '**if the group doesn't want to work, well, there's no hope**.' Zeta also noted the students' lack of engagement, especially in the third test, '*Guys, please, this is serious, you're going to get these style of questions when your unit 1 and 2 exams*.' The concerns raised by students in Zeta's class reflected the lack of seriousness evident in the last two tests due to the student's cognitive fatigue, which was not apparent in the other two classes. Social loafing occurred in both the first and second of Halo's tests. Halo's third high-stakes test demonstrated that all students actively participated in discussions without any reports of social loafing or time constraints, and most students found the

test rewarding. The analysis underscored the need for accountability and sufficient time to ensure active participation and deeper conceptual understanding. The difference was the lack of accountability in all other tests, which led to the social loafing seen by some students. McKay and Sridharan (2024) note that for collaborative work to be effective and mitigate social loafing, individual and group accountability must be addressed by rubrics or peer assessment. Additionally, Strom et al. (2018) assert that negative group dynamics may stem from students' lack of prior experience in collaborative work, which affects their sense of personal responsibility and ability to communicate when discussing problems effectively. Adjusting the group composition after the first tests, based on students' personalities and abilities, gender, and how the students intra-twined with their group, helped mitigate some of the issues, highlighting the need for careful group management.

Implementing collaborative testing within content-heavy curricula presents significant challenges. Teachers expressed constraints on their capacity to allocate time to innovative assessment methods, aligning with research in the United States (Acosta et al., 2020) and Australia (Appel, 2020) highlighting the pressures of preparing for high-stakes examinations such as the Higher School Certificate (HSC) and Queensland Certificate of Education (QCE). Halo noted, '*limitation is finding time to give students this big discussion… while also there's this explicit instruction and deliverables, you've got to show data for this that and the other*.' These concerns were also seen with Beta and Zeta, remarking that '*the focus is on the HSC, preparing them for that*' and '*the value and emphasis placed on external exams for the purpose of teaching … opportunity for collaboration has been cut short*.'

The variations in teacher engagement and attitudes towards collaborative testing influenced student perceptions and engagement in distinct ways. Halo and Zeta demonstrated a proactive and enthusiastic attitude towards the collaborative tests, recognising the increased student engagement as the tests progressed; their active involvement and positive attitudes likely facilitated a more supportive learning environment, contributing to the success of this approach. With Halo attesting, '*it's been really good, interesting to see the evolution from the first to last test*.' In contrast, Beta acknowledged the potential benefits of collaborative testing for students but remained committed to traditional views that testing should primarily be an individual effort. Beta expressed concerns about the extent of group work being implemented, which manifested in less supportive behaviours during the testing sessions. Unlike the other two cases, Beta urged the students to finish quickly, emphasising the need to '*do some work*,' subtly devaluing the collaborative process. This behaviour diminished the perceived importance of collaboration in learning and may have negatively impacted students' engagement and attitudes towards the testing method. Such divergent approaches among the

teachers underscore the critical role of teacher enthusiasm and engagement in shaping student experiences and outcomes in collaborative testing environments (Baeten et al., 2010). As Gillies and Boyle (2010) highlight, positive teacher involvement is crucial to effective learning outcomes. It is clear from this part of the cross-case analysis that students' and teachers' experiences provide critical insights into the dynamics of collaborative testing. The various reactions, ranging from initial hesitancy to a growing appreciation of the collaborative process, underscore the transformative potential of continued exposure to collaborative testing. The feedback consistently highlighted collaborative testing's capacity to enrich understanding, boost engagement, and develop essential skills.

Building Argumentation Skills Through Active Peer Discussions

The collaborative testing environment cultivated the development of students' argumentation skills, allowing them to articulate, defend, and refine their ideas through active peer discussions. Many students acknowledged this reflection and communication process in their comments, as they actively listened to other views, comparing them to their existing knowledge, which helped clarify their learning progress. Some students expressed this as '**confirming my thoughts**,' '**I could understand things more easily because all the ideas were voiced out loud**,' and '**I enjoyed being able to discuss the answers and find common ground**.' Furthermore, constructive disagreements among group members, built students' argumentation skills by challenging other group members to argue their points and defend their positions. As one student noted, '**not agreeing on certain points**' helped them to '**get their ideas across**,' while another remarked that '**it helped [them] to think clearly**' when discussing the questions. These dynamics are essential for deepening student understanding, as the intra-action between peers facilitates the active exchange of ideas, solidifying learning.

Beta and Zeta, in post-test conversations, observed the positive enhancements as students '*bounced ideas around'* and '*students being really engaged in discussions, sharing and clarifying their ideas*.' Active participation in peer discussions is crucial for developing argumentation skills and improving understanding during collaborative testing. In contrast to individual testing, where students answer questions without the benefit of verbalising their knowledge to their peers, this verbalisation positively impacts students as they identify gaps or misconceptions in their knowledge base (Hattie & Donoghue, 2016). Tullis and Goldstone (2020) support this, contending that engaging in peer discussions facilitates the verbalisation of ideas, leading to a shared understanding of the problem, essential for addressing misconceptions and enhancing subject matter understanding.

It is essential to acknowledge that not all intra-actions were positive. Some students, mainly those more introverted, found the argumentation aspect of collaborative testing challenging. Comments like '**Conflicting ideas or methods made it difficult**' and '**I didn't like how some people didn't get to have their input accounted for**' highlight the importance of carefully considering group dynamics and individual differences to ensure positive outcomes. Beta also observed that peer discussions where students debate questions may not be ideal for students who '*might be shy and [have] a fear of getting it wrong; they may not be forthcoming with the group*.' Miller et al. (2022) similarly emphasise the importance of group composition, noting that negative intra-actions can lead to communication breakdowns and hinder the learning process. The mixed perceptions expressed by students in this study are similar to findings from existing higher education literature in different collaborative testing disciplines, such as undergraduate science courses (Gilley & Clarkston, 2014; Leight et al., 2012; Siegel et al., 2015); chiropractor (Meseke et al., 2010); medicine (Levine et al., 2018), nursing (Green et al., 2018), and statistics (Eastridge & Benson, 2020), indicating that positive group dynamics can enhance student understanding through active peer discussions utilising argumentation skills.

In collaborative testing, the structure of the questions is crucial to the quality of the discussions and subsequent application of argumentative skills. In this study, I used higher-order thinking questions (HOTQ) based on Bloom's taxonomy (Krathwohl, 2002), which included explaining, analysing and evaluating. For example, *Describe the similarities and differences in your answers to D, E, and F. Explain what this demonstrates to you about vector quantities* and *Explain the environmental factors that could have generated the very high pasture production in Queensland compared to South Australia*. Reading the student transcripts, reflections, and observations, I found that many students were actively engaged in the discussions. For example, in Beta's class, on the third test, which has predominately calculation-based questions, each member of the group participated in the discussion:

> **But how long does it take for them to,' 'hey are travelling in opposite,' 'Oh wait … let's see a 1.5 t plus 1.25 t is equal to 55. Cause the velocity times the time is how much they travel, right? If they both travel until it is,' 'Wait, but that's just this is 55 divided by 1.5,' 'The same thing. Yeah.' 'T is equal to 20,' and 'Yeah, makes sense. Sounds right**

When undertaking HOTQ, students and teachers observed increased engagement and improved problem-solving skills. This approach mirrors the work of Eastwood et al. (2020) and Ting et al. (2019), who demonstrated that HOTQs prompt more sophisticated reasoning and analysis, leading

to a deeper understanding of the material. Additionally, research by Eastridge and Benson (2020), Siegel et al. (2015), and Taggart and Wheeler (2023) found that when faced with more straightforward questions, students tended to divide up the questions, limiting the depth of collaboration and critical engagement. In the context of higher education studies, Tullis and Goldstone (2020) found that a balanced mix of HOTQs and more straightforward questions can stimulate comprehensive discussions, noting that this balance is crucial to ensuring all students, regardless of ability, can engage in meaningful peer discussions, thereby enhancing collective learning. Rivas et al. (2022) emphasise that for collaborative testing to be effective, the questions must be aligned within or beyond the range of students' abilities. According to Levine et al. (2018) and Reiger and Heiner (2014) in their undergraduate studies, argumentation in peer discussions deepened students' understanding of content through the reasoned justification of claims and the resolution of differing viewpoints while developing 21st-century skills such as critical thinking, communication and collaboration. These principles, established in higher education studies, also apply to high school science settings. Halo, Beta, and Zeta's students reported that argumentation processes in peer discussions enhanced their understanding.

This study's insights into the role of argumentation during collaborative testing reinforce and extend the findings of Allchin and Zemplén (2020) and Guilfoyle et al. (2023), which demonstrate the role of argumentation in fostering deep understanding and critical engagement among tertiary education students. The structured use of HOTQs prompted students to engage in more sophisticated reasoning and analysis, leading to a deeper understanding of the content. Additionally, the balanced mix of questions recommended by Tullis and Goldstone (2020) effectively stimulated discussions that encouraged active participation. By detailing these findings and aligning them with existing research, this study demonstrates the critical role of argumentation in collaborative testing in enhancing secondary students' learning (Miller et al., 2020; Tullis & Goldstone, 2020). Iordanou and Rapanta (2021, p. 8) attest that argumentation is '*learning to argue, arguing to learn*' (emphasis added). Building on argumentation and gaining multiple perspectives through peer discussions introduces another layer of complexity and opportunity for deep learning. The cross-case analysis provided evidence of how peer discussions enhance secondary students' argumentation skills and deepen their understanding. Active engagement in discussions strengthened their ideas and enabled them to gain a diverse perspective, enriching the student's grasp of the content. These findings reinforce my thesis' key argument that collaborative testing through peer discussions can enhance student understanding in high school settings.

Enriching Understanding Through Peer Discussions and Multiple Perspectives

Active peer discussions during collaborative testing provide the opportunity for students to encounter multiple perspectives, thereby enhancing their comprehension of the subject matter. Students develop a multilayered understanding of the content as they exchange ideas and draw on their different life experiences. This nuanced understanding encourages reflection as students contemplate various viewpoints, some of which may or may not align with their conceptual understanding. One student summarised the benefits of multiple perspectives: '**we learn differently from different people**.' Such intra-actions enable students to 'tap' into their potential to generate a more layered response to the question, cultivating a deeper understanding. Peer discussions within groups play a significant role in refining students' understanding by identifying knowledge gaps and facilitating the correction of errors, which promotes problem-solving skills and conceptual understanding. The dynamics of intra-acting within positive groups allow for immediate feedback during these peer discussions, which is crucial for enhancing learning through 'error correction' (Nokes-Malach et al., 2015, p. 650). Students corrected their misconceptions and filled knowledge gaps through immediate feedback as they listened to different viewpoints, agreeing with Mende et al. (2021) and Taggart and Wheeler's (2020) studies, which emphasised the importance of considering multiple perspectives to enhance understanding through retrieval and immediate feedback.

Despite the benefits, several challenges arose when incorporating multiple perspectives in peer discussions. In collaborative testing environments, considering diverse opinions for each question can overwhelm some students, leading to cognitive overload and preventing students from recalling information (Tang et al., 2020). One student said, '**there were too many people talking, and I lost my train of thought**.' Additionally, listening to multiple perspectives often leads to prolonged discussions, during which dominant students may overly assert their opinions. In contrast, some students may be more reluctant to rely on peer answers, as they may be incorrect, preventing them from considering their peers' perspectives. Encouraging structured collaborative activities and fostering a positive and inclusive classroom environment that promotes respect and trust among peers can help mitigate these negative dynamics in collaborative testing. In this section, I analysed how peer discussions in collaborative testing encourage students to voice their points of view and ways of thinking, facilitating a deeper understanding of the subject matter. Moreover, encouraging different perspectives enabled immediate feedback, illustrating gaps in student knowledge and misconceptions. Sufficient discussion time must be allocated to encourage students to voice their viewpoints and prevent information

overload. In the next section, I discuss how peer discussions in collaborative testing can activate students' prior knowledge to enhance their understanding.

Activating Prior Knowledge Through Peer Discussions in Collaborative Testing

Collaborative testing promotes activating and integrating students' prior knowledge, enhancing learning outcomes. Through peer discussions, students actively use retrieval processes that bring previously learned information to the forefront. Once specific information is retrieved, students activate related knowledge, further developing their understanding. For instance, one student remarked, '**I don't even remember learning this**,' to which another student replied, '**oh, we did it a week ago**.' Moreover, many students across the three classes stated, '**I liked that it helped you remember stuff**.' During discussions, as students retrieve their knowledge and listen to their peers' explanations, they integrate new information, strengthening their memory and understanding.

Careful consideration needs to be given to the wording of the collaborative test questions to activate students' prior information. Test questions should transition from descriptive to synthesis and other HOTQs to improve student comprehension. If the questions utilise too many lower-order thinking verbs, such as 'identify' and 'state,' students may not engage in deep discussions, lessening the potential for active retrieval and integration of related knowledge. HOTQs, on the other hand, can encourage retrieval across different domains of prior knowledge, promoting understanding. An example from Halo's second test illustrates this: Assess the impact of earthquakes on society. Utilising open-ended questions with verbs like 'assess' encourages discussion as there are multiple answers, where students can retrieve prior information and incorporate multiple perspectives to build a multilayered answer. These findings illustrate how peer discussion aids in activating and integrating knowledge and confirm the importance of active engagement in learning processes.

These findings are consistent with existing literature. Studies by Fisher et al. (2012), Schell and Butler (2018), and Shing and Brod (2016) stress the necessity of actively engaging prior knowledge to intra-twine new and old information, thereby enriching students' comprehension. Dancy et al. (2016) add that retrieval processes pull prior knowledge from long-term memory into working memory, which is then processed again before constructing a more complex understanding of the content. Similarly, research by Bodie et al. (2015) and Paine and Knight et al. (2020) found that more cognitively challenging test questions engaged students in deeper discussion, and they used more sophisticated language when faced with these questions. Collaborative testing's impact on the retrieval of prior information adds to the existing literature on the effectiveness of peer discussions and how they

promote effective retrieval practices. This study emphasises the importance of carefully designing collaborative tests to encourage meaningful engagement, offering a critical perspective on how collaborative testing can activate prior knowledge through discussion.

Summary

The preceding analysis has provided evidence of how collaborative testing impacts student learning by encouraging environments that promote active learning, peer discussions and critical engagement with course material. Collaborative testing enhanced student learning by creating an active learning environment where students engaged in peer intra-actions, augmenting their critical thinking skills while promoting the activation and application of prior knowledge, enriching understanding and aiding knowledge construction. Argumentation plays a crucial role in students' reasoning through discussing the questions and learning to justify their thoughts, promoting students' ability to articulate and defend viewpoints. To enhance the efficacy of collaborative testing, students should have prior experience in structured collaborative activities that focus on team-building, communication, and conflict resolution. Furthermore, adjustments in test design and collaborative activities, such as including questions that require collaborative problem-solving, could cultivate more effective collaboration. Collaborative testing improves student understanding through peer discussions and argumentation, attaining multiple perspectives and activating their prior knowledge. While peer discussions form an integral foundation for enhancing learning in collaborative testing, their effectiveness is influenced by various material factors. I next explore the catalysts that drive the quality of peer discussions, focusing on group dynamics, classroom environments and the broader educational context.

High-Stakes Testing and its Impact on Peer Discussions in Collaborative Testing

High-stakes testing exerts considerable influence on collaborative testing in senior secondary education. The pressure of preparing students for critical exams like the HSC in NSW and the QCE often pushes teachers to focus on content delivery, limiting opportunities for innovative practices. Zeta, Beta and Halo reported in our conversations that curriculum demands in years 11 and 12 prioritise a focus on content. Their insights reflect a common struggle: '*time being pushed to get through the content*' and '*having to focus on the stage 6 [years 11 and 12] content,*' highlighting the challenge of balancing content coverage with broader educational goals. This struggle aligns with Hickey and Riddle (2023) and Masters (2019), who argue that accountability and performance outcomes inhibit teachers' ability to integrate more dynamic student-centred learning approaches.

Focusing on high-stakes tests in the classroom narrows the curriculum and shapes the assessment design. For example, Halo, Beta, and Zeta acknowledged that their in-class assessments often mirror the structure of external exams. In our post-test conversation, Halo mentioned they would '*embed more open-ended and cross-content questions to address the challenges students faced with similar questions in the HSC examination*.' This trend is consistent with Masters (2020) and Wilson et al. (2021), who suggest that teachers under external pressure design tests that resemble these exams, limiting diverse forms of classroom assessments. During my observations, Halo's collaborative test questions provided students with practice in HSC-style questions, encouraging peer discussions that improved both metacognitive skills and conceptual understanding. The reluctance to assign grades in collaborative testing, driven by external accountability concerns, diminished its perceived importance. Ungraded tests led to social loafing, as students saw less value in the collaborative tests. A student remarked, '**they would take the test more seriously if it counted for grades.' At the same time,** another instance saw students becoming distracted: '**M2[61], let me draw a Ute**,' '**M3, focus**,' prompting Zeta to remind them, '*this is serious*.' Although some groups remained highly engaged, these variations in behaviour could be attributed to different motivational levels (Gilley & Clarkston, 2014), anxiety related to peer discussions (Levine et al., 2018; Mahoney & Harris-Reeves, 2019; Meseke et al., 2010), and students' views on collaborative activities may also affect their involvement (Eastridge & Benson, 2020).

Despite acknowledging the clear benefits of collaborative testing for improving communication and collaborative skills, Halo, Beta and Zeta expressed reservations about its feasibility under current accountability pressures. Zeta highlighted the challenge of differentiating students' contributions: '*The problem with summative collaborative testing is how to differentiate the students, how to measure/quantify that, and their contribution*.' Halo addressed this concern by implementing a grading scheme of 80% for the individual test and 20% for the collaborative section, a practice consistent with Reiger and Reiger's (2020) review. This grading approach minimised issues like social loafing, which were absent in students' mind maps, transcriptions, or my observations during this third test. Recognising the challenges posed by this focus on curriculum, recent changes by the Australian Curriculum and Assessment Reporting Authority (ACARA, 2020) aim to grant teachers more autonomy in assessment practices within junior years. However, Zeta notes that this '*autonomy does not extend to senior assessments*.' Zeta observed a shift from in-class assessments, which contributed to final grades, towards adherence to '*assessments dictated by the Queensland Curriculum and Assessment Authority, with little flexibility for alternative classroom assessments*.' High-stakes testing in senior secondary education significantly limits teachers' ability to implement collaborative testing, as they

[61] M2 and M3 represent two specific male students in a group.

must prioritise content coverage to meet exam needs. Despite recognising the benefits of collaborative testing for developing essential skills and improving conceptual understanding, teachers face challenges with grading and accountability pressures, which hinder its broader application.

The Role of Group Dynamics in Shaping Peer Discussions During Collaborative Testing

Group dynamics shaped the quality of peer discussions during collaborative testing, with positive and negative intra-actions directly influencing student engagement and learning outcomes. Positive group dynamics and interdependence enhance students' testing experience and understanding by encouraging peer discussions where students can voice their opinions, share knowledge, and constructively debate the questions. As students commented, '**I liked sharing my knowledge**' and '**I liked having different people share their opinions in my group**.' Such positive group dynamics encourage self-expression, improving students' perceptions of collaborative testing and how they view their peers. Moreover, this engagement facilitates learning as students feel they can '**collaborate with new people**,' '**helped them to work as a team**,' and '**it takes the stress away**.' Positive group dynamics in collaborative testing encourage active peer discussions that enhance understanding and improve metacognitive skills. Through these discussions, students review, adapt and apply each other's approaches to problem-solving (Stanton et al., 2021). Positive collaborative groups can stimulate ideas and generate a deeper conceptual understanding through peer discussion.

However, when group dynamics are negative due to differences in personality, work ethic, or engagement levels, they can disrupt group cohesiveness and negatively affect students' experiences with this testing method. The success of collaborative testing depends on the group's interdependence; when this breaks down, individual differences can hinder the group's effectiveness. Some students reported, '**Some people would kind of take over**,' '**some group members don't listen**,' and '**talked over them, so I can't make a point**,' illustrating how personality differences can constrain peer discussions, undermining collaborative testing outcomes. Assertive personalities may dominate conversations, pressuring peers and discouraging students from sharing their opinions, impacting their overall experience. Additionally, students who lack motivation may contribute to social loafing, where less effort is made to engage in the discussions, leading to resentment and negative views of collaborative work (Stanton et al., 2021). One group's intra-action exemplifies this dynamic: '**M1—who gives a … man, I can't be bothered**,' '**M2—come on man**,' and '**M1—I don't want to do this**.'

The significant role of peer influences in collaborative testing underscores the importance of carefully considering how groups are formed to optimise group dynamics and the perceptions of this

assessment method. These findings align with research undertaken by Levy et al. (2018) and Miller et al. (2022), who attest that negative group dynamics diminish the benefits of collaborative testing, while Larrson (2017) and Premo et al. (2018) attest that differences in personalities, peer influences, and motivation are critical factors in group behaviour and the efficacy of collaborative strategies. Group dynamics in collaborative testing play a crucial role in shaping peer discussions, and positive dynamics can enhance understanding and engagement. In contrast, negative dynamics may disrupt participation, impacting the quality of the group discussions.

Impact of Group Formation on Peer Discussions During Collaborative Testing

Group formation influences the effectiveness of collaborative testing outcomes, primarily through two methods: student choice and teacher-assigned groups. In Halo's class, students chose their groups for tests 2 and 3, forming a variety of configurations, including equal-gender, single-sex, and groups with a predominant gender, that generally performed well. However, challenges arose among some student-selected groups in Halo's class, potentially exacerbated by the tests occurring at the beginning of the school year, when some students may not know each other, impacting some groups' cohesion and causing awkwardness among group members. Zeta's class also displayed these challenges as students formed all single-sex groups, where the effectiveness of the group was affected by personality differences. For example, while one all-female group quickly became productive, another struggled with distractions before engaging with the task. Both all-male groups were distracted, and disagreements occurred, leading some students to prefer '**maybe not picking our groups**.' As demonstrated by this comment and the actions of the groups, some students who choose their groups may not account for their friends' differing personalities and motivations. This sentiment reflects Bonem et al.'s (2020) and Le et al.'s (2018) findings on the difficulties of working with friends, which often lead to off-topic discussions and ineffective collaboration.

Teacher-assigned groups demonstrated mixed results, with some students appreciating the diversity, noting that '**new people helped them to solve questions**.' Nevertheless, others felt '**the groups weren't chosen; therefore, some people had less input**,' leading to uneven participation in groups across the classes. Positive experiences in heterogeneous groups were common, enhancing students' understanding and skills. A student commented, '**I liked hearing how other people interpreted the question. A lot of people have different skills, and they have different backgrounds, so it helps**.' Heterogenous groups create an environment that stimulates peer discussions as students use their abilities and interests to understand the concept being tested. Beta noted this and asserted they could see their students answering the questions more deeply. These positive results are in line

with Meseke et al. (2008), Miller et al. (2022) and Nafziger et al. (2011), who support the idea that heterogeneous groups bring together a more comprehensive range of cognitive resources and experiences, enriching collaboration.

Addressing Gender Imbalances on Group Dynamics in Collaborative Testing

In Halo's class, two female students in a group of four articulated their male peer's inattention to their ideas by stating, '**we need to talk about this. You guys aren't listening**,' with the second female saying, '**of course, they're not**.' This example highlights how gender imbalance can impact group dynamics. Strategically forming groups to balance personalities, genders, and abilities has shown promising results in enhancing the inclusivity and participation of female students (Miler et al., 2020). By carefully considering these factors, I arranged groups that worked well collaboratively.

Nevertheless, if group dynamics are not managed well, it can lead to dominant personalities (often male) overshadowing female contributions (Fernández-García et al., 2019). Beta's physics class had a significant gender imbalance, and the random assignment of groups in the first test impacted one female student's engagement. After reading the transcript for test one, this student did not say a word during the test, nor did the male students ask her questions. After observing these dynamics, I adjusted the group compositions, deliberately separating the four female students into different groups from male students. This change led to a noticeable improvement in engagement and participation by all-female students.

These findings agree with those by Miller et al. (2022), who note that for female students in collaborative testing, it is not the gender composition of the group that is important; it is the differences in personalities that influence the success of the group. This finding highlights how gender differences and personalities impact student perceptions of collaborative testing. Fernández-García et al. (2019) and Miles and Naumann (2021) agree and contend that not being included in discussions with their male peers can lead to a lack of confidence in their physics ability, reinforced by stereotypes that challenge female capabilities in math and science, lowering their self-efficacy in these subjects. This critical issue explains why girls often get 'lost' in science during high school, with group dynamics playing an integral role in collaborative testing outcomes. It is equally important to consider how these dynamics interact with other elements, such as the physical environment, to shape students' learning experiences. Collaborative testing through positive peer relationships can improve learning experiences for many students, particularly females in STEM. These insights into group dynamics during collaborative testing align with and extend the findings of Giuliodori et al. (2008) and Gilley and

Clarkston (2014) regarding the influence of group dynamics on the efficacy of this testing method. By exploring student-selected and instructor-assigned groups, I expand our understanding of how diverse group compositions influence peer discussions and collaborative testing outcomes. Furthermore, the findings offer a nuanced perspective on the effectiveness of gender-homogenous groups, suggesting that the most suitable group formation strategy may vary depending on the specific educational context and objectives.

Cultivating 21st-Century Skills Through Collaborative Testing

In this subsection, I explore how critical 21st-century skills like collaboration and communication are prerequisites for successful collaborative testing. I observed significant improvements in peer-to-peer communication within Halo, Zeta, and Beta's classes directly linked to this testing method. For example, one student remarked, '**My skills improved because I got to discuss more with other people when usually in class. I don't normally get to do that**.' To ensure productive communication within the group, students need prior experience in critical thinking skills to argue, listen and problem-solve during argumentation. Without this preparation, as I observed in this study, group dynamics can break down, leading to conflicts and reducing the perceived benefits of this method. Comments like, '**I don't really know what to contribute to this**' and '**I didn't know the group well, so I didn't feel like I could talk [to them] as well**' illustrate the challenges faced without adequate preparation.

During collaborative testing, actively practising communication skills can promote positive intra-actions that build trust and support between students while reducing conflicts. Students learn to formulate arguments, listen effectively, and communicate their ideas, improving their ability to engage constructively. For example, one student commented, '**In the first task, I did not talk much, but now I introduce all my ideas, even though they could be wrong**.' Another student added, '**My skills have improved since the first day. Because it's a lot easier to explain my opinions, and I feel more confident**.' Improvements in students' communication and collaborative skills were also observed by Halo, Beta, and Zeta, who, during our post-test conversations, said it was '*interesting to see the students really discussing and questioning each other*.' Effective communication within collaborative testing is not just about sharing ideas, navigating differences, and respecting other perspectives but also involves 'social negotiation,' as attested by Taggart and Wheeler (2023, p. 3). However, competitive or individualistic classroom environments can weaken this social negotiation between students, leading to negative intra-dependence and ineffective communication within groups. The effect of this results in '**conflicting ideas**' and '**arguments**,' similar to issues identified in undergraduate

courses (Levine et al., 2018; Mahoney & Harris-Reeves, 2019; Reiger & Heiner, 2014; Siegel et al., 2015).

The practice of these skills in collaborative testing environments encourages positive intra-group dynamics and prepares students to handle disagreements constructively, a necessary skill for reaching consensus. Both students and teachers noted that disputes and conflicts over opinions and who was 'right' impacted the ability of the group to reach a consensus, making some students feel their '**opinions didn't count**.' The students' comments align with Eastwood et al. (2020), who contend that the main drawbacks of collaborative testing include dealing with uncertainty about correct answers and the risk of being incorrect. Students must have critical thinking skills scaffolded through diverse activities to address these challenges. The success of collaborative testing in enhancing students' communication and critical thinking skills depends on effective, scaffolded practice in low-stakes environments. My observations note that students who have had practice undertaking multiple collaborative testing sessions improve their collaboration and communication skills, which leads to more effective group dynamics and a deeper understanding of the content. Teachers play a crucial role in this process by structuring collaborative activities, modelling clear communication, and encouraging a supportive classroom environment, enabling students to debate, argue, and listen effectively, enhancing their conceptual understanding.

Shaping Peer Discussions: The Material Environment's Role in Collaborative Testing

The critical aspect of this study was determining how the material environment influenced students' learning experiences through collaborative testing. The material environment encompasses the teachers and students and includes all human and material elements involved in the intra-actions.[62] Moreover, my role positioned me as an integral element within the entanglement of students, teachers, tests and other elements within that time-space context. I used multiple apparatuses to focus on emergent differences and similarities to delineate each collaborative test's inclusions, exclusions and environmental differences. In the discussion below, I focus on the three material constraints that emerged from the cross-case diffractive analysis. In Beta's class, uniquely, the ambient temperature impacted student learning and perceptions of collaborative testing. The second constraint, the time allocated for completing the tests, impacted all three classes. Likewise, the third constraint, noise levels caused by student discussions that affected students' concentration ability, appeared across all cases. Material and environmental factors, such as heat in Beta's class, cognitive

[62] Intra-actions refer to the intra-connectedness between students', teachers', myself, and the phenomena of the classroom environment (Barad, 2007).

fatigue in Zeta's class, and time constraints and lack of accountability in all three classes, impacted group dynamics, causing disengagement, lack of motivation and social loafing for some students.

On the day of Beta's second test, the outside temperature rose above 35°C, heating the classroom. Temperature affected the students, impacting their learning and perceptions of collaborative testing, where one student exclaimed, '**it is too hot**.' Moreover, as the test progressed, I observed students' flushed faces and decreased concentration ability. These findings align with research from Cho (2017) and Wargocki et al. (2019), who noted similar effects of heat on student learning in the USA. Moreover, Park et al. (2020) observed that rising temperatures in New York City schools adversely affected students' cognitive processing during critical Regents[63] examinations, confirming my observations that increased temperatures can lead to diminished concentration and greater distractibility through heat-related fatigue. Roach and Whitney (2022) support these findings, contending that 'heat and thermal stress cause heightened brain stress to fulfil cognitive processes like attention, memory, verbal learning, information processing, and concentration' (p. 31). These effects are particularly concerning due to the impacts of climate change, predicting changes in temperature across the globe. Climate change effects are not uniform worldwide; while there is a general trend towards hotter temperatures, there are geographical differences (Johnston et al., 2021). In Australia, Johnston et al. (2021) argue that regional cooling trends can negatively impact learning due to inadequate insulation and clothing for colder temperatures (Moore et al., 2019). The uneven effects of climate change are exacerbated in many schools that do not have access to air conditioners to either heat or cool the classrooms, impacting student learning (Johnston et al., 2021; Park et al., 2020a). Zeta and Halo's classrooms were modern and equipped with air conditioning, creating a comfortable environment where students could undertake collaborative testing. In contrast to Halo's room, Beta's classroom did not have air conditioning, impacting students throughout the test. Such disparity in resource allocation in the same school underlines the importance of the material environment to student learning while raising questions about equity and access to resources across schools.

In collaborative testing, the time allocated for completion is a critical factor affecting students' perceptions. Collaborative tests often require more time to complete than individual tests due to the need for critical thinking, argumentation, and reaching a consensus on the answer. The anxiety associated with insufficient time was observable across the classes, with students commenting, '**the time limit was a bit stressful**.' This stress can impair information retrieval, increasing students' stress

[63] In New York State, Regents Examinations are state-wide standardised examinations in core high school subjects, which students are required to pass to graduate (Park, 2020).

as they '**feel they may let the group down**.' Additionally, inadequate time affected students' ability to add depth to their answers, as they '**rushed through the questions**,' which caused some students to dominate the discussions, influencing group dynamics and their perceptions of collaborative testing. Correspondingly, Efu (2019), Levy et al. (2018), and Meaders and Vega (2022) note that adequate time allocation is necessary to encourage deep discussions and consensus among students. Time constraints can prevent deep discussions, essential for students to attain multiple perspectives and enhance their conceptual knowledge. Efu (2019), however, noted that research reporting minimal gains in understanding failed in their study to provide additional time for collaborative questions, suggesting a direct link between time allocation and the depth of understanding achieved.

Subsequent test formats were adjusted to address these concerns by reducing the number of questions and extending the test duration. This strategy agrees with research by Gilley and Clarkston (2014), who recommend reducing the number of test questions to mitigate time constraints, allowing students to delve deeper into each question. Therefore, these time and test format adjustments can enhance the effectiveness of collaborative testing by enabling more in-depth discussions and reducing student anxiety. Additionally, there are constraints on the amount of time teachers can allocate for collaborative testing. Zeta stated they did not '*feel with so much content that they could give the time needed to do collaborative tests*.' Nevertheless, the benefits of enhanced conceptual understanding and critical thinking skill development are important considerations. Therefore, the number of questions on the test and the teachers' available time determine the time allocated for collaborative tests.

Noise levels during collaborative tests influenced some students' ability to concentrate. The noise levels rose as students discussed the questions and debated their responses, which was particularly evident during the high-stakes test, with students reporting, '**It got kinda loud with all the groups collaborating**, '**It's too noisy**,' and '**It is difficult to hear what they are saying**.' Elevated noise levels prevent students from hearing questions and responses, impacting their ability to participate in the group. Additionally, noise can distract students, affecting their concentration and leading to a breakdown in group dynamics. The spatial arrangement of the classroom can significantly influence noise levels. In Beta and Zeta's classrooms, the spacious room allowed for desks to be spread out, reducing noise interference between groups. However, due to the number of students and the classroom layout, the proximity of student groups in Halo's classroom escalated noise levels, interfering with students' ability to concentrate. Thompson et al. (2022) and Massonnié et al. (2020) point out that noise disrupts concentration and students' productivity in collaborative environments. Excessive noise levels can impede students' ability to participate constructively in collaborative testing.

This section underscores the impact of the material environment on students as they undertake collaborative testing. I observed that lacking adequate climate control can impede students' cognitive functions and, therefore, their ability to participate effectively in their group. The allocation of sufficient time emerged as a critical factor, influencing students' anxiety and depth of answers. Noise levels due to overcrowding distracted some students and decreased their ability to concentrate. Understanding how these material factors intra-twine during collaborative testing helps present a more multifaceted understanding of student and teacher perceptions. These observations highlight the importance of the material environment, not just as a backdrop to collaborative testing but as an active participant in shaping the dynamics and effectiveness of peer discussions in this method. Addressing these environmental factors is crucial for creating conditions that support meaningful intra-actions to enhance students' and teachers' perceptions of collaborative testing.

Summary

In these sections, I presented evidence that high-stakes testing markedly affected the implementation of collaborative testing in high schools in myriad ways. The managerial education systems interpret 'accountability' in very narrow ways, primarily valuing standardised assessment forms. This focus on performance outcomes constrains teachers from incorporating collaborative test grades, which can lead to a perceived lack of accountability in students' eyes. As a result, students may not engage seriously in peer discussions during collaborative testing. Additionally, the pressure of an extensive curriculum and the alignment of assessments with high-stakes tests further limit teachers' capacity to implement collaborative testing. I have also evidenced key elements influencing the quality of peer discussions during collaborative testing, such as group dynamics and the physical and social classroom environment. These factors play a role in facilitating or constraining effective collaborative testing. These findings are situated within the ongoing debate in the literature that emphasises the complex balance between the benefits and drawbacks of performance-based outcomes in high-stakes environments. While the literature may argue that high-stakes testing may limit innovative practices, research in higher education (Newton et al., 2019; Miller et al., 2020; Siegel et al., 2015) demonstrates that, when implemented effectively, collaborative testing can promote peer discussions and serve as a valuable supplement to traditional assessments.

Implementing Collaborative Testing: A Research-informed Guide for Teachers

The findings of this research demonstrate that effective implementation of collaborative testing requires more than group work; it demands preparation that includes opportunities for students to learn skills and attitudes associated with effective collaboration, thoughtful question design, and a supportive learning environment. In the following section, I outline practical strategies for incorporating collaborative testing into classroom assessments, drawing from this study's findings and utilising supporting research.

Preparing Students for Collaborative Testing

Before introducing collaborative testing, creating a positive and productive classroom environment is essential. Teachers should provide opportunities for students to practice structured collaborative tasks and low-stakes tests beforehand. In this study, Halo and Zeta observed that their students had '*minimal skills*' during the first test but noted significant improvement with practice. For example, Beta reported that their students' communication and collaboration abilities increased through '*practising through the three tests*.' Start by incorporating structured collaborative activities into regular class time to set the stage for successful collaborative testing. Teachers can write open-ended questions, assign students to groups based on personality, gender, and ability, and guide them in collaboratively answering the questions. Practical activities in structured groups can help students learn how to communicate with their peers, promoting interdependence. Including peer evaluations or a rubric can help keep students accountable and reduce social loafing.

Students could also collaboratively retake a previous individual test as they become more comfortable over time. This approach aids comprehension by allowing students to discuss the questions and understand why they were incorrect, thereby solidifying their understanding. As discussed earlier, marks for a collaborative test could be assigned by adding a percentage point to the student's test scores. In the graded test, Halo used an individual test followed by a collaborative test for the high-stakes collaborative test, typically found in many undergraduate courses. Students first completed the individual test, which accounted for 80% of the total grade. Subsequently, they formed collaborative groups and undertook a 20% collaborative test containing different materials related to the individual test. This collaborative test included two open-ended questions with a stimulus and thirty minutes for completion, considered ample time.

Designing Effective Collaborative Test Questions

Various question types must be included in the test to encourage deep discussions and critical thinking. In this study, I used a mix of open-ended, structured short responses and mathematical problem-solving questions in Beta's class. Other collaborative testing studies have succeeded with well-structured multiple-choice questions with similar answer options, which can spark meaningful discussions (Mahoney & Harris-Reeves, 2019; Vogler & Robinson, 2016). Jang et al. (2017) used an online team-based scoring program in their calculus-based mechanics class, which used multiple-choice and long-response questions that students attempted in groups with whiteboards to work through the problems. They would submit an answer, and the program would immediately state if that question were right by awarding four points; the group had three tries before it said the correct answer. When designing the test questions, focus on higher-order thinking prompts, such as 'explain,' 'compare and contrast,' 'synthesise,' and 'critically analyse,' to ensure engagement and deep discussions across the group. For instance, I used a test question: 'Use the stimulus to describe why some people are concerned about new technologies and analyse the ethical considerations of using therapeutic cloning to treat medical conditions.' Research by Kaendler et al. (2016) and Taggart and Wheeler (2023) support this approach, demonstrating that complex problem-solving facilitates critical thinking skills.

Creating a Supportive Learning Environment

The classroom environment has a significant impact on the success of collaborative testing. Effective group dynamics are critical, so thinking carefully about how groups are formed is required. I typically formed groups of three or four students, in agreement with the literature that notes that larger groups can struggle to maintain group interdependence. This issue was seen in a large group of five students, who commented that there were '**many disagreements because there were too many of them**.' While collaborative testing can be done in pairs, it may limit the diversity of perspectives on the questions, which is crucial to a deeper understanding of the content. In this study, groups assigned by the teacher or me generally performed better than those chosen by the students. While still competitive, Halo and Zeta's classrooms encouraged students to collaborate and participate in testing. A collaborative, positive, and supportive environment promotes positive interdependence, which balances individual and collaborative objectives. This approach can motivate students to contribute to the group's success and enhance their conceptual understanding by encouraging and supporting their peers to take chances. A student commented, '**we learned to work as a team to learn more**.'

However, the timing of the test is essential; if students have multiple tests on the same day, cognitive fatigue can impact their engagement with their peers. Practising collaborative testing in a

non-competitive environment can also promote students' social skills, encouraging them to work together more effectively. Conversely, a competitive environment focusing on individual grades might discourage collaboration and lead to social loafing, where some students do not contribute but still benefit from the group's efforts. Halo observed, '*the students in this school are typically competitive with each other over marks,*' while one student mentioned that '**some people don't share their knowledge and still get credit for the work**.' To avoid these concerns, students must engage in structured collaborative activities and teachers' should carefully consider the marking system for accountability.

Optimising the Physical Classroom Environment

The physical environment of the classroom can affect the efficacy of collaborative testing. Initially, all desks were arranged in rows facing the front of the classroom, which is not conducive to collaborative testing. Desks should be spread out, with chairs on either side to facilitate better communication when collaborating. This setup allows students to hear and engage with each other more easily. Additionally, consideration needs to be given to factors such as ambient temperature and noise levels, as they can impact students' concentration and engagement. Most importantly, collaborative testing typically requires more time than individual testing, so there must be enough time to engage in deep discussions and complete the test. In this study, tests lasted between thirty and forty minutes, but the duration should be adjusted based on the number and complexity of the questions.

Summary

The findings from this study support the importance of preparing students by practising collaborative activities in a supportive environment where group dynamics are carefully considered. Effective collaborative testing also relies on well-designed test questions that promote deep discussions and critical thinking. Additionally, the physical arrangement of the classroom should facilitate student comfort and communication, and sufficient time should be allocated to allow active argumentation. By following these guidelines, teachers can deepen their students' conceptual understanding while improving their 21st-century collaborative testing skills.

Conclusion

This cross-case analysis of collaborative testing in high school settings has illustrated the complex dynamics and diverse perspectives of students and teachers. Through a detailed examination of three distinct educational environments, this analysis highlights the transformative potential of collaborative testing in developing critical thinking and collaborative skills, which are essential for 21st-century education. Key findings underscore the critical role of quality peer discussions in deepening students' conceptual understanding and their involvement in the learning process. The analysis of student and teacher reflections underscores the importance of group dynamics, the structure of collaborative activities, and the influence of contextual factors on the success of collaborative testing. Students reported that collaborative testing enabled them to appreciate diverse perspectives and engage in deeper conceptual understanding, bolstering their communication and critical thinking skills. However, challenges such as time constraints, group compatibility, and the physical environment significantly impacted students' and teachers' experiences. Teachers, on their part, recognised the value of collaborative testing in fostering active learning environments but expressed concerns about its practical implementation, especially regarding grading and a content-heavy syllabus. The findings suggest that for collaborative testing to be effectively integrated into high school curriculums, it requires not only the alignment of assessment strategies with educational objectives but also the cultivation of a supportive classroom culture. Such a supportive environment entails preparing students through structured collaborative activities that build familiarity and trust among peers, designing tasks encouraging meaningful intra-action and reflection, and adjusting the physical and material conditions to optimise the learning environment. This study finds that while collaborative testing presents a promising avenue for educational innovation, its success is contingent upon carefully considering pedagogical objectives and logistical realities. By addressing these factors, teachers can take advantage of the potential for collaborative testing to enrich students' learning experiences and equip them with 21st-century skills.

Chapter 9: Conclusion

Introduction

In this chapter, I synthesise the key findings of this study in response to the three central research questions, detailing how collaborative testing reshaped students' conceptual understanding and collaboration and communication skills in high school science classes. By undertaking a diffractive analysis, this research has demonstrated the significant influence of material factors on teachers' practice and student learning.

Research Questions and Key Findings

This study has provided a deep understanding of the complexities of the intra-actions surrounding collaborative testing in a high school science setting. By applying diffractive analysis to both students' and teachers' perceptions alongside observations of their practices, this thesis has accounted for factors, including the materiality of the intra-actions, that significantly impact student learning. This approach aligns with the core aim of this thesis, as outlined by the title: *Engaging diffractive ethnography to explore student and teacher perceptions of collaborative testing*.

The research questions guiding this study were:

1. How did students and teachers perceive collaborative testing?

2. How does collaborative testing impact student learning?

3. How does implementing collaborative testing unfold in a real classroom?

I will elaborate on the key findings by answering the research questions before drawing further conclusions about how this study contributes to knowledge about assessment *as, for and of learning* (NESA, 2022) and theoretically to the application of diffractive ethnography.

Q1: How Did Students and Teachers Perceive Collaborative Testing?

This thesis established that many students and teachers predominately perceived collaborative testing as beneficial through promoting positive classroom dynamics and peer experiences. Collaborative testing through group intra-actions encouraged students to work collaboratively to answer the test questions, enhancing the development of students' teamwork, critical thinking, and communication skills and facilitating positive peer intra-actions. Students and teachers highlighted the value of peer discussions in developing communication and social skills. For many students, this increased their confidence in sharing their thoughts within their groups. However, some students initially expressed discomfort due to concerns about accountability, group dependency, and time constraints. According to Halo, Beta, and Zeta, the benefits included increased student collaboration and communication skills and how students intra-acted with each other. Despite their reservations about grading, time constraints, and uneven group participation, the data suggests that many students' and teachers' maintained an overall positive view of collaborative testing, with the benefits outweighing the challenges. My research demonstrates how collaborative testing can promote positive intra-actions between students, contributing to transforming traditional assessments to prioritise 21st century skills and inform classroom practice.

Q2: How Does Collaborative Testing Impact Student Learning?

Collaborative testing enhanced student learning by sparking active, constructive peer discussions. As students navigate the testing process, they did not just answer the questions; they engaged in deep exchanges that enhanced their conceptual understanding through recalling and connecting knowledge, argumentation and integrating multiple perspectives. The benefits of these peer discussions were notable, with students and teachers observing the positive impact of 'bouncing' ideas around to their understanding. As the testing progressed, students' social and communication skills improved, enhancing the discussions' clarity and further aiding in peer-to-peer learning. This peer-to-peer intra-action, in which students explain concepts in relatable language, strengthened their collective understanding. There were some challenges, such as cognitive overload from 'too many opinions' and some reported negative group dynamics. Peer discussions enhance conceptual understanding, illustrating the influence of collaborative testing in constructing knowledge through active engagement and multiple perspectives. Through a diffractive analysis, I have demonstrated that student learning is entangled with the material-discursive environment, emphasising the intra-connectedness of group dynamics, classroom environment, and learning processes.

Q3: How Does Implementing Collaborative Testing Unfold in a Classroom?

Teachers' perspectives on implementing collaborative testing reveal a balance between recognising the educational benefits and the practical constraints. All three teachers, Halo, Beta and Zeta, acknowledged the significant benefits of collaborative testing in improving their students' 21st-century skills. They noted that peer discussions during the tests provided immediate feedback, encouraged students to defend their opinions, and facilitated the construction of new knowledge. A significant benefit the teachers identified was collaborative testing's ability to enable students to 'learn while testing' (Jang et al., 2017), allowing the students to fill knowledge gaps, correct misconceptions and analyse each other's understanding. Despite the acknowledged benefits of collaborative testing, all three teachers expressed reservations, primarily due to external constraints when evaluating this method.

The demands of final graduation examinations and education policies emphasising performance outcomes constrain teachers' ability to implement collaborative testing fully. These concerns were especially pronounced regarding the potential impact on students' grades in the senior years, leading teachers to exclude collaborative testing results from overall assessments. Additionally, Zeta and Beta described the challenges of assessing individual contributions within collaborative testing, especially in the critical years of 11 and 12. However, in Halo's class, where the summative test occurred, there were increased active peer discussions and less social loafing as students undertook the test, with many students noting positive perceptions of the summative test while stating few concerns regarding the grading of the collaborative section. The teachers' views on collaborative work shaped their perceptions of collaborative testing. Beta, while acknowledging the benefits to their students, argued that individual testing is crucial in senior school while viewing collaborative testing as a distraction from essential instruction time. In contrast, Zeta and Halo positively perceived collaborative testing, appreciating its intra-active nature. While the teachers considered collaborative testing a valuable tool for developing critical skills, implementing it is impeded by concerns over assessment practices and the pressure of high-stakes testing. However, this study has demonstrated that collaborative tests can be implemented as both a form of formative assessment and a summative test for students, with the associated learning and skills benefits.

The Material Impacts on Student Learning

A major insight of this research was the need to consider how the material environment intra-twines with assessments in classroom environments, a factor often overlooked in previous collaborative testing studies. This study identified a growing need to understand how material elements, such as temperature, test length, group composition, and the broader school context, affect students' learning during collaborative tests. These elements influence not only students' understanding but also their engagement and overall learning experiences, which is critical in the senior years of high school when outcomes have significant implications for students' future directions. By employing a diffractive method, this study revealed how material factors shaped the testing experience. High ambient temperatures reduced focus and engagement in Beta's class; across all classes, limited time heightened stress and restricted deep discussions; and noise from peer discussions disrupted concentration. These factors affected group dynamics, leading to social loafing and negative perceptions of collaborative testing. By highlighting the critical role of material elements in shaping outcomes in collaborative testing, this research offers a nuanced perspective on how educational realities and knowledge are continuously (re)constructed, illustrating how the material environment affects student learning.

Demonstrating the Impact of Collaborative Testing in High School Science Classes

The central aim of this thesis, as introduced in Chapter 1, was to investigate the perceptions of collaborative testing among high school science students and teachers. While collaborative testing has been explored in various higher education science fields, such as biology, nursing, statistics, and medicine, there is a notable scarcity of research on its application in high school settings. This study addressed this literature gap by comprehensively examining the effects of collaborative testing in three high school science classes, providing new insights into how this method can impact student learning and teaching practices in this context.

A principal finding throughout this research was that collaborative testing was a key driver in creating an active learning environment where peer discussions enhanced student conceptual understanding. For example, in Halo's class, during the third test, students engaged in discussions about cloning, which allowed them to challenge and refine each other's thinking and deepened their engagement with complex ethical and scientific concepts. Similarly, in Beta's class, students debated different strategies for solving complex physics problems, enhancing their understanding of the material and encouraging them to articulate their reasoning. These instances illustrate the value of

peer discussions in developing understanding and critical thinking skills, extending existing knowledge on collaborative testing observed in higher education to a high school setting.

Significantly, this thesis provides substantive evidence that collaborative testing in high school settings promoted critical thinking, argumentation, and collaborative skills in science subjects as diverse as agricultural science and physics. Many students commonly commented on the improvement in these essential 21st-century skills. For example, in Beta's class, students demonstrated improved listening and argumentation skills due to peer discussions and exposure to their peers' perspectives. Similarly, Zeta and Halo observed growth in their students' debating, communication, and collaborative skills throughout the testing process, validating the positive impact of collaborative testing. Despite these benefits, it is essential to note that the effectiveness of peer discussions and the resulting learning gains are generally tied to a positive classroom environment and a sense of positive interdependence among group members, emphasising that the classroom environment is crucial in creating the right conditions for successful collaborative testing. Positive group dynamics were contingent on carefully considering the group's personalities and abilities, an area less documented in the collaborative testing literature.

Despite these positive outcomes, it is essential to note that the effectiveness of group discussions and the consequent learning gains are tied to positive classroom environments and positive interdependence within the groups. This thesis extends the literature on collaborative testing by highlighting the complexities and challenges of group formation in high school settings. Initially, student groups struggled due to several factors, such as a lack of prior experience with collaborative activities and testing, differing personalities, and the awkwardness of working with unfamiliar peers. Furthermore, I contend that students' prior experiences, especially in new environments, can shape their views on collaborative testing, with some students noting unfamiliarity with their peers, leading to negative group intra-actions. I deepen the discussion around collaborative testing by demonstrating its potential as a powerful tool in high school science education to promote active peer discussions, critical thinking, and collaborative skills while also providing insight into the complexities of group dynamics. Moreover, these findings challenge the assumption that in-class tests are incompatible with innovative methods such as collaborative testing.

The Influence of High-Stakes Testing on Collaborative Testing in High Schools

A notable insight from this thesis is that teachers often perceive a lack of flexibility in implementing collaborative testing due to the emphasis on performance outcomes. Although curriculum documents do not specify instructional methods, teachers still hesitate to undertake alternative assessments. The teachers' hesitation stems from their beliefs about what constitutes valid assessments and the pressure to meet performance outcomes. For example, Zeta and Beta expressed concerns about performativity, leading them to work closely with the curriculum and favour individual testing, which provides a more precise indicator of students' performance and class rankings. This preference further discourages the exploration of alternative practices.

However, the successful implementation of a summative test in Halo's class demonstrates the feasibility of this method, not only to assess students but also to promote learning during the testing process. The prevalence of managerial practices in education, characterised by accountability, standardisation, and outcome metrics, further influences teachers' perceptions and implementation of assessments. For example, students' underperformance on a class test prompted Beta to emphasise individual work, reinforcing a trend towards individualisation in the senior years of schooling. This environment often constrains teachers' ability to implement innovative practices like collaborative testing. Nonetheless, the vignettes outlined in the three case studies suggest there is room for teachers to attempt different strategies to assess and improve their students' understanding and engagement. Teachers can align the test with curriculum goals by employing collaborative testing as a high-stakes test while enhancing student understanding and collaborative and communication skills. These insights underline the potential of this method to coexist within a performance culture that challenges the assumption that there is only one way to assess student understanding.

In the senior years of high school, the emphasis shifts from a broader educational perspective that cultivates various learning strategies and assessments to a concentration on end-of-school examinations. For example, in the United States (ACT and SAT), the United Kingdom (A-levels), and France (Baccalaureate), these exams play a crucial role in determining higher education prospects and career paths. A focus on high-stakes testing promotes an individualistic, competitive, and content-driven atmosphere, as highlighted by Halo and supported by Masters (2020), who contend that in senior years, the emphasis is on content-driven strategies over skill development. The high-stakes nature of these tests not only impacts students but also has ramifications for teachers and schools, as Halo, Beta, and Zeta expressed regarding their teaching and assessment practices. Zeta's concerns

highlight a common challenge: the pressure to 'teach to the test' limiting the value of collaborative activities, implying that classroom activities that do not mirror individual tests have less value and are, therefore, less important. Despite these challenges, teachers and students generally found collaborative testing to be a positive experience and an effective method to enhance learning.

While high-stakes tests dominate teacher practice and student perceptions of alternative assessment methods, I argue that collaborative testing can enhance learning and students' 21st-century skills while underscoring the need for a more balanced approach to assessment. Through the teachers' and students' perspectives of collaborative testing, particularly the positive perceptions of Halo's summative test, this study details the successful integration of collaborative testing within performance-driven contexts. These findings further support the broader argument of my thesis that collaborative testing, when aligned with diffractive analysis, can offer a balanced method of assessment that promotes conceptual understanding and 21st-century skill development while accommodating the pressures of standardised testing.

Innovative Methodological Contribution: Diffractive Ethnography in Educational Research

The significant contribution of this thesis to the methodological literature is the application of diffractive ethnography, which moves beyond human-centred methods to explore how all material conditions influenced learning and shaped students' and teachers' perceptions of collaborative testing in three high school science classes. Seeking a qualitative method that acknowledges the significance of all elements in the learning process, I adopted diffractive ethnography grounded in Karen Barad's (2003, 2007) concept of agential realism, a theoretical approach that acknowledges the inseparability of our reality, where we exist in spacetimematter (Barad, 2007) alongside all components of that environment. This study moves beyond traditional separations between human and nonhuman, object and subject, outside and inside.

Diffractive ethnography recognises that educational phenomena, such as collaborative testing, are emergent practices that come into being through complex material-discursive entanglements, extending beyond the focus on discursive elements. This thesis offers a unique perspective by engaging in diffractive ethnography, acknowledging all elements involved in the intra-action: high-stakes testing, temperature, test duration, group dynamics, students, and teachers (Halo, Beta and Zeta) to cultivate a more nuanced understanding of the impact of collaborative testing on learning and perceptions. Unlike other methodologies, a diffractive approach recognises that the classroom is not a static environment but is shaped and reshaped continuously throughout the class and day by the

entanglement between students, teachers and the classroom environment. This recognition enabled a deep contextual understanding of how students and teachers came to view collaborative testing, revealing insights that may remain out of focus in more traditional educational research practices.

Furthermore, a diffractive methodology articulates specific inclusions and exclusions that contextualise each class in that space and time, which are crucial to understanding the collaborative testing experience. By deliberately selecting the inclusions and exclusions, such as the school (co-educational), year (11 and 12), and subject area (senior science), I explored how these factors collectively shaped the collaborative testing experience. Through their intra-actions with collaborative tests and the classroom environment, the teachers and students in this study demonstrated the impact of assessments on students and how the environment influenced their learning and engagement. Thus, diffractive ethnography is a transformative practice as it contributes to the ongoing (re)configuration of what it means to teach and learn, acknowledging and embracing the intra-connectedness of all elements within the educational landscape. This synthesis contributes to the growing body of research advocating for more holistic approaches to teaching and assessment while extending the theoretical application of diffractive ethnography in educational settings.

Implications and Possibilities for Future Research

I critically advance the field by promoting the value of using a diffractive analysis to gain a deep, contextual understanding of how students and teachers experience collaborative testing within a high school science environment. Through three rich descriptive case studies, it became clear that both students and teachers viewed collaborative testing as a generally positive experience that enhanced conceptual understanding and developed critical and collaborative skills. Teachers and students, however, acknowledged the barriers posed by external high-stakes testing, pointing to a need to address the tension between innovative practices such as collaborative testing and the demands of high-stakes testing, particularly prevalent in the senior years of high school.

The entanglements between these two testing methods explored through a diffractive analysis, reveal how pressures within a managerial education system shape perceptions and practices. As students' progress into their senior years, as discussed in Chapter 8, the emphasis shifts towards performance and accountability, which influences teaching and assessment strategies. The focus on accountability measures can deter teachers' willingness to try innovative practices due to concerns over how such methods might impact individual grading systems and overall performance measures

(Hickey & Riddle, 2023). In an Australian context, the dominance of end-of-school exams and the Australian Tertiary Admission Rank (ATAR) often overshadow the potential considerations of other methods that deviate from external exams. It is essential to broaden the recognition of class-based assessments as valid and realisable measures of students' performance to facilitate the inclusion of collaborative testing in this performance-driven environment. Although there has been some movement in Australia towards reviewing assessment practices in junior years, this shift is limited in senior years, where external assessments predominate. However, recognising that there are diverse ways to measure the performance of schools, teachers, and students can lead to the introduction of innovative practices to improve outcomes, underscoring the need to reconsider established practices of assessment to promote a balanced approach to enhance student outcomes and align with modern workplace demands.

In light of the findings from this thesis, the application of diffractive ethnography in high school settings is worthy of further exploration. Future studies could extend collaborative testing to a broader array of schools, enriching our understanding of the diverse impacts and challenges posed by collaborative testing across varied educational landscapes. Research could also investigate the long-term effects of collaborative testing over an entire school year, revealing how this practice evolves and what the long-term consequences of this method may be on students' conceptual understanding and influence on high-stakes testing outcomes.

The rich descriptive case studies of Halo, Beta and Zeta contextualise their unique environments and demographic characteristics. They provide a road map for future researchers and the explicit inclusions and exclusions in Chapter 4. Extending the benefits identified in higher education collaborative testing studies, such as a deeper conceptual understanding, critical thinking, and collaborative skills, to a high school setting introduces a new perspective to collaborative testing literature and science education, confirming the need for educational policies and practices to adapt to the changing needs of students and the workforce. This study illustrates that while collaborative testing enhances student conceptual understanding and develops collaborative skills, its successful implementation is contingent upon several factors that influence peer discussions: active engagement, a supportive environment, prior collaborative experience, clear accountability and positive group dynamics. Furthermore, it is crucial to articulate the material elements that constitute the environment to fully capture all factors influencing the collaborative testing experience. Careful consideration of these elements is necessary for collaborative testing to be effectively embraced within high school settings.

Conclusion

I embarked on this thesis to demonstrate that collaborative testing could be a viable supplementary method to in-class testing while enhancing high school students' conceptual understanding and collaborative skills. Returning to my opening remark in Chapter 1, 'if we keep doing the same thing, then we will get the same outcomes,' I have found evidence supporting the pedagogical philosophy that employing different assessment methods can lead to better learning outcomes. My research highlights the positive outcomes of collaborative testing, which suggests a need to reconsider current assessment practices governed by education departments. Although NESA encourages diversity in end-of-school assessments, aiming to broaden the range of methods beyond traditional exams, this intention has not been realised in classroom practice. As the first in-depth study on collaborative testing in high schools, this research provides practical evidence of collaborative testing as an assessment method and a synthesis of practices to guide teachers on its implementation. Through the participation of Beta, Halo, Zeta and their students, this study has illustrated the benefits, challenges and feasibility of collaborative testing in high school science classes. The complex intra-actions that emerged through the three classes and tests revealed how material factors influence student learning. This analysis has deepened our understanding of collaborative testing's significance as a pedagogical tool and outlined effective strategies for its implementation. Through these intra-actions and the insightful comments from students and teachers, it is clear that developing collaborative testing as a viable supplement to high-stakes tests supports transformative learning for students.

References

Abrams, L. S. (2010). Sampling 'hard to reach' populations in qualitative research: The case of incarcerated youth. *Qualitative Social Work: Research and Practice, 9*(4), 536–550. https://doi.org/10.1177/1473325010367821

Acosta, S., Garza, T., Hsu, H.-Y., Goodson, P., Padron, Y., Goltz, H., & Johnston, A. (2020). The accountability culture: A systematic review of high-stakes testing and English learners in the United States during No Child Left Behind. *Educational Psychology Review, 32*, 327–352.

Ahmed, S. (2006). *Orientations toward objects. In queer phenomenology* (pp. 25–63). Duke University Press., https://doi.org/10.1215/9780822388074-002

Akiba, M., & Liang, G. (2016). Effects of teacher professional learning activities on student achievement growth. *The Journal of Educational Research, 109*(1), 99–110. https://doi.org/10.1080/00220671.2014.924470

Alexander, R. J. (2010). *Children, Their World, Their Education: Final Report of the Cambridge Primary Review*. Routledge.

Allchin, D., & Zemplén, G. Á. (2020). Finding the place of argumentation in science education: Epistemics and whole science. *Science Education, 104*, 907–933. https://doi.org/10.1002/sce.21589

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice, 27*(4), 350–372. https://doi.org/10.1080/0969594X.2019.1571992

Appel, M. (2020). Performativity and the demise of the teaching profession: The need for rebalancing in Australia. *Asia-Pacific Journal of Teacher Education, 48*(3), 301–315. https://doi.org/10.1080/1359866X.2019.1644611

Arroba, T., & Bell, L. (2001). *Staying sane: Managing the stress of caring*. Age Concern.

Asikainen, H., & Gijbels, D. (2017). Do students develop towards more deep approaches to learning during studies? A systematic review on the development of students' deep and surface approaches to

learning in higher education. *Educational Psychology Review, 29*, 205–234. https://doi.org/10.1007/s10648-017-9406-6

Au, W. (2007). High-Stakes testing and curricular control: A qualitative meta synthesis. *Educational Researcher, 36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

Au, W. (2008). Devising inequality: A Bernsteinian analysis of high-stakes testing and social reproduction in education. *British Journal of Sociology of Education, 29*(6), 639–651. https://doi.org/10.1080/01425690802423312

Au, W. (2011). Teaching under the new Taylorism: high-stakes testing and the standardisation of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25–45. https://doi.org/10.1080/00220272.2010.521261

Au, W. (2022). *Unequal By Design: High-Stakes testing and The Standardisation of Inequality* (2nd ed.). Routledge., https://doi.org/10.4324/9781003005179

Australian Curriculum Assessment and Reporting Authority. ACARA. (2015). General capabilities. http://www.australiancurriculum.edu.au/GeneralCapabilities/Pdf/Critical-and-creative-thinking

Baeten, M., Kyndt, E., Struyven, K., & Dochy, F. (2010). Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review, 5*(3), 243–260. https://doi.org/10.1016/j.edurev.2010.06.001

Baker, E. L., Chung, G. K. W. K., & Cai, L. (2016). Assessment gaze, refraction, and blur: The course of achievement testing in the past 100 years. *Review of Research in Education, 40*(1), 94–142. https://doi.org/10.3102/0091732X16679806

Ball, S. J. (2016). Subjectivity as a site of struggle: Refusing neoliberalism? *British Journal of Sociology of Education, 37*(8), 1129–1146. https://doi.org/10.1080/01425692.2015.1044072

Ball, S. J. (2017). *Edu.net: Globalisation and education policy mobility* (1st ed.). Routledge., https://doi.org/10.4324/9781315630717

Ball, S. J. (2021). *Policy, inequality and advantage. In the education debate* (4th ed., pp. 161–208). Bristol University Press.

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

Barad, K. (2003). Posthumanist Performativity: Toward an understanding of how matter comes to matter. *Signs (Chicago, Ill.), 28*(3), 801–831. https://doi.org/10.1086/345321

Barad, K. M. (2007). *Meeting the universe halfway quantum physics and the entanglement of matter and meaning*. Duke University Press., https://doi.org/10.2307/j.ctv12101zq

Barad, K. M. (2010). Entanglements and Hauntological relations of inheritance: dis/continuities, spacetime enfolding's, and justice-to-come. *Derrida Today, 3*(2), 240–268. https://doi.org/10.3366/drt.2010.0206

Barad, K. (2014). Diffracting diffraction: Cutting together-apart. *Parallax, 20*(3), 168–187. https://doi.org/10.1080/13534645.2014.927623

Barad, K., & Gandorfer, D. (2021). Political desiring's: Yearnings for mattering (,) differently. *Theory & Event, 24*(1), 14–66. https://doi.org/10.1353/tae.2021.0002

Barla, J. (2023). Beyond reflexivity and representation: diffraction as a methodological sensitivity in science studies. *Distinktion: Journal of Social Theory*, 1-23. https://doi.org/10.1080/1600910X.2021.1934506

Belcastro, A. T. (2015, March). Rebalancing quality education in a democratic society. *Creative Education, 6*(4), 428–439. https://doi.org/10.4236/ce.2015.64043

Beninghof, A., & Leensvaart, M. (2016). Co-teaching to support ELLs. *Educational Leadership, 73*(5), 70–73.

Bennett, J., Tognolini, J., & Pickering, S. (2012). Establishing and applying performance standards for curriculum-based examinations. *Assessment in Education: Principles, Policy & Practice, 19*(3), 321–339. https://doi.org/10.1080/0969594X.2011.614219

Bergstedt, B. (2021). The ontology of becoming: To research and become with the world. *Education Sciences, 11*, 491. https://doi.org/10.3390/educsci11090491

Berliner, D. (2011). Rational responses to high stakes testing: The case of curriculum narrowing and the harm that follows. *Cambridge Journal of Education, 41*(3), 287–302. https://doi.org/10.1080/0305764X.2011.607151

Bernstein, B. (1990). *The structuring of pedagogic discourse: Class, codes & control* (Vol. IV). Routledge.

Bernstein, D. A. (2018). Does active learning work? A good question, but not the right one. *Scholarship of Teaching and Learning in Psychology, 4*(4), 290–307. https://doi.org/10.1037/stl0000124

Bhagat, A. (2017). Inquiry-based learning, Assessing students' science inquiry skills. *Teacher, 11*(4), 34–37.

Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability, 21*(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9

Biesta, G. (2015). Resisting the seduction of the global education measurement industry: Notes on the social psychology of PISA. *Ethics and Education, 10*(3), 348–360. https://doi.org/10.1080/17449642.2015.1106030

Biesta, G. (2015). On the two cultures of educational research, and how we might move ahead: Reconsidering the ontology, axiology and praxeology of education. *European educational research journal EERJ, 14*(1), 11-22. https://doi.org/10.1177/1474904114565162

Biesta, G. (2015b). What is education for? On good education, teacher judgement, and educational professionalism. *European Journal of Education, 50*(1), 75–87. https://doi.org/10.1111/ejed.12109

Biesta, G. (2020). Risking ourselves in education: Qualification, socialisation, and subjectification revisited. *Educational Theory, 70*, 89–104. https://doi.org/10.1111/edth.12411

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Biggs, J. (1999). What the Student Does: Teaching for enhanced learning. *Higher Education Research & Development, 18*, 57–75. https://doi.org/10.1080/0729436990180105

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17–66). Springer Netherlands., https://doi.org/10.1007/978-94-007-2324-5_2

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice, 17*(2), 215–232. https://doi.org/10.1080/09695941003696016

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, 25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Black, S., & Allen, J. D. (2018). Part 5: Learning is a social Act. *The Reference Librarian, 59*(2), 76–91. https://doi.org/10.1080/02763877.2017.1400932

Blazer, C. (2011). *Unintended Consequences of High Stakes testing*

Blom, S. (2022). ReTurning Learning: exploring childhoodnature through posthumanism [Southern Cross University]. https://doi.org/10.25918/thesis.250

Bodie, G. D., Vickery, A. J., Cannava, K., & Jones, S. M. (2015). The role of "Active Listening" in informal helping conversations: Impact on perceptions of listener helpfulness, sensitivity, and supportiveness and discloser emotional improvement. *Western Journal of Communication, 79*(2), 151–173. https://doi.org/10.1080/10570314.2014.943429

Board of Studies Teaching and Education Standards. BOSTES. (2016a). *Stronger HSC Standards: Blueprint*. NSW Government. https://educationstandards.nsw.edu.au/wps/wcm/connect/d210fd41-8c61-4754-aa45-7476b9305b1d/stronger-hsc-standards-bostes-blueprint.pdf?MOD=AJPERES&CVID

Boghossian, P. (2006). Behaviourism, constructivism, and Socratic pedagogy. *Educational Philosophy and Theory, 38*(6), 713–722. https://doi.org/10.1111/j.1469-5812.2006.00226.x

Bonem, E. M., Fedesco, H. N., & Zissimopoulos, A. N. (2020). What you do is less important than how you do it: The effects of learning environment on student outcomes. *Learning Environments Research, 23*(1), 27–44. https://doi.org/10.1007/s10984-019-09289-8

Bowers, J. M., & Moyer, A. (2017). Effects of school start time on students' sleep duration, daytime sleepiness, and attendance: A meta-analysis. *Sleep Health, 3*, 423–431. https://doi.org/10.1016/j.sleh.2017.08.004 PMID:29157635

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review, 42*(1), 39–55. https://doi.org/10.1080/02796015.2013.12087490

Bozalek, & Fullagar. (2021). *Agential Cut, Navigating the Postqualitative, New Materialist and Critical Posthumanist Terrain Across Disciplines: An Introductory Guide* (K. Murris Ed.). Milton: Taylor & Francis Group.

Bozalek, V., & Zembylas, M. (2017). Diffraction or reflection? Sketching the contours of two methodologies in educational research. *International Journal of Qualitative Studies in Education : QSE, 30*(2), 111–127. https://doi.org/10.1080/09518398.2016.1201166

Braidotti, R. (2019). A theoretical framework for the critical posthumanities. *Theory, Culture & Society, 36*(6), 31–61. https://doi.org/10.1177/0263276418771486

Breedlove, W., Burkett, T., & Winfield, I. (2004). Collaborative testing and test anxiety. *The Journal of Scholarship of Teaching and Learning, 4*(2), 33–42.

Bremert, H., Stoff, A., & Boesdorfer, S. (2020). Collaborative assessments: Learning science and collaborative skills during summative testing. *Science Teacher (Normal, Ill.), 87*(9), 32–37. https://doi.org/10.1080/00368555.2020.12293541

Brinkmann, S. (2011). *Interviewing and the production of the conversational self.* In N. Denzin & M. Giardina (Eds.), Qualitative inquiry and global crises (pp. 56–75). Left Coast Press.

Brooks, D. C. (2012). Space and Consequences: The impact of different formal learning spaces on instructor and student behaviour. *Journal of Learning Spaces, 1*(2).

Bryant, L. R. (2016). Phenomenon and thing: Barad's performative Ontology. *Rhizomes, 30*, 1–11. https://doi.org/10.20415/rhiz/030.e11

Burgess-Allen, J., & Owen-Smith, V. (2010). Using mind mapping techniques for rapid qualitative data analysis in public participation processes. *Health Expectations, 13*(4), 406–415. https://doi.org/10.1111/j.1369-7625.2010.00594.x PMID:20550595

Burkett, T. (2018). Norm-Referenced Testing and Criterion-Referenced Testing. In J. I. Liontas (Ed.), *The TESOL Encyclopedia of English Language Teaching.* T. International Association and M. DelliCarpini., https://doi.org/10.1002/9781118784235.eelt0351

Burton, K. D., Lydon, J. E., D'Alessandro, D. U., & Koestner, R. (2006). The differential effects of intrinsic and identified motivation on well-being and performance: Prospective, experimental, and implicit approaches to self-determination theory. *Journal of Personality and Social Psychology, 91*(4), 750–762. https://doi.org/10.1037/0022-3514.91.4.750 PMID:17014297

Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice, 2*, 14. https://doi.org/10.1186/s42466-020-00059-z PMID:33324920

Cairns, R. (2021). Exams tested by Covid-19: An opportunity to rethink standardised senior secondary examinations. *Prospects, 51*, 331–345. https://doi.org/10.1007/s11125-020-09515-9 PMID:33110277

Calvert-Minor, C. (2014). Epistemological misgivings of Karen Barad's "Posthumanism.". *Human Studies, 37*(1), 123–137. https://doi.org/10.1007/s10746-013-9285-x

Cantwell, R., Sousou, J., Jadotte, Y., Pierce, J., & Akioyamen, L. (2016). Collaborative testing for improving student learning outcomes and test-taking performance in Higher education: A systematic review. Campbell Collaboration.

Cao, Y., & Porter, L. (2017b). *Impact of performance level and group composition on student learning during collaborative exams*. Proceedings of the 2017 ACM Conference on innovation and technology in computer science education, 2017, 152–157. https://doi.org/10.1145/3059009.3059024

de Carvalho Filho, M. K. (2010). Assessing changes in performance and monitoring processes in individual and collaborative tests according to students' metacognitive skills. *The European Journal of Cognitive Psychology, 22*(7), 1107–1136. https://doi.org/10.1080/09541440903336555

Charteris, J., Nye, A., & Jones, M. (2019). Posthumanist ethical practice: Agential cuts in the pedagogic assemblage. *International Journal of Qualitative Studies in Education : QSE, 32*(7), 909–928. https://doi.org/10.1080/09518398.2019.1609124

Chia, C. K., Ghavifekr, S., & Razak, A. Z. A. (2020). Online interview tools for qualitative data collection during Covid-19 pandemic; Review of web conferencing platforms' functionality. *Malaysian Journal of Qualitative Research, 7*(1), 95–106.

Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: A systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review, 33*, 1499–1522. https://doi.org/10.1007/s10648-021-09613-w

Chen, F., Lui, A. M., Andrade, H., Valle, C., & Mir, H. (2017). Criteria-referenced formative assessment in the arts. *Educational Assessment, Evaluation and Accountability, 29*, 297–314. https://doi.org/10.1007/s11092-017-9259-z

Cheng, Y. (2021). Cross-Cultural Differences in Collaborative Learning and Relevant Factors. *Advances in Social Science, Education and Humanities Research, 594* (Proceedings of the 2021 3rd International Conference on Literature, Art and Human Development (ICLAHD 2021)), 689-696.

Centre for Education Statistics and Evaluation. (2023), Schools and students: 2022 Statistical Bulletin, NSW Department of Education, education.nsw.gov.au/cese

Cheryan, S., Ziegler, S. A., Plaut, V. C., & Meltzoff, A. N. (2014). Designing classrooms to maximise student achievement. *Policy Insights from the Behavioural and Brain Sciences, 1*(1), 4–12. https://doi.org/10.1177/2372732214548677

Cho, H. (2017). The effects of summer heat on academic achievement: A cohort analysis. *Journal of Environmental Economics and Management, 83*, 185–196. https://doi.org/10.1016/j.jeem.2017.03.005

Chubb, S. R. (2020). The impact of a new national curriculum on subject leaders in primary and secondary schools. ProQuest Dissertations & Theses.

Clark, U. (2005). Bernstein's theory of pedagogic discourse: Linguistics, educational policy and practice in the UK English/literacy classroom. *English Teaching, 4*(3), 32–47.

Cobussen, M. (2022). *Engaging with Everyday Sounds*. Open Book Publishers., https://doi.org/10.11647/OBP.0288

Comber, B. (2012). Mandated literacy assessment and the reorganisation of teachers' work: Federal policy, local effects. *Critical Studies in Education, 53*(2), 119–136. https://doi.org/10.1080/17508487.2012.672331

Cooke, J. E., Weir, L., & Clarkston, B. (2019). Retention following two-stage collaborative exams depends on timing and student performance. *CBE Life Sciences Education, 18*(2), ar12. https://doi.org/10.1187/cbe.17-07-0137 PMID:31008674

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research : planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology, 11*(1), 100–100. https://doi.org/10.1186/1471-2288-11-100 PMID:21707982

Crowley, S. J., Wolfson, A. R., Tarokh, L., & Carskadon, M. A. (2018). An update on adolescent sleep: New evidence informing the perfect storm model. *Journal of Adolescence (London, England.), 67*(1), 55–65. https://doi.org/10.1016/j.adolescence.2018.06.001

Cumming, J. J. (2010). Classroom assessment in policy context *(Australia)*. In P. Peterson, E. Baker, & B. McGaw (Eds.*), International Encyclopedia of Education* (3rd ed., pp. 417–424). Elsevier., https://doi.org/10.1016/B978-0-08-044894-7.01101-5

Cumming, J. J. (2020). Senior secondary school assessment and standard-setting in Queensland, Australia: Social context and paradigmatic change. *Assessment in Education: Principles, Policy & Practice, 27*(2), 160–177. https://doi.org/10.1080/0969594X.2019.1684877

Cumming, J., Jackson, C. A., Day, C., Maxwell, G. S., Adie, L., Lingard, B., . . . Heck, E. (2018). *2018 Queensland NAPLAN review: School and system perceptions report and literature review*.

Cumming, J., Wyatt-Smith, C. M., & Colbert, P. J. (2016). Students at risk and NAPLAN: the collateral damage. (2015). In *National Testing in Schools*. Routledge. https://doi.org/10.4324/9781315659312

Curşeu, P. L., & Pluut, H. (2013). Student groups as learning entities: The effect of group diversity and teamwork quality on groups' cognitive complexity. *Studies in Higher Education, 38*(1), 87–103. https://doi.org/10.1080/03075079.2011.565122

Curşeu, P. L., & Sari, K. (2015). The effects of gender variety and power disparity on group cognitive complexity in collaborative learning groups. *Interactive Learning Environments, 23*(4), 425–436. https://doi.org/10.1080/10494820.2013.788029

Curşeu, P. L., Chappin, M. M. H., & Jansen, R. J. G. (2018). Gender diversity and motivation in collaborative learning groups: The mediating role of group discussion quality. *Social Psychology of Education, 21*(2), 289–302. https://doi.org/10.1007/s11218-017-9419-5

Dacey, A. (2020). Come now, let us reason together: Cognitive bias, individualism, and interactionism in critical thinking education. *Informal Logic, 40*(1), 47–76. https://doi.org/10.22329/il.v40i1.6024

Daliri-Ngametua, R., & Hardy, I. (2022). The devalued, demoralised and disappearing teacher: The nature and effects of datafication and performativity in schools. *Education Policy Analysis Archives*, *30*. https://doi.org/10.14507/epaa.30.6174

Dancy, M., Henderson, C., & Turpen, C. (2016). How faculty learn about and implement research-based instructional strategies: The case of Peer Instruction. *Physical Review. Physics Education Research, 12*(1), 010110. https://doi.org/10.1103/PhysRevPhysEducRes.12.010110

DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation, 61*, 159–169. https://doi.org/10.1016/j.stueduc.2019.03.012

Demir-Yildiz, C., & Tatik, R. S. (2019). Impact of Flexible and Non-Flexible Classroom Environments on Learning of Undergraduate Students. *European Journal of Educational Research, 8*(4), 1159–1173. https://doi.org/10.12973/eu-jer.8.4.1159

Denzin, N. K. (2001). The reflexive interview and a performative social science. *Qualitative Research, 1*(1), 23–46. https://doi.org/10.1177/146879410100100102

Dikker, S., Haegens, S., Bevilacqua, D., Davidesco, I., Wan, L., Kaggen, L., McClintock, J., Chaloner, K., Ding, M., West, T., & Poeppel, D. (2020). Morning brain: Real-world neural evidence that high school class times matter. *Social Cognitive and Affective Neuroscience, 15*(11), 1193–1202. https://doi.org/10.1093/scan/nsaa142 PMID:33068110

Ding, N., Bosker, R. J., & Harskamp, E. G. (2011). Exploring gender and gender pairing in the knowledge elaboration processes of students using computer-supported collaborative learning. *Computers & Education, 56*, 325–336. https://doi.org/10.1016/j.compedu.2010.06.004

Dixson, D., & Worrell, F. (2016). Formative and summative assessment in the classroom. *Theory into Practice, 55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Dolmans, D. H. J. M., Loyens, S. M. M., Marcq, H., & Gijbels, D. (2016). Deep and surface learning in problem-based learning: A review of the literature. *Advances in Health Sciences Education : Theory and Practice, 21*(5), 1087–1112. https://doi.org/10.1007/s10459-015-9645-6 PMID:26563722

Doolittle, P., Wojdak, K., & Walters, A. (2023). Defining active learning: A restricted systemic review. *Teaching & Learning Inquiry, 11*. Advance online publication. https://doi.org/10.20343/teachlearninqu.11.25

Earl, S. R., Taylor, I. M., Meijen, C., & Passfield, L. (2023). Trajectories in cognitive engagement, fatigue, and school achievement: The role of young adolescents' psychological need satisfaction. *Learning and Individual Differences, 101*, 102248-. https://doi.org/10.1016/j.lindif.2022.102248

Eastridge, J. A., & Benson, W. L. (2020). Comparing two models of collaborative testing for teaching statistics. *Teaching of Psychology, 47*(1), 68–73. https://doi.org/10.1177/0098628319888113

Eastwood, J. L., Kleinberg, K. A., & Rodenbaugh, D. W. (2020). Collaborative testing in medical education: student perceptions and long-term knowledge retention. *Medical Science Educator, 30*(2), 737–747. https://doi.org/10.1007/s40670-020-00944-x PMID:34457732

Education Council. (2020). *Looking To The Future, Report of the Review of Senior Secondary Pathways into Work. Further Education and Training*. https://www.education.gov.au/quality-schools-package/resources/looking-future-report-review-senior-secondary-pathways-work-further-education-and-training

Efu, S. I. (2019). Exams as learning tools: A comparison of traditional and collaborative assessment in higher education. *College Teaching, 67*(1), 73–83. https://doi.org/10.1080/87567555.2018.1531282

Ellery, K. (2017). Framing of transitional pedagogic practices in the sciences: Enabling access. *Teaching in Higher Education, 22*(8), 908–924. https://doi.org/10.1080/13562517.2017.1319812

Emslander, V., Levy, J., Scherer, R., & Fischbach, A. (2022). Value-added scores show limited stability over time in primary school. *PLoS One, 17*(12), e0279255. https://doi.org/10.1371/journal.pone.0279255 PMID:36576932

Evans, K. (2016). Beyond a logic of quality: Opening space for material-discursive practices of 'readiness' in early years education. *Contemporary Issues in Early Childhood, 17*, 65–77. Advance online publication. https://doi.org/10.1177/1463949115627904

Farvis, J., & Hay, S. (2020). Undermining teaching: How education consultants view the impact of high-stakes test preparation on teaching. *Policy Futures in Education, 18*(8), 1058–1074. https://doi.org/10.1177/1478210320919541

Fearnley, C. J. (2022). Mind mapping in qualitative data analysis: Managing interview data in interdisciplinary and multi-sited research projects. *Geo : Geography and Environment, 9*(1), e00109. https://doi.org/10.1002/geo2.109

Fernández García, C. M., Torío-López, S., García-Pérez, O., & Inda-Caro, M. (2019). Parental support, self-efficacy beliefs, outcome expectations and interests in science, technology, engineering and mathematics (STEM). *Universitas Psychologica, 18*(2), 1–15. https://doi.org/10.11144/Javeriana.upsy18-2.psse

Fernández García, C.-M., Maulana, R., Inda-Caro, M., Helms-Lorenz, M., & García-Pérez, O. (2019). Student perceptions of secondary education teaching effectiveness: General profile, the role of personal factors, and educational level. *Frontiers in Psychology, 10*, 533. https://doi.org/10.3389/fpsyg.2019.00533 PMID:30915009

Fisher, D., Frey, N., & Lapp, D. (2012). Building and activating students' background knowledge: It's what they already know that counts: Teachers must assess and build on the background knowledge students possess. *Middle School Journal, 43*, 22–31. https://doi.org/10.1080/00940771.2012.11461808

Flyvbjerg, B. (2011). Case study. In N. K. Denzin & Y. S. Lincoln (Eds.), The *Sage Handbook of Qualitative Research* (4th ed., pp. 301–316). Sage.

Flick, U. (2018). *The Sage handbook of qualitative data collection*. SAGE Publications Ltd., https://doi.org/10.4135/9781526416070

Fournier, K. A., Couret, J., Ramsay, J. B., & Caulkins, J. L. (2017). Using collaborative two-stage examinations to address test anxiety in a large enrollment gateway course. *Anatomical Sciences Education, 10*, 409–422. https://doi.org/10.1002/ase.1677 PMID:28135034

Fournier, M., d'Arripe-Longueville, F., Rovere, C., Easthope, C. S., Schwabe, L., El Methni, J., & Radel, R. (2017). Effects of circadian cortisol on the development of a health habit. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association, 36*(11), 1059–1064. https://doi.org/10.1037/hea0000510

Fox, N. J., & Alldred, P. (2017). *Sociology and the new materialism : theory, research, action.* SAGE Publications Ltd., https://doi.org/10.4135/9781526401915

Fray, L., Jaremus, F., Gore, J., Miller, A., & Harris, J. (2023). Under pressure and overlooked: The impact of COVID-19 on teachers in NSW public schools. *Australian Educational Researcher, 50*(3), 701–727. https://doi.org/10.1007/s13384-022-00518-3 PMID:35431405

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics.

Proceedings of the National Academy of Sciences of the United States of America, 111(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111 PMID:24821756

Fullan, M. (2016). The elusive nature of whole system improvement in education. *Journal of Educational Change*, *17*(4), 539–544. https://doi.org/10.1007/s10833-016-9289-1

Fuller, K., & Stevenson, H. (2019). Global education reform: Understanding the movement. *Educational Review, 71*(1), 1–4. https://doi.org/10.1080/00131911.2019.1532718

Geisinger, K. F. (2016). 21st century skills: What are they and how do we assess them? *Applied Measurement in Education, 29*(4), 245–249. https://doi.org/10.1080/08957347.2016.1209207

Gilley, R. & Clarkston, (2014). The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction. 14*(2), 197–213. https://doi.org/info:doi/

Gillies, R. M. (2014). Cooperative learning: Developments in research, International. *Journal of Educational Psychology, 3*(2), 125–140.

Gillies, R. (2016). Cooperative learning : Review of research and practice. *The Australian Journal of Teacher Education, 41*(3), 39–54. https://doi.org/10.14221/ajte.2016v41n3.3

Gillies, R. M., & Boyle, M. (2010). Teachers' reflections on cooperative learning: Issues of implementation. *Teaching and Teacher Education, 26*(4), 933–940. https://doi.org/10.1016/j.tate.2009.10.034

Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high- and low-performing students. *Advances in Physiology Education, 32*(4), 274–278. https://doi.org/10.1152/advan.00101.2007 PMID:19047503

Göloğlu Demir, C., & Kaplan Keleş, Ö. (2021). The impact of high-stakes testing on the teaching and learning processes of mathematics. *Journal of Pedagogical Research, 5*(2), 119-137. https://doi.org/10.33902/JPR.2021269677

Gonzalez, A., Peters, M. L., Orange, A., & Grigsby, B. (2017). The influence of high-stakes testing on teacher self-efficacy and job-related stress. *Cambridge Journal of Education, 47*(4), 513–531. https://doi.org/10.1080/0305764X.2016.1214237

Gonski, D., Arcus, T., Boston, K., Gould, V., Johnson, W., O'Brien, L., … Roberts, M. (2018). *Through growth to achievement: The report of the review to achieve educational excellence in Australian schools*. Canberra: Commonwealth of Australia.

Gottfried, M. A. (2014). The achievement effects of tardy classmates: Evidence in urban elementary schools. *School Effectiveness and School Improvement, 25*(1), 3–28. https://doi.org/10.1080/09243453.2012.728135

Gray, L., Wong-Wylie, G., Rempel, G., & Cook, K. (2020). Expanding qualitative research interviewing strategies: Zoom video communications. *The Qualitative Report, 25*, 9. Advance online publication. https://doi.org/10.46743/2160-3715/2020.4212

Green, J., Absoud, M., Grahame, V., Malik, O., Simonoff, E., Le Couteur, A., & Baird, G. (2018). Pathological demand avoidance: Symptoms but not a syndrome. *The Lancet. Child & Adolescent Health, 2*(6), 455–464. https://doi.org/10.1016/S2352-4642(18)30044-0 PMID:30169286

Groppo, V., & Kraehnert, K. (2017). The impact of extreme weather events on education. *Journal of Population Economics, 30*(2), 433–472. https://doi.org/10.1007/s00148-016-0628-6

Gullion, J. S. (2018). *Diffractive ethnography: social sciences and the ontological turn*. Routledge., https://doi.org/10.4324/9781351044998

Guilfoyle, L., Hillier, J., & Fancourt, N. (2023). Students' argumentation in the contexts of science, religious education, and interdisciplinary science-religious education scenarios. *Research in Science & Technological Education, 41*(2), 759–776. https://doi.org/10.1080/02635143.2021.1947223

Guetterman, T. C. (2015). Descriptions of sampling practices within five approaches to qualitative research in education and the health sciences. *Forum Qualitative Social Research, 16*(2), 25.

Gurr, D. (2020). Australia: The Australian Education System. In H. Ärlestig & O. Johansson (Eds.), *Educational Authorities and the Schools. Educational Governance Research* (Vol. 13). Springer., https://doi.org/10.1007/978-3-030-38759-4_17

Hämäläinen, R., & Vähäsantanen, K. (2011). Theoretical and pedagogical perspectives on orchestrating creativity and collaborative learning. *Educational Research Review, 6*(3), 169–184. https://doi.org/10.1016/j.edurev.2011.08.001

Hammersley, M. (2018). What is ethnography? Can it survive? Should it? *Ethnography and Education, 13*(1), 1–17. https://doi.org/10.1080/17457823.2017.1298458

Hannaway, J., & Hamilton, L. S. (2008). Performance-based accountability policies: implications for school and classroom practices. *Political Science (Wellington, N.Z.)*. Advance online publication. https://doi.org/10.1037/e722482011-001

Hardy, I., & Boyle, C. (2011). My school? Critiquing the abstraction and quantification of education. *Asia-Pacific Journal of Teacher Education, 39*, 211–222. https://doi.org/10.1080/1359866X.2011.588312

Haraway, D. (1991). *Simians, Cyborgs, and Women: The Reinvention of Nature*. Routledge.

Haraway, D. J. (1997). *Modest Witness Second Millennium.FemaleMan Meets OncoMouse: Feminism and technoscience*. Routledge.

Harrison, H., Birks, M., Franklin, R., & Mills, J. (2017, January). Case study research: Foundations and methodological orientations. In *Forum qualitative Sozialforschung/Forum: qualitative social research* (Vol. 18, No. 1).

Hammond, M. (2017). Online collaboration and cooperation: The recurring importance of evidence, rationale and viability. *Education and Information Technologies, 22*, 1005–1024. https://doi.org/10.1007/s10639-016-9469-x

Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education: Principles, Policy & Practice, 14*(2), 185–199. https://doi.org/10.1080/09695940701478594

Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change, 21*(3), 393–420. https://doi.org/10.1007/s10833-020-09380-5 PMID:38624328

Hattie, J. A. C., & Donoghue, G. M. (2016). Learning strategies: A synthesis and conceptual model. *NPJ Science of Learning, 1*(1), 16013. https://doi.org/10.1038/npjscilearn.2016.13 PMID:30792898

Hayes, D., Talbot, D., & Mayes, E. (2020). Diffractive accounts of inequality in education: Making the effects of differences evident. *International Journal of Qualitative Studies in Education : QSE, 33*(3), 357–371. https://doi.org/10.1080/09518398.2019.1676481

Headrick, J., Harris-Reeves, B., & Daly-Olm, T. (2021). Collaborative testing in Sport and Exercise Degrees: A comparison of first and third year students' perceptions. (2021). *Journal of University Teaching & Learning Practice, 18*(6), 155–173. https://doi.org/10.53761/1.18.6.11

Hendrikx, W. (2021). Managerialism and its consequences for professional identity: A comparative analysis of GPs and secondary school teachers. *Administration & Society, 53*(8), 1178–1202. https://doi.org/10.1177/0095399721993744

Hernandez, K. (2023). *Activate: Deeper learning through movement, talk, and flexible classrooms*. Routledge. https://doi.org/10.4324/9781032680538 JISC (2009). *Designing spaces for effective learning: A guide to 21st century learning space design* (pp. 1–33). Bristol: HEFCE. http://www.jisc.ac.uk/media/documents/publications/learningspaces.pdf

Hickey, A., & Riddle, S. (2023). Performative enactments of pedagogy in the classroom: Strategies and tactics of relationality. *Pedagogy, Culture & Society*, 1–16. Advance online publication. https://doi.org/10.1080/14681366.2023.2187438

Hmelo-Silver, C. E., Chinn, C. A., Chan, C., & O'Donnell, A. M. (Eds.). (2013). The *International Handbook of Collaborative Learning*. https://doi.org/10.4324/9780203837290

Högberg, B., & Horn, D. (2022). National high-stakes testing, gender, and school stress in Europe: A difference-in-differences analysis. *European Sociological Review, 38*(6), 975–987. https://doi.org/10.1093/esr/jcac009

Holford, W. D. (2020). Knowledge and Power Across the Material-Discursive Practice of Agential Realism. In W. D. Holford (Ed.), *Managing Knowledge in Organizations: A Critical Pragmatic Perspective* (pp. 109–131). Springer International Publishing. https://doi.org/10.1007/978-3-030-41156-5_7

Howell, K. E. (2012). An introduction to the philosophy of methodology. *An Introduction to the Philosophy of Methodology*, 1-248.

Howie, P. & Bagnall, R. (2013). A critique of the deep and surface approaches to learning model. *Teaching in Higher Education, 18*(4), 389–400. https://doi.org/10.1080/13562517.2012.733689

Hultman, K., & Lenz Taguchi, H. (2010). Challenging anthropocentric analysis of visual data: A relational materialist methodological approach to educational research. *International Journal of Qualitative Studies in Education : QSE, 23*(5), 525–542. https://doi.org/10.1080/09518398.2010.500628

Hyett, N (PhD Candidate), Amanda Kenny Dr & Virginia Dickson-Swift Dr (2014) Methodology or method? A critical review of qualitative case study reports, International Journal of Qualitative Studies on Health and Well-being, 9:1, 23606, https://doi.org/10.3402/qhw.v9.23606

Hyndman, B., & Button, B. (2023). The Influences of Extreme Cold and Storms on Schoolchildren. In The Impact of Extreme Weather on School Education (pp. 60–77). Routledge. https://doi.org/10.4324/9781003103165-4

Ingham-Hill, E., Hewitt, A., Lester, A., & Bond, B. (2024). Morning compared to afternoon school-based exercise on cognitive function in adolescents. *Brain and Cognition, 175*, 106-135. https://doi.org/10.1016/j.bandc.2024.106135 PMID:38306762

Iordanou, K., & Rapanta, C. (2021). "Argue With Me": A Method for Developing Argument Skills. *Frontiers in Psychology*, *12*, 631203-. https://doi.org/10.3389/fpsyg.2021.631203

Jang, H., Lasry, N., Miller, K., & Mazur, E. (2017). Collaborative exams: Cheating? Or learning? *American Journal of Physics, 85*(3), 223–227. https://doi.org/10.1119/1.4974744

Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK? *Review of Education, 9*, e3270. https://doi.org/10.1002/rev3.3270

Jerrim, J., Allen, R., & Sims, S. (2024). High stakes assessments in primary schools and teachers' anxiety about work. *Educational Assessment, 29*(2), 59–74. https://doi.org/10.1080/10627197.2024.2350961

Johansson, L. (2016). Post-qualitative line of flight and the confabulative conversation: A methodological ethnography. *International Journal of Qualitative Studies in Education : QSE, 29*(4), 445–466. https://doi.org/10.1080/09518398.2015.1053157

Johnson, B. (2003). Teacher collaboration: Good for some, not so good for others. *Educational Studies, 29*(4), 337–350. https://doi.org/10.1080/0305569032000159651

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2000). *Active learning: Cooperation in the college classroom* (2nd ed.). Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2005). New developments in social interdependence theory. *Genetic, Social, and General Psychology Monographs, 131*(4), 285–358. https://doi.org/10.3200/MONO.131.4.285-358 PMID:17191373

Johnson, D., & Johnson, R. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher, 38*(5), 365–379. https://doi.org/10.3102/0013189X09339057

Johnston, D. W., Knott, R., Mendolia, S., & Siminski, P. (2021). Upside-down down-under: Cold temperatures reduce learning in Australia. *Economics of Education Review, 85*, 102172 https://doi.org/10.1016/j.econedurev.2021.102172

Juceviciene, P., & Vizgirdaite, J. (2012). Educational empowerment of collaborative learning at the university. *Social Sciences, 75*(1), 41–51. https://doi.org/10.5755/j01.ss.75.1.1589

Juelskjær, M., Plauborg, H., & Adrian, S. W. (2021). *Dialogues on agential realism : engaging in worldings through research practice.* Routledge.

Kaendler, C., Wiedmann, M., Leuders, T., Rummel, N., & Spada, H. (2016). Monitoring student interaction during collaborative learning: Design and evaluation of a training program for pre-service teachers. *Psychology Learning & Teaching, 15*(1), 44–64. https://doi.org/10.1177/1475725716638010

Kanevsky, L., Lo, C. O., & Marghelis, V. (2022). Individual or collaborative projects? Considerations influencing the preferences of students with high reasoning ability and others their age. *High Ability Studies, 33*(1), 87–119. https://doi.org/10.1080/13598139.2021.1903842

Kapitanoff, S. H. (2009). Collaborative testing: Cognitive and interpersonal processes related to enhanced test performance. *Active Learning in Higher Education, 10*(1), 56–70. https://doi.org/10.1177/1469787408100195

Kelly, A. V. (2004). *The Curriculum : Theory and Practice* (5th ed.). Sage Publications Ltd.

Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner; Doing critical participatory action research*. Springer Singapore., https://doi.org/10.1007/978-981-4560-67-2

Kennette, L. N., & Myatt, B. (2018). How the post-secondary classroom can benefit from positive psychology principles. *Psychology Teaching Review, 24*, 63–66. https://doi.org/10.53841/bpsptr.2018.24.1.63

Kennedy, T. J., & Sundberg, C. W. (2020). 21st Century Skills. In B. Akpan & T. J. Kennedy (Eds.), *Science Education in Theory and Practice: An Introductory Guide to Learning Theory* (pp. 479–496). Springer International Publishing., https://doi.org/10.1007/978-3-030-43620-9_32

Klenowski, V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation, 37*(1), 78–83. https://doi.org/10.1016/j.stueduc.2011.03.003

Klenowski, V., & Wyatt-Smith, C. (2012). The impact of high stakes testing: The Australian story. *Assessment in Education: Principles, Policy & Practice, 19*(1), 65–79. https://doi.org/10.1080/0969594X.2011.592972

Klykken, F. H. (2023). 'Are we going to do that now?' Orientations and response-abilities in the embodied classroom. *Pedagogy, Culture & Society, 31*(5), 1079–1095. https://doi.org/10.1080/14681366.2021.1998794

Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE Life Sciences Education, 12*(4), 645–654. https://doi.org/10.1187/cbe.13-05-0090 PMID:24297291

Koro-Ljungberg. Maclure, M., & Ulmer, J. (2018). D...a…t…a…, Data++, Data, and some problematics. In (pp. 462–484).

Koro-Ljungberg, M. (2016). *Reconceptualizing Qualitative Research: Methodologies without Methodology. Sage (Atlanta, Ga.)*. Advance online publication. https://doi.org/10.4135/9781071802793

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Krispenz, A., & Dickhäuser, O. (2018). Effects of an inquiry-based short intervention on state test anxiety in comparison to alternative coping strategies. *Frontiers in Psychology, 9*(201), 201. https://doi.org/10.3389/fpsyg.2018.00201 PMID:29515507

Kulasegaram, K., & Rangachari, P. K. (2018). Beyond "formative": Assessments to enrich student learning. *Advances in Physiology Education, 42*(1), 5–14. https://doi.org/10.1152/advan.00122.2017 PMID:29341810

Kuntz, A. M., & Presnall, M. M. (2012). Wandering the tactical: From interview to intraview. *Qualitative Inquiry, 18*(9), 732–744. https://doi.org/10.1177/1077800412453016

Lakes, R. D., & Carter, P. A. (2011). Neoliberalism and education: An introduction. *Educational Studies (Ames), 47*(2), 107–110. https://doi.org/10.1080/00131946.2011.556387

Larsson, K. (2017). Understanding and teaching critical thinking—A new approach. *International Journal of Educational Research, 84*, 32–42. https://doi.org/10.1016/j.ijer.2017.05.004

Lather, P. (2013). Methodology -21: What do we do in the afterward? *International Journal of Qualitative Studies in Education : QSE, 26*(6), 634–645. https://doi.org/10.1080/09518398.2013.788753

Le, H., Janssen, J., & Wubbels, T. (2017). Collaborative learning practices: Teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education, 48*(1), 103–122. https://doi.org/10.1080/0305764X.2016.1259389

Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional Science, 43*(5), 561–590. https://doi.org/10.1007/s11251-015-9348-7

van Leeuwen, A. V., & Janssen, J. (2019). A systematic review of teacher guidance during collaborative learning in primary and secondary education. *Educational Research Review, 27*, 71–89. https://doi.org/10.1016/j.edurev.2019.02.001

Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE Life Sciences Education, 11*(4), 392–401. https://doi.org/10.1187/cbe.12-04-0048 PMID:23222835

Lenz Taguchi, H. (2010). An intra-active pedagogy and its dual movements. In *Going Beyond the Theory/Practice Divide in Early Childhood Education* (pp. 112–138). Routledge. https://doi.org/10.4324/9780203872956-10

Lenz Taguchi, H. (2010). Learning and becoming in an onto-epistemology. In *Going Beyond the Theory/Practice Divide in Early Childhood Education* (pp. 64–84). Routledge. https://doi.org/10.4324/9780203872956-8

Lenz Taguchi, H. (2010). *Going Beyond the Theory/Practice Divide in Early Childhood Education: Introducing an Intra-Active Pedagogy (1st ed.). Routledge. https://doi.org/10.4324/9780203872956*

Levine, R. E., Borges, N. J., Roman, B. J. B., Carchedi, L. R., Townsend, M. H., Cluver, J. S., Frank, J., Morey, O., Haidet, P., & Thompson, B. M. (2018). High-stakes collaborative testing: Why not? *Teaching and Learning in Medicine, 30*(2), 133–140. https://doi.org/10.1080/10401334.2017.1365719 PMID:29220581

Levy, D., Svoronos, T., & Klinger, M. (2023). Two-stage examinations: Can examinations be more formative experiences? *Active Learning in Higher Education, 24*(2), 79–94. https://doi.org/10.1177/1469787418801668

Lewis, S., & Hardy, I. (2015). Funding, reputation and targets: The discursive logics of high-stakes testing. *Cambridge Journal of Education, 45*(2), 245–264. https://doi.org/10.1080/0305764X.2014.936826

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage Publications., https://doi.org/10.1016/0147-1767(85)90062-8

Lingard, B., & Sellar, S. (2016). The changing organizational and global significance of the OECD's education work. In K. Mundy, A. Green, B. Lingard, & A. Verger (Eds.), *Handbook of Global Education Policy* (pp. 357–373). John Wiley & Sons., https://doi.org/10.1002/9781118468005.ch19

Linnenbrink-Garcia, L., Patall, E. A., & Pekrun, R. (2016). Adaptive motivation and emotion in education. *Policy Insights from the Behavioral and Brain Sciences, 3*, 228–236. Advance online publication. https://doi.org/10.1177/2372732216644450

Llorens-Gámez, M., Higuera-Trujillo, J. L., Omarrementeria, C. S., & Llinares, C. (2022). The impact of the design of learning spaces on attention and memory from a neuroarchitectural approach: A systematic review. *Frontiers of Architectural Research, 11*(3), 542–560. https://doi.org/10.1016/j.foar.2021.12.002

Lobascher, S. (2011). What are the potential impacts of high-stakes testing on literacy Education in Australia? *Literacy Learning: The Middle Years, 19*(2), 9–19.

LoGiudice, A. B., Pachai, A. A., & Kim, J. A. (2015). Testing together: When do students learn more through collaborative tests? *Scholarship of Teaching and Learning in Psychology, 1*, 377–389. http://dx.doi.org.cupdx.idm.oclc.org/10.1037/shttps://doi.org/10.1037/stl0000041 https://doi.org/10.1037/stl0000041

Lok, B., McNaught, C., & Young, K. (2016). Criterion-referenced and norm-referenced assessments: Compatibility and complementarity. *Assessment & Evaluation in Higher Education, 41*(3), 450–465. https://doi.org/10.1080/02602938.2015.1022136

Lombardi, D., Shipley, T. F., Bailey, J. M., Bretones, P. S., Prather, E. E., Ballen, C. J., Knight, J. K., Smith, M. K., Stowe, R. L., Cooper, M. M., Prince, M., Atit, K., Uttal, D. H., LaDue, N. D., McNeal, P. M., Ryker, K., St. John, K., van der Hoeven Kraft, K. J., & Docktor, J. L. (2021). The Curious Construct of Active Learning. *Psychological Science in the Public Interest, 22*(1), 8–43. https://doi.org/10.1177/1529100620973974

Loyens, S. M. M., & Gijbels, D. (2008). Understanding the effects of constructivist learning environments: Introducing a multi-directional approach. *Instructional Science, 36*(5-6), 351–357. https://doi.org/10.1007/s11251-008-9059-4

Luke, A., Weir, K., & Woods, A. (2021). Development of a set of principles to guide a P–12 syllabus framework A report to the Queensland Studies Authority Queensland.

Lumsden, K., Bradford, J., & Goode, J. (2019). Reflexivity: theory, method and practice. Routledge., https://doi.org/10.4324/9781315692319

Lyons, T., & Quinn, F. (2010). *Choosing Science: Understanding the declines in senior high school science enrolments*. University of New England: National Centre of Science, ICT and Mathematics Education for Rural and Regional Australia (SiMERR Australia).

Mahon, K., Heikkinen, H. L. T., & Huttunen, R. (2019). Critical educational praxis in university ecosystems: Enablers and constraints. *Pedagogy, Culture & Society, 27*(3), 463–480. https://doi.org/10.1080/14681366.2018.1522663

Mahon, K., Francisco, S., & Kemmis, S. (2017). *Exploring education and professional practice: Through the lens of practice architectures.* Springer., https://doi.org/10.1007/978-981-10-2219-7

Mahoney, J., & Harris-Reeves, B. (2019). The effects of collaborative testing on higher order thinking: Do the bright get brighter? *Active Learning in Higher Education, 20*(1), 25–37. https://doi.org/10.1177/1469787417723243

Manca, S., Cerina, V., Tobia, V., Sacchi, S., & Fornara, F. (2020). The effect of school design on users' responses: A systematic review (2008–2017). *Sustainability (Basel), 12*(8), 3453. https://www.mdpi.com/2071-1050/12/8/3453https://doi.org/10.3390/su12083453 https://doi.org/10.3390/su12083453

Marin, S. V., Schwarz, L. & Sabarwal, S. (2024) IMPACT OF CLIMATE CHANGE ON EDUCATION and What To Do About It, Report, International Bank for Reconstruction and Development / The World Bank, Washington, World Bank Publications, April.

Marklund, B. B., & Taylor, A. A. (2016). Educational games in practice the challenges involved in conducting a game-based curriculum. *Electronic Journal of e-Learning, 14*, 122–135.

Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2019). The impact of national standardised literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Language Testing, 36*(2), 265–287. https://doi.org/10.1177/0265532218775758

Martin, A. P. (2018). A quantitative framework for the analysis of two-stage exams. *International Journal of Higher Education, 7*(4), 33–54. https://doi.org/10.5430/ijhe.v7n4p33

Mason, L. L., Sawchak, A., Scott Curtis, S., Andrews, A., Arriaga, A., & Pena, H. (2018). A comparison of outcomes from criterion-referenced and experimental evaluations of verbal behaviour. *Behavioral Development Bulletin, 23*(2), 118–129. https://doi.org/10.1037/bdb0000080

Massonnié, J., Frasseto, P., Mareschal, D., & Kirkham, N. Z. (2022). Learning in noisy classrooms: children's reports of annoyance and distraction from noise are associated with individual differences in mind-wandering and switching skills. *Environment and Behaviour, 54*(1), 58–88. https://doi.org/10.1177/0013916520950277

Masters, G. N., Australian Council for Educational Research, issuing body, & Australian Council for Educational Research, issuing body. (2013). *Reforming educational assessment : imperatives, principles and challenges*. Australian Council for Educational Research.

Masters, G. (2020). *Nurturing wonder and igniting passion: designs for a new school curriculum: NSW Curriculum Review*(9781743012215). Retrieved from Sydney: https://nswcurriculumreview.nesa.nsw.edu.au/

McKay, J., & Sridharan, B. (2024). Student perceptions of collaborative group work (CGW) in higher education. *Studies in Higher Education, 49*(2), 221–234. https://doi.org/10.1080/03075079.2023.2227677

Mead M. (1928). *Coming of age in Samoa: A psychological study of primitive youth for western civilisation*. Harper Perennial (2001).

Meaders, C. L., & Vega, Y. (2022). Collaborative two-stage exams benefit students in a biology laboratory course. *Journal of Microbiology & Biology Education, 24*, e00138–e22. https://doi.org/10.1128/jmbe.00138-22PMID:37089218PMID:37089218 PMID:37089218

Meador, D. (2023, April 5). Solutions for Teaching in an Overcrowded Classroom. Retrieved from https://www.thoughtco.com/teaching-in-an-overcrowded-classroom-3194352

van Mechelen, M., Zaman, B., Laenen, A., & Abeele, V. V. (2015). *Challenging group dynamics in participatory design with children: lessons from social interdependence theory.* Paper presented at the Proceedings of the 14th International Conference on Interaction Design and Children (IDC '15), New York, NY. https://doi.org/10.1145/2771839.2771862

Meijer, H., Hoekstra, R., Brouwer, J., & Strijbos, J. W. (2020). Unfolding collaborative learning assessment literacy: A reflection on current assessment methods in higher education. *Assessment & Evaluation in Higher Education, 45*(8), 1222–1240. https://doi.org/10.1080/02602938.2020.1729696

Mende, S., Proske, A., & Narciss, S. (2021). Individual preparation for collaborative learning: Systematic review and synthesis. *Educational Psychologist, 56*(1), 29–53. https://doi.org/10.1080/00461520.2020.1828086

Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction, 1*(1), 12–21. https://doi.org/10.1016/j.lcsi.2012.03.001

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Meseke, C. A., Nafziger, R., & Meseke, J. K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *The Journal of Chiropractic Education, 24*(1), 19–29. https://doi.org/10.7899/1042-5055-24.1.19 PMID:20480011

Mik-Meyer, N. (2020). *Multimethod qualitative research.* Qualitative research. *Sage (Atlanta, Ga.)*.

Miles, J. A., & Naumann, S. E. (2021). Science self-efficacy in the relationship between gender & science identity. *International Journal of Science Education, 43*(17), 2769–2790. https://doi.org/10.1080/09500693.2021.1986647

Miller, K., Kestin, G., & Miller, O. (2022). How gender composition and group formation impact the effectiveness of group work in two-stage collaborative exams. *Physical Review. Physics Education Research, 18*(2), 020137. https://doi.org/10.1103/PhysRevPhysEducRes.18.020137

Minges, K. E., & Redeker, N. S. (2016). Delayed school start times and adolescent sleep: A systematic review of the experimental evidence. *Sleep Medicine Reviews, 28*, 86–95. https://doi.org/10.1016/j.smrv.2015.06.002 PMID:26545246

Mitchell, K. M., & Clark, A. M. (2018). Five steps to writing more engaging qualitative research. *International Journal of Qualitative Methods, 17*(1). Advance online publication. https://doi.org/10.1177/1609406918757613

Moore, E. J., & Bell, S. M. (2019). Is instructor (faculty) modeling an effective practice for teacher education? Insights and supports for new research. *Action in Teacher Education, 41*(4), 325–343. https://doi.org/10.1080/01626620.2019.1622474

Moore, T., Berry, S., & Ambrose, M. (2019). Aiming for mediocrity: The case of Australian housing thermal performance. *Energy Policy, 132*, 602–610. https://doi.org/10.1016/j.enpol.2019.06.017

Morales-Bravo, J., & Navarrete-Hernandez, P. (2022). Enlightening wellbeing in the home: the impact of natural light design on perceived happiness and sadness in residential spaces. *Building and Environment*, 223. 109317. ISSN03601323 https://doi.org/10.1016/j.buildenv.2022.109317

Morgan, B. (2013). *Designing an online course to promote deep learning outcomes*. Paper presented at the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government, Las Vegas, NV.

Morgan, H. (2016). Relying on high-stakes standardised tests to evaluate schools and teachers: A bad idea. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 89*(2), 67–72. https://doi.org/10.1080/00098655.2016.1156628

Morgan, H. (2020). Best practices for implementing remote learning during a pandemic. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 93*(3), 135–141. https://doi.org/10.1080/00098655.2020.1751480

Morreale, S. P., Valenzano, J. M., & Bauer, J. A. (2017). Why communication education is important: A third study on the centrality of the discipline's content and pedagogy. *Communication Education, 66*(4), 402–422. https://doi.org/10.1080/03634523.2016.1265136

Morris, J. E., Imms, W., & Dehring, A. (2023). The relationship among classroom furniture, student engagement and teacher pedagogy. In J. E. Morris & W. Imms (Eds.), *Teachers as Researchers in Innovative Learning Environments.* Springer., https://doi.org/10.1007/978-981-99-7367-5_9

Morse, J. M., & Coulehan, J. (2015). Maintaining confidentiality in qualitative publications. *Qualitative Health Research, 25*(2), 151–152. https://doi.org/10.1177/1049732314563489 PMID:25527378

Moss, J., Godinho, S. C., & Chao, E. (2019). Enacting the Australian Curriculum: Primary and secondary teachers' approaches to integrating the curriculum. *The Australian Journal of Teacher Education, 44*(3), 24–41. https://ro.ecu.edu.au/ajte/vol44/iss3/2. https://doi.org/10.14221/ajte.2018v44n3.2

Muckenthaler, M., Tillmann, T., Weiß, S., & Kiel, E. (2020). Teacher collaboration as a core objective of school development. *School Effectiveness and School Improvement, 31*(3), 486–504. https://doi.org/10.1080/09243453.2020.1747501

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning, 6*, 421–443. https://doi.org/10.1007/s11412-011-9122-z

Murris, K. (2017). Reconfiguring educational relationality in education: The educator as pregnant stingray. *Journal of education (Durban),* (69), 117–138.

Murris, K., & Bozalek, V. (2019). Diffraction and response-able reading of texts: The relational ontologies of Barad and Deleuze. *International Journal of Qualitative Studies in Education : QSE, 32*(7), 872–886. https://doi.org/10.1080/09518398.2019.1609122

Murris, K. (2021). *A Glossary for Doing Postqualitative, New Materialist and Critical Posthumanist Research Across Disciplines* (1st ed.). Routledge., https://doi.org/10.4324/9781003041153

Murris, K. (2022). *Karen Barad as Educator. In Karen Barad as Educator, Agential Realism and Education* (1 ed., pp. XV, 95): Springer Singapore. https://doi.org/10.1007/978-981-19-0144-7

Mutereko, S. (2018). Marketisation, managerialism and high-stake testing. *International Journal of Educational Management, 32*(4), 568–579. https://doi.org/10.1108/IJEM-04-2017-0096

Nafziger, R., Meseke, J. K., & Meseke, C. A. (2011). Collaborative testing: The effect of group formation process on overall student performance. *The Journal of Chiropractic Education, 25*(1), 11–15. https://doi.org/10.7899/1042-5055-25.1.11 PMID:21677868

Nathan, S., Newman, C., & Lancaster, K. (2019). Qualitative Interviewing. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 391–410). Springer Singapore., https://doi.org/10.1007/978-981-10-5251-4_77

New South Wales Education Assessment Authority (NESA). (2022) https://education.nsw.gov.au/teaching-and-learning/professional-learning/teacher-quality-and-accreditation/strong-start-great-teachers/refining-practice/aspects-of-assessment/approaches-to-assessment

Newton, G., Rajakaruna, R., Kulak, V., Albabish, W., Gilley, B., & Ritchie, K. (2019). Two-stage (collaborative) testing in science teaching: Does it improve grades on short-answer questions and retention of material? *Journal of College Science Teaching, 48*(4), 64–73. https://doi.org/10.2505/4/jcst19_048_04_64

Nguyen, K. A., Borrego, M., Finelli, C. J., DeMonbrun, M., Crockett, C., Tharayil, S., Shekhar, P., Waters, C., & Rosenberg, R. (2021). Instructor strategies to aid implementation of active learning: A systematic literature review. *International Journal of STEM Education, 8*(1), 9. https://doi.org/10.1186/s40594-021-00270-7

Nichols, S. L., & Harris, L. R. (2016). Accountability assessment's effects on teachers and schools. *In Handbook of Human and Social Conditions in Assessment* (1st ed., pp. 40–56). Routledge., https://doi.org/10.4324/9781315749136-4

Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review, 27*, 645–656. https://doi.org/10.1007/s10648-015-9312-8

Norton (2001) *Diffraction Diagram*, published in 'The Quantum Theory of Waves and Particles', Pittsburgh.

NSW Education Standards Authority. NESA. (2018) *Understanding the Curriculum, Assessment Principles*, https://educationstandards.nsw.edu.au/wps/portal/nesa/k-10/understanding-the-curriculum/assessment

O'Connell, M., Milligan, S. K., & Bentley, T. (2019). *Beyond ATAR: a proposal for change.* Retrieved from Melbourne, Victoria: Office of Human Research Protections. (2018). Informed consent FAQs. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/informed-consent/index.html#4145

O'Neill, O. (2013). Intelligent accountability in education. *Oxford Review of Education, 39*(1), 4–16. https://doi.org/10.1080/03054985.2013.764761

Ong, A. (2006). *Neoliberalism as exception: Mutations in citizenship and sovereignty*. Duke University Press.

O'Nyumba, T., Wilson, K., Derrick, C. J., & Mukherjee, N. (. (2018). The use of focus group discussion methodology: Insights from two decades of application in conservation. *Methods in Ecology and Evolution, 9*(1), 20–32. https://doi.org/10.1111/2041-210X.12860

Orlikowski, W. J., & Scott, S. V. (2015). Exploring Material-Discursive Practices. *Journal of Management Studies, 52*(5), 697–705. https://doi.org/10.1111/joms.12114

Ozkan Bekiroglu, S., Ramsay, C. M., & Robert, J. (2022). Movement and engagement in flexible, technology-enhanced classrooms: Investigating cognitive and emotional engagement from the faculty perspective. *Learning Environments Research, 25*(2), 359–377. https://doi.org/10.1007/s10984-021-09363-0

Paine, A. R., & Knight, J. K. (2020, December). Student behaviours and interactions influence group discussions in an Introductory Biology lab setting. *CBE Life Sciences Education, 19*(4), ar58. https://doi.org/10.1187/cbe.20-03-0054 PMID:33259280

Palmer V. M. (1928). *Field studies in sociology: A students' manual*. University of Chicago Press.

Park, R. J., Behrer, A. P., & Goodman, J. (2021). Learning is inhibited by heat exposure, both internationally and within the United States. *Nature Human Behaviour*, *5*(1), 19–27. https://doi.org/10.1038/s41562-020-00959-9

Park, R. J., Goodman, J., Hurwitz, M., & Smith, J. (2020). Heat and learning. *American Economic Journal. Economic Policy, 12*(2), 306–339. https://doi.org/10.1257/pol.20180612 PMID:33020588

Park, R. J. (2022). Hot Temperature and high-stakes performance. *The Journal of Human Resources, 57*(2), 400–434. https://www.muse.jhu.edu/article/848531. https://doi.org/10.3368/jhr.57.2.0618-9535R3

Parkinson, P. (2014). Collaborative assessment: Middle school case study. *Current Issues in Middle Level Education, 19*(1), 43–49.

Pascale, C.-M. (2010). *Cartographies of Knowledge: Exploring Qualitative Epistemologies*. *Sage (Atlanta, Ga.)*.

Peralta, L. M. (2021). *Applying Barad's ontology to reconceptualise teaching and learning mathematics for social justice* (1st ed.). Routledge.

Perkins, G., Wellman, N., null., & null. (2008). The impact of high-stakes testing: Recommendations for professional school counsellors and other support personnel. *Academic Leadership: The Online Journal, 6*(4), 15. https://doi.org/10.58809/GEGV1815

Phillippi, J., & Lauderdale, J. (2018). A guide to field notes for qualitative research: Context and conversation. *Qualitative Health Research, 28*(3), 381–388. https://doi.org/10.1177/1049732317697102 PMID:29298584

Placier, M., Walker, M., & Foster, B. (2002). Writing the "Show–Me" standards: Teacher professionalism and political control in US state curriculum policy. *Curriculum Inquiry, 32*(3), 281–310. https://doi.org/10.1111/1467-873X.00229

Plank, S. B., & Condliffe, B. F. (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal, 50*(5), 1152–1182. https://doi.org/10.3102/0002831213500691

Plauborg, H. (2018). Towards an agential realist concept of learning. *Subjectivity, 11*, 322–338. https://doi.org/10.1057/s41286-018-0059-9

Pluim, G., Nazir, J., & Wallace, J. (2020). Curriculum integration and the semicentennial of Basil Bernstein's classification and framing of educational knowledge. *Canadian Journal of Science, Mathematics and*

Technology Education = Revue Canadienne de l'Enseignement des Sciences, des Mathématiques et de la Technologie, 20(4), 715–735. https://doi.org/10.1007/s42330-021-00135-9

Pickering, A. (2017). The Ontological turn: Taking different worlds seriously. *Social Analysis, 61*(2), 134–150. https://doi.org/10.3167/sa.2017.610209

Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy, 29*(5), 640–657. https://doi.org/10.1080/02680939.2013.865082

Polesel, J., Gillis, S., Suryani, A., Leahy, M., & Koh, S. (2021). The Australian senior certificates: After 50 years of reforms. *Australian Educational Researcher, 48*(3), 565–584. https://doi.org/10.1007/s13384-020-00403-x

Premo, J., Cavagnetto, A., Davis, W. B., & Brickman, P. (2018). Promoting collaborative classrooms: The impacts of interdependent cooperative learning on undergraduate interactions and achievement. *CBE Life Sciences Education, 17*(2), ar32. https://doi.org/10.1187/cbe.17-08-0176 PMID:29799312

Premo, J., Wyatt, B. N., Horn, M., & Wilson-Ashworth, H. (2022). Which group dynamics matter: Social predictors of student achievement in team-based undergraduate Science classrooms. *CBE Life Sciences Education, 21*(3), ar51. https://doi.org/10.1187/cbe.21-06-0164 PMID:35900893

Priestley, M., & Drew, V. (2016) Teachers as agents of curriculum change: closing the gap between purpose and practice. *European Conference for Educational Research, Dublin, 23-26 September 2016*, 23.08.2016. http://www.eera-ecer.de/ecer-programmes/conference/21/contribution/39307/

Priya, A. (2021). Case study methodology of qualitative research: Key attributes and navigating the conundrums in its application. *Sociological Bulletin, 70*(1), 94–110. https://doi.org/10.1177/0038022920970318

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education, 26*(1-4), 37–41. https://doi.org/10.1152/advan.00032.2001 PMID:11850326

Ravitch, D. (2010). *The death and life of the great American school system : how testing and choice are undermining education*. Basic Books.

Reeves, P. M., Pun, W. H., & Chung, K. S. (2017). Influence of teacher collaboration on job satisfaction and student achievement. *Teaching and Teacher Education*, 67, 227–236. . tate.2017.06.016 https://doi.org/10.1016/j.tate.2017.06.016

Reeves, S., Kuper, A., & Hodges, B. D. (2008). Qualitative research methodologies: Ethnography. *BMJ (Clinical Research Ed.), 337*(7668), a1020. https://doi.org/10.1136/bmj.a1020 PMID:18687725

Reeves, S., Peller, J., Goldman, J., & Kitto, S. (2013). Ethnography in qualitative educational research: AMEE Guide No. 80. *Medical Teacher, 35*(8), e1365–e1379. https://doi.org/10.3109/0142159X.2013.804977 PMID:23808715

Reid, A. (2019). National curriculum: An Australian perspective. *Curriculum Perspectives, 39*(2), 199–203. https://doi.org/10.1007/s41297-019-00077-1

Rempel, B., McGinitie, E., & Dirks, M. (2023). The influence of two stage collaborative testing on peer relationships: A study of first year university student perceptions. *The Canadian Journal for the Scholarship of Teaching and Learning, 14*(2), 1–19. https://doi.org/10.5206/cjsotlrcacea.2023.2.14252

Resnick, L. B., & Schantz, F. (2017). Testing, teaching, learning: Who is in charge? *Assessment in Education: Principles, Policy & Practice, 24*(3), 424–432. https://doi.org/10.1080/0969594X.2017.1336988

Rice, S., Dulfer, N., Polesel, J., & O'Hanlon, C. (2015). NAPLAN and student wellbeing: Teacher perceptions of the impact of NAPLAN on students. In *National Testing in Schools* (1st ed., pp. 72–85). Routledge. https://doi.org/10.4324/9781315659312-6

Rickard, K., Lu, L., & Wales, N. S. Centre for Education Statistics and Evaluation, issuing body. (2014). *Family Occupation and Education Index (FOEI) 2013*Retrieved August 30, 2024, from http://nla.gov.au/nla.obj-2575671888

Rieger, G. W., & Rieger, C. L. (2020). Collaborative assessment that supports learning. In *Active Learning in College Science* (pp. 821–837). Springer International Publishing., https://doi.org/10.1007/978-3-030-33600-4_51

Rieger, G., & Heiner, C. (2014). Examinations that support collaborative learning: The students' perspective. *Journal of College Science Teaching, 43*(4), 41–47. https://doi.org/10.2505/4/jcst14_043_04_41

Rivas, S. F., Saiz, C., & Ossa, C. (2022). Metacognitive strategies and development of critical thinking in higher education. *Frontiers in Psychology, 13*, 913219. https://doi.org/10.3389/fpsyg.2022.913219 PMID:35783800

Rowell, P. C., Shannonhouse, L., Galloway, A., & Garner, Z. D. (2021). Mining the inner experience: A qualitative study of utilising mind maps in group work. *Journal for Specialists in Group Work, 46*(3), 272–289. https://doi.org/10.1080/01933922.2021.1942345

Roach, T., & Whitney, J. (2022). Heat and learning in elementary and middle school. *Education Economics, 30*(1), 29–46. https://doi.org/10.1080/09645292.2021.1931815

Rust, J., & Altman, S. (2020). Vlogging as sense making: Fostering diffractive practitioners. In *Fieldnotes in Qualitative Education and Social Science Research* (1st ed., pp. 217–233). Routledge. https://doi.org/10.4324/9780429275821-18

Rutakumwa, R., Mugisha, J. O., Bernays, S., Kabunga, E., Tumwekwase, G., Mbonye, M., & Seeley, J. (2020). Conducting in-depth interviews with and without voice recorders: A comparative analysis. *Qualitative Research, 20*(5), 565–581. https://doi.org/10.1177/1468794119884806 PMID:32903872

Sackstein, S. (2017). Peer feedback in the classroom: Empowering students to be the experts. *Association for Supervision & Curriculum Development* (ASCD).

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*(2), 191–209. http://www.jstor.org/stable/1050133. https://doi.org/10.1080/0305498870130207

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education, 30*(2), 175–194. https://doi.org/10.1080/0260293042000264262

Sadovnik, A. R. (1991) Basil Bernstein's theory of pedagogic practice: A structuralist approach, Sociology of Education, Jan., Vol. 64, No. 1, Special Issue on Sociology of the Curriculum pp. 48-63

Salas, E., Reyes, D. L., & McDaniel, S. H. (2018). The science of teamwork: Progress, reflections, and the road ahead. *The American Psychologist, 73*(4), 593–600. https://doi.org/10.1037/amp0000334 PMID:29792470

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In P. Sahlberg, A. Verger, B. Lingard, A. Green, & K. Mundy (Eds.), *The Handbook of Global Education Policy* (pp. 128–144). John Wiley & Sons, Ltd., https://doi.org/10.1002/9781118468005.ch7

Sandahl, S. S. (2010). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives, 31*(3), 142–147. PMID:20635617

Saunders, A., Say, R., Visentin, D., & McCann, D. (2019). Evaluation of a collaborative testing approach to objective structured clinical examination (OSCE) in undergraduate nurse education: A survey study. *Nurse Education in Practice, 35*, 111–116. https://doi.org/10.1016/j.nepr.2019.01.009 PMID:30776722

Saxvig, I. W., Evanger, L. N., Pallesen, S., Hysing, M., Sivertsen, B., Gradisar, M., & Bjorvatn, B. (2021). Circadian typology and implications for adolescent sleep health. Results from a large, cross-sectional, school-based study. *Sleep Medicine, 83*, 63–70. https://doi.org/10.1016/j.sleep.2021.04.020 PMID:33990068

Schell, J. A., & Butler, A. C. (2018). Insights from the Science of Learning can inform evidence-based implementation of peer instruction. *Frontiers in Education, 3*, 33. Advance online publication. https://doi.org/10.3389/feduc.2018.00033

Schmidt, W. H., Wang, H. C., & McKnight, C. C. (2005). Curriculum coherence: An examination of US mathematics and science content standards from an international perspective. *Journal of Curriculum Studies, 37*(5), 525–559. https://doi.org/10.1080/0022027042000294682

Serin, H. (2018). A comparison of teacher-centered and student-centered approaches in educational settings. *International Journal of Social Sciences & Educational Studies, 5*(1), 164–167. https://doi.org/10.23918/ijsses.v5i1p164

Shaviro, S. (2015). Consequences of Panpsychism. In R. Grusin (Ed.), *The Nonhuman Turn* (pp. 19–44). University of Minnesota Press., http://www.jstor.org/stable/10.5749/j.ctt13x1mj0.5

Shay, S. (2008). Beyond social constructivist perspectives on assessment: The centring of knowledge. *Teaching in Higher Education, 13*(5), 595–605. https://doi.org/10.1080/13562510802334970

Shergold, P., Calma, T., Russo, S., Walton, P., Westacott, J., Zoellner, D., & O'Reilly, P. (2020). *Looking to the Future: Report of the review of senior secondary pathways into work, further education and training*. (1 ed.) Education Services Australia. https://apo.org.au/node/307138

Shepherd, S. (2017). Managerialism: An ideal type. *Studies in Higher Education, 43*(9), 1668–1678. https://doi.org/10.1080/03075079.2017.1281239

Shing, Y. L., & Brod, G. (2016). Effects of prior knowledge on memory: Implications for education. *Mind, Brain and Education : the Official Journal of the International Mind, Brain, and Education Society, 10*(3), 153–161. https://doi.org/10.1111/mbe.12110

Siegel, M., Roberts, T., Freyermuth, S., Witzig, S., & Izci, K. (2015). Research and teaching: aligning assessment to instruction: Collaborative group testing in large-enrollment science classes. *Journal of College Science Teaching, 44*(6). Advance online publication. https://doi.org/10.2505/4/jcst15_044_06_74

Silm, G., Pedaste, M., & Täht, K. (2020). The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educational Research Review, 31*, 100335. https://doi.org/10.1016/j.edurev.2020.100335

Sim, R. H. L., Zhang, Y., Chan, M. C., & Low, B. K. H. (2020, November). Collaborative machine learning with incentive-aware model rewards. In *International conference on machine learning* (pp. 8927-8936). PMLR.

Simon, M., & Tierney, R. 2011. "Use of Vignettes in Educational Research on Sensitive Teaching Functions as Assessment." 24th International Congress for School Effectiveness and Improvement 4th-7th Jan 2011 pp.1-13, Limassol, Cyprus http://www.icsei.net/icsei2011/Full%20Papers/0153.pdf

Simzar, R. M., Martinez, M., Rutherford, T., Domina, T., & Conley, A. M. (2015). Raising the stakes: How students' motivation for mathematics associates with high- and low-stakes test achievement. *Learning and Individual Differences, 39*, 49–63. https://doi.org/10.1016/j.lindif.2015.03.002 PMID:27840563

Skilling, K., & Stylianides, G. J. (2020). Using vignettes in educational research: A framework for vignette construction. *International Journal of Research & Method in Education, 43*(5), 541–556. https://doi.org/10.1080/1743727X.2019.1704243

Skinner, J. (2012). *The interview: An ethnographic approach*. Bloomsbury Academic.

Skinner, B., Leavey, G., & Rothi, D. (2019). Managerialism and teacher professional identity: Impact on well-being among teachers in the UK. *Educational Review, 73*, 1–16. https://doi.org/10.1080/00131911.2018.1556205

Skukauskaite, A., Noske, P., & Gonzales, M. (2018). Designing for discomfort: Preparing scholars for journeys through qualitative research. *International Journal of Qualitative Research, 11*(3), 334–349. https://doi.org/10.1525/irqr.2018.11.3.334

Slavin, R. (1995). Cooperative learning in middle and secondary schools. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 69*(4), 200–204.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43–69. https://doi.org/10.1006/ceps.1996.0004

Slavin, R. E. (1999). Comprehensive approaches to cooperative learning. *Theory into Practice, 38*(2), 74–79. https://doi.org/10.1080/00405849909543835

Slavin, R. E. (2014). Making cooperative learning powerful. *Educational Leadership, 72*(2), 22–26.

Smith, D. V. (2011). One brief, shining moment? The impact of neo-liberalism on science curriculum in the compulsory years of schooling. *International Journal of Science Education, 33*(9), 1273–1288. https://doi.org/10.1080/09500693.2010.512368

Smith, T. E., Thompson, A. M., & Maynard, B. R. (2022). Self-management interventions for reducing challenging behaviour's among school-age students: A systematic review. *Campbell Systematic Reviews, 18*(1), e1223. https://doi.org/10.1002/cl2.1223 PMID:36913185

Smith, W. C., & Benavot, A. (2019). Improving accountability in education: The importance of structured democratic voice. *Asia Pacific Education Review, 20*(2), 193–205. https://doi.org/10.1007/s12564-019-09599-9

Smith, W. C., & Holloway, J. (2020). School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability, 32*, 461–479. https://doi.org/10.1007/s11092-020-09342-8

Snaza, N. (2013). Bewildering Education. *Journal of Curriculum and Pedagogy, 10*, 38–54. https://doi.org/10.1080/15505170.2013.783889

Snaza, Nathan., & Weaver, J. A. (2015). *Posthumanism and educational research*. Routledge. https://doi.org/10.4324/9781315769165

Soares, D., & Woods, K. (2020). An international systematic literature review of test anxiety interventions 2011-2018. *Pastoral Care in Education, 38*(4), 311–334. https://doi.org/10.1080/02643944.2020.1725909

Springgay, S. (2015). Approximate rigorous abstractions: Propositions of activation for posthumanist research in education. In N. Snaza & J. Weaver (Eds.), *Posthumanism and Educational Research* (pp. 76–88). Routledge.

Springer, S. (2015). Postneoliberalism? *The Review of Radical Political Economics, 47*(1), 5–17. https://doi.org/10.1177/0486613413518724

Spyrou, S. (2019). An Ontological turn for childhood studies? *Children & Society, 33*(4), 316–323. https://doi.org/10.1111/chso.12292

Srivastava, P., & Hopwood, N. (2009). A practical iterative framework for qualitative data analysis. *International Journal of Qualitative Methods, 8*, 76–84. Advance online publication. https://doi.org/10.1177/160940690900800107

Staller, K. M. (2021). Big enough? Sampling in qualitative inquiry. *Qualitative Social Work: Research and Practice, 20*(4), 897–904. https://doi.org/10.1177/14733250211024516

Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE Life Sciences Education, 20*(2), fe3. https://doi.org/10.1187/cbe.20-12-0289 PMID:33797282

Stewart, D. W. (2018). Focus groups. In B. B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation* (Vol. 2, pp. 687–692). Sage Publications.

Strom, P. S., Hendon, K. L., Strom, R. D., & Wang, C.-H. (2018). How peers support and inhibit learning in the classroom: Assessment of high school students in collaborative groups. *School Community Journal, 29*(2), 183–202.

Swain, J., & King, B. (2022). Using Informal Conversations in Qualitative Research. *International Journal of Qualitative Methods, 21*, 16094069221085056. Advance online publication. https://doi.org/10.1177/16094069221085056

Taggart, J., & Wheeler, L. B. (2023). Collaborative learning as constructivist practice: An exploratory qualitative descriptive study of faculty approaches to student group work. 1-18. *Active Learning in Higher Education, 14697874231193938*, 14697874231193938. Advance online publication. https://doi.org/10.1177/14697874231193938

Talbert, R., & Mor-Avi, A. (2019). A space for learning: An analysis of research on active learning spaces. *Heliyon, 5*(12), e02967. Advance online publication. https://doi.org/10.1016/j.heliyon.2019.e02967 PMID:32368631

Talbot, D. (2023). Using institutional ethnography and students' voices to explore capacity to learn. *Australian Educational Researcher*. Advance online publication. https://doi.org/10.1007/s13384-023-00667-z

Tang, T., Vezzani, V., & Eriksson, V. (2020). Developing critical thinking, collective creativity skills and problem solving through playful design jams. *Thinking Skills and Creativity, 37*, 100696. https://doi.org/10.1016/j.tsc.2020.100696

Tanujaya, B. (2016). Development of an instrument to measure higher order thinking skills in senior high school mathematics instruction. *Journal of Education and Practice, 7*(21), 144–148.

Tanujaya, B., Mumu, J., & Margono, G. (2017). The relationship between higher order thinking skills and academic performance of student in mathematics instruction. *International Education Studies, 10*, 78–85. https://doi.org/10.5539/ies.v10n11p78

Taylor, C. A. (2013). Objects, bodies and space: Gender and embodied practices of mattering in the classroom. *Gender and Education, 25*(6), 688–703. https://doi.org/10.1080/09540253.2013.834864

Taylor, C., & Hughes, C. (2016). *Edu-crafting a Cacophonous Ecology: Posthumanist Research Practices for Education.* Palgrave Macmillan UK.

Taylor, C. A. (2018). What can bodies do? En/gendering body-space choreographies of stillness, movement and flow in post-16 pedagogic encounters. *International Journal of Educational Research, 88*, 156–165. https://doi.org/10.1016/j.ijer.2018.02.001

Taylor, C. A. (2019). Unfolding: Co-Conspirators, contemplations, complications and more. In C. A. Taylor & A. Bayley (Eds.), *Posthumanism and Higher Education.* Palgrave Macmillan., https://doi.org/10.1007/978-3-030-14672-6_1

Thelwall, M., & Nevill, T. (2021). Is research with qualitative data more prevalent and impactful now? Interviews, case studies, focus groups and ethnographies. *Library & Information Science Research, 43*(2), 101094. https://doi.org/10.1016/j.lisr.2021.101094

Theobald, E. J., Hill, M. J., Tran, E., Agrawal, S., Arroyo, E. N., Behling, S., Chambwe, N., Cintrón, D. L., Cooper, J. D., Dunster, G., Grummer, J. A., Hennessey, K., Hsiao, J., Iranon, N., Jones, L., II, Jordt, H., Keller, M., Lacey, M. E., Littlefield, C. E., . . . Freeman, S. (2020). Active learning narrows achievement gaps for underrepresented students in undergraduate science, technology, engineering, and math.

Proceedings of the National Academy of Sciences of the United States of America, 117(12), 6476–6483. https://doi.org/10.1073/pnas.1916903117 PMID:32152114

Thiel, J. (2018). Student feedback apparatuses in higher education: An agential realist analysis. *Discourse (Abingdon), 41*(3), 471–483. https://doi.org/10.1080/01596306.2018.1494544

Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative Inquiry, 17*(6), 511–521. https://doi.org/10.1177/1077800411409884

Thomson, K. E., & Trigwell, K. R. (2016). The role of informal conversations in developing university teaching? *Studies in Higher Education*, *43*(9), 1536–1547. https://doi.org/10.1080/03075079.2016.1265498

Thompson, G., & Harbaugh, A. G. (2013). A preliminary analysis of teacher perceptions of the effects of NAPLAN on pedagogy and curriculum. *Australian Educational Researcher*, *40*(3), 299–314. https://doi.org/10.1007/s13384-013-0093-0

Thompson, J., & Burkholder, C. (2020). *Fieldnotes in Qualitative Education and Social Science Research: Approaches, Practices, and Ethical Considerations*. Routledge.

Thompson, R., Smith, R. B., Bou Karim, Y., Shen, C., Drummond, K., Teng, C., & Toledano, M. B. (2022). Noise pollution and human cognition: An updated systematic review and meta-analysis of recent evidence. *Environment International, 158*, 106905. https://doi.org/10.1016/j.envint.2021.106905 PMID:34649047

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence, 11*(3), 54. https://doi.org/10.3390/jintelligence11030054 PMID:36976147

Thurston, A., Keere, K., Topping, K., Kosack, W., Gatt, S., Marchal, J., Mestdagh, N., Schmeinck, D., Sidor, W., & Donnert, K. (2007). Peer learning in primary school science: Theoretical perspectives and

implications for classroom practice. *Electronic Journal of Research in Educational Psychology, 5*, 477–496.

Tight, M. (2022). Internationalisation of higher education beyond the West: Challenges and opportunities – the research evidence. *Educational Research and Evaluation, 27*, 239–259. https://doi.org/10.1080/13803611.2022.2041853

Ting, F. S. T., Lam, W. H., & Shroff, R. H. (2019). Active learning via problem-based collaborative games in a large mathematics university course in Hong Kong. *Education Sciences, 9*(3), 172. https://doi.org/10.3390/educsci9030172

Tognolini, J., & Stanley, G. (2007). Standards-based assessment : A tool and means to the development of human capital and capacity building in education. *Australian Journal of Education, 51*(2), 129–145. https://doi.org/10.1177/000494410705100203

Tognolini, J., & Davidson, M. (2013). Assessment, standards-referencing and standard setting. In M. M. C. Mok (Ed.), *Self-directed Learning Oriented Assessments in the Asia-Pacific* (pp. 23–41). Springer Netherlands.

Tolmie, A. K., Topping, K. J., Christie, D., Donaldson, C., Howe, C., Jessiman, E., Livingston, K., & Thurston, A. (2010). Social effects of collaborative learning in primary schools. *Learning and Instruction, 20*(3), 177–191. https://doi.org/10.1016/j.learninstruc.2009.01.005

Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education, 59*(1), 134–144. https://doi.org/10.1016/j.compedu.2011.10.009

Törmänen, T., Järvenoja, H., & Mänty, K. (2021). Exploring groups' affective states during collaborative learning: What triggers activating affect on a group level? *Educational Technology Research and Development, 69*, 2523–2545. https://doi.org/10.1007/s11423-021-10037-0

Tran, D. (2013). Theoretical perspectives underlying the application of cooperative learning in classrooms. *International Journal of Higher Education, 2*(4). Advance online publication. https://doi.org/10.5430/ijhe.v2n4p101

Tran, D., & O'Connor, B. R. (2024). Teacher curriculum competence: How teachers act in curriculum making. *Journal of Curriculum Studies, 56*(1), 1–16. https://doi.org/10.1080/00220272.2023.2271541

Trinh, T. M. (1988). Not you/like you: Post-colonial women and the interlocking questions of identity and difference. In A. McClintock, A. Mufti, & E. Shohat (Eds.), *Dangerous liaisons: gender, nation, and postcolonial perspectives* (pp. 415–419). University of Minnesota Press.

Tsing, A. (2015). *The mushroom at the end of the world*. Princeton University Press.

Tullis, J. G., & Goldstone, R. L. (2020). Why does peer instruction benefit student learning? *Cognitive Research: Principles and Implications, 5*(1), 15. https://doi.org/10.1186/s41235-020-00218-5 PMID:32274609

Tufford, L., & Newman, P. (2012). Bracketing in Qualitative Research. *Qualitative Social Work: Research and Practice, 11*(1), 80–96. https://doi.org/10.1177/1473325010368316

Vangrieken, K., Grosemans, I., Dochy, F., & Kyndt, E. (2017). Teacher autonomy and collaboration: A paradox? Conceptualising and measuring teachers' autonomy and collaborative attitude. *Teaching and Teacher Education, 67*, 302–315. https://doi.org/10.1016/j.tate.2017.06.021

Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review, 71*(1), 5–30. https://doi.org/10.1080/00131911.2019.1522045

Verlie, B., & Blom, S. M. (2022). Education in a changing climate: Reconceptualising school and classroom climate through the fiery atmos-fears of Australia's Black summer. *Children's Geographies, 20*(5), 618–632. https://doi.org/10.1080/14733285.2021.1948504

Vogler, J. S., & Robinson, D. H. (2016). Team-based testing improves individual learning. *Journal of Experimental Education, 84*(4), 787–803. https://doi.org/10.1080/00220973.2015.1134420

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wan, Z. (2019). Participant selection and access in case study research. *Challenges and Opportunities in Qualitative Research: Sharing Young Scholars' Experiences*, 47-61.

Wang, X., Su, Y., Cheung, S., Wong, E., & Kwong, T. (2013). An exploration of Biggs' constructive alignment in course design and its impact on students' learning approaches. *Assessment & Evaluation in Higher Education, 38*(4), 477–491. https://doi.org/10.1080/02602938.2012.658018

Wargocki, P., Porras-Salazar, J. A., & Contreras-Espinoza, S. (2019). The relationship between classroom temperature and children's performance in school. *Building and Environment, 157*, 197–204. https://doi.org/10.1016/j.buildenv.2019.04.046

Webb S., Webb B. (1932). *Methods of social study*. Longman Green.

Westbury, I. (2008). *Making Curricula: Why do States Make Curricula, and How*? SAGE Publications, Inc.

Wheeldon, J., & Ahlberg, M. (2019). Mind Maps in Qualitative Research. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences*. 68-83 Springer., https://doi.org/10.1007/978-981-10-5251-4_7

Wheaton, A. G., Chapman, D. P., & Croft, J. B. (2016). School start times, sleep, behavioral, health, and academic outcomes: A review of the literature. *The Journal of School Health, 86*, 363–381. https://doi.org/10.1111/josh.12388 PMID:27040474

Whetton, P., Holper, P., Clarke, J., Webb, L., Hennessy, K., Colman, R., Moise, A., Power, S., Braganza, K., Watterson, I., Murphy, B., Timbal, B., Hope, P., Dowdy, A., Bhend, J., Kirono, D., Wilson, L., Grose, M., Ekstrom, M. Li, Y. Climate change in Australia. Projections for Australia's NRM regions. Melbourne: CSIRO and Bureau of Meteorology; 2015. csiro:EP154327. https://doi.org/10.4225/08/58518c08c4ce8

Wilson, K., Copeland-Solas, E., & Guthrie-Dixon, N. (2016). A preliminary study on the use of mind mapping as a visual-learning strategy, in general education science classes for Arabic speakers in the United Arab Emirates. *The Journal of Scholarship of Teaching and Learning, 16*(1), 31–52. https://doi.org/10.14434/josotl.v16i1.19181

Wilson, R., Piccoli, A., Hargreaves, A., Ng, P. T., & Sahlberg, P. (2021). Putting students first: Moving on from NAPLAN to a new educational assessment system. *The Gonski Institute Policy Paper, 2*.

Wininger, S. R., Redifer, J. L., Norman, A. D., & Ryle, M. K. (2019). Prevalence of learning styles in educational psychology and introduction to education textbooks: A content analysis. *Psychology Learning & Teaching, 18*(3), 221–243. https://doi.org/10.1177/1475725719830301

Wright, R., & Froehlich, H. (2012). Basil Bernstein's theory of the pedagogic device and formal music schooling: Putting the theory into practice. *Theory into Practice, 51*(3), 212–220. https://doi.org/10.1080/00405841.2012.690307

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309 PMID:33683913

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. https://doi.org/10.1080/0969594X.2017.1295020

Yin, R. K. (2014). *Case study research: Design and methods*. *Sage (Atlanta, Ga.)*.

Zacharis, N. Z. (2010). Innovative assessment for learning enhancement: Issues and practices. *Contemporary Issues in Education Research, 3*(1), 61–69. https://doi.org/10.19030/cier.v3i1.162

Zajda, J., & Rust, V. (2021). Globalisation and Comparative Education: Changing Paradigms. Globalisation, Comparative Education and Policy Research. Volume 24. *Globalisation and Comparative Education Changing Paradigms*. https://doi.org/10.1007/978-94-024-2054-8

Zapatero, E. G., Maheshwari, S. K., & Chen, J. (2012). Effectiveness of active learning environments: Should testing methods be modified? *Academy of Educational Leadership Journal, 16*(4), 101–114.

Zepeda, S. J. (2014*). Job-Embedded Professional Development: Support, Collaboration, and Learning in Schools* (1st ed.). Routledge., https://doi.org/10.4324/9781315719693

Appendices

Appendix A: University of Sydney Ethics Approval

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Wednesday, 6 October 2021

Dr Debra Talbot
School of Education and Social Work Research Operations; Faculty of Arts and Social Sciences
Email: debra.talbot@sydney.edu.au

Dear Debra,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that after consideration of your response, your project has been approved.

Details of the approval are as follows:

Project No.:	**2021/699**
Project Title:	**Exploring student and teacher perceptions of collaborative testing in secondary school science.**
Authorised Personnel:	**Talbot Debra; Bremert Helen; Brunker Nicole;**
Approval Period:	**06/10/2021** to **06/10/2025**
First Annual Report Due:	**06/10/2022**

Documents Approved:

Date Uploaded	Version Number	Document Name
29/09/2021	Version 2	Clean Parent Consent Statement
29/09/2021	Version 3	Clean teacher Consent Statement
27/09/2021	Version 2	Clean Parent Information Statement
27/09/2021	Version 2	Clean Student Information Statement
27/09/2021	Version 2	Clean Teacher Information Statement
24/09/2021	Version 2	Clean Student Consent Form
06/08/2021	Version 1	Principal Consent Form
06/08/2021	Version 1	Principal Information Sheet
06/08/2021	Version 1	Email to Principals
26/07/2021	Version 1	Teacher Focus Group Questions
26/07/2021	Version 1	Teacher Interview Questions
26/07/2021	Version 1	Mind Map exemplar
26/07/2021	Version 1	Student Focus Group Questions

Special Condition/s of Approval

- It is a condition of approval that permission is obtained from the Principals of all participating schools, and as appropriate, the Department of Education (SERAP) and the Catholic Education Office. Evidence of permission should be obtained and kept on file prior to the relevant part of the research being conducted.

- It will be a condition of approval that all people listed on this project (including researchers, students, university staff and volunteers) have a current/valid Working with Children Check (WWCC) prior to undertaking any research activity with children. The WWCC must be valid for the duration of the research activity. Evidence must be kept on file as part of your study records and be made available to the University when requested.

  You do not need to provide a copy to the Ethics Office.

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

THE UNIVERSITY OF SYDNEY

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.
- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or clarification.

Sincerely,

Associate Professor Mark Arnold
Chair, Human Research Ethics Committee (HREC 2)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2018) and the NHMRC's Australian Code for the Responsible Conduct of Research (2018)

Appendix B: Approval to Conduct Research in NSW Public Schools (SERAP)

| NSW Department of Education

Ms Helen Bremert
321/8 Point Street
PYRMONT NSW 2009

DOC22/232243
SERAP 2021360

Dear Ms Bremert

I refer to your application to conduct a research project in NSW government schools entitled *Exploring student and teacher perceptions of collaborative testing in secondary school science.*. I am pleased to inform you that your application has been approved.

You may contact principals of the nominated schools to seek their participation. **You should include a copy of this letter with the documents you send to principals.**

This approval will remain valid until 5 April 2023.

The following researchers or research assistants have fulfilled the Working with Children screening requirements to interact with or observe children for the purposes of this research for the period indicated:

Researcher name	WWCC	WWCC expires
Helen Bremert	WWC2223290E	30-Dec-2025

I draw your attention to the following requirements for all researchers in NSW government schools:

- The privacy of participants is to be protected as per the NSW Privacy and Personal Information Protection Act 1998.
- School principals have the right to withdraw the school from the study at any time. The approval of the principal for the specific method of gathering information must also be sought.
- The privacy of the school and the students is to be protected.
- The participation of teachers and students must be voluntary and must be at the school's convenience.
- Any proposal to publish the outcomes of the study should be discussed with the research approvals officer before publication proceeds.
- All conditions attached to the approval must be complied with.

When your study is completed please email your report to: det.serap@det.nsw.edu.au.
You may also be asked to present on the findings of your research.

I wish you every success with your research.

Yours sincerely

[redacted]

Brianna McCourt
Senior Policy Analyst, Strategic Analysis and Research
Centre for Education Statistics and Evaluation
5 April 2022

education.nsw.gov.au
Strategic Analysis And Research
Centre For Education Statistics And Evaluation
NSW Department Of Education
105 Phillip Street, Parramatta NSW 2150 | GPO Box 33, Sydney NSW 2001
Email: det.serap@det.nsw.edu.au

Appendix C: Participant Information Sheets

THE UNIVERSITY OF SYDNEY

School of Education
Faculty of Arts and Social Science

ABN 15 211 513 464

CHIEF INVESTIGATOR Dr Debra Talbot
Senior Lecturer

Room 817
Education Building, A35
The University of Sydney
N.S.W. 2006 AUSTRALIA
Telephone: +61 2 93516287
Email: debra.talbot@sydney.edu.au
Web: http://www.sydney.edu.au/

Exploring student and teacher perceptions of collaborative testing in secondary school science.

STUDENT PARTICIPANT INFORMATION STATEMENT

(1) What is this study about?

You are invited to participate in a research study on determining whether topic tests completed in a collaborative group rather than individually can impact learning for science students in years 10, 11 and 12.
The researcher, Helen Bremert, will collaborate with science teachers to develop collaborative testing opportunities to explore your perceptions of collaborative testing and its impacts on your learning. The participatory design of this project holds the potential to benefit NSW Department of Education school leaders, teachers, and students as the researcher works alongside classroom teachers to develop an understanding and implementation of collaborative testing. The study explores student and teacher perceptions of collaborative testing as a logical extension of collaborative learning. Effective collaboration has been shown to positively impact student engagement, foster positive peer relationships, reduce test anxiety, and move students from a surface to a deep understanding of concepts. Collaborative testing may improve teachers' practice through augmenting teachers' professional practices in utilising diverse assessment strategies and engaging in professional learning to enhance student outcomes.

You have been invited to participate in this study because your school and teachers have indicated a willingness to determine if collaborative testing can impact student learning. This Participant Information Statement tells you about the research study, and knowing what is involved will help you decide if you want to participate in the research. Please read this sheet carefully and ask questions about anything you don't understand or want to know more about.

Participation in this research study is voluntary.

By giving your consent to take part in this study, you are telling us that you:
- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

(10) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(11) What if I would like further information about the study?

When you have read this information, *Helen Bremert,* who will be available at the time of consent, will discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact Helen Bremert, PhD. Student, hbre8322@uni.sydney.edu.au.

(12) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the box on the student consent form. This feedback will be in the form of a one-page summary of the research findings. You will receive this feedback after the study is finished.

(13) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney *[INSERT* protocol number once approval is obtained*].* As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007).* This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

This information sheet is for you to keep

THE UNIVERSITY OF SYDNEY

School of Education
Faculty of Arts and Social Science

ABN 15 211 513 464

CHIEF INVESTIGATOR Dr Debra Talbot
Senior Lecturer

Room 817
Education building, A35
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 93516287
Email: debra.talbot@sydney.edu.au
Web: http://www.sydney.edu.au/

Exploring student and teacher perceptions of collaborative testing in secondary school science.

TEACHER INFORMATION STATEMENT

(1) What is this study about?

You are invited to participate in a research study on determining whether topic tests completed in a collaborative group rather than individually can impact learning for science students in years 9, 10, 11 and 12. The researcher, Helen Bremert, will collaborate with science teachers to develop collaborative testing opportunities to explore students' perceptions of collaborative testing and its impacts on their learning. Additionally, the study will look at teachers' insights into collaborative testing and the approach of this method in their classrooms to determine the benefits and limitations of collaborative testing.

The participatory design of this project holds the potential to benefit NSW Department of Education school leaders, teachers, and students as the researcher work alongside classroom teachers to develop understanding and implementation of collaborative testing. The study explores student and teacher perceptions of collaborative testing as a logical extension to collaborative learning. Effective collaboration has been shown to positively impact student engagement, foster positive peer relationships, reduce test anxiety, and move students from surface to deep understanding of concepts. Collaborative testing may improve teachers' practice through augmenting teachers' professional practices in utilising diverse assessment strategies and engaging in professional learning to enhance student outcomes.

You have been invited to participate in this study on collaborative testing. This Participant Information Statement tells you about the research study. Please read this sheet carefully and ask questions about anything you don't understand or know more about. Knowing what is involved will help you decide if you want to take part in the research.

Participation in this research study is voluntary.

If you decide to take part in the study and then change your mind later, you are free to withdraw at any time. You can do this by informing the researcher, hbre8322@uni.sydney.edu.au.

You are free to stop the interview at any time, and your interview recordings will be deleted from the data set. You may also refuse to answer any questions that you do not wish to answer during the interview.

If you decide to withdraw from the study, we will not collect any more information from you. However, any data we have already collected will be kept in our study records and may be included in the study results.

(7) Are there any risks or costs associated with being in the study?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

(8) Are there any benefits associated with being in the study?

Benefits may include meals provided such as sandwiches, cakes, tea, and coffee, dependent on the interview and focus groups' time.

(9) What will happen to information about me that is collected during the study?

This study will collect audio recordings of the initial and post-study interview and optional focus group. The audio recordings will enable the researcher to analyse the teacher's perceptions of collaborative assessment and how to undertake this method in your classroom. In addition, I will also collect audio recordings of the reflection discussions. Reflections will be used to determine any issues with the method or groupings. All three data sets will be used for analysis in the study and publication in peer-reviewed journal publications.

The researcher will adhere to the university's data management protocols to protect research materials and data from breaches and loss during the study. The hard copies of the assessments, focus group audio transcripts and mind maps will be stored in a secured filing cabinet in the School of Education & Social Work, University of Sydney. The audio recordings and student information will be held in the university's One Drive (Office 365) in a password-protected folder only accessible to the researchers. The data will be kept in a secured location for a minimum of 20 years or until subjects are 25 years old as per the State Records Authority of NSW.

By providing your consent, you agree to us collecting information about you for this research study. Your information will only be used for the purposes outlined in this Participant Information Statement unless you consent otherwise.

Your information will be stored securely, and your identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but you will not be individually identifiable in these publications. However, as you will participate in the focus groups, participants may be known to each other and, therefore, may be able to identify one another in the final dissemination of the research.

We will keep the information we collect for this study, and we may use it in future projects. By providing your consent, you are allowing us to use your information in future projects. We don't know what these other projects will involve at this stage, and we will seek ethical approval before using the information in these future projects.

(10) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(11) What if I would like further information about the study?

When you have read this information, Helen Bremert will be available to discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact Helen Bremert, PhD. Student, hbre8322@uni.sydney.edu.au

(12) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by informing Helen Bremert and ticking the relevant box on the consent form. This feedback will be in the form of a one page lay summary, and you will receive this feedback after the study is finished.

(13) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney (2021/699). As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007).* This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted, or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

This information sheet is for you to keep

THE UNIVERSITY OF SYDNEY

School of Education
Faculty of Arts and Social Science

ABN 15 211 513 464

CHIEF INVESTIGATOR Dr Debra Talbot
Senior Lecturer

Room 817
Education Building, A35
The University of Sydney
N.S.W. 2006 AUSTRALIA
Telephone: +61 2 93516287
Email: debra.talbot@sydney.edu.au
Web: http://www.sydney.edu.au/

Exploring student and teacher perceptions of collaborative testing in secondary school science.

PARENT INFORMATION STATEMENT

(1) What is this study about?

Your child is invited to participate in a research study determining whether topic tests completed in a collaborative group rather than individually can impact learning for years 10, 11 and 12 science students. The researcher, Helen Bremert, will collaborate with science teachers to develop collaborative testing opportunities to explore students' perceptions of collaborative testing and its impacts on their learning. Additionally, the study will look at teachers' insights into collaborative testing and the approach of this method in their classrooms to determine the benefits and limitations of collaborative testing.

The participatory design of this project holds the potential to benefit the Department of Education school leaders, teachers, and students as the researcher works alongside classroom teachers to develop an understanding and implementation of collaborative testing. The study explores student and teacher perceptions of collaborative testing as a logical extension of collaborative learning. Effective collaboration has been shown to positively impact student engagement, foster positive peer relationships, reduce test anxiety, and move students from a surface to a deep understanding of concepts. Collaborative testing may improve teachers' practice through augmenting teachers' professional practices in utilising diverse assessment strategies and engaging in professional learning to enhance student outcomes.

Your child has been invited to participate in this study. This Participant Information Statement tells you about the research study. Please read this sheet carefully and ask questions about anything you don't understand or want to know more about. Knowing what is involved will help you decide if you want your child to participate in the research.

Participation in this research study is voluntary.

By giving your consent for your child to take part in this study, you are telling us that you:

- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

You will be given a copy of this Participant Information Statement to keep.

(2) Who is running the study?

The study is being conducted by Helen Bremert and will form the basis for the degree of Doctor of Philosophy at The University of Sydney under the supervision of Dr Debra Talbot, Senior lecturer, School of Education & Social Work.

(3) What will the study involve for your child?

This study will involve your child being placed in assigned groups to work on three collaborative tests. Your child's teacher, alongside the researcher, will write the three topic tests. The assessments will take place in their classroom and with their teacher. Therefore, this study should not take time away from the student's learning, as the assessment will be the same; they will undertake them as a collaborative group instead of individually.

While taking the test, your child will be audio recorded to determine how they discuss the questions with their group members. This dialogue will be transcribed, any identifying information will be removed, and they will be assigned a letter and number to protect their identity.

After the last collaborative test, your child will be asked to make a mind map of their thoughts on this collaborative method. Any identifying information will be removed from the mind map.

At the end of the study, your child may volunteer to join a focus group to discuss their perceptions of this testing method. This focus group will be recorded and transcribed; however, all identifying markers will be removed.

Information on your child's performance level will be collected from the teacher to determine if this factor impacts your child's perceptions of collaborative testing. However, all identifying markers will be removed to maintain student privacy.

(4) How much time will the study take?

The study will take four and half hours over two NSW school terms, twenty weeks. The participant will undertake three tests throughout terms one and two. The assessments will be the standard length of previous tests undertaken with the teacher. Additionally, if your child volunteers for the focus group, they will attend a thirty to forty minute meeting conveniently during or after school hours.

(5) Does your child have to be in the study? Can they withdraw from the study once it has started?

Being in this study is entirely voluntary, and your child does not have to take part. The decision whether to participate will not affect your current or future relationship with the researchers or anyone else at the University of Sydney.

If you decide to allow your child to participate in the study and then change your mind later, they are free to withdraw at any time. You can do this by contacting your child's teacher or Ms Bremert, hbre8322@uni.sydney.edu.au

If your child participates in a focus group, they are free to stop participating at any stage or refuse to answer any of the questions. However, it will not be possible to withdraw their comments from our records once the group has started, as it is a group discussion.

(6) Are there any risks or costs associated with being in the study?

We do not expect that there will be any risks or costs associated with taking part in this study.

(7) Are there any benefits associated with being in the study?

Benefits for the students who participate in the focus group may include school appropriate snacks and drinks, dependent on when the focus group occurs. This refreshment should not unduly influence your child's participation in the study.

(8) What will happen to information about your child that is collected during the study?

This study will collect three hard-copy collaborative assessments undertaken by your child and a mind map. The assessments will be used to determine if collaborative assessments impact their learning. I will also collect audio recordings of your child working in their group and if they volunteer for the focus group. The audio recording will help analyse your child's understanding of the questions asked, and the mind map and focus group will reveal their views of collaborative testing.
Information on your child's performance level will be collected from the teacher to determine if this factor impacts student learning and their perceptions of collaborative testing. However, all identifying markers will be removed to maintain student privacy.
All data sets will be used for analysis in the study and publication in peer-reviewed journal publications.

The researcher will adhere to the university's data management protocols to protect research materials and data from breaches and loss during the study. The hard copies of the assessments, focus group audio transcripts and mind maps will be stored in a secured filing cabinet in the School of Education & Social Work, University of Sydney. The audio recordings and student information will be held in the university's One Drive (Office 365) in a password-protected folder only accessible to the researchers. The data will be kept in a secured location for a minimum of 20 years or until subjects are 25 years old, as per the State Records Authority of NSW.

By providing your consent, you agree to us collecting personal information about your child for this research study. Unless you consent otherwise, the information will only be used for the purposes outlined in this Participant Information Statement.

The information will be stored securely, and your child's identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but your child will not be individually identifiable in these publications.

We will keep and archive the information we collect for this study, and we may use it in future projects. By providing your consent, you are allowing us to use your child's information in future projects. We don't know what these other projects will involve at this stage, and we will seek ethical approval before using the information in these future projects.

(9) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(10) What if I would like further information about the study?

When you have read this information, Helen Bremert, who will be available at the time of consent, will discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact Helen Bremert, PhD. Student, hbre8322@uni.sydney.edu.au.

(11) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the box on the consent form. This feedback will be in the form of a one-page summary of the research findings. You will receive this feedback after the study is finished.

(12) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney (2021/699**)**. As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007).* This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

This information sheet is for you to keep

Appendix D: Example of Test Questions

Examples of the Test Questions from the Three Classes

Halo

Assess the impact of earthquakes on society. 8 marks

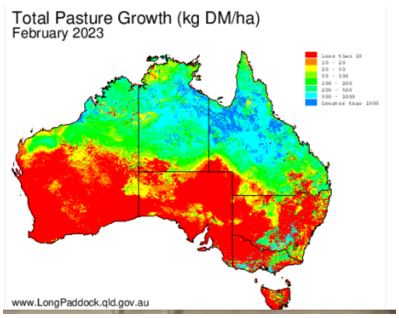
Criteria	Marks
• Clearly and coherently describes how earthquakes impact society, including individuals and economies • Describes at least two examples in detail • Suitable scientific language is used • Makes a suitable judgement linked to the examples	7-8

Analyse the ethical considerations of using therapeutic cloning to treat medical conditions. 8 marks

Criteria	Marks
• Describes how therapeutic cloning is used in medicine • Describes ethical issues related to cloning • Uses relevant scientific language • Is clear and coherent	7-8

Zeta

The map below shows the total pasture growth in Australia.

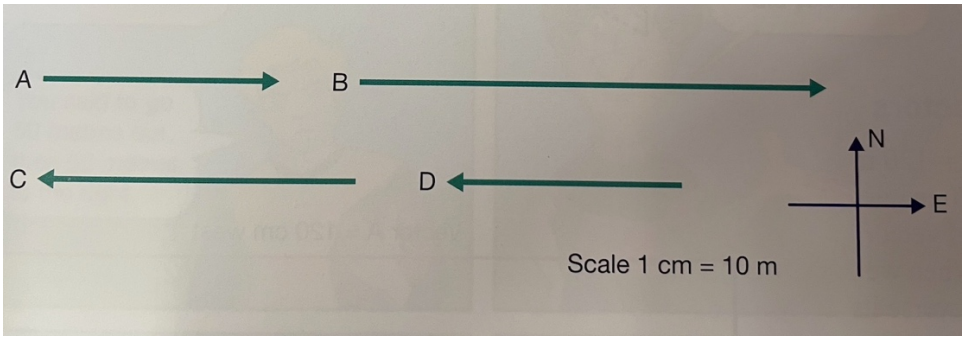


Explain the environmental factors that could have generated the very high pasture production in Queensland compared to South Australia.

Compare the advantages and disadvantages of using different genetic tools, including breed plans and estimated breeding values (EBV) to improve animal production.

Beta

The diagram below demonstrates several displacement vectors. Use them to answer the questions below. 6 marks



1. Use the above diagram to find

   a. A + D – C

   b. D + B – A

   c. C – A + B + D

   d. A + D – C – B

   e. D – C – B + A

   f. A – C + D - B

Describe the similarities and differences in your answers to D, E, F. **Explain** what this demonstrates to you about vector quantities. 4 marks

This is a motion graph of a butterfly that starts going left.
Describe what is occurring in this motion graph and **justify** your answers using distance, velocity, speed, acceleration and total displacement. (12 marks)