RETHINKING ASSESSMENT:

STUDENTS' EXPERIENCES, LEARNING APPROACHES, AND MOTIVATIONS IN UNIVERSITY EXAMS

A Thesis Submitted to the Committee on Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Faculty of Arts and Science

TRENT UNIVERSITY

Peterborough, Ontario, Canada

Psychology M.Sc. Graduate Program

January 2025

Abstract

Rethinking Assessment:

Students' Experiences, Learning Approaches, and Motivations in University Exams

Thomas Edward Bodrug

This thesis aimed to explore students' perspectives on university exams in two studies. Using a phenomenological approach, Study 1 recruited 10 senior undergraduate students to participate in semi-structured interviews. A thematic analysis revealed key themes related to exam purpose, effectiveness, fairness, and emotional responses. These insights informed the development of a sequence model explaining how students form exam perceptions. Study 2 was a concurrent, embedded, correlational mixed-methods case study of sophomore Psychology students' (N = 35) experiences with a case-based take-home exam designed according to motivational design principles. Findings from this study highlight the associations between students' motivation orientation and motivational assessment features and students' positive experiences and perceptions of the exam. This thesis offers an informative framework for instructors aiming to promote student buy-in, while also meeting course learning outcomes and facilitating deeper engagement with assessments.

Keywords: assessment design, alternative assessment, university exams, learning outcomes, student motivation

Acknowledgements

I would like to extend my thanks to Dr. Keefer, my thesis advisor, for helping transform my abstract ideas on the future of assessment and higher education into coherent, constructive scientific arguments. Under your guidance, I have become a better researcher and critical thinker, and I consider myself lucky to have had you as my supervisor. Your influence has ignited my love for Educational Psychology research. To committee members, Dr. Kevin Peters and Dr. Ann Celestini, thank you for sharing both of your expertise for this thesis. KP, specifically for challenging my thinking about scientific research findings. I would like to thank Daniel Ghorayeb and Caylin Graham, whose contributions during their research practicums greatly assisted in crafting literature reviews for background information and qualitative work.

I am grateful to various members of the Trent community who have taken a special interest in this project, including Sue Beckwith, Erin Stewart Eves, Dana Capell, Fergal O'Hagan, and the folks at the CTL. Special thanks to my friends in this program; I now realize the importance of connecting with similar-minded individuals passionate about education and research. To my family and close friends, including Cathy and Dave Bodrug and Jessica Draaistra, thank you for tolerating my rants and unwavering support in every possible way. Lastly, a big thank you to the participants who generously dedicated time to complete questionnaires and interviews. Your invaluable contributions will greatly improve the exam experiences for future students, making them more meaningful and effective for learning.

This study was supported by the Canada Graduate Research Scholarship from the Social Sciences and Humanities Research Council of Canada and the Trent Teaching Fellowship grant from the Trent Centre for Teaching and Learning.

Table of Contents

List of Figures

List of Tables

Rethinking Assessment: Students' Experiences, Learning Approaches, and Motivations in University Exams

Standardized testing and examinations are a staple of the North American higher education system (Harlen & Deakin Crick, 2003). Examination results have important career, financial, social, and personal implications for students, as they are often used in scholarship rankings, advanced placement decisions, and professional licensure. Given these high stakes, careful consideration must go into exam design (Kibble, 2017; Norcini et al., 2011). Educational literature on assessment design has identified many relevant criteria, including test validity and reliability (Kibble, 2017; Norcini et al., 2011), assessing appropriate learning outcomes (Anderson & Krathwohl, 2001; Bloom et al., 1956), and exam fairness (Brown & Hirschfeld, 2008; Murillo & Hidalgo, 2020; Nisbet & Shaw, 2019). However, assessment design criteria vary depending on the assessment's purpose. For example, *formative* assessments are lower stakes and include constructive feedback from the instructor, and they allow students to make more errors during their learning (Glazer, 2014). In contrast, *summative* assessments are typically high stakes and used to evaluate performance but may not include any feedback. This thesis specifically focuses on design considerations around summative assessments, namely university exams.

In a review of current trends in research on university exams, Buckley (2023) identified a gap between the research on what exams should be like, and the reality of what exams are like as practiced. Many university instructors continue to use traditional exam formats, such as proctored memory-based tests, despite the longstanding concerns about their undesirable effects on students' motivation and learning (Harlen & Deakin Crick, 2003; Struyven et al., 2005). Traditional exams have been criticized for relying too heavily on rote memorization and lacking

testing students' critical thinking skills (Bengtsson, 2019; Gardner, 1992; Litchfield & Dempsey, 2015). As a result, traditional exam formats tend to promote more surface-level study strategies in students, as opposed to deeper-level thinking (Struyven et al., 2005). Moreover, when students do not see personal relevance to an assessment beyond the extrinsic goal of grades, they have little intrinsic motivation to engage in the process, and as a result the quality of their learning, and even their wellbeing, may suffer (Howard et al., 2021).

In response to these criticisms, educators have been exploring alternative assessment formats that are explicitly designed to target higher-order learning outcomes while promoting more meaningful student engagement (Struyven et al., 2005). These student-centred approaches include pedagogies such as inquiry-based learning, case-based learning (CBL), and other forms of authentic assessments (e.g., Gulikers et al., 2008; Segers et al., 2008; Way et al., 2021). By asking students to apply their learning to real-world problems or cases, alternative formats can add intrinsic value to the learning process and promote deeper learning strategies in students (Struyven et al., 2005). However, whether the promise of these alternative approaches is fully realized depends on how these assessments are designed, implemented, and perceived (Gulikers et al., 2004; Villarroel et al., 2018). Most existing research on CBL and authentic assessment has been done with formative assessments, whereas their applications to summative exams remain under-researched. Lack of empirical research and specific guidelines on how to apply these alternative formats to summative exams may be one of the reasons why educators continue to rely on traditional exams.

If we, as educators, wish to increase students' buy-in into the assessment process, and to design exams that would facilitate rather than undermine students' motivation, wellbeing, and desirable learning behaviours, we must understand the students' phenomenological perspective

on exams (Lynam & Cachia, 2018). Students' perceptions and past experiences with exams matter because they influence how students approach future learning (Gerritsen-van-Leeuwenkamp et al., 2019). Because different stakeholders may view exams as serving different uses and purposes, students may hold different views from those of educators as to what constitutes a valid, fair, and effective exam (Norcini et al., 2011). Therefore, assessment design models derived from the educators' perspective may not be representative of students' views, which is why qualitative methods are considered more suitable for this research purpose (Lynam & Cachia, 2018). Presently, there remains little qualitative research into students' experiences with diverse types of high-stakes assessments, particularly within specific disciplines such as Psychology.

The present thesis aimed to explore undergraduate Psychology students' perceptions of summative exams, guided by the following broad questions: What purpose do students believe exams serve or hope exams would serve? What attributions do students make about exams' fairness and validity? What exam aspects evoke positive and negative emotional responses in students? What exam features motivate students to put in more effort and adopt deeper learning strategies? Two related studies were conducted to answer these questions. Study 1 was a broad phenomenological exploration of students' general perceptions of exams' purpose, validity, fairness, study strategies, and affective responses towards exams, based on qualitative interviews with a select group of upper-year Psychology students. Study 2 was a mixed-methods case study of students' experiences with an alternative case-based exam implemented in a sophomore Psychology course. In addition to students' perceptions and learning approaches towards the new case-based exam, Study 2 also examined students' motivational orientations and how these related to specific exam features and perceptions of the exam's motivational properties.

Understanding students' perspective on exams will help inform the design of student-centered assessments that are pedagogically sound, beneficial, and acceptable to students. By focusing specifically on student-centered approaches to assessment, this research also aimed to derive a thematic framework based on the thematic analysis of students' perceptions of exams, which would offer constructive suggestions for educators looking to make the exam experience more positive for students while encouraging desirable learning outcomes.

To provide the reader with a deeper understanding of the author's personal and professional background, reflexivity pieces will be integrated throughout this thesis, as per the recommendations of Olmas-Vega et al. (2023) and Sabnis and Wolgemuth (2023). These reflexivity pieces will highlight how my experiences relate to the writing and interpretation of the data that make up this thesis. I intend that my transparency and positionality will help contextualize this thesis' findings.

Considerations in Assessment Design

Educator's Concerns

Before looking at students' perspectives on exams, it is useful to first understand the industry standards for effective assessment design and educators' conceptions of summative assessments. Kibble (2017) and Norcini et al. (2011) explicitly state several criteria educators should keep in mind when designing good assessments. These criteria are concerned about various pedagogical and non-pedagogical aspects that shape assessment design. These considerations include test validity, reliability, equivalence, feasibility, educational effect, catalytic effect, and acceptability (Norcini et al., 2011). Meaning, tests should be designed in a way that accurately measures course learning outcomes, can be reproduceable, and an equivalent

and appropriate level of challenge across different institutions. Moreover, tests should also be cost effective and practical given grading resources and level of study, have motivational impact, prepare students for future classes, and produce credible results. Kibble (2017) and Norcini et al. (2011) recommend that measures used to test students maximize their learning and recommend that educators use test results to improve the quality of future teaching and learning.

Buckley's (2023) review of the most recent research trends on university exams identified five areas of active ongoing inquiry: the relationship between exams and students' learning, a related issue of authenticity of traditional exams, considerations related to assessment fairness and inclusivity, concerns about cheating and academic integrity, and the emotional impact of exams on students' wellbeing. Most of the areas identified by Buckley (2023) overlap with those mentioned above, suggesting that the same major considerations remain highly relevant today. Buckley (2023) also identified one area where research was lacking, and that is on how university instructors make assessment-related decisions in practice, including what additional environmental constraints they may face, and how they balance those constraints against all these other considerations. As will be discussed below, limitations such as budget and time constraints may lead instructors to choose more rational assessment methods, even if they differ from the ideal approach.

Alignment with Learning Outcomes

From an educator's perspective, assessments should be designed in a way that reflects educational objectives set out in the course outline, as well as accurately capture learning gains from the course content (Kibble, 2017; Norcini et al., 2011). This is because summative exams are achievement tests meant to assess students' level of knowledge and skills relative to target learning outcomes of the program and course. Moreover, these concepts should have been taught

in the course. Thus, the validity of an exam hinges on its alignment with learning outcomes and instruction.

Learning outcomes can be classified through Bloom's Taxonomy of Higher Educational Objectives (Bloom et al., 1956). Bloom's taxonomy is a pedagogical tool rooted in a theoretical understanding of how students move from surface-level thinking to deep-level thinking. In the original model, Bloom proposed six levels in the hierarchy of cognitive skills, with each level building upon one another and evaluating different areas of students' learning. Knowledge is the simplest tier in the cognitive domain; at this level, students focus on remembering, recalling, and recognizing facts. Comprehension refers to students' understanding of information; at this level, students can interpret and extrapolate meaning from information. According to Bloom, the comprehension tier is traditionally considered "the largest general class of intellectual abilities and skills emphasized in schools and colleges" (Bloom et al., 1956, p. 89). Knowledge and comprehension are often referred to as lower-order thinking skills, in reference to their lower position within Bloom's taxonomy. Further up the hierarchy are the higher-order skills. Application involves utilizing comprehension of knowledge and being able to apply it and use it in the right context. Analysis refers to students' ability to connect, differentiate, and organize information. Synthesis involves drawing these ideas together to produce a new and coherent knowledge product. Finally, at the top of the hierarchy, students can evaluate their new knowledge product by judging it and comparing it to other ideas.

An updated model proposed by Anderson and Krathwohl (2001) modified the levels of Bloom's taxonomy by changing the labels from nouns to verbs to emphasize what thinking action students are expected to do at each tier. In addition, they merged the analysis and synthesis tiers into one tier (Analyze) and added a new tier at the very top, Create, which involves

generating, planning, and producing new knowledge. In either case, the tiers still range from lower-level cognitive learning outcomes (Remember, Understand) to higher-level cognitive learning outcomes (Apply, Analyze, Evaluate, Create). Anderson and Krathwohl (2001) also added additional layers of thinking involved with each tier - called the knowledge dimensions. The knowledge dimensions include factual knowledge (knowledge of details), conceptual knowledge (knowledge of classifications and generalizations), procedural knowledge (knowledge of skills and procedures), and metacognitive knowledge (self-knowledge). Simply put, each cognitive process tier in this revised taxonomy includes all the knowledge dimensions. For example, a student would need to demonstrate factual, conceptual, procedural and metacognitive knowledge to complete an assessment with an emphasis on the Apply tier of the taxonomy. Bloom's taxonomy, and its more recent revision, is a useful framework for generating assessments that best align with educational objectives of post-secondary curricula.

Educational objectives can vary across programs, disciplines and levels of study. One relevant example of this comes from the degree-level expectations put forth by the Ontario Universities Council on Quality Assurance (2010). At the undergraduate level, degree-level expectations describe the acquisition of skills students are expected to demonstrate by the end of the Bachelor's degree, regardless of their major. These areas encompass areas of students' depth and breadth of knowledge, knowledge of methodologies, application of knowledge, communication skills, awareness of limits of knowledge, and autonomy and professional capacity. According to these guidelines, the learning expectations are higher for students as they progress with increasing levels of education. For example, an undergraduate student is expected to understand the limitations of their own knowledge and how these limits might shape their interpretation of novel problems, whereas an Honours undergraduate student is expected to

appreciate the uncertainty and ambiguity in addition to these prior learning outcomes. In general, with increasing levels of education, the learning expectations are often higher and assume students have engaged in more higher-order thinking. Thus, the council recommends that assessments match with students' level of intellectual development.

Applying degree-level expectations to Bloom's Taxonomy (1956), a first-year level course could be to focus on building theory-understanding skills to prepare for upper-year courses. Meaning, a first-year student could be more likely to read, memorize, and understand a range of ideas in their discipline (i.e., the lower tiers). Once students have built this foundation, they are expected to think about these ideas in-depth (i.e, the higher tiers) in upper-level courses (application of knowledge). Moreover, undergraduate students are more likely to focus on consuming knowledge, whereas graduate students are more likely to focus on generating knowledge (Ontario Universities Council on Quality Assurance, 2010). Bloom's taxonomy is only meant to be a guide for categorizing higher-order and lower-order learning outcomes in the entire learning process. The learning objectives may vary greatly from one course to another, from one discipline to another, and from one stage of learning to another, and the content and format of exams needs to be aligned with the intended outcomes (Kibble, 2017; Norcini et al., 2011).

Fairness

Educators are also concerned with maintaining fairness and reducing bias in the assessment and grading process. Exam results are commonly used to rank students in high-stakes decisions in admissions to graduate programs, scholarships, and professional licensures. Therefore, educators consider methods to ensure fair and unbiased assessments to mitigate

potential negative consequences or injustice (Darabi Bazvand & Rasooli, 2022; Rasooli et al., 2018)

Educators conceptualize fairness in assessment in two competing ways: equality and equity (Murillo & Hidalgo, 2020; Nisbet & Shaw, 2019). The elements of assessment that relate to equality include perceptions that a fair assessment should be equal across all conditions of the assessment process, including the materials, time limits, and resources (Murillo & Hidalgo, 2020, Nisbet, 2019). For example, ensuring the same administration, relevant content, transparent scoring, and objective interpretation of results for each student is a common equality-building practice in assessment design. Ensuring that tests and test questions are reliable and valid measures of learning is also important for equality-building practices (Nisbet & Shaw, 2019; Worrell, 2016), and often involves re-assessing assessment design after analyzing students' scores.

In the other realm, a fair assessment that is based on equity practices includes supporting students' individual and unique needs and characteristics (Tierney, 2014). This includes factors that may influence subjectivity in grading practices, such as avoiding stereotyping or attributing labels to students, or include more objective features in the assessment design itself (e.g., cultural references), that present as barriers to students with various backgrounds. Promoting equity in assessment may look like considering students' needs and context, and accurately measuring students' efforts through multiple means of assessment. For example, one equity-based practice is to implement accessibility-related accommodations for learners with disability- and health-related needs (Kibble, 2017). However, given increases in class sizes and diversity of students' backgrounds and abilities, individual accommodations may not be feasible.

A more inclusive alternative approach is to use the universal design for learning (UDL) framework, which has recently garnered traction to address equity-related concerns in education (Almeqdad et al., 2023). UDL is a tool educators use to facilitate an accessible learning environment for all learners. The framework holds that there are multiple means of engagement, representation, and action and expression in the learning process, and provide students with multiple options to access, build and internalize learning. By allowing students flexibility in how they approach assessments, in comparison to a one-size fits all equality-focused approach, the UDL model helps keep things equitable while still meeting course learning outcomes (Almeqdad et al., 2023).

So, while some educators believe equality in assessment process shapes fairness, others maintain that equity in assessment (making sure educators reduce bias in testing process) is paramount to keeping assessments fair (Murillo & Hidalgo, 2020). These current conflicting pressures have been at the forefront of educator's conceptions of assessment fairness. The consensus among assessment theorists is that fairness in assessment boils down to an absence of construct-irrelevant bias (Nisbet & Shaw, 2019). This means that assessments are fair only when they measure the intended construct (e.g., learning outcome), without being influenced by external factors unrelated to what is being assessed.

Academic Integrity

Maintaining academic integrity also remains a concern for educators – specifically facilitating assessment environments that prevent cheating and making sure students do their due diligence completing assessments honestly. Academic integrity is related to both validity and fairness of assessment (Salehi & Gholampour, 2021). When students feel unprepared, they are more likely to cheat on exams (Salehi & Gholampour, 2021). In a study examining factors

related to cheating behaviour in online examinations, Henderson and colleagues (2023) found students who reported exam experiences as stressful were more likely to cheat, with exam systems and proctoring services contributing to this stress. To reduce the potentiality of cheating on exams, instructors tend to opt for proctored and closed-book exams rather than open-book or take-home exams (Bengtsson, 2019). Therefore, concerns over academic integrity must be weighed against other considerations in assessment design (Kibble, 2017).

Pragmatic Considerations

Very little research has examined the influence of pragmatic factors and constraints on the university instructors' exam-related choices and practices (Buckley, 2023). In one such rare study, Bearman et al. (2017) interviewed a diverse group of university educators and identified a number of environmental influences. They contended that some educators inherit courses, must be consistent with organizational requirements and culture, and adapt to different student cohort abilities. In some cases, the content being assessed is part of the larger curriculum and going through the process of changing this content through approvals can be a burden. Other non-pedagogical considerations included resource influences, such as time constraints, money, technology, and logistics. Ideally, assessments should capture course learning objectives, however these environmental influences may not make this feasible. For example, an instructor may have limited access to grading support and opt for assessments that are easier to grade but are not ideal for capturing the target learning outcomes. These ideas imply that pragmatic issues can outweigh pedagogical ones in educators' choice of exam design.

Students' Conceptions of Assessment

Given the changing landscape of education research and practice, it becomes equally pressing to gain a phenomenological understanding of students' needs in education. Historically

there's been little interest in students' perceptions of exams (Tozoglu et al., 2004; Zeidner, 1987), but this has been changing. There is growing recognition among educators that students' perceptions and past experiences of assessments matter because they influence how students approach learning and studying (Gerritsen-van-Leeuwenkamp et al., 2019; Struyven et al., 2005). In this section, we briefly define the learning approaches and then review quantitative and qualitative research on different aspects of students' conceptions of assessment, and how those may relate to different learning approaches and outcomes.

Learning Approaches

Three general approaches to learning are distinguished in learning literature: surface, deep, and strategic (Asikainen & Gijbels, 2017). Surface approaches to learning refer to mechanistic study strategies aimed at memorizing facts and information with little reflection or desire to understand, typically for extrinsic reasons such as to meet task requirements or to get a passing mark (Asikainen & Gijbels, 2017). An example could include memorizing a list of key theorists' names and key dates before a test. In this approach, a student is only superficially engaged with course material (Struyven et al., 2005). Deep approaches to learning refer to students' active thinking about the meaning, relationships, and relevance of facts and information (Asikainen & Gijbels, 2017). Building on the example above, this could include building a deeper understanding of how and when each theorist developed their theory and how these theorists' contributions compare to other theorists. Essentially, deep approaches to learning focus students' internal efforts to build a comprehensive understanding of course information (Struyven et al., 2005). The surface and deep approaches to learning map on to Bloom's taxonomy of learning outcomes described above, where surface approaches are more consistent with the lowest tier of the hierarchy (remember), and deep approaches are more consistent with

the higher tiers (understand, apply, analyze). Learning science has demonstrated that deep study strategies, such as practice testing, elaborative interrogation, and self-explanation, lead to better long-term retention of the material, compared to surface study strategies such as re-reading, highlighting, and rote memorization (Dunlosky et al., 2013). Therefore, a deep approach to learning is widely regarded as desirable because it leads to higher quality learning outcomes (Struyven et al., 2005).

A third, strategic approach to learning has also been identified, which refers to how selectively students organize and put effort into their learning and studying to maximize exam performance and good grades (Asikainen & Gijbels, 2017). That is, students with a strategic approach may selectively study course material that they know will be tested on the exam and ignore other course content that will not be on a final test. Strategic learners complete assignments or assessments in accordance with assessment instructions and rubrics, rather than following curiosity, and will exert effort into thinking deeply about the material only when that's required for a good grade. That is why strategic learning tends to be more strongly correlated with university grades than pure deep learning (Richardson et al., 2012). In fact, Struyven et al. (2005) argued that most students are strategic about their learning, because students change their learning approaches in accordance with their perceptions of assessments.

Multi-Dimensional Models of Student Perceptions

Several models have been developed to capture multiple dimensions of students' perceptions of assessments in general (Brown & Hirschfeld, 2008; Gerritsen-van Leeuwenkamp et al., 2018). To derive these models, researchers first generated a pool of items based on previous research on student perceptions and the criteria that educators consider important (as outlined in the previous section). Then these items were administered to various groups of

students, and their responses were factor analyzed, to determine the core dimensions underlying students' perceptions.

One such model, the Students' Conceptions of Assessment (SCoA) inventory, was proposed by Brown and Hirschfeld (2008) based on a survey of over 3400 secondary school students after completing a reading comprehension assessment. Factor analysis identified four conceptions of assessment held by students: *assessment makes students accountable*; *assessment keeps schools accountable*; *assessment is enjoyable*; and *assessment is ignored*. Student accountability was the most endorsed dimension, which referred to students' belief that the purpose of assessments is to evaluate their work and progress toward learning objectives. School accountability was the second highest held perception, which referred to students' belief that assessments provide information on how well schools are doing at teaching students. Next was assessment as enjoyable, which referred to favourable perceptions that assessment can be a positive and enjoyable experience for the student and classroom culture. Finally, assessment as ignored referred to students' unfavourable perceptions that assessments can be irrelevant, unfair, and produce invalid results (Brown & Hirschfeld, 2008). This was the least endorsed dimension. From both instructor and student angles, the belief that assessments are designed to hold students accountable and classify student performance is deeply engrained (Brown & Hirschfeld, 2008; Sambell et al., 1997).

Brown and Hirschfeld (2008) entered the four SCoA dimensions into a model predicting students' reading comprehension test grades. They found that viewing assessment as essential for constructive growth and student accountability was the strongest predictor of higher test grades. In contrast, believing that assessments keep institutions accountable, perceptions of assessment as fun, and dismissive attitudes toward assessment were all weakly and negatively correlated

with test grades. Interestingly, students who enjoyed assessments often assumed that assessments were meant to evaluate school curriculum. Further research reviewed in Brown (2011) worked to validate and update the SCoA model, linking newly refined dimensions of conceptions (improvement, external attributions, affective benefit, and irrelevance) to self-regulation and academic achievement. From this updated scale, they again found the improvement dimension (i.e., assessment makes students accountable) was the only dimension to positively predict academic achievement (Brown, 2011). In discussing the implications of these results, Brown and Hirschfeld (2008) suggested that these results have important implications for assessments, particularly in how instructors inform students of the goals of assessment. Based on these findings, if students believe that assessments are intended to measure student learning, their grades are likely to go up, more so than if they believe they are only intended to evaluate school and teacher accountability.

Brown and Hirschfeld's (2008) SCoA model was based on research with primary and secondary students and was not directly generated from post-secondary students. To address this limitation, another model of university students' perceptions of assessment quality was developed by Gerritsen-van Leeuwenkamp et al. (2018), the Students' Perceptions of Assessment Quality Questionnaire (SPAQQ). Their research yielded six elements of perceptions of assessment: *effects on learning, fairness, conditions of assessment, interpretation of results, authenticity*, and *credibility*. Effects of assessment on learning referred to students' perceived learning outcomes, including relevance, motivation, and confidence in completing the activities. Fairness of assessment referred to students' perceptions that assessments are appropriate for education level and equal across students. Conditions of assessment referred to students' perceptions of assessments being well organized, timely, and carefully designed. Interpretation of

results referred to students' perceptions of validity and reliability, that scores on one test are consistent with scores on another. Authenticity referred to students' beliefs that the assessment aligns with knowledge and skills relevant to their future occupation. Lastly, credibility of assessment referred to perceptions that the assessment is of good quality, scored objectively and with integrity (Gerritsen-van Leeuwenkamp et al., 2018).

Taking the SPAQQ model a step further, Gerritsen-van Leeuwenkamp et al. (2019) examined how the six dimensions of students' conceptualizations of assessment quality related to learning behaviors and outcomes. Data were collected from 204 students from an applied sciences university in the Netherlands. Students' perceptions of assessment were measured using the newly developed SPAQQ. Students also completed a measure of deep, surface, and strategic learning approaches. In addition, students' grades were obtained from the university records, including an overall average, as well as separate averages across lower-order knowledge tests and higher-order assessments, to measure two levels of academic achievement. The researchers then ran multiple regressions, where all six SPAQQ dimensions were entered simultaneously to identify which dimensions uniquely predicted students' learning approaches and assessment grades.

Gerritsen-van Leeuwenkamp et al. (2019) found that students' perceptions of the effect of assessment on learning was the only significant predictor of all three learning approaches, positively for deep and strategic approaches and negatively for surface approach to learning. However, there was also significant shared variance, so all perceptions of assessment quality were associated with approaches to learning, but only one was a significant unique predictor. As for the overall assessment performance, as well as for assessments that tested higher-order thinking (as opposed to knowledge tests), student perceptions of assessment conditions appeared

to be a unique predictor of grades, but this was a marginally significant effect. These results indicate that students' perceptions of assessments are associated with their learning approaches: positive perceptions are associated with a deeper and more strategic learning approach, while negative perceptions are associated with a more surface-level approach. This effect was attributed by Gerritsen-van Leeuwenkamp et al. (2019) to how students view the challenges and value of assessments: if students think it is worthwhile, they will deepen their approach. Gerritsen-van Leeuwenkamp et al. (2018, 2019) recommended further research into what students' perceptions are based on, including students' expectations and past experiences with assessments. They also recommended gathering qualitative data to supplement quantitative surveys, to work toward a consensus about what assessment quality means between educators and students (Gerritsen-van Leeuwenkamp et al., 2019).

Qualitative Themes in Student Perceptions

Most qualitative research on students' perceptions of assessment has been done in the context of case studies evaluating specific assessment types (these will be reviewed later). Otherwise, few qualitative studies explored student's conceptions of assessments in general. Some notable examples include studies by Murillo and Hidalgo (2017) and Lynam and Cachia (2018). Murillo and Hidalgo (2017) focused on students' perceptions of fairness in assessment, exploring which dimensions of assessment students care about most and examples of specific instances of fairness and unfairness in assessments. Using a phenomenological approach, Murillo and Hidalgo (2017) conducted 32 interviews with primary and secondary school students, all from a range of diverse backgrounds (public vs. private; various socio-economic statuses). Interviews ranged from 15 to 20 minutes, beginning by asking participants what they perceived to be a fair assessment. The researchers uncovered two themes: *Fair assessment as an*

egalitarian approach and *fair assessment as equitable*. Consistent with the equality perspective, students explained that fairness in assessment should involve equal treatment of all students in the classroom, ensure equal testing conditions, include appropriate preparation, be assessed using a variety of assessment methods, and be graded on performance rather than behaviour (Murillo & Hidalgo, 2017). Another group of students expressed a desire to have assessments adapt to the needs of different students and have assessments more integrated with the learning process rather than fragmented, which is in line with the equity perspective on fairness (Murillo & Hidalgo, 2020).

A qualitative research study conducted by Lynam and Cachia (2018) aimed to understand psychology students' perceptions of how assessments in higher education affected their learning. This study intended to gain a deep understanding into how students view assessments, so that educators can see how these views can relate to their own. Lynam and Cachia (2018) conducted three focus groups with 23 sophomore and junior undergraduate psychology students. Participants were high achieving A-B grade students. Focus groups were led by two researchers using a semi-structured interview protocol. Questions asked participants to discuss their assessment experiences and approaches toward assessment. Focus groups lasted around one hour. The data revealed two main themes: Teaching factors and Student factors. Teaching factors included those aspects that were within the instructors' control, such as timeliness of assessments, sufficient guidance and instructions, and scaffolded assignments leading up to the assessment, all of which contributed to students' confidence going into an exam. Another set of teaching factors related to the type of assessment, such as predictability of the assessment content and format. Lynam and Cachia (2018) found that students' perceptions of predictable assessments were related to better preparation and grades, whereas low predictability was related

to higher workload and stress. In addition, student-focused assessments, which allowed choice, creativity, and authentic application of skills, were perceived as more interesting and enjoyable by students (Lynam & Cachia, 2018).

Lynam and Cachia's (2018) focus groups also brought up a number of student factors, including academic maturity that reflected students' metacognitive awareness of their own preparedness (self-evaluations) and whether their learning strategies were effective (academic awareness). Students also understood that grades do not always reflect learning levels, and their academic motivation appeared to stem from intrinsic sources such as career ambitions. Students also talked about how their emotions influenced their engagement with assessments. Specifically, stress and anxiety hindered engagement, while satisfaction and excitement helped motivate students. A key mediator that shaped these perceptions was found to be a tutor-student relationship, shaping perceptions of support, consistent instruction, and good quality feedback. In discussing the implications of their findings, Lynam and Cachia (2018) emphasized the need for institutions to be cognizant of assessment factors that create unnecessary stress for students and that help improve students' self-regulation strategies and intrinsic motivation related to assessments. One question that was not explored in Lynam and Cachia's (2018) focus groups is how students perceive the purpose of assessments in higher education. Although quantitative models like the SCoA and the SPAQQ (described above) include this aspect, they originate from the educators' perspective, which may differ from students' phenomenological view.

Section Summary

To conclude, research has identified several dimensions central to students' perceptions of assessments, namely alignment with learning outcomes, relevance and authenticity, fairness, and affective responses in the exam process. Some of these aspects directly overlap with

educators' concerns. For instance, both instructors and students appear to agree that the chief role of assessment is to promote and measure students' learning outcomes (Brown & Hirschfeld, 2008; Gerritsen-van Leeuwenkamp et al., 2018; Kibble, 2017; Norcini et al., 2011); this conception is an important predictor of students' learning approaches and exam performance (Brown & Hirschfeld, 2008; Gerritsen-van Leeuwenkamp et al., 2019). However, the extent to which students adopt a deep approach to learning also depends on how relevant the learning outcomes and assessment is to them (Lynam & Cachia, 2018). This is evidenced by students' perceptions of relevance in shaping their learning approaches; if the assessment is of value to them, they will exert more effort, whereas if students have low regard for the assessment process, they will adopt a surface approach (Gerritsen-van Leeuwenkamp et al., 2019). Both instructors and students think about fairness the same way, through equity or equality lens (Murillo & Hidalgo, 2017, 2020), and both care that the assessment results are reliable and valid (Brown & Hirschfeld, 2008; Gerritsen-van Leeuwenkamp et al., 2018; Kibble, 2017).

However, concerns about academic integrity appear to be more salient for instructors than for students (Bengtsson, 2019; Henderson et al., 2023). In contrast, affective responses such as confidence and anxiety are of greater importance to students, whereas this aspect may be overlooked by instructors. While educators may be aware of the negative impacts of stress and anxiety on students' learning and performance (Buckley, 2023), they tend to underestimate the motivational impacts of positive emotions and mindsets such as confidence, interest, and enjoyment (Brown & Hirschfeld, 2008; Lynam & Cachia, 2018). Lastly, both students and educators care about how assessments are formatted, timed, and administered, but for different reasons. Instructors often find themselves weighing both pedagogical and pragmatic considerations when they design and administer assessments (Bearman et al., 2017). For

students, assessment properties and conditions affect their levels of perceived stress, confidence, and motivation, which in turn shape their approaches to learning and ultimately learning outcomes (Gerritsen-van Leeuwenkamp et al., 2019; Lynam & Cachia, 2018).

To understand how specific assessment formats may be perceived differently by students, we next review the literature on traditional versus alternative exams in general, as well as two categories of alternative assessments, take-home exams and authentic assessments.

Traditional versus Alternative Exams

Traditional and alternative assessments are two overarching classes of summative assessment commonly distinguished in the literature. Different assessment types tend to evaluate different levels of cognitive skills (Bengtsson, 2019; Bloom et al., 1956), thus both styles have their desirable aspects as well as drawbacks. This section reviews the key features that distinguish traditional from alternative exam formats and summarizes background knowledge on their perceived advantages and limitations.

Traditional Exam Formats

The term “traditional assessment” is often used in reference to exams that are proctored, in-class, memory based (closed-book), and have a hard time limit of a few hours (Bengtsson, 2019). Traditional exams often take the form of standardized multiple-choice tests, although other types of questions can also be used, including fill in the blank, short- or long-answer questions, problem sets, in-class essays, or a mix of any of these question formats (Kibble, 2017). Some traditional assessments may not be entirely closed-book and may allow limited access to reference materials (e.g., formula sheet, one-page “cheat sheet” of notes), but access to

external sources such as websites or search engines is heavily restricted (Permzadian & Cho, 2023).

Traditional exams remain the dominant form of assessment in post-secondary education, with roots dating back to Binet's development of the first standardized intelligence test as a convenient method to assess students' likelihood of succeeding in the education system (Gardner, 1992). This approach has influenced educators' preference for traditional assessments in higher education due to its ease in administration (i.e., large set of questions with few grading resources needed), objective scoring, and ability to yield quantitative merit scores which are useful for ranking and comparing students (Kibble, 2017; Polat, 2020). Importantly, proctored exams are less vulnerable to academic integrity violations than non-proctored exams (Bengtsson, 2019). The standardized content and administration of traditional exams also aligns with the equality perspective on fairness (Murillo & Hidalgo, 2020). Students also often prefer traditional exams because they are highly familiar and perceived to be easier to prepare for (Struyven et al., 2005).

Despite their widespread use and many advantages, traditional exams have also received heavy criticism (Buckley, 2023). Students find the timed, proctored testing conditions too constraining, and many find testing in a class setting with other anxious students unnecessarily stressful (Polat, 2020). Traditional assessments, with their emphasis on factual recall under time pressure, have also been criticized by both students and educators for not being realistic relative to how learning will actually be used vocationally (Struyven et al., 2005). Indeed, the biggest criticism is that traditional exams primarily assess students' retention and comprehension of content knowledge as opposed to students' higher-order cognitive skills (Bengtsson, 2019; Gardner, 1992; Litchfield & Dempsey, 2015). That is, traditional assessments are effective at assessing the lower cognition tiers of Bloom's taxonomy (*Remember*, *Understand*), and may be

tailored to assess the middle tiers (*Apply*, *Analyze*) through carefully designed questions and problem sets. However, traditional exams are not suitable for assessing higher-order learning outcomes (*Evaluate*, *Create*), which require independent, divergent, and generative thinking that cannot be adequately tested with predefined answer options and problem sets (Bengtsson, 2019). As a result, traditional assessments tend to promote more of a surface and superficial approach to learning and studying among students (Struyven et al., 2005). To compensate for these limitations, educators have been exploring alternative assessment formats that are better aligned with higher-level learning outcomes (Struyven et al., 2005).

Alternative Exam Formats

Alternative assessments (also known as non-traditional assessments) are juxtaposed to traditional exams in that they tend to occur outside of the classroom, are not proctored, not memory-based, and have longer timelines to complete (days or weeks rather than hours; Bengtsson, 2019). One commonly used alternative format in place of traditional exams is take-home exams, where students get to complete the exam on their own time (Bengtsson, 2019). However, the category of alternative assessments is highly heterogeneous, encompassing many varied formats. All alternative assessments are designed with the common goal of testing higher-level learning outcomes, with many additionally having an explicit applied focus – to test students' ability to use the acquired knowledge in authentic ways that simulate real-world or professional settings (Struyven et al., 2005). Examples of such authentic assessments include final term research or capstone projects, portfolios (Segers et al., 2008), and case-based learning assessments (Baeten et al., 2012; Yoo & Park, 2015). Some alternative assessments may also incorporate groupwork and collaborative elements to enhance their authenticity. Thus, the authentic assessment framework is not merely an alternative format for summative exams, it is a

qualitatively different way of approaching the very purpose of assessments (Litchfield & Dempsey, 2015).

Research on the learning outcomes and student perceptions of alternative assessments usually pits these formats against traditional exams. The following two sections review the relevant research with respect to take-home exams and authentic assessments, specifically focusing on individually written exams.

Take-Home Exams

Take-home exams are a subset of alternative assessments in which students are given a series of questions or tasks to complete on their own time, non-proctored, over an extended period of time, with unlimited access to sources (Bengtsson, 2019). This assignment-style format allows assessing a wide range of learning outcomes, including those at the higher tiers of Bloom's taxonomy (Jacobs, 2021). Another related term is "open-book" exams. In the literature, some researchers use this term and "take-home exams" interchangeably, however not all open-book exams are take-home, some are still proctored and timed like traditional exams, but with access to reference notes (Permzadian & Cho, 2023). Similarly, during the COVID-19 pandemic the use of "online" exams was widely practiced, but not all exams conducted online were truly take-home, many were still remote-proctored, closed-book, and timed like traditional exams (Henderson et al., 2023). Here, we review research that falls under the category of alternative assessment – mainly research that uses take-home exams that are non-proctored, not memory-based, and have longer timelines to complete.

Affordances and Criticisms

Research findings on the advantages and disadvantages of take-home and open-book exams have been summarized in several qualitative reviews (Bengtsson, 2019; Durning et al., 2016; Johanns et al., 2017; Permzadian & Cho, 2023). Most of these reviews are limited in scope, either focusing on a single discipline (Durning et al., 2016), a narrow learning outcome (Johanns et al., 2017) or proctored rather than take-home open-book exams (Permzadian & Cho, 2023). Here, we will focus on Bengtsson's (2019) review, as this is the most inclusive evaluation of take-home exams to date.

Bengtsson's (2019) review was based on 35 studies published between 1968 and 2017. The studies used a variety of research methods including randomized controlled comparisons of take-home versus traditional exams, student surveys, and qualitative interviews. Therefore, the studies represent a mix of findings based on both educators' and students' perspectives. In the review, Bengtsson (2019) identified two advantages of take-home exams that were frequently cited across studies. The first included reduced stress and anxiety and more positive exam experience among students, who reported liking the flexibility regarding choice in location and the extended time frame. The second advantage, espoused by both educators and students, was that take-home exams are suitable for testing higher-order learning outcomes, such as students' abilities to apply, analyze, synthesize, and create knowledge.

Despite the reported advantages of take-home exams, members of the academic community remain resistant to implement these assessments. The most commonly cited disadvantage, particularly among educators, is academic misconduct concerns, where the non-proctored environment will lead to increased instances of academic dishonesty, plagiarism, or answer hunting (Bengtsson, 2019; Henderson et al., 2023). Educators' main concern was that

take-home exams are too open for cheating and plagiarism when they are used to assess students' knowledge of facts and definitions (Bengtsson, 2019). Given these trade-offs, Bengtsson (2019) concluded that testing lower-order learning outcomes on take-home exams is counter-advised because of academic integrity issues. However, these concerns can be mitigated when capitalizing on the strengths of take-home exams - by testing higher-level learning outcomes.

Bengtsson (2019) also identified some pragmatic considerations expressed by educators. Most agreed that take-home exams saved class time and resources compared to traditional exams where they needed staff and facilities to proctor them. However, there was some disagreement regarding whether take-home exams were more time-consuming to write and mark for instructors compared to traditional tests. Last but not least, Bengtsson (2019) reported that the most controversial aspect appears to be with respect to the impact of take-home exams on students' learning and study approaches and long-term retention, noting that instructor's opinions were polarized on this topic, and the research findings were also mixed.

Impact on Learning

Earlier comparative studies examining the impact of take-home exams on students' learning, study approaches, and long-term retention found that take-home exams were inferior to traditional exams. For example, in a seminal experiment, Marsh (1984) concluded that traditional tests were better for long term retention. Based on ten classes from five universities, they found that students who were randomly assigned to take an in-class test performed better on a surprise follow-up multiple-choice quiz one week later compared to those who had taken the same test but as a take-home exam. Marsh (1984) also surveyed students about their study behaviours and found that students tended to invest more effort in studying for the in-class test compared to the take-home test.

Similar to Marsh (1984), Haynie (2003) conducted an experimental design to explore how differences in testing format affected retention of learning at a longer follow-up of three weeks. They wanted to determine if initial testing, whether in-class or take-home, influenced retention and if students' study behaviours differed based on the anticipated test type. The study involved 279 technology education students randomly divided into four groups, two of which received an initial test and two control conditions that were not tested. Of the two tested groups, one received an in-class multiple-choice test, and the other received a take-home short-answer test based on the same questions. After the initial test, all participants took a surprise multiple-choice retention test three weeks later. The results showed that students who took the take-home test did better on a subtest containing repeat questions but scored significantly lower on a subtest containing novel questions compared to the other groups. Haynie (2003) concluded that take-home exams might encourage students to search for answers rather than adequately studying, unlike the groups who likely prepared more thoroughly for in-class tests, resulting in better retention of material.

Moore and Jensen (2007) also found evidence in favor of closed-book tests. In their experiment, they split a sample of 351 biology students into two groups: a control group section (n = 172) where students completed three closed-book exams in-class throughout the semester, and an experimental section (n = 179) where students completed two of three exams as open-book take-homes. Students were also asked to complete a survey asking about predictions for which areas of the exams or preparatory behaviour will impact grades. For example, students were asked the extent to which they agreed with the statement "I will earn higher grades on open-book tests than on closed-book tests", and both groups agreed (89 percent agreement for control group, 88 percent for experimental group). Results indicated that grades on exams 2 and

3 were indeed significantly higher in the experimental (open-book) section than the control section, while in course assignment grades were not significantly different across conditions. Class attendance rates were the same up to exam one but dropped for the experimental condition for exams 2 and 3, which was attributed to the exams being open-book. So, based on this evidence, Moore and Jensen (2007) concluded that open-book/take-home exams promoted poorer academic behaviour which resulted in worse long-term learning.

The three studies described above all contend that traditional in-class exams are better than take-home exams for long-term knowledge retention, partly because students studied more and prepared better for in-class exams (Haynie, 2003; Marsh, 1984; Moore & Jensen, 2007). However, a major limitation of these studies is that the in-class, take-home, and retention tests they used were all aimed at the lower tiers of Bloom's taxonomy, often using very similar questions. Therefore, their conclusion regarding the inferiority of take-home exams only holds when testing low-order cognitive skills (e.g., memorization, comprehension). Caution is advised against generalizing this conclusion to all types of take-home exams, as research suggests that take-home exams are usually designed with higher-order learning objectives in mind (e.g., application, analysis, evaluation; Bengtsson, 2019). Thus, the methods in these studies may not have been the best approach to measuring the effect of testing condition on learning outcomes, when the overarching learning outcomes between testing conditions should emphasize different levels of thinking. This would be like comparing apples to oranges.

When the operationalizations of closed-book and take-home exams are aligned with their respective learning outcomes, evidence shows that take-home exams lead to more engaged studying and improved retention. For example, Rich (2011) found evidence in favor of take-home exams leading to deeper processing and greater retention of course material and test

scores, as well as reduced anxiety. In their within-subjects experiment, they collected data from 35 students enrolled in two sections of an experimental psychology course, who completed four chapter tests throughout the course. Different chapters were randomly assigned to be assessed with either an in-class test or a take-home test (counter-balanced across sections). The take-home tests required students to understand, synthesize, and analyze the material in order to generate and answer their own questions (Rich, 2011). At the end of the course, students were given a surprise extra credit activity in lieu of a final exam, where 15 questions from each of the four previous tests were compiled to make a new in-class exam. After the final exam, participants were given a survey asking about their study strategies and habits when preparing for in-class versus take-home tests. Results showed a significant difference in performance on the items from in-class versus take-home tests. Specifically, the number of take-home test items answered correctly was higher than the number of in-class test items answered correctly, suggesting better retention of the material that had been tested through the take-home format. Participants also reported reading the chapter more times, creating more extensive notes, using self-explanations more often, studying more before and after class, being more engaged with the course content, and participating in more group study sessions when preparing for take-home tests than for in-class tests. In short, take-home exams did prompt students to study harder, use deeper learning strategies, and learn more than in-class exams (Rich, 2011).

Similarly, Theophilides and Koutselini (2000) surveyed 201 education majors asking them to respond to statements regarding their study behaviours for open- and closed-book exams, both leading up to and during the assessment. All students had some experience with open-book exams that tested higher-order skills, including analysis and synthesis of information from multiple sources (Theophilides & Koutselini, 2000). Consistent with the surface approach,

students tended to prepare for closed-book exams by focusing on memorizing knowledge and information, selectively studying what will be on the exam, and cramming in studying at the end of the semester. In contrast, students' preparation strategies for open-book exams included studying course material in more depth, going beyond the assigned readings, and applying higher-order thinking skills. Moreover, during the examination period, participants rated the majority of survey items (16 of 17) in favor of open-book tests, with the top three being that students showed more creativity, made better use of material organized in the course, and applied more critical thinking when taking open-book exams (Theophilides & Koutselini, 2000). One limitation of this study is that it was a retrospective survey and did not test knowledge retention.

A more recent study investigated whether take-home exams led to better retention over in-class exams using a longer time lag of 4-6 months (Spiegel & Nivette, 2023). They compared two consecutive cohorts of undergraduates who completed the same course with either an in-class exam or a take-home exam. The in-class exam contained a mix of lower-order multiple-choice and higher-order essay questions, whereas the take-home exam contained several complex higher-order essay questions. Four to six months after the exam, students in both cohorts completed a brief multiple-choice retention test of core course facts. In terms of long-term retention, there was no evidence for the advantage of one exam method over another, as both cohorts achieved comparable scores on the retention test. There was also no difference in exam grades between the two exam types. Interestingly, Spiegel and Nivette (2023) found students' exam grades in both take-home and in-class conditions were positively correlated to knowledge retention test scores at the 4–6-month follow-up, which they attributed to the importance of deeper learning approaches in general, regardless of exam type.

The discrepant findings between all of these studies could be due to several reasons. For one, the measures testing learning outcomes between conditions tend to be the same (same material, treated as same level of thinking), and comparing two groups with the same outcome variable (usually academic achievement generated from the same retention task). Given the design of each testing method and its alignment with Bloom's taxonomy (1956), it would not be theoretically sound to actually judge an assessment method's efficacy when test conditions have different learning outcomes by nature. Because in-class tests and take-home tests require different cognitive processes, Permzadian and Cho (2023) argued that application-based questions are ideal for take-home exams. Thus, the target learning outcomes are worse when the take-home exam is a multiple-choice test or aimed at lower-level learning outcomes, as opposed to essay which tests higher-level learning outcomes. The difficulty of an exam could be altered in this way (e.g., having a closed-book exam asking application-based questions would be difficult; an open-book exam assessing memory would be easier). This begs the question as to whether retention quizzes are appropriate for testing higher-level learning outcomes.

Similarly, other researchers have argued that take-home exams can lead to better learning outcomes when they include features that encourage deeper processing. For example, Roelle and Berthold (2017) tested whether using complex questions facilitated better retention in open-book tests. They compared retention performance following open-book versus closed-book testing conditions with low-complexity (summarization) versus high-complexity (inference) questions. Their findings showed that while low complexity questions did not significantly affect retention performance in closed- or open-book tests, high-complexity questions were more beneficial in an open-book format (Roelle & Berthold, 2017). Moreover, Hagström and Scheja (2014) argued in favour of using metacognition – thinking about thinking – to promote deeper processing in

alternative exams. They introduced meta-reflection activities to address high failure rates on a take-home exam in a politics of power course. These activities involved students reflecting on problems posed in the take-home exam and on their post-exam strategies and learning in general. After implementing these activities, Hagström and Scheja (2014) found the students were better able to grasp complex concepts, leading to better performance and reducing failed take-home exams.

Cahill-Ripley (2015) argued in favor of open-book exams, especially for law students. They noted that open-book tests themselves serve as a learning experience and are integral to the entire curriculum design, not just the end-point assessment. At the curriculum level in law programs, there is a need to assess two different areas of learning - legal writing skills and comprehension of terminology and critical thinking skills. Cahill-Ripley (2015) noted that closed-book exams may create a false impression of what skills will be tested on qualifying exams that are open-book, where the latter is more authentic and mimics real life situations a lawyer would go through in their career. In a survey of 47 law students about their assessment preferences, many students suggested using more problem-based questions and a variety of assessment approaches that align with curriculum design to achieve higher-level learning outcomes (Cahill-Ripley, 2015).

Students' Perceptions of Take-Home Exams

Research focusing on students' preferences, perceptions, and experiences of take-home exams has been limited and underrepresented (Tozoglu et al., 2004; Zeidner, 1987). Recently, there has been an increased interest in this topic as more instructors adopted take-home and open-book exams spurred by the COVID-19 pandemic. For example, Jacobs (2021) shared their experience using a take-home exam in a small upper-year Chemistry class during the shift to

remote learning. This was an open-book problem-based exam that required students to complete a series of scenarios and application questions in a week. Most students provided well elaborated well researched answers, however a small number of students still regurgitated course notes into the exam in a “kitchen-sink” approach in hopes for a good grade (Jacobs, 2021). In their end-of-course feedback, students reported they enjoyed taking this style of exam, particularly being given a scenario to work on. They also related that this approach helped remove the stress and anxiety associated with online proctoring services and was deemed a good way to be tested while also learning to apply course material. Jacobs (2021) recommended only using these styles of exams for smaller classes where grading is manageable, as they did not find it feasible to use this for their regular 200 student class throughout the normal school year. This speaks to the logistical constraints of administering alternative exams in practice.

Tam (2022) completed a qualitative research study with college students from Hong Kong to understand their perceptions of online take-home exams used at the onset of remote learning during the COVID-19 pandemic. They conducted nine semi-structured interviews with high-to-mid level achieving students enrolled in Arts, Business and Science programs. Each participant was interviewed twice for one hour about perceptions of take-home exams as well as study practices used in completing take-home exams and factors affecting their behaviours. Students’ perceptions of take-home exams were categorized into thematic areas around the types and level of difficulty of questions, question design according to Bloom’s taxonomy, technical issues, time span, and benefits and drawbacks of take-home exams (Tam, 2022). Participants expressed that take-home exams were more difficult than in-person exams due to the open-ended and applied nature of the questions, and that take-home exams promoted thinking at the higher tiers of Bloom’s taxonomy. This meant students put more effort into their exam and sought

research from external resources beyond rehashing the course content. Some participants expressed frustration due to the added difficulty associated with these aspects of take-home exams that they were not accustomed to in in-person exams. Perceptions of increased stress were associated especially with take-home exams that had tight time limits that caused rushing to upload documents to submission portals. Moreover, when the exams contained questions without an obvious "right" answer, this subsequently played into students' anxiety during the exam. Tam (2022) also found several aspects related to take-home exams that boosted students' confidence and motivation. Specifically, students appreciated it when instructors were actively involved throughout the take-home exam process in terms of providing feedback and answers to students' questions. Students also appreciated having rubrics and knowing the assessment criteria beforehand, so that could mitigate test anxiety by clarifying expectations and increasing predictability (Tam, 2022).

Slack and Priestley (2023) conducted a similar study examining factors influencing undergraduate students' experiences with online assessments during the COVID-19 pandemic and their implications for student well-being. A sample of 55 students from various institutions in the UK participated in focus groups of 5 to 15 students each. Focus groups asked participants about their experiences with online learning and online assessments during COVID-19. Particular to online assessment, three subthemes emerged: uncertainty over expectations, technology and environmental challenges, and confidence and flexibility. Students reported that unfamiliar instructor expectations in conjunction with the novelty of online testing environments elevated their stress and dissatisfaction with online exams. For instance, technological aspects of online exams including layout and knowing how to navigate the submission portal, added to exam stress. Online proctoring services and 24-hour exam systems also exacerbated exam stress.

Though, some students reported a liking toward the flexibility associated with online examinations, explaining that extending time limits helped boost confidence in their own performance. In some cases, participants noted that the university and instructors could have done more to recognize the additional emotional and cognitive loads during the abrupt shift to online learning. Slack and Priestley (2023) recommended that expectations need to be clearly communicated so that students know take-home online exams involve greater cognitive effort, and that online assessments need to be properly scaffolded to help students prepare and know what is expected of them.

Taken together, there are certain tradeoffs between traditional and take-home exam formats from the students' perspective. Take-home exams are perceived as less stressful because of their extended timeline and flexibility around when and where to do them (Bengtsson, 2019; Jacobs, 2021; Slack & Priestley, 2023). Whereas strict time limits are stressful for students regardless of the closed-book or open-book exam format (Slack & Priestley, 2023; Tam, 2022). However, students may be apprehensive of take-home exams if they are uncertain about expectations and have had little practice answering open-ended or applied questions (Slack & Priestley, 2023; Tam, 2022). As a result, students may prefer traditional exams due to their familiarity and predictability, or because they promote surface level processing and thus appear easier to study for (Struyven et al., 2005). At the same time, many students also value the opportunities to engage in higher-order thinking afforded by well-designed take-home exams (Bengtsson, 2019; Cahill-Ripley, 2015; Jacobs, 2021; Theophilides & Koutselini, 2000). In fact, Johnson (2015) reported an increase in students' preferences for take-home exams after completion, suggesting a better-than-expected experience. This preference contrasted with memory-based traditional exams, where preferences slightly decreased. Therefore, students'

preferences and attitudes may shift in favour of take-home exams after having had a positive experience with such exams (Johnson, 2015). Solutions to boost students' positive perceptions of the value of take-home exams include adding more application-based questions (Cahill-Ripley, 2015; Jacobs, 2021), adding high-complexity preparation tasks leading up to an exam, and having students reflect on their cognitions to think about their study process (Hagström & Scheja, 2014). Another area that can be used to increase student buy-in is ensuring the authenticity of the task is appealing and reflects students' interests – both vocationally and personally (Struyven et al., 2005).

Authentic Assessments

The term "authentic academic achievement" first appeared in Archibald and Newmann's (1988) *Beyond Standardized Testing*. In response to education reforms related to over-reliance on standardized testing, there was a growing awareness of the value of implementing innovative assessment pedagogy into curriculum to help capture more meaningful and transferrable forms of content mastery. In an education system that tends to value academic achievement and test scores as a gold-standard metric of learning, the movement toward authentic academic achievement gave rise to other more informative learning outcomes, such as individual learning experiences, cognitive growth, and critical thinking development. Archibald and Newmann (1988) hinted that authentic achievement can be promoted in instructional environments that utilize disciplined inquiry, application of knowledge, and value beyond evaluation. Part of the authentic assessment movement was built on ideas in line with Dale's (1969) Cone of Experience Learning Model. That is, students learn best when they actively engage in and "do" a task as opposed to learning in passive learning conditions, particularly when that learning activity has direct benefits to students' vocational lives. Authentic achievement is theoretically reached when students integrate

relevant course material into real-life contexts, within and beyond the classroom (Gulikers et al., 2004; Wiggins, 1991).

Goals and Features

Authentic assessment is a broad umbrella term used to describe alternative assessments with this applied focus. Early conceptualizations of authentic assessment conceded that authenticity is developed, or achieved, when the assessment material involves application to professional life – often a discipline-related career. Current conceptualizations have broadened the scope suggesting that authentic assessment is "a form of assessment where students engage in real-world activities where they can realistically apply knowledge and skills with fidelity" (Litchfield & Dempsey, 2015, p. 71). This definition is consistent with those early ideas presented above, but perhaps the key difference being authentic assessment is not limited to developing competencies in professional work conditions. This new wave of assessment design saw increases in authenticity in course material that extends to real-world scenarios experienced in professional and non-professional life to capture new student learning outcomes (Litchfield & Dempsey, 2015; Villarroel et al., 2018).

Frameworks exist to help guide instructors develop authentic assessments (Ashford-Rowe et al., 2014; Gulikers et al., 2004; Meyers & Nulty, 2009). One of the recent frameworks comes from Villarroel and colleagues (2018), who conducted a systematic literature review dating back to 1988. Reviewing 112 studies, they found a total of 12 concepts related to authentic assessment design, which were latently categorized into three main over-arching dimensions: *Realism, Cognitive Challenge,* and *Evaluative Judgement.* Realism in assessment can be enhanced when students engage in performance-based tasks that mimic assignments closely related to career aspirations (practical value), through simulated learning situations, or

other everyday situations. Assessments that promote cognitive challenge are assessments that target higher-order thinking skills, problem-solving and inquiry (i.e., the higher tiers on Bloom's taxonomy). Numerous studies also have focused on evaluative judgement (e.g., self-assessment, peer-assessment, instructor feedback) as a dimension of authentic assessment (e.g., Andrade, 1999; Ghosh et al., 2020; Ibabe & Jauregizar, 2010). Providing students with feedback on assessments, as done vocationally, helps regulate learning and lets students learn from mistakes. While also aligning with the job market, self-assessment (i.e., allowing students to evaluate themselves) is linked to metacognitive awareness and learning gains (Andrade, 1999; Graham et al., 2015; Sanchez et al., 2017).

Other frameworks in authentic assessment point toward a checklist of specific components of assessment design that must exist to be considered authentic. For example, Meyers and Nulty (2009) purport that authentic assessments need to be real-world and relevant to the students; be constructive; require use of higher-order cognitive processes; be aligned with the desired learning outcomes; and provide the students with challenge, interest and motivation to learn at a deep rather than surface level. Moreover, often in an authentic assessment design, a rubric is given to students a priori so they can understand educators' performance expectations, and as with any professional working environment, employees must know how to perform in advance (Gulikers et al., 2004). These features intentionally blur the line between summative and formative assessments in the service of learning (Meyers & Nulty, 2009).

Students' Perceptions and Impact on Learning

Authenticity in assessment is in part based on student perception (Gulikers et al., 2004; Kreber et al., 2007), and thus there is no clearcut blueprint to designing an authentic assessment. In fact, there may be a discrepancy between students' perceptions of what is authentic and

rewarding, and teachers' perceptions of what they believe to be authentic (Gulikers et al., 2004). That is, a student's individual goals for pursing post-secondary education may not align with other students' perceptions or with educators' learning objectives for the students. For example, a student wanting to pursue an academic career in psychology may consider writing a research paper more authentic or valuable than would a student pursuing a psychotherapy career in psychology (where a clinical case study would be preferred). Since authentic assessments can take on widely varied forms, students' perceptions and learning approaches to any particular alternative format should be examined on a case-study basis. To illustrate the diversity of this research, we review three case studies involving social-work case roleplays (Gulikers et al., 2008), science portfolios (Segers et al., 2008), and management simulations (Way et al., 2021). These case studies reflect how educators have taken creative and innovative approaches to promote realism and cognitive challenge in assessment design, paying closer attention to the context in which assessments are carried out (Herrington & Herrington, 1998).

Gulikers et al. (2008) studied how perceptions of assessment authenticity may differ between two groups of social work students (81 freshmen and 118 seniors) that differed in their level of practical work experience. The assessment involved roleplaying a social-work-related case where students had to demonstrate the skills and competencies learned in the course in front of judges. After the assessment, students completed a survey about their perceptions of authenticity in assessment design (including task, physical and social context, form, criteria), the alignment of the assessment with instructional goals, perceived generic skill development, and deep and surface study approaches. Gulikers et al. (2008) found that freshman students perceived the roleplay assessment as more authentic and worthwhile for generic skill development compared to senior students. There were no significant differences in study approaches or other

perceptions between the two groups. Follow-up focus groups with a subsample of participants revealed that senior students would have preferred more personalized cases related to their specific interests and areas of social work relevant to them, as well as more direct and specific prompts. First-year students did not have these preferences, most likely because they were new to the program and had no practical work experience yet. This provides further evidence that the degree of authenticity can be shaped by experience (Gulikers et al., 2008). Importantly, students who perceived the assessment to be more authentic were more likely to engage in deep learning and report higher perceived skill development than students who perceived it as less authentic, and these relationships held for both freshman and senior students (Gulikers et al., 2008).

Adding to Gulikers et al.'s (2008) findings, Segers et al. (2008) studied students' perceptions of authentic assessment and how these perceptions related to learning approaches, using science competency portfolios as assessments. Surveying 110 applied science students, the study asked them to rate their effort, motivation, learning gains, and use of feedback, as well as their deep and surface learning approaches. Their findings revealed that students employed more deep than surface learning approaches when completing the portfolio assessment and rated the use of feedback highly. There was a significant positive correlation between deep learning and use of feedback they receive, perceived learning gains, and the motivational impact of the portfolio assessment. Surface learning approach was negatively correlated with students' perceptions of adequacy and timing of feedback. It should be noted that although the portfolios were used instead of exams in this study, the way they were implemented was more formative (cumulative throughout the course, with regular feedback) than summative (Segers et al., 2008). Therefore, the findings about feedback may not generalize to other summative assessments.

In another study of authentic assessment that was more formative in nature, Way et al. (2021) implemented online authentic case-based simulations in a post-graduate management course. Throughout 12 weeks of instruction, students completed real-world activities that a manager would commonly experience in their job (e.g., a phone call with a distressed customer, handling a workplace emergency). At the end of the course, qualitative interviews were conducted with the participants. Students appreciated the authentic nature of the simulations, which made the assessments more challenging but also more rewarding and worthwhile. Students also expressed how the simulations contributed to increased perceived confidence and mastery of course content, transfer of knowledge, metacognition, engagement, and motivation (Way et al., 2021). The researchers also compared records of students' study behaviours (e.g., lecture views, discussion forum posts) in the simulation course with those of another group who took the same course without the simulations. They found that students in the simulation course were significantly more engaged with the course material (more lecture views) and with one another (more discussion posts), indicating more active learning (Way et al., 2021).

Aside from small case studies, there have also been larger comparative studies that have explored students' perceptions of "traditional" versus "alternative" assessment methods more generally (Flores et al., 2015; Iannone & Simpson, 2017; Pereira et al., 2022). In these studies, the "alternative" category specifically referred to authentic assessments such as portfolios, capstone projects, and case-based analyses. For instance, in the largest and most recent of these studies, Pereira et al. (2022) developed a scale to quantify students' perceptions of various uses of assessment types and analyzed their relationship with perceptions of effectiveness, fairness, and engagement. The study surveyed 5549 students across five universities from a range of programs, including STEM, health, medical and social sciences, and humanities. The assessment

use scale included 33 items related to preference for alternative and traditional methods, assessment features associated with perceptions of effectiveness and fairness, engagement with assessments, and extent of continuous assessment. Pereira et al. (2022) found that preference for alternative assessments was moderately positively correlated with perceptions of effectiveness, fairness, and engagement. Conversely, a preference for traditional methods was moderately negatively associated with levels of engagement and preference for alternative methods, and not associated with perceptions of effectiveness or fairness. These relationships held across disciplines and levels of study. These findings show that, overall, students tend to hold more favourable perceptions of authentic assessments relative to traditional exams, and that authentic assessments elicit greater student engagement in the learning process (Pereira et al., 2022).

Section Summary

The value of authentic assessment design mimics the benefits of alternative and take-home exams in terms of targeting higher-order cognitive skills (e.g., synthesis, critical thinking), with the addition of requiring students to integrate course material into real life contexts (i.e., the application of knowledge (Litchfield & Dempsey, 2015). Student buy-in is largely based on how they perceive the authenticity of assessment, in part dependent on the experience level and alignment with instructional goals (Gulikers et al., 2008). This approach fosters more positive behaviours and perceptions in students, including deeper learning approaches, expected knowledge transfer, as well as greater confidence and engagement (Gulikers et al., 2008; Pereira et al., 2022; Segers et al., 2008; Way et al., 2021). The evidence provided here shows that a movement toward more student-centered assessment approaches that target higher order learning outcomes indeed has promise (Struyven et al., 2005).

However, there are also some limitations regarding the use of authentic assessments in place of traditional exams. First, an authentic assessment framework might be better suited to formative rather than summative assessments, as ongoing feedback to the student is an integral feature of this approach (Meyers & Nulty, 2009; Segers et al., 2008). If authentic assessments are to be used for summative evaluation, ideally they should be scaffolded with similar formative assessments leading up to the final exam (Slack & Priestley, 2023). Moreover, authentic assessments might be better suited to upper levels of study. Since the nature of authentic assessment theory assumes that students already have a working knowledge of course content, they typically assess higher-order skills (Way et al., 2021). Novice learners, thus, may not be able to engage with authentic assessments effectively until they have a strong foundation with the concepts utilized in the authentic assessments. Novice learners completing authentic assessments have been found to experience cognitive overload and negative impact on learning (Sweller et al., 1998). In contrast, senior students tend to appreciate the challenge of tackling complex and novel authentic tasks (Gulikers et al., 2008; Way et al., 2021). There is a case to be made for both traditional and alternative assessments, as they both have pedagogical and non-pedagogical advantages and disadvantages in terms of their suitability for assessing certain learning outcomes and how these shape students' perceptions and approaches to learning.

Study 1 Objectives

The objective of this study was to build on the research discourse on assessments by exploring students' experiences and attitudes toward contemporary exams in higher education. We used a phenomenological approach to explore these perceptions, which captures the nuances in students' conceptions of assessment that may differ from those of educators (Gulikers et al., 2004; Lynam & Cachia, 2018; Struyven et al., 2005). Phenomenology is a qualitative research

approach that seeks to understand and describe the lived experiences of individuals with a specific phenomenon (Cypress, 2018). Rather than aiming for causal explanations, as in the Grounded Theory approach, Phenomenology focuses on uncovering the meanings participants attribute to their experiences. This method was appropriate for this study because our primary concern was not the factual accuracy of participants' accounts but rather identifying the key issues and perspectives that were meaningful to them in relation to the university exam process.

Traditional quantitative models of students' conceptions of assessment (e.g., Brown & Hirschfeld, 2008; Gerritsen-van Leeuwenkamp et al., 2018) are limited because they are constructed from the educators' perspective on assessments. There remains little qualitative research into students' experiences with diverse types of high-stakes assessment, particularly within specific disciplines such as Psychology. One notable exception was the focus groups by Lynam and Cachia (2018), which explored the role of instructor and student factors in shaping students' emotional and motivational responses to assessments. However, no qualitative studies have explored students' perceptions of the purpose of summative exams. Understanding this dimension, especially in the context of Psychology, is critical given that students in this discipline pursue diverse academic and applied career pathways. Students' appraisal of the assessment's relevance is one of the strongest predictors of deep learning engagement (Gerritsen-van Leeuwenkamp et al., 2019; Struyven et al., 2005). Therefore, we queried students not only about their perceptions of purpose based on past experiences, but also on what they thought the ideal purpose of exams should be.

Another limitation of existing research on students' conceptions of assessment is that much of it was conducted prior to the COVID-19 pandemic. Students have been exposed to many more alternative assessments since then, and an updated picture of their perceptions is

needed. Additionally, we aimed to fill a gap in current understanding of the sources and attributions of students' perceptions. Existing studies have heavily relied on quantitative methodologies that looked for predictors of students' perceptions, but these may not align with students' phenomenological attributions. Our research aims to create a thematic framework based on the thematic analysis of students' perceptions of exams, which would offer constructive suggestions that incorporate students' feedback with common instructional practices. The goal is to inform educators who are looking to enhance assessment design practices that balance both students' goals and course learning objectives.

We conducted qualitative interviews with upper-year Psychology students to inquire about students' perceived understanding of the purpose of exams, characteristics defining an effective exam, perceptions of fairness and unfairness, affective responses, and learning and preparation strategies. Thus, one research question was generated to guide Study 1: *What are current post-secondary students' perspectives, attitudes, and experiences with summative exams?*

Study 1 Methods

Participants

A total of 10 participants were recruited to participate in interviews. At the time of interviews (Spring/Summer 2023), all participants were upper-year undergraduate students enrolled at Trent University, a mid-sized suburban university in Ontario, Canada. Participants that completed interviews were senior students (third year of study or higher) majoring in Psychology or a related science discipline (see Table 1).

Participants for this study were recruited from a cohort of students who had previously (in Spring/Summer 2022) completed a second-year Introduction to Abnormal Psychology course

at Trent University taught by the faculty supervisor (Kateryna Keefer). This cohort was selected for this study because they had completed an alternative take-home exam as part of that course. In addition, they had also experienced remote learning during the COVID-19 pandemic the year before that. This way, we could ensure that all interviewed participants had experienced a variety of assessment methods including various traditional (in class, sit down, proctored) and alternative (take-home, open book, online) exams prior to the interview.

Procedure

This study was approved by the Trent University Research Ethics Board (REB protocol # 28469). In May of 2023, I sent an email invitation to participate in a research interview study titled "Authentic Assessment: Student Experiences and Learning Outcomes". Of the 29 invitations that were sent, 10 students accepted, resulting in a 34.48 percent response rate. I met one-on-one with each participant at a time that was convenient for them remotely via a secure Zoom connection. I followed the interview protocol (see Appendix A for Study 1 materials) to inquire about students' experiences and attitudes toward exams, including positive and negative experiences, exam preparation behaviours, fairness, and satisfaction. The questions were open-ended to facilitate a more natural conversation. I also gave myself leeway to ask additional follow-up questions not outlined in the protocol based on the flow of the conversation to encourage a more directed focus on the topics mentioned by the interviewee. Interviews ranged between 41 to 78 minutes, with an average time of approximately 55 minutes. The audio was initially recorded and transcribed using the enabled captioning function on Zoom, and I subsequently reviewed the audio to ensure the final transcript had clarity and was accurate. Transcripts were distributed within the research team for coding and analysis. All participants received a $20 Amazon.ca electronic gift card at the interview's conclusion.

My role in the research process was as the primary researcher. This meant that I took the lead in constructing the interview protocol, recruiting and interviewing participants, performing thematic analysis and writing the results. I was also a former teaching assistant for approximately half of the students interviewed, although all interviews were conducted one year after the course completion. Nevertheless, this dual role may have meant that even the follow-up questions, where I asked participants to elaborate on certain points, were likely influenced by my positionality. The faculty member of the research team (Kateryna Keefer) played an advisory role in this study and was not in direct contact with the participants, although her name was listed on the recruitment and consent documents.

Interview Protocol

During protocol construction, the faculty advisor and I regularly met and shared drafts of interview questions. We provided each other with feedback on how to rephrase questions that were leading, and to make questions more open-ended. After we agreed on an initial protocol, I completed two mock interviews with friends not part of the project, and I used my discretion to modify the protocol based on interview flow or leading questions identified in the mock interviews. The final interview protocol was semi-structured, and included warm-up questions, key questions and closing questions.

The goal behind using warm-up questions was to start building rapport with the participant and to get them talking. To do this, I asked more broad and personal questions that led into the research questions, such as why they chose to pursue a university degree and what their favourite course has been to take at Trent and why. These warm-up questions primed the participants to start thinking about their post-secondary education. I gave myself enough leeway during this phase to ask follow-up questions based on participants' responses. For example, if a

participant mentioned choosing to pursue a university degree with the goal of attaining a specific career, I would have asked them why they wanted to pursue that career. I gave myself this much leeway because the goal of this phase was to build rapport and taking an interest in this regard helped set participants' comfort levels for the remainder of the interview.

The key interview questions were used for the bulk of the analysis. Here, I asked participants questions such as what they thought the purpose of an exam was, how they felt going into exams, to describe a fair and valid assessment of learning, what makes an exam effective, and to describe their best and worst exam experiences. We also added a safeguard question near the beginning of the key questions, but after participants had already described their purpose: "what kind of exam formats are you familiar with". Based on past literature, I gathered that students' default answers tended to lean toward traditional assessments, even if they have completed alternative exams. Here, we clarified any misconceptions and offered examples of other types of exams to help our participants understand exams beyond their schemas of exams as memory-based tests. Some leeway was given for the key interview questions. Most of the time, I asked participants to elaborate and expand on certain points, asking them to provide examples to illustrate what they just described. Early in the interview process, I did make an inadvertently suggestive comment by asking a follow-up question prompting interviewees to think about the purpose of exams in terms of accountability of the instructor. This follow-up question was inspired by the first participant I interviewed, who brought these arguments up naturally. During analyses, I excluded these and other responses that followed other inadvertently suggestive comments or leading follow-up questions asked by me.

Lastly, closing questions were asked to end the interview. Questions such as "what is one kind of exam you wish you had but never had," and "if you could design your own ideal exam

for one of the courses you are taking, what would this look like" were asked. Unlike warm—up questions that brought participants' focus to the research topic, the closing questions asked future-oriented questions. Based on how these questions were answered, these features of ideal exams were used to help inform the main themes.

Data Analysis

Thematic analysis was conducted using Atlas.ti., following Braun and Clarke's (2006) six-stage process. During the qualitative data analysis, the faculty advisor and I both completed an initial code generation activity independently of each other, where we read two of the transcripts and inductively coded them. It was clear there were multiple ways these transcripts could have been coded. I approached my initial code list by evaluating, line by line, what each passage was trying to describe. This generated approximately 30 different codes, which were refined and whittled down after regular discussions with the faculty advisor until the final list was agreed upon. Convergent ideas led to a total of five codes, which were applied to the subsequent transcripts All transcripts were coded to reflect these changes: *Purpose, Effectiveness, and Fairness; Format; Preparation and Process; Emotions and Cognitions Around Exams; and Improvement* (refer to Appendix B for the code manual). Following the initial analysis of the code report, I identified a total of 10 potential themes. Upon reviewing the initial 10 themes, it became apparent that some of the larger themes encapsulated other, smaller themes. Based on prior literature and the structure of the questions, it made sense to develop themes that captured student-related aspects in the assessment process (emotions and cognitions, preparation), instructor-controlled technical aspects (exam format), and students' thinking about instructor-controlled technical aspects (purpose, effectiveness, fairness). When reviewing these themes further against the coded transcripts, I attempted to look at the bigger narrative of how

participants thought about assessment. This led to a further revision and narrowing down of the themes to three main themes related to *Purpose of Exams, Exam Fairness and Effectiveness*, and *Affective Responses,* all of which are described in detail below. These themes reflected the prompts used in the core interview questions, though the specific subthemes identified were different from the findings in academic literature in similar studies. This is because these over-arching areas were the aspects I was hoping to gain insight about.

Study 1 Results and Discussion

As shown in Table 2, the thematic analysis of participants' interview data yielded three main themes with additional subthemes. The first theme identified was related to the *Purpose of Exams*. The interviewer asked two questions about purpose of exams, one based on students' experiences of past exams, and one about students' ideas of what exams should be for. Interview participants highlighted a discrepancy between the ideal and desired purpose of assessments (exams as should be) with how assessments were actually used (exams as practiced), and these perceptions were framed in terms of the alignment between exams and program objectives. In the second theme, students discussed positive and negative appraisals related to *Exam Fairness and Effectiveness*, often using these terms interchangeably, even though the interview questions asked separately about effectiveness and fairness. The last theme highlighted the *Affective Responses* in students' exam experiences. To this end, several interview questions inquired about students' emotions, including how participants felt going into an exam, as well as their best and worst exam experiences. Participants explained various factors leading to their onset of test and grade anxiety and confidence during the exam process.

Theme 1: Purpose of Exams

One key question we aimed to investigate was related to how students form perceptions of the purpose of exams. Throughout the interviews, participants shared their perspectives on why and how university exams were currently used, or ought to be used, in their courses. These insights were categorized into two subthemes: students' understanding of *exams as should be*, and students' understanding of *exams as practiced*. In addition, running across both these subthemes was a third subtheme reflecting students' perspectives on how the *purpose of exams aligns with program goals and learning objectives*. The third subtheme is incorporated throughout the discussion of the other two subthemes.

Exams as Should Be

Several ideas were relevant to participants' perceptions of the ideal purpose of exams: accountability, meaningful learning, and continuous learning.

Accountability. The chief purpose of an exam being tied to the assessment of learning was a prevalent explanation within the interviews. Generally, all participants agreed that the final exam is intended to test students' knowledge and skills gained from the course, as exemplified in this quote:

> "*I think the on-paper focus of an exam is to display all of the knowledge that you've learned through the course in one spot. I feel like that is the bottleneck idea of what the final exam should be*" (P4).

This default answer is consistent with the dominant dimension in Brown and Hirschfeld's (2008) model of students' conceptions of assessment, namely that assessments make students accountable, where the intent of using assessments is to evaluate students' work and progress toward meeting learning objectives. This finding is not surprising, as the belief that assessments

are designed to hold students accountable and classify student achievement is deeply engrained in both students and instructor's perceptions (Brown & Hirschfeld, 2008; Kibble, 2017; Sambell et al., 1997).

Another, but less frequently mentioned explanation, aligned with a different dimension of Brown and Hirschfeld's (2008) model, namely that assessment is a means of keeping schools accountable. For instance, Participants 1 and 5 expressed the belief that students' performance on an exam also served as an indicator of the instructors' effectiveness in teaching the course. Participant 1 shared:

> *"I think it's also kind of reflective on the professor and how well they taught. Because obviously if they're not very effective at teaching then their average won't be that high"* (P1).

Although this particular participant did mention this perspective, it is worth pointing out they did so after discussing the default answer many other participants provided: that assessments are mostly intended to evaluate students' learning.

Meaningful Learning. Another often discussed topic that surfaced focused on how aligned exams are with program goals and learning objectives. Digging deeper into the question of what kinds of learning outcomes should be evaluated on exams, participants overwhelmingly agreed that exams should go beyond simple memory recall and instead test deeper understanding and application of course material. For example, Participant 8 shared that exams are

> "*supposed to test your comprehension. It's just sometimes it always depends on the class, but sometimes it just tests what you have memorized rather than what you understand"* (P8).

In fact, this explicit juxtaposition between assessments that tested for understanding versus assessments that tested for memorization was a common point of distinction in participants' perceptions of exams as should be versus exams as frequently practiced (more on the latter below).

Aside from demonstrating comprehension, several participants also talked about the added value of showing what they have taken away from the course or being able to apply what they have learned. For example, Participant 4 described how meaningful learning was not purely theoretical but also personally relevant:

> *"I find it nice when there's one question in there that's like, 'what did you gain from this course?' Because sometimes you really do learn things that, maybe not be as academic, but maybe more applicable to your life. For example… learning about a mental disorder that affects you personally or your family personally. It's obviously an academic gain, but also a personal gain as well" (P4).*

In a similar vein, Participant 3 appreciated that, in addition to their exams in Computer Science being an applied learning exercise directly relevant to their discipline, the conditions of the exam also mimicked real-life working conditions:

> "*there's something to be said about testing things like how you handle information under pressure… under anxiety situations, what can you actually recall or what can you do? And again … it's kind of hitting back to that whole practical job level thing" … you have to make these snap decisions, or you have to do your job immediately. Don't think, just do, sort of thing. And I think the exams help with that as well*" (P3).

In sum, our participants' perspectives on what learning outcomes should be tested on exams emphasized higher-order cognitive skills (e.g., comprehension, application) over

memorization, meaningful opportunities to apply course material in their lives, and alignment with professional skills related to their future careers. These views are consistent with the literature on take-home exams and alternative assessment designs that also prioritize higher-order learning outcomes such as application, analysis, and synthesis of knowledge (Bengtsson, 2019; Cahill-Ripley, 2015; Jacobs, 2021; Struyven et al., 2005). Moreover, our participants' preferences are closely aligned with the notion of authentic assessment and its key features of cognitive challenge and realism (Meyers & Nulty, 2009; Villarroel et al., 2018). Authenticity is also one of the dimensions in Gerritsen-van Leeuwenkamp et al.'s (2018) model of students' conceptions of assessment, and it emerged as one of the qualitative themes in Lynam and Cachia's (2018) focus groups with psychology students. In turn, students' positive experiences with alternative and authentic assessments have been associated with greater confidence and mastery of course content, perceived transfer of knowledge, deeper learning approaches, and enhanced engagement and motivation (Gulikers et al., 2008; Segers et al., 2008; Theophilides & Koutselini, 2000; Way et al., 2021).

Leading up to this point, our participants mentioned their preference for application-based questions as well as having assessment aligned with program objectives. It is worth raising how these ideas might be shaped by students' levels of experience. While our sample was comprised of upper-year students, Gulikers et al. (2008) found differences in how freshman versus senior students thought about authentic assessment differently. In their study, freshman students in a social work program perceived case roleplay assessments as more worthwhile for general skill development compared to senior students, likely due to the fact that senior students had already been exposed to these generic case studies. Alternatively, senior students expressed preference for more complex cases that were even more closely aligned with the realities of

professional practice (Gulikers et al., 2008). There is also research showing that students develop more positive perceptions of alternative assessments with greater experience with those assessments (Johnson, 2015). Since all of our participants had prior experience with a case-based take-home exam, their perspectives may be more favourable than students who haven't had that experience.

Continuous Learning. An additional aspect arising from students' perceptions of the ideal purpose of exams stems from more philosophical attitudes concerning how and when students should be assessed. Specifically, Participants 1, 3, 5, 6, and 8 all argued against the conventional notion that assessing students at one point in time holds value. According to Participant 6, the ideal exam would

> "*look like prioritizing learning over people's ability to retrieve information at a particular time… exploring different non-traditional methods of testing knowledge and making sure that people are actually learning*" (P6).

This perspective is in line with research on students' perceptions of take-home exams, showing that when an exam has an extended timeline (several days or weeks), students perceive it as more appropriate for testing higher-order learning skills and will adopt a deeper learning approach, which contrasts with the limitations of traditional timed exams (Bengtsson, 2019; Jacobs, 2021; Slack & Priestley, 2023). This comment also echoes the findings by Hagström and Scheja (2014), who had students complete metacognitive activities to promote deeper processing. These activities helped students grow and learn from assessments rather than just measuring their knowledge at one point in time. In addition to helping students become aware of what they are learning, it also helps them develop critical thinking strategies to transfer into new assessments.

A related idea was expressed by another participant, who mentioned the purpose of exams should be tied to helping students reach the next sequential step in their program:

> *"[Exams are] supposed to be like the building block. Because obviously when you do a final exam to end the course, it doesn't stop there. Usually it's something that will lead into something else*" (P7).

The notion that assessment should be an ongoing process that informs continuous and future learning is another key feature of authentic assessment (Meyers & Nulty, 2009; Villarroel et al., 2018). Ongoing opportunities for practice and feedback have been found to contribute to students' positive perceptions of alternative assessments (Segers et al., 2008; Tam, 2022).

Exams as Practiced

Some of the ideas for participants' ideal exams were based on the exams they had experienced, indicating a certain degree of consistency between the intended and actual uses of exams. However, there also existed a contrast between how the participants talked about the ideal and actual uses of exams.

Focus on Memorization. The discrepancy in perceptions of ideal versus actual exams was especially evident with respect to the assessed learning outcomes. While the emphasis in the first subtheme – exams as should be – predominately centered around comprehension, application, and relevance, the descriptions for the actual purpose leaned toward surface-level testing of memorization and recall (which was perceived as less relevant) with fewer participants mentioning comprehension in this context. This quote from Participant 10 captures the shared attitude among the participants that exams as practiced often fall short of the ideal:

> "*Sometimes it's more of just, you know, recall like spitting facts… like, you know, when was this? Blah, blah blah. Where is this?*" (P10).

This stance forms the subtheme, exams as practiced, as it highlights that students think about the purpose of a traditional exam in terms of testing the retention of course material with memory-based questions as the tool used to assess this retention. This perspective echoes the general dissatisfaction with traditional exams among educators as well (Buckley, 2023; Struyven et al., 2005). As discussed in the literature review, traditional exams tend to promote more of a surface and superficial approach to learning and studying among students (Bengtsson, 2019; Gardner, 1992; Litchfield & Dempsey, 2015; Struyven et al., 2005). This means that students often perceive these exams as a means to test retention and comprehension of content knowledge acquired over the semester. This might explain why participants' default answers in our interviews focused on traditional exams. Also, traditional tests that evaluate students' ability to remember and recall course material and recognize facts may have been most commonly used in our participants' education experience at the time of interview and thus most readily available on their mind when explaining their perceived purpose of exams.

Participants' responses juxtaposing actual versus ideal purpose of exams were closely linked to Bloom's taxonomy of learning objectives (Bloom, 1956). Using this framework, we can understand the dichotomy in students' thinking about exams. Our participants aligned the assessments as practiced perspective with the lowest tier in the taxonomy, memorization (knowledge) with fewer mentions of understanding (comprehension), and the assessments as should be perspective with the second and third tiers of the taxonomy, understanding (comprehension) and application.

Non-Pedagogical Reasons. Several participants brought up other explanations for why exams may not always align with the intended learning objectives, such as administrative barriers that could ultimately determine the type of exam given to students. For example,

Participant 7 noted the reality that not all exams can align with the course and program objectives:

> *"I know it's because of the lack of resources, and of course there's only one professor and a limited handhold of TAs. So yeah, I'm well aware of like the situation. I know it's a lot, but I understand that too" (P7).*

In addition to participants' appraisals of how they view the purpose of exams, they also acknowledged the practicality and appropriateness of certain exams in specific courses. For example, various university programs now have job training goals, leading to differences in the types of exams administered. Participants 2, 3, and 8, all in what would be considered more applied streams (Nursing, Computer Science, and Biomedical Sciences, respectively), highlighted the lack of standardization in assessment practices across courses and across disciplines.

> *"There is definitely a big disconnect or just teaching styles and what they're trying to prepare you for… you can have the exact same class taught by two different professors and have two wildly different exams, either in terms of difficulty or even just the general set up" (P3)*

Lastly, one participant shared a negative take related to exams, explicitly conveying they felt exams were designed to be intentionally challenging to lower students' grades. This viewpoint implies that exams can be used as a convenient method for instructors to designate a significant portion of students' grades to a single type of assessment:

> *"I do feel like they kind of in some ways, not set you up for failure, but they're not designed to be done well, unfortunately … Most people from what I've experienced with*

all of my friends and I never, ever, expect my grade to go up after a final exam … it's so easy to bring your final grade down" (P4)

To summarize the entire Theme 1, our participants generally perceived the purpose of exams as practiced as testing students' cumulative knowledge of course content at a specific point in time. Contrastingly, the same participants liked to have multiple opportunities for assessment throughout their courses, or at the very least, effectively build students' readiness to perform well on exams. These differing perspectives were tied to the participants' perceived ideas about the goals of their course and program goals. It is worth noting here that the interviews did prompt students to think about exams beyond the scope of traditional assessment. Default answers tended not to delve into other, alternative assessments, potentially because, again, students associate the umbrella term "exams" with exams that are sit down, in class, and timed. This semantic association may be why participants did not mention other, higher-order tiers of Blooms' taxonomy (analysis, evaluation, or creation) in their responses related to the purpose of exams as practiced.

Theme 2: Exam Fairness and Effectiveness

Two of our research questions aimed to understand how students evaluated the fairness and effectiveness of exams, as well as the criteria that were relevant to students when evaluating exams. The interviews prompted participants to think about exams in terms of fairness and effectiveness separately. In their responses, however, participants brought up similar ideas that could not be easily differentiated to warrant two separate themes. Thus, we collapsed these aspects of assessment together, breaking them into two areas that shaped participants' perceptions of effectiveness and fairness: *Alignment with Learning Objectives and Expectations*

and *Effectiveness and Fairness in Exam Format*. In addition, participants also expressed how *Exam Performance was Not a Factor* in their judgments of effectiveness and fairness.

Alignment with Learning Objectives and Expectations

Building off the purpose of exams discussed in Theme 1, some of the participants' ideas about exam effectiveness and fairness focused on the alignment between exams and students' expectations of what exams were supposed to test. Participants framed these ideas in terms of the general learning objectives and the specific expectations communicated by the instructor.

Alignment with Learning Objectives. The (mis)alignment between course learning objectives and assessment was one of the holistic factors that shaped students' perceptions of an exam's effectiveness and fairness. According to Participant 2, the effectiveness of exams should be judged in relation to the exam's goal:

> *"What makes an exam effective? Effective at what? Achieving its goal. What's its goal? What makes an exam effective? What is the goal of this exam we're talking about?" (P2).*

Similarly, Participants 2, 5, 7, and 8 all spoke to exams being tied to learning goals, getting at applied versus theoretical program learning outcomes and how assessment should be tied to this. Participant 5 suggested that each exam question should be relevant to the learning outcomes:

> *"I suppose the question needs to be justified. You know, why is the question being asked? Yeah, and If it's not really justified, then it shouldn't be on the exam" (P5).*

And, although many instructors consistently match learning outcomes with assessment questions, Participant 2 mentioned their frustrations with when they perceived an exam question was irrelevant to them:

> *"when I know it starts getting unfair is when I'm like I don't know the answer to this. I chose not to know the answer to this because I feel like this is not important to me in any way shape or form" (P2).*

The dimension of credibility is highlighted in both Brown and Hirschfeld's (2008) and Gerritsen-van Leeuwenkamp et al.'s (2018) models of students' conceptions of assessment, indicating that students will ignore or disengage from assessments if they perceive them to be irrelevant, unfair, or invalid.

Alignment with Instructor's Communication and Expectations. Most commonly, participants expressed that instructor's communication with students leading up to the exam determined their perceptions of exam effectiveness and fairness. Areas of communication (or lack thereof) that caused frustration included material promised to be on the exam that was not there, or inconsistent preparation guides, or the instructor's tendency to *"teach one thing and then test us on another" (P5).* Regarding the lack of communicated test expectations, Participants 7 and 8 both emphasized the importance of instructors consistently sharing with students their expectations of the exam:

> "*As long as what you would said matches up with what shows up in the exam, I feel it's a fair exam. Because at the end of the day I'm a student, I'm expected to know and know the content. I'm supposed to study this content. That's why I'm here. So as long as you set your expectations for me that I can actually obtain them, and that when I go to the exam, I see what you would ask of me... I'm fine" (P7).*

In one example of an exam that was perceived to be unfair, Participant 1 shared their experience with a new instructor, stating that the bulk of their course included many guest

lectures by external speakers in their field, none of which they were tested on, despite assuming they were going to be:

> "*A lot of these guest speakers were over zoom… I guess leading up there were quite a few hiccups… stuff that the guest lecturers had talked about wasn't really on [the exam]"* (P1).

In another instance, several participants mentioned the lack of clear expectations they experienced in a take-home exam. In situations where there is lots of writing, participants were worried they didn't fulfil these expectations at the time of writing:

> *"worrying that you didn't say something that you should that they wanted to know because I think you remember the final exam, it was like try to give as much as you can to show that you know the content". (P2)*

Related to test-prep guides, some participants felt the content covered in these guides either did not help them with the exam, or was misleading about what would be tested on the exam:

> *"That's what's frustrating about it, cause they take a hard topic, you learn about it, you spend a lot of time to write about it perfectly, because they said it's gonna be a long answer. And it's just not on there at all. And there is a question about something else. Would you spend less time studying because that question, which you studied for was supposed to be 15 points" (P8).*

Conversely, Participant 1 noted that while their friends experienced inconsistent prep guides and expectations, they usually had positive experiences with this:

> *"I think that my professors have been pretty fair on, you know, what's on and what's not. I think some of them are kind of smarter in that they just give a blanket statement. So obviously it's not necessarily just a 'know everything'. It's like focus on XYZ" (P1).*

However, Participant 1 offered a different example of an unfair assessment, when they shared their disappointment in the instructor using questions made by other students in the exam, which gave some students an unfair advantage over others:

> *"[they] used test questions that students had submitted. It was for an assignment… if you saw your question on the exam you're like, 'oh, perfect'. Check it off. You obviously know the answer. Whereas I know my questions weren't on it. I didn't love that whole experience". (P1)*

To sum up, participants wanted to know what will be covered on the test so they could prioritize their study time efficiently. This is consistent with Struyven's (2005) suggestion that many students adopt a strategic approach when preparing for exams, putting more effort into the material they expect to be tested on and passing over the rest. When this information was not communicated by the instructor, or there was a clear misalignment between course content and exam questions, our participants perceived this as frustrating and unfair, because it could lead to poor exam performance. The importance of predictability has been identified in other studies exploring students' experiences of exams, particularly during the pandemic. For example, Tam (2022) and Jacobs (2021) found that clear communication of expectations, including providing students with rubrics and assessment criteria beforehand, was crucial in reducing student anxiety and improving performance. Slack and Priestley (2023) noted that students often felt stressed and dissatisfied with online exams due to unclear expectations. In their focus groups with psychology

students, Lynam and Cachia (2018) also found that instructor-related factors played a key role in students' favourable perceptions of assessments.

Exam Performance is Not a Factor. Finally, it is worth mentioning that participants rarely mentioned grades themselves as an attribution of exams' ineffectiveness or unfairness. And those who did bring up grades explicitly acknowledged that the fairness of an exam was *not* necessarily tied to how well or how poorly they performed on the exam. As explained by Participant 5:

> *"as a student it's pretty clear when an exam is fair. Like say it's fair, but you do poorly, or unfair and you did poorly. Like it's pretty clear which, if it's fair or not, even if you did well or not" (P5).*

The same sentiment was expressed by Participant 6, who also understood an exam to be fair when it was aligned with course expectations, even if a student did poorly on it:

> "*there are times when I'm writing an exam and I can kind of tell that if I studied better, I'd be able to do better on the test. I recognize that it's a fair exam. The questions aren't out of pocket… it's very much things that we have learned. So I'm able to recognize at that moment that, OK, this is me not studying properly" (P6).*

Perhaps grades could be viewed as a trigger that prompts students to search for attributions of their poor performance. A student may make an internal attribution, something about their own preparation as expressed by Participant 6 above. Or they may make an external attribution, something about the exam or the instructor, as illustrated in this quote from Participant 4 explaining how they knew an exam was fair:

> *"If I felt good about it after… If I come out of an exam and I don't feel good about it. I have a tendency to blame it on the prof. Which is probably not fair, but I mean, a lot of*

the… I can't help but do that… I think that if I come out of an exam, I don't feel good about it, I just assumed that it was unfair if I studied everything and like felt like I understood everything and I was still sent with curveballs, I immediately think it was unfair." (P4)

Effectiveness and Fairness in Exam Format

Participants mentioned a number of specific factors related to the format of the exam that shaped their perceptions of exam effectiveness and fairness, including mixed-format exams, student autonomy, recall aids, and universal design for learning.

Mixed-Format Exams. As mentioned earlier, participants' interpretation of the ideal purpose of exams was tied to comprehension and application of course material in addition to memory and recall. In line with this, our participants expressed a preference for mixed-format exams that included a combination of multiple-choice, short-answer, and longer-answer questions. Participant 1, for example, mentioned that mixed-format exams were more effective not just for demonstrating learning, but also for promoting their learning:

"I feel like I can get the most from them or I can show the most from them" (P1).

Other participants shared how assessing both surface and deep level thinking through mixed format could prompt them to think about everything they have learned and how all the content could be integrated into a final longer answer question. For example, participant 10 reasoned that in addition to recognizing different students have different strengths, the inclusion of mixed-format questions helped encapsulate the course better:

"It makes you really think about everything of what you learn, not just little bits and pieces like kind of everything throughout your 12 weeks all intertwines into your final thing" (P10).

Bengtsson (2019) and Tam (2022) both discuss the benefits of incorporating a variety of question formats in exams, especially in take-home exams. As described here, this practice was recommended by our participants to help cater to different learning preferences and increase exam effectiveness by allowing students to demonstrate their knowledge in multiple ways. Moreover, Rich (2011) found evidence showing students engaged more deeply with course content when take-home exams included a variety of question types, suggesting that mixed-format exams might better align with learning outcomes and elicit greater student engagement.

Student Autonomy. Another frequently mentioned area was participants' preference for choice and autonomy during exams, specifically in demonstrating how they have met course learning objectives:

> "*Ideally [exams] should just be a way to reflect on the learning experience... You should be allowed to communicate that in your own way*" (P2).

Moreover, some participants spoke of autonomy in terms of taking ownership of their learning. Participant 4 supported this sentiment when they said:

> "*I'll learn a lot about certain things, but I don't know what to do with that information if I'm not given opportunities to apply it… but I guess that's also my job as a student to figure out, OK, why is this important? Like, why am I learning this?*" (P4).

Participants' preference for choice in assessment questions is consistent with previous findings. For instance, Jacobs (2021) found that students appreciated having some autonomy in take-home exams when they could choose scenarios or problems to work on. This choice not only reduced anxiety but also allowed students to demonstrate their learning in ways that aligned with their strengths. Tam (2022) noted that students valued being able to answer questions that used external resources outside of the course, as it allowed them to showcase their ability to

apply knowledge beyond the course material. These ideas further support student autonomy in assessment.

Recall Aids. In reference to traditional memory-based exams, participants mentioned how certain features of test design shaped their perceptions of an exam's effectiveness and fairness. Participants observed how the mechanics of test questions, and the progression of test questions can enhance their recall. For example, Participant 4 mentioned that using multiple-choice questions as warm-up questions kickstarted the process by jogging their memory as well as establish the tone of the test:

> "*It makes those other questions a little bit easier to answer because you're already kind of on a roll. So when the questions sort of start easier but less intense, I find that really effective*" (P4).

Participants 6 and 9 also noted that incorporating cue words used by instructors in exams, as used in-class, was another effective method that students appreciated:

> *"Buzzwords are very triggering, so it could really help the students remember and I don't think an exam should aim towards not providing buzzwords because… why are you trying to block the recall process?"* (P6).

Likewise, Participant 4 found it more effective when questions were presented similarly to the way instructors taught them:

> *"you're trying to figure out what to do in the question when you've never been presented a question even if you know how to do the question if it was worded differently. But it just adds an additional troubleshoot that is unnecessary and it's ineffective" (P4).*

Participants also brought up how the use of chunking in test questions, for example, presenting all questions from Unit 1 together, helped them with recall. Chunking in education

refers to breaking down information into smaller blocks to help with consolidation and recovery of information. For example, breaking down lectures using chunking (presenting material between 3 to 17 minutes) can improve students' attention, intrinsic motivation (Martínez-Huamán et al., 2023), and is preferred over didactic – single block lectures (Humphries & Clark, 2021). Though not extensively applied to the education literature, exams that display topic-related questions in small increments may help students recall information better than if questions were presented randomly.

Taken together, these comments reflect participants' general apprehension that the format of traditional exams could hinder rather than facilitate their demonstration of knowledge, which went against the purpose of exams and therefore was perceived as unfair and ineffective.

Universal Design for Learning (UDL). UDL emerged as a response among some participants to address their concerns about unfairness in assessments, framed from an equity perspective. Specifically, participants suggested that time constraints and subsequent stress experienced in the exam environment could be mitigated through the implementation of UDL practices. For example:

> "*I have this one professor who uses what [they] calls a universal design, which I thought was a beautiful idea. None of the [student accessibility services] students actually have to leave the room… We all got a certain amount of time and that amount of time was enough where a student who may need extra time should still be able to answer the questions*" (P2).

Participant 7 added that even for the average student, the time allotted for traditional exams was often not enough. This is a common grievance students express about traditional timed exams (Polat, 2020).

Theme 3: Affective Responses

Our final research question of interest focused on the various factors related to how students felt before and during exams. In general, students evaluated the spectrum of emotions related to test confidence and anxiety at every stage of the exam process.

Preparation Leading to Confidence

Participants found their best exam experiences were those they were most prepared for. Participants attributed their feelings of preparedness and confidence to two sources: their own study strategies and support from the instructor.

Study Strategies. Students' confidence in their exam performance appeared to be shaped by their proactive approach to studying the course material beforehand, involving effective studying practices. As one participant explained:

> "*The best [exams] were probably the ones that I was just the most prepared for. Where I went in to just sit down in the gym and was able to go clean through the whole thing without having to stop to second guess… I knew the information. It was all good. I just had to put it on paper" (P3).*

As for why students chose to initiate exam preparation, some participants mentioned they did so to mitigate feelings of anxiety:

> "*I do everything I can to avoid feeling [anxious]. I've never gone into an exam unprepared, and I've never gone into an exam without going over every single concept and making sure that I knew the important stuff or the main themes of every lecture, sort of thing" (P4).*

To enhance confidence, participants detailed commonly used preparation strategies for exams. These included weekly reviews of course material, the use of cue cards, and diagrams,

self-testing, repetitive study sessions with study buddies, and engaging in mock teaching scenarios. In another instance, one participant described beginning their exam study session with writing everything they knew before diving into course notes. Study strategies such as the ones used by our participants are active learning strategies identified in the research as highly effective for long-term retention (Dunlosky et al., 2013). Some participants mentioned they used self-testing as a method to study for an exam, which is known as the testing effect (Polack & Miller, 2022). The testing effect is a phenomenon in which a students' comprehension of course material is strengthened when they actively test themselves as opposed to only studying course material. The retrieval processes involved in pre-testing and post-testing (Latimier et al., 2019), have both been shown to enhance learning gains. Therefore, students' confidence that comes from using these strategies is justified.

On the flipside, some preparation practices, especially more surface-level ones such as re-reading, cramming, and rote memorization, may leave students feeling confident yet underprepared for exams (Dunlosky et al., 2013). For example, a "cheat sheet" could create an illusion of preparedness and a false sense of confidence, and the subsequent challenge of answering complex exam questions could lead to frustration. As one participant shared: *"I had a class I was really prepared for and we also were allowed a cheat sheet… I had written the whole course material on this paper, open my final exam, I started flipping the pages and realized that it looked like I didn't know anything… It's frustrating. It was really sad because I expected to get a really good grade in this class" (P8).*

Keeping in mind our participants were all in their third year of post-secondary studies or higher, they may have naturally learned the more effective preparation/ study strategies over the course of their studies. Participant 5 added that every student should experience a first exam

early in their undergraduate career with low stakes. This is because it could provide an opportunity to learn from any mistakes or fumbles, encouraging better adoption of preparation strategies to succeed in future exams.

Instructor Support. The role course instructors play in preparing students also has implications for shaping confidence. When reflecting on positive exam experiences, one participant thought back to the theories of the zone of proximal development and effective scaffolding practices. Describing their ideal type of example, one participant mentioned wanting oral exams, where the instructor can work with the student to build understanding:
"I think back to the zone of proximal development. All the scaffolding. That's what it was. So I think that exam does a good job at testing what you've learning, but also kind of pushing you to the next sequential course" (P5).

These instructor-facilitated strategies helped prepare them for a level of deeper comprehension. Scaffolding in education is an individualized process that refers to temporary support by a teacher to a student and helps the student master different stages of learning through small and incremental guidance (e.g., evaluative feedback). Scaffolding helps student achieve mastery that they might not have been able to achieve alone (Acosta-Gonzaga & Ramirez-Arellano, 2022). The Zone of Proximal Development refers to space between what a teacher knows and what a student is capable of, that is, the level of skill a student can master if given effective scaffolding (Acosta-Gonzaga & Ramirez-Arellano, 2022). Scaffolding techniques leading up to tests are associated with improved academic outcomes (Fong et al., 2023; Kim & Shakory, 2017).

In contrast, lack of adequate scaffolding and practice throughout the course undermined students' confidence, as Participant 9 shared:

> *"tests are way harder than what we learned in the class, and we just didn't have enough questions to, like, practice to. Like just five or like questions in the back of the chapter, I felt it wasn't enough". (P9)*

Instructor support and in-course scaffolding were also identified as important sources of students' positive exam experiences in Lynam and Cachia's (2018) focus groups with psychology students. Similarly, Tam (2022) found students appreciated instructors who were actively involved throughout a take-home exam because they provided feedback and answers to questions. Roelle and Berthold (2017) discussed having instructors actively involved in helping students prepare for exams by having them answer complex questions about course content rather than summarizing lecture notes. These instructor-facilitated strategies might help boost students' confidence by promoting deeper learning and better preparation.

Test Anxiety

Participants attributed their test anxiety to external factors over internal factors, focusing on aspects like the exam environment and format. For some students, no matter how well-prepared they felt, confidence was compromised when disruptive environmental factors pulled their attention away from writing the exam. Participant 4 in particular shared the repercussions of having their attention pulled:

> "*I'll be totally prepared for an exam. I'll know everything … but I'm thinking about the person shaking their leg beside me, [or] somebody opening the door over there, [or] I can hear the clock ticking… those things completely take my attention away" (P4).*

In the institution where this study was conducted, students in large classes frequently take exams in a giant testing room; Participant 4 also highlighted the inconvenience associated with these conditions:

> "The *door on the gym was super squeaky … And then also the person in the back corner of the gym decided to wear high heels to the exam and so when they were leaving it was just clicking. anyway" (P4). "Yeah, it was doing the squeak and then the click at the end and then when I left like I was like WD40. Like, please, like, just WD40 on that freaking door. Oh my God" (P4).*

In line with Participant 4's mention of hearing the sound of the clock ticking, Participants 2 and 3 also conveyed how time limits exacerbated their stress during exams:

> *"adding that two-hour/three-hour timer to everything just makes it that much more anxiety-inducing" (P3).*

The format of the test can evoke specific types of anxiety. Generally, participants' default association of exam anxiety was with the traditional exams, while take-home exams were seen as less stressful. However, participants acknowledged the trade-offs of these two assessment designs, recognizing that each format presented different challenges. For example, Participant 8 mentioned they:

> "*really enjoy my take-home exams... I sit at home, I write something, I can have a break whenever I want. It's not stressful. I'm not in a room with 300 other people sweating and being cold at the same time in the gym. The grad students are like hawks looking at you. Scary… it's also really cold in there" (P8).*

Participant 4 added that removing the time limit can make exam questions more challenging because they are open ended. Some of these accounts touch on the subjectivity of grading in take-home exams versus objectivity in traditional exams. For take-home exams, some participants highlighted the challenge of not knowing how much writing was sufficient to meet assessment expectations:

"worrying that you didn't say something that you should that they wanted to know because I think you remember the final exam, it was like try to give as much as you can to show that you know the content. And I'm like, oh, she has certain section she's looking for in this questions to demonstrate knowledge" (P2).

Extended word lengths, uncertainty about the approach to take in answering open ended questions, and ambiguity in assessment expectations in take-home exams or essay style exams, can also be mentally demanding. One participant shared difficulties with one such exam:

"I think the difference is not so much the confidence, but, just kind of like whittling away at my perseverance that I can do this. 'I got this' just eventually becomes 'OK, yes, let's just get this over with'" (P3).

These findings were consistent with the literature on students' wellbeing and anxiety during exams. For instance, Slack and Priestley (2023) found that students' well-being scores were significantly associated with perceived workload stress, with other students' behaviours adding to this stress. Slack and Priestley (2023) also discussed how environmental factors such as technological challenges and unfamiliarity with online testing formats elevated their participants' stress. Tam (2022) explained that tight time limits in take-home exams caused rushing and added to the anxiety of their students, especially when combined with the pressure of uploading documents to submission portals on time. All of these examples indicate that the format and associated demands of the exam environment contribute to test anxiety.

Participants 7 and 10 shared strategies that they use to reduce their anxiety during exams. For Participant 7, it was making a checklist to help keep them organized:

"Sometimes because I'm anxious I tend to make a little checklist of little goals that I make throughout the exam… for example, long answer, sometimes I'll write point form on the

actual paper of what topics I can talk about… I write down little points to remind myself…. If I go back I can continue writing" (P7).

In terms of environment, Participant 7 shared they *"really love writing in the same spot that I was taught or writing in the same environment that you were taught in. That really does help a lot" (P7)*, with Participant 10 adding *"When you're in the same room as when you learn it, it's easier. Your recall is a bit easier, however, that works" (P10).* While both participants mentioned feeling more comfortable writing the test in the same location where they learned the content, in instances where this was not possible, both mentioned the importance of sitting in particular seats in testing halls to ease anxiety:

"If I sit in the front, I feel like I can kind of block out the other people, like hundreds of kids in the gym, if that makes sense. I can kind of try and zone out. But like, as soon as you walk in, there's the rows of desks kind of like a prison. That's what my friends and I kind of picture it. And there's, like, all these papers, and the teachers are just standing at the front, just staring at you, right? That setting definitely causes added stress" (P10).

Grade Anxiety

Aligned with the external influences of test anxiety, many participants highlighted a specific area that warranted a separate subtheme: grade anxiety. Grades and marking, which are not directly tied to students' experience in the exam process (and are often beyond students' control), shaped students' anxiety toward exams. Participants' concern over grades – and the connection to how this contributed to anxiety – was often expressed as a desire to achieve a certain grade and frustrations around marking. For example, Participants 5, 7, 8 and 10 all mentioned their goal going into an exam was to receive a certain grade in line with their self-set

expectations and experienced anxiety while waiting for their grades. Participant 10 further explained that excelling in exams was crucial for them to maintain their scholarship:

> *"I have one of those like the renewable scholarships. So if I get an 80 average for the year, I get $2500 for the next [year]. Which is a lot of money, especially when you're a broke [student]. So that is a really big incentive to me to bust my [butt] pretty much the whole year to get the 80 so it basically takes off a third of my tuition cost" (P10).*

While participants acknowledged their pursuit of specific grades in exams, one identified source of frustration hindering them was the delay in receiving coursework grades leading up to the exam:

> "*You can imagine how frustrating that is... you don't know what you're getting going into the exam. You can obviously speculate, but again you're waiting for marks... which is a big thing… if you get the [course work grades] before the exam, you can prepare better" (P7).*

To alleviate this anxiety, several participants added that calculating their grades before exams helped calm their minds:

> "*I know this usually stresses more people out than it's worth, but I always calculate what my grade is going into the exam and how much I need on the exam to get my desired grade… That helps me kind of calm down my thinking" (P10).*

However, participants argued that not having access to all grades before the exam did not allow them to know their lowest-performing courses, which prevented them from making strategic decisions about where to focus study efforts.

Undergraduate students face high pressure to achieve certain grades due to financial and academic pressures (Slack & Priestley; Tam, 2022), and there is a common idea in education that

students tend to be fixated on grades (Horne et al., 2022). Even with intervention attempting to alter students' mindset toward grades, students continue to perceive they only lose grades as opposed to earning them (Horne et al., 2022). A potential silver lining, though, is that grades become less important for students over the school years, and career-related activities progressively become more important throughout the years (Jennings et al., 2013). Otherwise, an advantage of take-home exams is that they reduce stress and anxiety by offering flexibility in location and extended time limits, which could alleviate grade anxiety as students might feel more in control over the exam conditions (Bengtsson, 2019).

Thematic Map

Figure 1 represents an integrative thematic map of the qualitative results – a view of the assessment process from the students' perspective. This sequence model considers both instructor-facilitated and student-centered factors at every stage of the assessment process: planning, execution, and evaluation. The arrows in the figure indicate the attributional sources of students' perceptions and the sequential nature of the evolving assessment process, and how instructor-facilitated factors may (mis)align with student-centered factors.

The starting point includes beginning goals and attitudes about assessments held by students and instructors. At the starting point, instructors begin by conceptualizing their approach to assessments. This involves balancing between ideal educational objectives (e.g., higher-order learning outcomes, authentic learning) with pragmatic constraints (e.g., time, resources, academic integrity issues, class size). While instructors aim to design assessments that align with course learning outcomes, these practical limitations can influence the assessment chosen, resulting in an exam that may not be the ideal way of testing all learning outcomes appropriate for the course and curriculum objectives. A student enters into the assessment process with their

own preconceptions about the purpose of exams. These expectations often focus on evaluating lower-order thinking skills (memorization) but may also include desires for assessments that test a deeper understanding and application of course material. This creates a dichotomy between perceptions of actual purpose (exams as practiced) and ideal purpose (exams as should be). Both students' and instructors' perspectives on the ideal purpose of exams seem to align closely with the authentic assessment framework. However, both also have their own non-pedagogical pragmatic considerations. For students, these include concerns over grades to maintain funding or to advance to the next level, as well as competing priorities from other courses. These considerations may push students to adopt a strategic approach to exam preparation in the planning stage, instead of a deep approach which is closer to the ideal type of learning. Students also seem to be aware of the practical limitations instructors face and accept the need for the compromise of testing lower-order thinking skills.

The planning stage involves the preparation process leading up to an assessment. In the planning stage, instructor-facilitated factors involve an instructor's efforts to provide in-course scaffolding and practice opportunities, communicate expectations, provide study guides and grading rubrics, and maintain a teaching presence to ensure students are appropriately prepared for the exam. The messaging around the exam is also crucial here – ensuring the students understand the purpose, expectations, and content to focus on, as well as best study strategies to use. This is especially important for alternative or less familiar exam formats. These instructional decisions are shaped by the instructor's initial goals and the trade-offs they make. Students' receptivity toward the instructional design – mainly students' impressions of the suitability, credibility, and relevance of the assessment – is also crucial. Preparation strategies, time and effort invested in studying, and anxiety surrounding grades all play a role in shaping how

students will feel going into the exam. Effective communication and clear expectations from instructors help reduce students' anxiety, improve confidence, and increase buy-in.

The execution stage includes factors relevant to the exam experience itself, such as exam design elements and students' affective responses. During the execution phase, the exam format, structure, testing conditions, and level of difficulty are all factors. For instructors, these features are the vehicle to determining whether learning objectives have been met with validity, reliability, and academic integrity. For students, feelings of confidence and anxiety during the exam are heavily shaped by these features, specifically, how well the exam content and format align with the preparation (both instructor-facilitated and student-led) that took place during the planning phase. When there is alignment, students feel more confident, whereas misalignment can lead to increased anxiety and negative appraisals. For students, feelings of confidence were attributed primarily to factors surrounding exam preparation and the assessment experience, and more specifically how this was connected to instructor-provided exam preparation materials. Interaction with Instructor factors impacted how students felt before the exam: when students' exam prep was aligned or integrated with in-course activities, materials, and exam expectations, they felt more confident going into the exam. In contrast, when instructor messaging about exam expectations was lacking or unclear, or when students didn't have feedback on their in-course assignments, students felt less confident in their prep and more anxious ahead of the exam. Students' lack of confidence going into an exam went hand-in-hand with feelings of anxiety.

The evaluation stage post-assessment references how both instructors and students felt about the exam process. These areas may overlap with how instructors felt the assessment process went, although this information is used to inform future assessment modifications. Students' attributions of these areas also shaped their perceptions of the overall assessment

experience. In the evaluation stage, students reflect on the fairness, effectiveness, challenge, and relevance of the assessment based on their initial goals, communicated expectations, assessment design features, and their affective reactions. When exam properties were aligned with student's personal goals and expected learning outcomes, the exam was perceived to be fair and effective at testing learning. Factors that positively shaped these perceptions included having a variety of question formats, choice of which exam questions to answer, and having flexibility with when and where to complete the exam. When exam properties posed barriers to achieving student's personal goals, the exam was perceived to be less fair, less effective at testing learning, and evoked more anxiety. For example, assessments with time restrictions, environmental distractions, memory-based tests, and lack of recall cues are all factors that were perceived to prevent students from demonstrating learning, achieving desirable grades, and disadvantage some neurodiverse students.

These reflections all shape students' overall perception of the assessment experience. When instructor goals, instructional design, exam design, and student preparation are all well-aligned, students are more likely to view the assessment positively. Misalignments can lead to perceptions of unfairness, ineffectiveness, and increased anxiety, which in turn may contribute to student disengagement from the learning process. The model represents students' phenomenological experiences of exams. It is our hope that understanding students' perspective can help instructors design high-quality assessments that better align with both instructor and student goals.

Methodological Notes

While participants gave insight into the factors that shaped their perceptions of purpose, affective responses, and effectiveness and fairness of exams, the qualitative nature of these data

precludes drawing inferential or causal conclusions. Some themes in our qualitative data paralleled the existing quantitative models of students' conceptions of exams, particularly the dimensions of learning outcomes, accountability, authenticity, credibility, and conditions of assessment (Brown & Hirschfeld, 2008; Gerritsen-van Leeuwenkamp et al., 2018). However, several dimensions from those models, notably fairness, validity, and enjoyability, were not clearly represented in our data. In our participants' responses, the distinction between what was considered a fair/unfair and effective/ineffective exam was not well differentiated and they tended to speak to either concept in similar ways. In terms of affective responses, our participants focused primarily on confidence and anxiety, with very few mentions of other positive emotions in relation to exams. This could be due to the way the questions were framed. Our interview questions referred specifically to "exams", as our focus was on summative assessments. Whereas many previous studies used the broader term "assessments", which might be interpreted as inclusive of both summative and formative assessments.

The thematic map and findings presented above are only representative of the particular group of students interviewed, who were predominantly women, psychology majors, junior and senior undergraduate students who came from the same cohort enrolled at the same medium-sized Ontario university. A commonality in education research is that higher-achieving, conscientious students tend to participate more in research studies of this kind, as was also the case in other qualitative studies of students' perceptions of assessment (e.g., Lynam & Cachia, 2018; Tam, 2022). Students who receive higher grades tend to view assessment as essential for constructive growth and student accountability (Brown & Hirschfeld, 2008), to be more self-efficacious and self-regulated learners, and to adopt less surface and more strategic approaches to learning (Richardson et al., 2012). The themes derived from our participants' responses

exemplify this high-achieving profile, but they may not reflect the perspectives of lower-achieving students.

Interviews via televideo may have reduced potential rapport building. For instance, in one interview the participant lost internet midway through and rejoined later. In a different instance, another participant had audio and video issues, and the interview proceeded with no video and difficult audio. My own positionality as an interviewer may have unintentionally led participants to think about and say things they would have not otherwise said. Moreover, social desirability and demand characteristics pressures are stronger in an in-person interview than in an anonymous survey, which may have prompted students to respond in a way that is not normally characteristic of them, or to withhold voicing certain views. About half of the interview participants knew me from the prior course they were pooled from, where I was a graduate teaching assistant, although that course ended a year before the interviews took place. This prior relationship might have facilitated the rapport for some participants, or it might have made some participants less comfortable to express their honest opinions about assessments.

Conclusion

The primary objective of this study was to explore students' experiences and attitudes toward university exams. With the changing landscape of university education, we sought to understand current students' perspectives on various aspects of exams, including exam purpose, formats, and effectiveness, as well as their best and worst exam experiences and preparation strategies. Based on our findings, students recognize the contrast between how exams are practiced versus how they should be practiced, with the latter emphasizing testing higher-order skills, improved relevance and autonomy in assessment avenues, and more proactive preparation strategies to boost confidence and minimize anxiety. These perspectives are aligned with student-

centered approaches to assessment. One area that was not extensively tapped into were students' motivational reasons for completing assessments. That is, what drivers increase student buy-in for wanting to engage in student-centered assessments in the first place? Thus, an empirical look on the added value of motivational assessment design is warranted.

Study 2: Students' Motivations and Perceived Learning Outcomes in a Case-Based Take-Home Exam

Study 1 explored psychology students' perceptions of exams, study approaches, and affective responses to exams in general. Study 2 reports on a case study of a specific alternative exam, a take-home case-based exam, which was designed in line with recommendations from the literature on take-home exams and authentic assessment theory. Therefore, the literature review from Study 1 is also pertinent to Study 2 (it will not be repeated here). In addition to exploring students' perceptions and learning approaches towards the new case-based exam, Study 2 also explored students' motivational orientations and how these related to specific exam features and perceptions of the exam's motivational properties. Considering motivational factors in assessment is important, because students' learning approaches and outcomes are contingent on the degree and quality of their motivation and engagement in the exam process (Howard et al. (2021). In turn, how students perceive the motivational value of the assessment depends on the exam design itself (Keller, 1979).

According to the literature on motivational instructional design, assessments can theoretically be designed to enhance student motivation and engagement in deeper learning (Keller, 1979, 2010). To this end, educators have increasingly adopted student-centered approaches to learning, moving from passive traditional lecture style classrooms to interactive learning activities to boost student motivation and engagement (Baeten et al., 2012; Baeten et al.,

2013a; Cai et al., 2022; Gholami et al., 2021; Latif, 2014; Ma & Lee, 2021; Yoo & Park, 2015). One such student-centered approach is case-based learning (CBL), where students practice applying course material while analyzing real-world cases and scenarios (Mayo 2002; 2004). Although CBL has been shown to improve students' motivation and learning outcomes when used with learning activities (Harman et al., 2015), very few studies have applied CBL to summative exams. The aim of this study was to fill this gap by exploring the pedagogical value of building a CBL-based summative assessment with motivation-driven characteristics. Below, we review the motivational and CBL literatures that informed Study 2.

Motivational Framework

Two motivational theories informed the present study: Ryan and Deci's (2020) Self-Determination Theory (SDT) on the nature and antecedents of different types of motivation, and Keller's (1979) Attention, Relevance, Confidence, Satisfaction (ARCS) model of motivational instructional design. The following sections provide brief background on each theory.

Self-Determination Theory (SDT)

Motivation is defined as "the conscious or unconscious stimulus for action towards a desired goal provided by psychological or social factors; that which gives purpose or direction to behaviour" (Simpson & Weiner, 1989, p. 1131). Ryan and Deci's Self-Determination Theory (SDT) is one of the most researched and widely utilized motivation theories in education research (Guay, 2022; Ryan & Deci, 2020). SDT focuses on the quality, sources, and learning outcomes of different motivation orientations.

Types of Motivation Orientation

SDT proposes that individuals are motivated by either intrinsic or extrinsic reasons, which in turn determine their motivation orientation (Deci & Ryan, 2000). Motivation orientation according to SDT is categorized into three qualitatively distinct types: intrinsic motivation, extrinsic motivation, and amotivation. *Intrinsic* motivation refers to doing something purely because one finds satisfaction in engaging in and completing activities that provide inherent enjoyment. Those with intrinsic motivation are genuinely interested in an activity for the experience itself. In contrast, *extrinsic* motivation involves the pursuit of activities for any number of reasons other than intrinsic interest, which could include external consequence or instrumentally related rewards (e.g., money). *Amotivation* refers to a lack of intentionality, where motivation is rarely present because there are neither intrinsic nor relevant extrinsic sources of motivation (Ryan & Deci, 2020).

The category of extrinsic motivation is further unpacked into four subtypes, with varying levels of internalization: external regulation, introjection, identification, and integration (Ryan & Deci, 2020). *External regulation* refers to doing something explicitly and exclusively because of external rewards or punishments, with zero degree of internalization. *Introjected regulation* involves partially internalized reasons such as to boost self-esteem or gain approval from others, or to avoid feelings of shame, guilt, or anxiety. *Identified regulation* occurs when a person internalizes the value of an activity, doing it because they see it as worthwhile even if it may not be enjoyable. *Integrated regulation* occurs when an activity is not only perceived as worthwhile, but when it also aligns with an individual's other interests and values central to their identity. These four subtypes of extrinsic motivation orientation fall along a continuum of perceived locus of causality, from fully externally driven to more internalized and finally fully internally driven.

The continuum is book-ended by intrinsic motivation on the internal end and by amotivation on the external end (Ryan & Deci, 2020).

The most commonly used measure to assess these motivation orientations is the Self-Regulation Questionnaire (SRQ; Ryan & Connell, 1989), which has many versions for specific domains of regulated behaviour (e.g., exercise, schoolwork). In educational settings the Academic Self-Regulation Questionnaire (SRQ-A) is used, which assesses the reasons why students engage in learning behaviours along four dimensions: external, introjected, identified (merged with integrated), and intrinsic (Ryan & Connell, 1989). The SRQ items are meant to be adapted to suit the specific context of each study, so there are many derivatives of the original SRQ-A. The SRQ-A scores are frequently categorized into two "super" categories: autonomous and controlled motivation (Ryan & Deci, 2000), depending on the extent to which reasons for completing a task are internalized. *Autonomous* motivation includes intrinsic and identified/integrated subtypes; it stems from an internal locus of causality and is characterized by autonomy and ownership over behaviour. In contrast, *controlled* motivation includes external and introjected subtypes; it stems from an external locus of causality and occurs when actions are perceived as induced rather than willed. Simply put, autonomous motivation is synonymous with a more internalized goal orientation, and controlled motivation reflects a more externalized goal orientation.

Associated Learning Outcomes

To better understand how students' motivation orientations relate to their approach to their studies, Howard et al. (2021) conducted a meta-analysis examining the associations between specific motivation orientation subtypes with several outcomes, including academic achievement, persistence, well-being, goal orientations, and self-evaluations. Academic

performance was typically measured through GPA, while persistence reflected a students' effort, drive, and engagement in their studies. Well-being reflected both negative affect, such as anxiousness and boredom, and positive affect, such as enjoyment and vitality. Goal orientation referred to students' pursuit of learning (mastery goals) versus pursuit of good grades (performance goals), and self-evaluation included students' beliefs in their ability to accomplish goals (self-efficacy), and the positive regard they have for themselves (self-esteem). For their meta-analysis, Howard et al. (2021) searched for articles that included at least one validated motivation scale based on SDT, were conducted in an educational context, and presented at least one correlation between motivation type and an outcome. The final sample of articles was 344, encompassing 223,209 participants across all studies, ranging from primary school through to post-secondary education, with an average participant age of 16.19 years.

First, Howard et al. (2021) split up the analyses into adaptive (positive) and maladaptive (negative) outcome categories. They found a consistent pattern of results where the correlations between motivation styles and adaptive outcomes became stronger and positive as the degree of internalization increased, whereas the reverse was found for maladaptive outcomes. For example, adaptive outcomes, such as higher academic performance, mastery goals, self-efficacy, satisfaction, effort, engagement, and enjoyment, were moderately negatively associated with amotivation (r = -.24), unrelated with external regulation (r = -.01), weakly positively associated with introjected regulation (r = .17), and strongly positively associated with identified (r = .38) and intrinsic motivation (r = .41). Conversely, maladaptive outcomes, including anxiety, boredom, and absenteeism, were moderately positively associated with amotivation (r = .28), weakly positively associated with external (r = .18) and introjected regulation (r = .19), unrelated with identified regulation (r = -.04), and weakly negatively associated with intrinsic motivation

(r = -.13). This suggests that higher levels of internalized autonomous motivation are linked to better adaptive outcomes, whereas higher levels of controlled motivation are linked to more maladaptive outcomes. The degree of amotivation plays a significant role in both adaptive and maladaptive outcomes.

When Howard et al. (2021) meta-analyzed the associations with each specific adaptive and maladaptive outcome, they found the same general pattern of results for all variables except performance goal orientation, which showed positive associations with all motivation types but was most strongly positively associated with controlled regulation variables. Students who hold performance goals are more concerned about achieving good grades than about learning, which explains their controlled motivation style. In contrast, students who hold mastery goals are more concerned about meaningful learning than about grades, and this was strongly correlated with autonomous motivation.

Howard et al. (2021) also conducted relative weight analyses, to estimate the unique variance in outcomes accounted for by each motivation type. Averaged across all outcomes, intrinsic motivation came out as the strongest unique predictor, closely followed by identified regulation (the strongest form of internalized extrinsic motivation). Amotivation also played a unique role in many outcomes, and it was the strongest predictor of lower grades and higher absenteeism. Compared to other motivation types, controlled forms of motivation (external and introjected) carried less weight. The approach taken by Howard et al. (2021), focusing on the weight of separate motivation orientation types, may conceal important interaction effects between controlled and autonomous categories. SDT recognizes that people can hold multiple motivation orientations simultaneously for the same behaviour (Ryan & Deci, 2020), and several studies have explored their combined and interactive effects through a person-oriented approach.

Ratelle et al. (2007) conducted a person-oriented analysis of 410 first-year college students to examine their motivation profiles. They measured students' academic motivations using an academic motivation scale similar to the SRQ-A, which assesses students' intrinsic motivation, identified regulation, introjected regulation, external regulation, and amotivation. The researchers also measured participants' academic achievement via grades, and academic persistence using second-year enrollment status (enrolled in program vs. dropped out). Ratelle et al. (2007) found three distinct motivational profiles. The most prevalent profile was characterized by a combination of high levels of autonomous and controlled motivation, with low levels of amotivation (High AU-C profile, about 39% of the sample), supporting the idea that both motivation orientations can exist simultaneously. Another profile showed solely high autonomous motivation, with low controlled motivation and low amotivation (AU profile, 36% of the sample). The third profile was characterized by low to moderate levels of autonomous and controlled motivation and a moderate degree of amotivation (Low AU-C profile, 25% of the sample). Students with a Low AU-C profile had the lowest grades and the highest odds of dropout, compared to the other two profiles. Students with a combined High AU-C profile had similarly high academic achievement to those with a purely autonomous profile (AU). However, students with high autonomous and low controlled motivation (AU) were twice as likely to persist in academic studies compared to the combined High AU-C group. These findings underscore the chief role of internal motivators in adaptive academic outcomes, but also show that adding external motivators to the mix may undermine certain outcomes (Ratelle et al., 2007).

In another person-oriented study conducted by Vansteenkiste et al. (2009), researchers examined motivation profiles to identify the types of students that would exhibit greater learning

outcomes. The study sampled 484 first year education students using the SRQ-A to measure students' motivation for an educational sciences course, as the authors argued that assessing motivation at a course-specific level is more appropriate than at a global level (Vansteenkiste et al., 2009). Learning outcomes were measured using the Motivated Strategies for Learning Questionnaire (Pintrich et al., 1991, as cited in Vansteenkiste et al., 2009), which assesses depth of cognitive processing (elaboration, critical thinking, organization), metacognitive self-regulation (test anxiety, time and environment management, metacognitive strategy use), and determination (effort regulation and procrastination). They also measured students' perceptions of the instructor's teaching style, focusing on subscales targeting autonomous support, structure, and involvement.

Vansteenkiste et al. (2009) identified four clusters of motivational profiles, three of which were similar to Ratelle et al. (2007). "Good quality" cluster was similar to the AU profile (high autonomous, low controlled; 23% of the sample). "High quantity" cluster was similar to the High AU-C profile (high autonomous and high controlled; 22% of the sample). "Low quantity" cluster was similar to Low AU-C profile (low autonomous and low controlled; 35% of the sample). The fourth profile was "poor quality" cluster, characterized by high controlled but low autonomous motivation (19% of the sample). In the comparisons, significant differences in outcomes were found between the motivation groups, with good quality and high quantity motivation generally associated with better learning outcomes than poor quality and low quantity motivation. Poor quality motivation was associated with the highest levels of procrastination, test anxiety, and lowest effort regulation, even when compared to low quantity motivation. High quantity motivation showed more adaptive functioning than low quantity motivation, but it showed higher test anxiety and lower perceived teacher support compared to good quality motivation. These

findings are aligned with Ratelle et al.'s (2007) results, indicating that good quality motivation shows the best display of learning outcomes, closely followed by high quantity motivation. In addition, Vansteenkiste et al. (2009) found that poor quality motivation (high external control in the absence of internal motivators) was associated with the worst learning outcomes, worse even than the amotivated (low quantity) group. These reliable differences raise an important question: what determines students' motivation orientation in an educational environment?

Antecedents of Autonomous Motivation

According to SDT, people are intrinsically motivated to grow and reach their goals when they perceive satisfaction of three universal psychological needs: autonomy, competence, and relatedness (Ryan & Deci, 2020). Within a classroom environment, autonomy refers to the need for intellectual freedom, relatedness to a sense of connection and belonging with others, and competence to confidence in one's own abilities for achieving a difficult task. To understand which of these needs were most strongly associated with autonomous motivation and how to foster these adaptive education outcomes within a classroom, Bureau et al. (2022) conducted a meta-analysis linking the universal needs posited by Ryan and Deci (2020) to different types of student motivation and to autonomy support from teachers and parents. After conducting a literature search focusing on these antecedents and broad categories linked to academic motivation, they meta-analyzed 144 research studies, sampling a combined 79,079 participants, ranging from primary school through to university. Their hypothesized path model proposed that teacher and parental autonomy support would predict the perceived fulfillment of needs for autonomy, competence and relatedness, and that these basic needs, in turn, would predict students' motivation type.

Bureau et al. (2022) revealed several key findings. First, similar to Howard et al. (2021), each psychological need for autonomy, competence, and relatedness, as well as autonomy support from both teachers and parents, all had stronger and more positive associations with more internalized types of motivation. Specifically, each of these variables was moderately negatively associated with amotivation (r = -.23 to -.38), unrelated to external regulation (r = -.10 to .05), weakly positively associated with introjected regulation (r = .15 to .23), and moderately to strongly positively associated with identified (r = .28 to .53) and intrinsic motivation (r = .23 to .58). Second, relative weight analyses revealed that satisfaction of competence needs came out as the strongest unique predictor of autonomous motivation, closely followed by autonomy, whereas relatedness carried less weight in the prediction of motivation types. Third, autonomy support from teachers was a stronger predictor of students' perceived need satisfaction than autonomy support from parents. Other notable findings showed the contributions of autonomy support from parents all diminished as the sample of students grew older, suggesting that parental involvement plays a stronger role in primary school than in university. What's more, the relationships between needs satisfaction and amotivation became more negative as mean age of sample increased, indicating that needs fulfillment is even more important for student motivation in university than in primary school (Bureau et al., 2022).

To promote autonomous motivation in students, several strategies can be employed by instructors based on the SDT research (Ryan & Deci, 2020). First, teachers should aim to support and encourage student autonomy in their learning, as this is a reliable predictor of needs fulfillment and autonomous motivation (Bureau et al., 2022; Vansteenkiste et al., 2009). For example, explaining why a topic or task is important or relevant and encouraging independent thinking can help students see the value in what they are learning and feel more in control of

their education (Steingut et al., 2017). Another autonomy supporting strategy is to provide students with choice to shift the initiative onto the student (Ryan & Deci, 2020). Additionally, instructors can foster a sense of competence by providing frequent, constructive feedback, and acknowledging effort and improvement, helping students to build confidence in their abilities (Ryan & Deci, 2020). Similarly, competence can be supported through provision of structure in the form of clear expectations, consistent rules, and scaffolded learning activities that are challenging but doable (Ryan & Deci, 2020).

While relatedness is often considered an important universal need in SDT (Ryan & Deci, 2020), Bureau et al. (2022) found that it is a relatively weak factor across all motivation types. This suggests that facilitating social connections might not be as crucial for outcomes related to independent learning and mastery, even though it may still be important for promoting persistence and reducing dropout rates. Crafting instructional materials that students enjoy can also significantly enhance their intrinsic motivation (Ryan & Deci, 2020). Instructors can interact with students to seek feedback on these materials, ensuring that they are engaging and relevant. Indeed, satisfaction is a strong predictor of students' autonomous motivation (Bureau et al., 2022). By understanding the antecedents of students' motivation, educators can create a more motivating and engaging learning environment that fosters autonomous motivation, ultimately leading to better educational outcomes.

Motivational Instructional Design

Traditional approaches to instructional design often focus directly on learning behaviours (e.g., effective study strategies, distributed practice) and cognitive processes (e.g., higher-order thinking). These methods aim to change a students' behaviour or cognitions. Recognizing the complementary relation between motivation and instructional design in research, Keller (1979)

developed the Attention, Relevance, Confidence, Satisfaction (ARCS) model of motivation to address "the heart" of students' learning, adding the affective dimension to the focus on behaviour and cognition in traditional models. Keller argued that motivational design, which involves making instructional materials that appeal to students' intrinsic motives and basic needs, can engage students' interests, thereby deepening their engagement and thinking about course material, while also maximizing learning outcome gains (Keller, 1979).

The ARCS Model

Keller (2010) defined motivational design as "the process of arranging resources and procedures to bring about changes in people's motivation" (p. 22). He drew inspiration for the ARCS model from multiple theories of motivation, including SDT and Eccles and Wigfield's Value-Expectancy Theory (2002, as cited in Keller, 2010). According to the latter, students' perceived value of an instructional task can be divided into their impact of interest (curiosity) and relevance (perceived utility; goal directed), their perceived expectancy of achieving the instructional goal (self-confidence; feelings of control over environment; expectations for success; how likely one thinks they will succeed), and the likely outcomes of the behaviour (positive/reinforcing or negative/punishing). The ARCS model incorporates these elements to provide educators with a framework to design, test, and modify instructional materials that would promote and maintain student motivation (Keller, 2010). Specifically, it emphasizes four dimensions of instructional design that contribute to autonomous motivation and self-directed learning: Attention, Relevance, Confidence, and Satisfaction (Li & Keller, 2018).

Attention involves capturing and sustaining students' interest and curiosity throughout the learning process (Keller, 1987), tapping into sources of intrinsic motivation. Some strategies to capture students' attention include using eye-catching materials presented through different

modalities, case studies and examples that are surprising or thought-provoking, or humour and positive emotions to stimulate students' participation in course discussions (Keller, 2010).

Relevance refers to students' perceptions of the importance of the learning material as it relates to their personal values, goals, and lived experiences (Keller, 1987), tapping into sources of identified regulation. Instructors can emphasize relevance through the content itself, such as by highlighting its real-world applications or importance for practical skill development, or through the teaching process, such as incorporating group work for those with a need for relatedness. Examples also include modelling enthusiasm when presenting course material, providing choice to support students' autonomy needs, and explicitly stating how the instructional technique/design relates to students' current or future careers or education goals (Keller, 2010).

Confidence refers to a students' positive expectations of success, feelings of persistence in completing a task, as well as their perceived competence that they can complete the task satisfactorily (Keller, 1987). Many of the SDT strategies for supporting students' competence needs are also relevant here, notably providing many and varied formative opportunities for students to tackle learning tasks and experience success (Ryan & Deci, 2020). Other strategies to boost confidence include using self-evaluation tools for clearly stated learning outcomes, which help the student gain confidence by having a reference point to check their understanding of course material (Keller, 2010).

Satisfaction relates to students' positive feelings about their learning experience after completing the instructional design, and their desire to continue learning (Keller, 1987). This sense of accomplishment can come from internal sources (e.g., alignment with valued actions) or it can come from external rewards or reinforcers that are meaningful to students. Strategies to

enhance satisfaction include providing opportunities to showcase newly acquired skills in practice, giving verbal praise and motivating feedback, and awarding fair grades (Keller, 2010).

Instructors can implement ARCS-informed strategies into the design of instructional materials from the start, following the recommendations outlined by Keller (2010). This approach ensures that the materials are designed to enhance student motivation and engagement. The ARCS model can also be used to evaluate the motivational properties of existing instructional materials, by measuring students' levels of engagement with those materials. This method allows educators to assess which ARCS dimensions need improvement and then use specific ARCS strategies to modify the materials accordingly (Keller, 2010).

Keller's (2010) Instructional Materials Motivation Survey (IMMS) is frequently used to measure the attention, relevance, confidence, and satisfaction dimensions of ARCS. The IMMS is a 36-item situationally based questionnaire (rated on a 5-point Likert scale) that is meant to be adapted to the particular instructional material or method. IMMS scores are interpreted through descriptives and average ratings across each dimension. According to Huang and Hew's (2016) Range of Motivation Levels guidelines, a score between 4.0-5.0 indicates a "High" level of motivation, 3.50-3.99 indicates an "Upper Medium" level of motivation, 3.0-3.49 indicates a "Medium" level of motivation, and under 3 indicates a "Low" level of motivation. Higher total and subscale scores representing each of four dimensions (averaged across each subscale items) represent higher perceived motivational properties for the instructional design element being evaluated.

ARCS Applications

In an overview of the ARCS literature, Li and Keller (2018) analyzed 27 articles from peer-reviewed sources to examine which educational contexts, research designs, and research

outcomes the ARCS model has been applied to. They found that researchers have used ARCS globally, with 40% of the studies conducted in Asia, 30% in North America, and 26% in Europe. The application of ARCS spans various education levels, from K-12 to higher education, and has been applied to both academic and vocational programs. Li and Keller (2018) observed that most ARCS applications were in web-based and computer assisted instruction, with fewer instances of applications to in-person instruction. The types of instructional materials varied widely, including face-to-face instruction, email messaging, instructional texts, lecture videos, and various course activities. Out of 27 articles reviewed, 20 employed quantitative methods using the IMMS, including correlational analyses (e.g., associations with academic performance), experimental, or quasi-experimental design (e.g., assessing perceived motivational properties pre-to-post design modification). Fewer qualitative studies tend to explore students' phenomenological perceptions toward instructional elements. Typical outcomes incorporated in the ARCS studies included affective, cognitive, and behavioral variables. Affective variables, found in 26 of 27 studies, encompassed concepts such as motivation, feelings and attitudes toward the course. Cognitive variables, included in 20 studies, often related to academic achievement. Behavioural variables, included in 8 studies, covered engagement indicators like study time, assignment submission rates, and course retention. However, Li and Keller (2018) did not estimate effect sizes for these outcomes.

Another comprehensive systematic review by Fang et al. (2023) extended the work of Li and Keller (2018) by analyzing the contents of 55 ARCS research studies published between 2011 and 2022. They noted a steady increase in ARCS usage after 2014, a dip in 2021, and a rise again in 2022. Similar to Li and Keller (2018), Fang et al. (2023) found that most ARCS studies were based in Asia (49% of sample), with Europe and North America following (16% each).

While the range of education levels was still varied, more recent studies have applied ARCS to higher education settings, particularly in social and natural science courses. Fang et al. (2023) also found that ARCS was increasingly applied to learner-centered environments. For example, the IMMS has been applied to instructional designs centered around active learning pedagogy (e.g., Gholami et al., 2021; Isa et al., 2023; Ma & Lee, 2021), with recent applications to blended learning, gamification, and flipped-classroom design (Cai et al., 2022; Chang et al., 2018; Durrani et al., 2022; Juan & Chao, 2015). The analysis by Fang et al. (2023) was not quantitative, so they did not estimate the effects of ARCS design on students' outcomes.

Effects on Motivation and Learning

To examine the effects of ARCS-informed material design on students' motivation, Dinçer (2020) conducted a meta-analysis of prior experimental or quasi-experimental studies that compared the IMMS scores between materials designed according to ARCS model and typical/control materials. This analysis aimed to establish benchmarks for evaluating the effectiveness of strategies intended to enhance students' engagement with course materials. Dinçer (2020) identified 26 studies (with 28 effect sizes) published between 2004 and 2017 to be included in the meta-analysis. Across all studies, the analysis revealed medium effect sizes for all IMMS dimensions: Attention ($g = 0.55$), Relevance ($g = 0.48$), Confidence ($g = 0.49$), and Satisfaction dimension ($g = 0.53$). The overall effect of ARCS-informed materials on motivation was also medium ($g = 0.54$). A follow-up moderator analysis based on duration of implementation indicated that effect size for total motivation was larger ($g = 0.97$) for longer programs (8 weeks or more). This suggests that prolonged exposure to ARCS-informed materials enhances motivation and improves engagement with course content.

In another meta-analysis, Goksu and Islam Bolat (2021) evaluated the effect of the ARCS instructional design on students' motivation and academic achievement, while taking into consideration differences between learning environments, disciplines, and educational levels. Their analysis included 38 experimental or quasi-experimental studies comparing ARCS-informed versus typical/control learning environments, involving 8690 students ranging from K-12 to university level. Consistent with Dinçer (2020), Goksu and Islam Bolat (2021) found that the experimental effect of ARCS-informed design on students' motivation was medium for Relevance, Confidence, and Satisfaction (g = 0.70 to 0.78) but large for Attention (g = 1.20; Cohen, 1988; Sawilowsky, 2009, as cited in Goksu & Islam Bolat, 2021). Moderation analyses by discipline revealed no significant differences in the effects of ARCS design on motivation; there were insufficient studies for comparisons by learning environment and education level. Goksu and Islam Bolat (2021) also found a medium overall effect of the ARCS-informed design on students' academic achievement (g = 0.74). The effects of ARCS design on academic achievement did not differ by education level. When grouped by learning environment, the effect size was large for blended learning (g = 1.57), medium for game-based and class-based learning (g = 0.70 and 0.64), and small for specific strategies like print materials and email messages (g = 0.35 and 0.34). In terms of discipline, courses in architecture, science, and computer technology showed the highest effect sizes (g = 1.31 to .81), while social science and health sciences had very small effect sizes (g = .10 and -.05). In summary, the findings of controlled experiments indicate that when the learning environment is enhanced with the ARCS model, it significantly increases both motivation and academic achievement in some disciplines.

Case-Based Learning (CBL)

One student-centered approach with high motivational potential is Case-Based Learning (Keefer et al., 2021; Keefer & Bradley, 2020). This constructivist approach prompts students to think deeply about course content by analyzing and applying it to real-world case scenarios. CBL is aligned with authentic assessment principles of realism and cognitive challenge (Villarroel et al., 2018). CBL focuses on building understanding through knowledge construction and critical thinking rather than right or wrong answers, so long as students provide evidence and theory-based justifications to support their case analysis. Thus, CBL is widely regarded as effective for training and testing higher-order learning outcomes, such as the application, analysis, and synthesis tiers of Bloom's taxonomy (Baeten et al., 2012; Yoo & Park, 2015).

CBL has gained traction in programs with an applied focus due to its suitability to train students for professional programs. For example, a nursing course may present scenarios where students need to interact with difficult patients (Yoo & Park, 2015). A law student might be given a written scenario of a complex workplace injury incident requiring them to apply course concepts to finding a probable solution for filing an insurance claim (e.g., Cahill-Ripley, 2015). Whether emphasizing theoretical or practical justifications, these examples all share the component of applying learned knowledge to real-world scenarios relevant to the higher-education program or career field. Thus, CBL can take various forms depending on the training's objectives and context, though its goals remain static.

Effects of CBL on Learning Outcomes

A number of studies have supported the use of the CBL method through its link to improved academic achievement (Atwa et al., 2019; Latif, 2014; Mayo, 2002, 2004) and higher-order thinking (Baeten et al., 2012; Harman et al., 2015; Yoo & Park, 2015). However, other

studies have highlighted certain challenges associated with CBL (Amiri & Heidari, 2014; Rhodes et al., 2020). In this section, we review selected research that represents these two perspectives.

Benefits of CBL

In two studies with undergraduate psychology students, Mayo (2002, 2004) demonstrated positive effects of CBL on students' learning outcomes. They used CBL to address the "trade-off" between the amount of content covered in typical foundational psychology courses and students' depth of understanding of the content covered in those courses, suggesting that CBL could help psychology students think deeper about course content. They conducted two similar experiments, one with a first-year introductory psychology course (Mayo, 2002) and one with a second-year psychology of adjustment course (Mayo, 2004). Different sections of each course were randomly assigned to either CBL or non-CBL condition, while keeping all lectures, readings, and assessments the same. The first-year case study involved a character who was struggling with emotional and behavioural problems, which could be analyzed from multiple theoretical perspectives. Students in an experimental group were asked to read the case study and to complete an applied learning activity in small groups in class, applying course content to the case. Students in the second-year course completed CBL exercises for each week of content, with the exercises including stories of characters, asked to brainstorm and write about adjustment-related implications (based on course content, applied to the case study and their own lives). The dependent variables in both experiments were students' exam performance and end-of-course survey ratings of the CBL approach. In both studies, students who completed CBL activities outperformed students who only had traditional lecture instruction on course exams, suggesting that CBL methods helped them retain and comprehend the material more deeply.

Moreover, most students (69% in first year and 75% in second year) perceived the CBL approach to be helpful for their learning (Mayo 2002; 2004).

There is also complementary qualitative research on students' positive experiences with CBL. For example, Harman et al. (2015) used CBL as an instructional method to promote higher-order cognitive learning by exposing students in two upper-year undergraduate nutrition courses to clinical scenarios. In this case study, a research assistant introduced the study and collected data through audiotaped focus group discussions and booklets asking for individual written responses. Across the courses, there were 426 written responses from 69 students and around 2 hours of verbal focus-group responses from 46 students recorded. Through thematic analysis of students' comments, the research team found that participants noted analyzing the information in more depth and being able to elaborate in their own words more and apply it to new situations. Participants mentioned the interactivity of the CBL design, agreeing it was better for active learning and taking the knowledge outside of the classroom setting. Moreover, they mentioned using more resources in gathering more information to help with solving the scenarios and the case study helped contextualize a critical awareness that played a role in shaping students' critical thinking skills related to their future practice (Harman et al., 2015). Overall, upper-year students in this study held favourable perceptions of the CBL approach.

Limitations of CBL

On the other hand, others have reported mixed experiences with CBL. For example, Hodges (2005) discussed their varied uses of CBL across three biochemistry courses, reflecting on what worked and what didn't based on their own teaching experience. In their large lecture-based course, Hodges (2005) utilized case studies as homework or in-class exercises. In their smaller foundational course, they used group-based casework to supplement lectures. In their

senior course, instead of traditional lectures, the course was entirely done with CBL, structured around ongoing class discussions of a case study of a real-world problem relevant to the course content. These discussions built on one another, where it took several classes to find a solution to the problem highlighted in the case study. Hodges (2005) observed that the more familiar students were with active and self-directed learning formats, the more comfortable they were with CBL, and that senior students were more receptive to CBL pedagogy than sophomores. Hodges (2005) recommended to take different approaches to CBL depending on education level, by using more structured case studies with novice learners and then increasing the degree of independent inquiry for more advanced learners.

In a similar vein, Rhodes et al. (2020) argued that cognitive load theory plays a role in shaping how effective CBL approaches are at enhancing students' learning. Specifically, Rhodes et al. (2020) compared the learning gains between and within two experimental groups: a conventional learning format (traditional lecture), and CBL. Third year students at a large university were randomly assigned to complete a physiology case study outlining a patient's progression through insulin resistance to type II diabetes; students then applied what they have read about, and then moved on to the next stage of the case study. All participants completed a pre- to post test, with 10 multiple choice questions with multiple correct answers. They found no difference in learning gains between the CBL condition and the traditional lecture condition. However, they found that certain student characteristics moderated the effect. Specifically, the CBL method was more effective at shaping learning gains for students with stronger academic backgrounds and more post-secondary experience. Rhodes et al. (2020) concluded that case-based methods are only effective when students are given scaffolded instruction to complete the studies, matched to their preparation level.

Cognitive load theory could also explain the null findings by Amiri and Heidari (2014), who compared the performance of midwifery students on a quiz immediately following the completion of either a CBL workshop or a standard lecture. These researchers found no evidence to indicate CBL was more effective at improving performance on the quiz, even though the quiz contained a mix of knowledge, comprehension, application, and analysis questions (Amiri & Heidari, 2014). In their feedback following the CBL workshop, participants indicated that the class time allotted to the case study was insufficient to fully process the material. These results suggest that student-centered CBL approaches that lack time or structure may not realize their full learning benefits. For cognitive load, gradual approaches to learning would theoretically help strengthen the amount of information in students' memory to boost positive perceptions of learning.

Effects of CBL on Motivation

In a meta-analysis, Wijnia et al. (2024) examined the motivational impact of CBL along with two other student-centered approaches, problem-based learning (PBL) and project-based learning (PjBL). These approaches share the same focus on active, student-centered, problem solving, but are different in that CBL focuses on case analysis, PBL around problem-solving, and PjBL around completing comprehensive projects. The meta-analysis assessed the combined and unique effects of CBL, PBL, and PjBL on various motivation dimensions related to general attitudes, competence beliefs (can I do it?), values (do I want to do it?), and autonomous versus controlled reasons (why am I doing it?). The meta-analysis was based on 132 experimental or quasi-experimental studies that compared student-centered approaches versus traditional teacher-centered approaches. The authors also explored how differences in implementation, such as type

of approach, whether it involved product creation, discipline, education level (K-12 to higher education), and duration and level of implementation, moderated these effects.

Wijnia et al.'s (2024) results showed a medium positive combined effect of student-centered approaches on motivation ($d = 0.50$), meaning these approaches promoted greater motivation to learn compared to traditional teacher-centered approaches. Looking at specific motivation dimensions, medium positive effects were found for students' attitudes ($d = 0.59$), competence beliefs ($d = 0.48$), and autonomous reasons ($d = 0.60$), but the impact on perceived value was minimal ($d = 0.19$). Moderation analyses showed that the effects of CBL, PBL, and PjBL on motivation were similar, and approaches that involved creating a product didn't affect motivation any more than the basic CBL or PBL approaches did. Moreover, the effects did not differ by education level (secondary vs. post-secondary) or by duration of implementation. However, there were significant differences by discipline, where student-centered approaches boosted motivation in STEM ($d = 0.60$) and healthcare fields ($d = 0.54$), but not in the "other domains" category ($d = 0.08$). The effects also differed by implementation type, where single-course implementations had a moderate positive effect on motivation ($d = 0.54$), while curriculum-wide implementations had only a small positive effect ($d = 0.19$). Wijnia et al. (2024) concluded that, on the whole, student-centered approaches were more effective at promoting students' autonomous motivation, perceived competence, and positive attitudes toward learning compared to traditional teacher-centered approaches. One limitation noted by Wijnia et al. (2024) is that multiple motivation measures were used across studies, and this meta-analysis incorporated multiple types of studies. However, when they analyzed the effects by type of motivation measure and study quality, they found no significant differences. Therefore, the conclusions appear to replicate across different motivation measures and study designs.

CBL and the ARCS Model

Several studies evaluated the CBL approach using the ARCS framework. Gholami et al. (2021) used the IMMS to assess learning motivation levels in a third-year nursing course, where the first half of the course was taught in a lecture-based format, and the second half of the course was taught with CBL involving weekly discussions of emergency nursing cases. The IMMS was administered after the lecture-based portion and again after the CBL portion, alongside a self-report measure of perceived problem-solving ability. Gholami et al. (2021) found significantly higher ratings of Attention, Relevance, Confidence, and Satisfaction for the CBL classes compared to lecture-based classes. In addition, students' perceived problem-solving ability was found to be stronger after CBL classes compared to lecture-based classes.

Similarly, Yoo and Park's (2015) study was particular for nursing students developing communication skills. This was a quasi-experimental design with an intervention group, which completed a CBL enhanced course, and a control group, which completed the same course as a traditional lecture-based format. Nursing students in the CBL enhanced group were given five situationally based case-studies of scenarios commonly experienced in the nursing profession. At the start and end of the course, Yoo and Park (2015) measured nursing students' communication skills (rated by clinical instructors), self-reported problem-solving ability, and learning motivation assessed with the IMMS. They found that all outcome variables significantly increased for the invention group from pre to post test. Notably, learning motivation increased from "low" to "medium" in the CBL group (based on the IMMS scores). The control group also showed a significant increase in communication skills, but the effect was significantly smaller than for the CBL group. Learning motivation did not significantly change for the control group (remaining in the "low" range), and their perceived problem-solving skills significantly

decreased from pre to post test. The higher perceived problem-solving ability in the CBL group might be because CBL elicits greater inquiry and questioning. Yoo and Park (2015) concluded that CBL methods have a positive effect on students' learning motivation, perceived cognitive competencies, and objectively measured course-specific learning outcomes.

CBL and the SDT Framework

To understand the link between CBL, learning outcomes, and motivation orientations within the SDT framework, Baeten and colleagues (2012) explored students' perceptions and academic performance in lecture-based, case-based, and combined learning environments, as well as whether these outcomes depended on students' motivation and learning approach profiles. At the start of semester, they measured college students' motivation orientation using an adapted version of the SRQ-A, and deep, surface, and strategic approaches to learning using the Approaches to Learning and Studying Inventory (Entwistle, McCune & Hounsell, 2002, as cited in Baeten et al., 2012). The study was done with 1098 first-year education students assigned to one of four different conditions of the same child development course: lecture-based, case-based, alternating lecture/case-based, and a gradual CBL implementation design. The gradual implementation started off with lectures for the first several weeks, followed by lectures combined with CBL, and ending the course with several weeks of only CBL. At the end of the course, students completed a case-based take-home exam that involved application of course material to a new case study. Finally, students completed a questionnaire about their course experiences.

Consistent with SDT research (Howard et al., 2021), Baeten et al. (2012) took the stance that autonomous motivation is more synonymous with deeper approaches to learning, while controlled motivation aligns with surface approaches to learning. Using K-means clustering, they

identified two broad motivation/learning approach profiles. The first profile was labeled "autonomously motivated deep-strategic thinkers" (55% of the sample), who scored high on autonomous motivation, deep approach, and strategic approach, but low on controlled motivation and surface approach. The second profile was labelled "little motivated and less pronounced deep-strategic learners" (45% of the sample), who scored lower on autonomous motivation, deep approach, and strategic approach, and higher on controlled motivation and surface approach. The two profiles differed in their course experiences, such that the autonomously motivated students had more positive perceptions of the instructor, course workload, and skills gained, regardless of the learning condition. Surprisingly, Baeten et al. (2012) found no significant differences in exam grades between the two motivation/learning approach profiles. Instead, exam performance differed by learning condition. Students in the gradual CBL and lecture-based groups achieved the highest exam grades, followed by the alternating lecture/CBL group, with the CBL-only group achieving the lowest exam grades. This finding suggests that a base knowledge may need to be established before engaging in CBL activities to improve achievement (Baeten et al., 2012).

In a parallel study with the same sample and course design, Baeten et al. (2013a) sought to examine whether CBL improved students' autonomous motivation. They re-administered the SRQ-A at post-test (end of course). They found that autonomous motivation increased in the gradual CBL condition, decreased in the CBL-only condition, and remained unchanged in the other two conditions. Moreover, the level of autonomous motivation at post-test positively correlated with students' exam grades, whereas the initial level of motivation at the start of term did not predict exam grades. This suggests that the effects of CBL on the quality of students' motivation depend, again, on how CBL is implemented (Baeten et al., 2013a).

In another parallel study with the same sample and course design, Baeten et al. (2013b) sought to answer the extent to which CBL approach leads to deeper thinking. They re-administered the Approaches to Learning and Studying Inventory at post-test (end of course), along with an open-ended question about students' perceptions of the instructional method. Results showed a statistically significant decrease in the deep and strategic approaches to learning in all conditions except the gradual CBL condition. This contradicted their initial hypothesis that CBL would enhance deep learning. Their model also indicated that the gradual CBL and alternating lecture/CBL conditions showed significant decreases in surface approaches, whereas this was not the case for the lecture-based and CBL-only conditions. In students' open-ended feedback, the greatest number of dislikes was expressed in the CBL-only condition, with students saying they felt they did not learn anything by solving case studies without theory from the instructor. These perceptions are consistent with the lowest exam performance in this group (see above, Baeten et al., 2012). However, the number of dislikes was lowest and the number of likes was highest in the gradual CBL and lecture-based conditions. Students in the gradual CBL condition appreciated the variation in instructional methods and the theoretical foundation received prior to engaging with cases (Baeten et al., 2013b). Based on their findings, Baeten et al. (2012, 2013a, 2013b) argued that CBL needs to be introduced to novice learners gradually and in combination with lectures, to avoid potential backfire effects on motivation and learning.

Whether students see the value in CBL or not, student-centered approaches to learning and assessment are more valuable to the student experience in terms of developing critical thinking skills, higher-order thinking, and understanding of course material (Harman et al., 2015; Mayo 2002; 2004). Despite students reporting dislike toward some CBL implementations, perceptions can be improved by implementing measures such as establishing instructor supports

and grading expectations early on, scaffolding and gradual implementation, removing time limitations, and sustaining and maintaining students' intrinsic motivation and engagement throughout the instructional design process (Baeten et al., 2013a; 2013b; Wijnia, 2024). These considerations are especially relevant when incorporating CBL into summative assessments.

Case-Based Exams

The majority of CBL research is limited to in-course learning activities and formative assessments, whereas the use of CBL for summative exams is very rarely evaluated. Although Baeten et al. (2012, 2013a) used a case-based exam as a culminating exam in their child development course, it was only intended to explain differences in academic achievement between different instructional groups; they did not examine students' perceptions or experiences of the exam itself. There are, however, a few exceptions in medical education. For example, Durak et al. (2007) developed a case-based stationary exam to assess clinical problem-solving skills. Clinical stations prioritized health information important to the community and were relevant to the goals of training healthcare practitioners and the educational objectives of the course. Each station presented a clinical case of a patient, and the student had to use a protocol taught in class to assess the patient. Durak et al.'s (2007) exploratory findings based on post-exam surveys showed the students were very receptive to the case-based exam, indicating they liked it and found it fair, and the majority (97.64) of students completed the exam satisfactorily.

Fortun et al. (2017) used a similar approach, where medical students completed several CBL activities in addition to regular lectures, and then wrote an essay-style case-based exam, followed by a national multiple-choice exam six months later. For the case-based exam, students could choose between a case that was related to topics covered in CBL activities or a case-based essays on topics covered in lectures only. They found that students who completed their case-

based essays based on the material covered in preparatory CBL activities performed better than students who completed their case-based essays based on the material covered in lectures only. However, there was no significant difference in achievement between CBL-based and lecture-based essay groups on the subsequent national multiple-choice exam six months later. The findings of these studies suggest that integrating CBL into summative assessment may provide valuable insights in some areas, but investigations into students' experiences and learning outcomes are essential to fully understand its efficacy in various educational contexts.

One point worth raising is that these applications were limited to the medical training programs. The use of case-based exams is more common for practical or professional programs like law and medicine, but less so for theory-based programs like psychology. This raises a generalizability question whether case-based exams would be similarly acceptable to students in psychology. Moreover, no studies to date have examined students' motivation orientations, motivational properties, or learning approaches specifically associated with case-based exams. As highlighted in Study 1, students' perceptions of past exams are an important determinant of their motivations and learning approaches towards future exams (Struyven et al., 2005).

Study 2 Exam Design and Objectives

Capitalizing on the motivational value of the CBL approach, the present study examined students' motivations, learning approaches, and perceived learning outcomes with a newly developed take-home case-based exam in a sophomore Abnormal Psychology course. Previous studies have investigated CBL as a valuable pedagogical and motivational approach, however these studies have limited their scope to instructional design and not summative assessment design. The present study applied CBL theory to a high-stakes, summative exam, with the intention that it would not only stimulate students' deeper approaches to learning, but also

promote students' motivation and engagement with course material while taking this exam. Therefore, the novelty of the current research is that it is the first study to explicitly apply the motivation theory lens to a CBL-based exam. Keller's (1979) ARCS model has been applied to many student-centered instructional designs in tandem with perceived learning outcomes such as critical thinking and problem-solving ability. The ARCS model has not, however, been used to evaluate students' learning motivation toward a case-based exam. Thus, the overall goal of Study 2 was to describe students' experiences and perceptions with the case-based exam, including students' motivation orientations towards the exam based on SDT, perceptions of the ARCS motivational properties of the case-based exam design, learning approaches, emotional responses and perceived learning outcomes, and the perceptions of fairness and suitability of the case-based exam.

The take-home exam used in Study 2 was designed in line with CBL and ARCS principles, as well as recommendations from the take-home exam literature reviewed in Study 1 (e.g., Bengtsson et al., 2019). As such, the case-based exam involved essay-style open-ended questions, and was built around a specific real-life case assigned by the instructor. The case-based exam targeted a range of learning outcomes by asking students to answer three types of questions: Theory, Application, and Discussion. Theory-based questions aimed at testing lower-level learning outcomes, which prompted students to describe and explain their understanding of course concepts/theories in their own words. Application-based questions aimed at testing mid-level learning outcomes, which prompted students to connect the course concepts/theories to the case by identifying specific examples and counter-examples, and to analyze the case drawing on the entire course material. Discussion-based questions aimed at testing higher-level learning outcomes, which prompted students to demonstrate their independent and critical thinking about

the case and the course material. Students were given an extended time frame for completing the take-home exam (10-14 days), which is recommended to alleviate anxiety and give more opportunity to engage in deeper thinking about the course material (Bengtsson, 2019). Based on the recommendations by Rhodes et al. (2020) and Hodges (2005), we used a scaffolded CBL practice where students completed several practice cases throughout the course before the exam. This practice is also recommended in the SDT and ARCS literatures for building students' sense of competence and confidence going into the exam (Keller, 2010; Ryan & Deci, 2020).

The present study used a concurrent, embedded, correlational mixed-methods research design, where a small qualitative component was integrated into the predominantly quantitative survey (Creswell & Plano Clark, 2011). In this study design, quantitative data served as the primary focus, while qualitative data provided secondary insights. The rationale behind using this design is to supplement the correlational findings by capturing participants' experiences with the novel exam design in their own words, while also guarding against assumptions inherent in generic quantitative surveys that were adapted for this study.

Hypotheses

Descriptive Hypotheses

The primary descriptive objective of the present study was to examine students' evaluations of the motivational properties and perceptions of anxiety, fairness, and validity of the case-based exam. Additionally, we were interested in students' motivational orientations and learning approaches adopted towards this type of exam.

H1. In line with previous studies demonstrating higher motivational properties of CBL design (Gholami et al., 2021; Yoo & Park, 2015), we predicted students' evaluations of the

ARCS motivational properties of the case-based exam will be elevated (medium to high range, on average).

H2. Based on previous findings of students' favourable perceptions of case-based exams (Durak et al., 2007), as well as related literatures on alternative and take-home exams reviewed in Study 1 (e.g., Bengtsson, 2019; Jacobs, 2021; Pereira et al., 2022), we predicted students will have favourable perceptions of the case-based exam, including moderate to high average levels of fairness and validity (H2a) and low to moderate average levels of anxiety (H2b).

H3. Based on previous research demonstrating improved autonomous motivation in CBL design (Baeten et al., 2013a; Wijnia et al., 2024), and the negative association between autonomous motivation and amotivation (Ratelle et al., 2007), we predicted moderate to high average levels of autonomous motivation (H3a) and low to moderate average levels of amotivation (H3b) towards the case-based exam. No specific prediction was made about controlled motivation, as it can coexist alongside both high and low levels of autonomous motivation (Ratelle et al., 2007; Vansteenkiste et al., 2009).

H4. Based on previous research showing that CBL design promotes less surface-level and more deep learning (Baeten et al., 2013b; Harman et al., 2015), we predicted moderate to high average levels of deep-level processing (H4a) and low to moderate average levels of surface-level approaches (H4b) towards the case-based exam in the present study.

These descriptive hypotheses were examined using both the quantitative scales (including IMMS and SRQ-A) and students' qualitative feedback about their experiences with and opinions of the case-based exam. Students responded to one open-ended question prompting them to share their thoughts about the case-based exam, and their responses were analyzed for common themes.

Inferential Hypotheses

Another objective was to replicate previous findings that students' perceptions of motivational design (from the ARCS perspective) and internalized motivation orientations (from the SDT perspective) predict deeper learning approaches and better learning outcomes in the context of a case-based exam.

H5. Based on previous research showing that ARCS-enhanced instruction promotes greater higher-order thinking and academic achievement (Gholami et al., 2021; Goksu & Islam Bolat, 2021; Yoo & Park, 2015), we predicted that students who rate the case-based exam higher on the IMMS would also self-report engaging in deeper-level processing (H5a), lower surface-level approaches (H5b), and receive higher exam grades (H5c).

H6. Based on previous research showing that autonomous motivation is associated with deeper learning approaches, better academic performance, and less academic anxiety (Baeten et al., 2012; Howard et al., 2021; Ratelle et al., 2007), we predicted that Autonomous motivation orientation would positively correlate with deep-level processing (H6a) and exam grades (H6b), and negatively correlate with surface-level processing (H6c) and anxiety (H6d).

H7. Previous research on controlled motivation found positive but weak associations with both adaptive and maladaptive outcomes (Baeten et al., 2012; Howard et al., 2021; Vansteenkiste et al., 2009). Accordingly, students' controlled motivation orientation was predicted to weakly positively correlate with deep-level processing (H7a), surface-level processing (H7b), anxiety (H7c), and exam grades (H7d).

H8. Previous research on amotivation found that it was associated with more surface learning approaches, higher academic anxiety, and lower academic performance (Howard et al., 2021; Ratelle et al., 2007). Accordingly, amotivation orientation was predicted to be negatively

associated with deep-level processing (H8a) and exam grades (H8b), and positively associated with surface-level processing (H8c) and anxiety (H8d).

H9. For our last hypothesis, we predicted that exam grades would be positively associated with deeper approaches to learning.

We also explored the linear relationships between students' perceptions of exam fairness, validity, and preference with the other study variables, without making specific predictions due to lack of prior research on case-based exams.

Study 2 Methods

Participants

The current study was conducted with students enrolled across three sections of a second-year Introduction to Abnormal Psychology course at Trent University, taught by the faculty supervisor. The instructor was the same for each section (two sections offered in the Winter 2022 term and one offered in the Summer 2022 term). Of the 169 students enrolled across the three sections, a total of 38 students consented to participate in the study (response rate of 22.5%). Of these respondents, 34 consented to the analysis of their exam grades, and 35 responded to at least one of the questionnaires. Majority of participants identified as female (80.56%), were between 19-21 years old (70.2%; $M = 22.49$, $SD = 7.55$), and pursing a major or joint-major in Psychology (73.68%). Majority of the sample (56.8%) indicated they were a third-year student at the time of completion, 24.3% of the sample indicated being a second-year student, and the remaining 18.9% were in their fourth year or above. Over half of the participants were high academic achievers, as 54% of the sample self-reported an overall GPA of an A (80% or higher), 35.1% self-reported GPA of a B (70-79%), and 10.8% self-reported GPA of a C or lower (50-

69%). Based on the self-reported final course grade, 64.8% had a final course grade of an A (80% or higher), 24.3% had a final course grade of a B (70-79%), 10.8% had a final course grade of a C or D (50-69%). A substantial number of participants (86.5%) indicated they had some prior experience with a case-based exam in a previous course.

Case-Based Take-Home Exam

As part of regular course requirements, students completed the case-based Take-Home Exam during the final examination period, worth 25% of the course grade. Students in the Winter 2022 sections were given 17 days to complete the Take-Home Exam, whereas students in the Summer 2022 section had 10 days due to the compressed nature of the summer term. The exam was the same for all sections. Students were presented with a real-life case study involving a patient experiencing an anxiety disorder – it was an article from *The Atlantic* magazine (also available as an audio podcast). Students were given a choice of four out of five long-answer questions/topics, each with a choice of one of two sub-topics, to be answered in a 10-page essay, using the course study materials as references (extra sources were allowed but not required). Students were instructed that the objective of the exam was "to demonstrate your knowledge, understanding, and independent thinking about core course material, while applying your knowledge of the course material to a specific case". The case study itself was new to students, but the case-based format was familiar to students, as students had completed a series of similar case-inquiry exercises throughout the course. Winter section students completed half of those exercises in writing and half as in-class discussions, whereas Summer section students completed all of them as in-class discussions. Students were allowed to refer to the notes and rubric from these earlier case-inquiry assignments while completing the Take-Home Exam. The course instructor maintained an active question-and-answer discussion board throughout the

examination period, to clarify expectations for the Take-Home Exam. Exam grading was based on three components: Theory (students' ability to define and explain course concepts in their own words), Application (student's analysis of the case study through the application of course concepts), and Discussion (students' independent and critical thinking about course concepts, such as understanding of the benefits, limitations, challenges, alternative explanations, and additional considerations).

Measures

Following course completion, participants completed an online survey containing the following sections (in this order): Demographics form, Open Ended feedback, Instructional Materials Motivation Survey (IMMS), Perceived Impact Questionnaire (PIQ), Academic Self-Regulation Questionnaire (SRQ-A), and two other questionnaires that were part of a larger project not included in the present study. The full questionnaire is enclosed in Appendix A. Participants were also asked for permission to access their Take-Home Exam grade from the course records (graded on a 0-100 scale).

Demographics Form

Demographic information was collected to describe the sample. Participants were asked to self-report their gender, age, academic program, cumulative GPA, number of completed course credits, prior exposure to a case-based exam, and self-reported final course grade.

Open Ended Feedback

To better understand students' experiences and thoughts about the Case-Based Exam, participants were presented with an open-ended feedback box prompting participants to "share their thoughts about the Case-Based Exam".

Instructional Materials Motivation Survey (IMMS)

The Instructional Materials Motivation Survey (IMMS; Keller, 2010) is a situational self-report measure of students' motivational reactions to a specific instructional design. The IMMS contains 36 items rated on a 5-point Likert-type scale (1 = not true; 5 = very true). Students' motivational reactions are measured along four distinct dimensions, including Attention (12 items, e.g., "There was something interesting about this exam that got my attention from the beginning"), Relevance (9 items, e.g., "It was clear to me how the content of this exam was related to things I already knew"), Confidence (9 items, e.g., "When I first looked at the Take-Home exam, I had the impression that it would be easy for me"), and Satisfaction (6 items, e.g., "Completing the Take-Home exam gave me a satisfying feeling of accomplishment"). The wording of items was altered in accordance with the scale developer's (Keller, 2010) recommendation (e.g., "Lesson" was changed to "Take-Home Exam"). Higher total and subscale scores (averaged across items) represent higher perceived motivational properties of the Take-Home Exam.

For this study, we added 7 additional items at the end of the IMMS, rated on the same 1-5 scale, to inquire about students' perceptions of the validity, appropriateness, fairness, challenge, and length of the Case-Based Exam (e.g., "The Take-Home exam was a valid way to test my learning in this course"). One of these extra items asked about students' preference for a traditional memory-based test over the take-home case study. These extra items were not included in the total or subscale IMMS scores, their responses were examined individually in exploratory analyses.

Academic Self-Regulation Questionnaire (SRQ-A)

The Academic Self-Regulation Questionnaire (SRQ-A) is a 17-item questionnaire that helps to identify motivational reasons for why students put time and effort into completing their

schoolwork, in this case specifically the Take-Home Exam. The items for this scale were adapted from analogous items found in several adaptations of the Self-Regulation Questionnaire (SRQ; Ryan & Connell, 1989), as done elsewhere (Gagné & Deci, 2005). The SRQ and its derivatives, including the SRQ-A, measure five different forms of motivation based on the types of regulation found in Self-Determination Theory: External (3 items, e.g., "I wanted to get a good grade"), Introjected (4 items, e.g., "I wanted the professor to think well of me as a student"), Identified (3 items, e.g., "I found the exam material to be useful for me"), Intrinsic (3 items, e.g., "The exam was fun to do"), and Amotivation (4 items, e.g., "I felt the exam was a waste of my time"). Items are rated on a 5-point Likert Scale (ranging from 1 = not at all true, to 5 = very true). In the present study, Identified and Intrinsic items were combined into a higher-order Autonomous regulation subscale, and the External and Introjected items were combined into a higher-order Controlled regulation subscale, as done elsewhere (Baeten et al., 2012; 2013a). A higher subscale score (averaged across items) represents a stronger endorsement of that particular regulatory style.

Perceived Impact Questionnaire (PIQ)

The Perceived Impact Questionnaire (PIQ) is designed to measure students' perceptions of how completing specific coursework impacted their learning experience. The PIQ was developed based on previous qualitative data obtained from our research lab; items for this scale were generated based on several themes that emerged from students' comments to open-ended questions about similar case-based assessments (Keefer et al., 2021). For this study, participants were asked to respond to a list of statements following the phrase "The Take-Home Exam prompted me to...," and to rate each statement on a 5-point Likert scale (ranging from 1 = not at all true, to 5 = very true). There is no standard scoring for the PIQ items. Therefore, exploratory

principal component analysis was conducted to identify groups of items that could be combined into subscales (see Appendix E for full results). This analysis yielded five factors for this measure: Anxiety (3 items), Preparation (3 items), Cursory Approach (3 items), Divergent Thinking (3 items), and Understanding (5 items). Anxiety subscale assessed exam-related stress and concerns about doing poorly on the exam (e.g., "Stress over whether I was doing the task correctly"). Preparation subscale assessed how thoroughly students studied and used course material during the exam process (e.g., "Read the textbook chapters thoroughly"). Cursory Approach subscale assessed students' tendency for surface-level engagement while completing the exam (e.g., "State some things without fully understanding them"). Divergent Thinking subscale assessed students' original and self-reflective thinking while completing the exam (e.g., "Think originally and outside the box"). Understanding subscale assessed students' deep comprehension and higher-order thinking about the course material (e.g., "Think critically about the course material"). A higher subscale score (averaged across items) represents stronger endorsement of that particular perceived impact of the Take-Home Exam.

Procedure

Students enrolled across three sections of the Introduction to Abnormal Psychology course (two sections in Winter 2022, one section in Summer 2022) were invited to participate in this study via email following the completion of the Take-Home Exam and submission of final grades. Students interested in this study were directed to a secure survey link on the Qualtrics portal and provided with information about the goal of the research project, the implications of the project, and the risks and benefits of their participation. Students were assured that their participation or non-participation would not influence any future relationship with the course instructor or any other researcher conducting this study. In addition to consent for the Qualtrics

survey, participants were also asked to provide itemized consent for permission to access their Take-Home Exam grade from the course records, and to contact them at a later date for a follow-up interview. Consenting participants completed the Qualtrics survey which took approximately 30 minutes of their time, and then were sent a debriefing form explaining the purpose of the study. All data were anonymized following the matching of the surveys to the exam grades. The name of each participant was entered into a raffle to win one of six Amazon.ca gift certificates (four valued at $25 and two valued at $50). The Letter of Information and Consent form are enclosed in Appendix A. This study was approved by the Trent University Research Ethics Board (protocol #28028).

Data Analysis

For hypothesis one, we followed Huang and Hew's (2016) motivation levels guidelines to interpret students' overall level of learning motivation toward the case-based exam, based on the total IMMS scores and for each dimension of ARCS (Keller, 1987). As previously stated, these guidelines illustrate which scores fall into High, Upper Medium, Medium, and Low levels of motivation. These scores are interpreted through descriptives and average ratings across each dimension. By calculating the scale mean, each IMMS dimension was compared to these guidelines. Higher subscale scores represent higher perceived levels of engagement for that instructional design element. This helped us determine how motivated and engaged students were across the attention, relevance, confidence, and satisfaction areas of the case-based exam.

For hypotheses 2-4, we calculated the measures of central tendency and variability on continuous scale variables, and frequencies for ordinal item variables. Distributions of scale variables were assessed for elevated skewness or kurtosis z-values outside the range of -2 to +2, which would indicate deviations from normality (George & Mallery, 2010).

Assuming there were no major deviations, a bivariate Pearson's correlation matrix was computed to identify any significant relationships relevant to hypotheses 5-9. Scale variables that violated assumptions of normality were analyzed with Spearman's rank correlations instead. Given the ordinal nature of the additional items added to the IMMS scale (e.g., validity, fairness, appropriateness), these associations were also explored through Spearman's correlations. Because of the small sample size, these analyses may lack power to detect the hypothesized relationships of weak-to-moderate magnitude at the conventional significance level of $p < .05$. Therefore, we also considered non-trivial effect sizes in addition to significance ($r > .20$). Correlation effect sizes were interpreted according to the empirically derived guidelines recommended by Gignac and Szodorai (2016), who suggested using 0.10 for trivial/small effect, 0.20 for typical/medium effects, and 0.30 for large effects, in personality research.

Students' qualitative comments were thematically analyzed following Braun and Clarke's (2006) six-stage process. Codes were independently generated by two members of the research team. Once agreed on, the convergent ideas were re-applied to students' comments, and codes were combined into overarching themes.

Study 2 Results

Table 6 presents the descriptive results for the sample of study participants. Most study variables were normally distributed. Only two variables had elevated skewness and kurtosis indicators (z-values well outside the range of -2 to +2), suggesting violations of normality (George & Mallery, 2010). The SRQ-A Amotivation subscale was positively skewed and leptokurtic, and the Exam Grade was negatively skewed and leptokurtic. Thus, Spearman's rho was calculated for these two variables in addition to Pearson's correlations to verify reliability of findings.

Exam Grades

The average exam grade in the sample was 78.82 ± 9.11 (on a 0-100 scale), which fell in the B to A range. Compared to the full-class average of 73.35 ± 13.57 (which fell in the C to A range), the sample mean was significantly higher, $t(33) = 3.51$, $p < .01$. This means that the students in our sample were not representative of the entire class, but rather of the high achieving students in the class, as indicated by higher exam grades.

IMMS Variables

According to Huang and Hew's (2016) Range of Motivation Levels guidelines, mean IMMS scores for Attention, Relevance, and Confidence all fell into the top "High" level of motivation (4.00 – 5.00), and Satisfaction fell into the "Upper Medium" level of motivation (3.50 – 3.99). The mean for the Total IMMS scale fell in the "High" level of motivation (see Table 1). These results are consistent with Hypothesis 1, which predicted medium to high average ARCS levels. Looking at inter-scale correlations among the IMMS subscales, all ARCS factors were strongly positively interrelated (ranging from .52 to .84), in line with their shared motivational properties and in support of the internal consistency of the Total IMMS scale. However, the IMMS subscales also showed different patterns of correlations from one another with other study variables (see Table 7). Therefore, in subsequent analyses the IMMS subscales were treated as separate variables in addition to the planned analyses with the composite IMMS scale. Treating the subscales separately can highlight the differences in correlational patterns between the specific motivational properties of the Case-Based Exam.

Descriptives for the seven extra items added to the IMMS can be found in Table 8. Based on the sum of "mostly true" and "very true" responses, the majority of the sample perceived the Case-Based Exam to be a valid test of learning (82.8%), a fair assessment of knowledge (80%),

appropriate for the learning objectives (80%), and an opportunity to demonstrate one's strongest knowledge (74.3%), while being sufficiently challenged by the exam (71.5%). A third of the sample (34.3%) indicated the exam took them longer to complete than anticipated, and only a small minority (20%) expressed preference for a traditional memory-based test over the Case-Based Exam. These results are consistent with Hypothesis 2a, which predicted overall favourable perceptions of fairness and validity of the case-based exam.

Spearman's correlations among the seven extra IMMS items were computed due to the ordinal nature of item scores (see Table 8). Items pertaining to perceptions of the exam's validity (#37), fairness (#41), alignment with learning objectives (#38), and strongest knowledge (#43) were very strongly positively inter-correlated (r_s = .69 to .80), indicating substantial variance overlap. Therefore, these four items were combined into a single composite variable (Valid/Fair) for subsequent analyses. Perceptions of being sufficiently challenged (#39) were significantly strongly positively correlated with perceptions of validity/fairness (r_s = .42 to .64), but neither were associated with perceptions of the exam being too long (#42). Preference for traditional memory-based exam (#40) was significantly strongly negatively correlated with perceptions of validity/fairness (r_s = -.44 to -.49), and moderately negatively (but not significantly) correlated with perceptions of sufficient challenge (r_s = -.23), but it was not associated with the subjective lengthiness of the exam. Taken together, these correlations indicate that students who perceived the Case-Based Exam as more valid/fair also tended to perceive the exam as more challenging and tended to prefer the take-home exam to a traditional exam. Conversely, students who perceived the exam as too easy also tended to rate the exam as less fair and less appropriate assessment of their learning, in addition to preferring a traditional exam over a take-home exam. Perceived lengthiness of the exam was not correlated to any of the other added IMMS items,

which suggests that students could judge the exam as longer or shorter, regardless of how they perceived validity/fairness, suitability, challenge, and preference for a traditional exam.

SRQ-A Variables

On average, students endorsed high levels of both Autonomous regulation (4.00 ± 0.71) and Controlled regulation (4.02 ± 0.58), which fell between the "mostly true" and "very true" response categories. Both ratings imply that, as a sample, students put effort and time into the case-based exam both from external and internal motivational reasons. The average rating for Amotivation was very low (1.21 ± 0.47), which fell between the "not at all true" and "mostly not true" rating categories. Amotivation was also positively skewed and leptokurtic, with the min/max ranging from 1.00 to 3.00 on a 1-5 scale. This suggests that majority of participants were motivated to complete the exam, and only a select few students were moderately unmotivated to complete the exam. These results are consistent with Hypothesis 3, which predicted moderate to high levels of autonomous motivation (H3a) and low to moderate levels of amotivation (H3b).

Correlational analysis revealed one statistically significant relationship among the three SRQ variables (see Table 7). Autonomous orientation was significantly strongly correlated with lower Amotivation (r_s = -.54), suggesting that participants who lacked intrinsic or identified reasons for completing the exam were more likely to report no motivation for completing the exam. However, Controlled orientation was not associated with either Autonomous orientation or Amotivation. This suggests that students who put time and effort into the take-home exam for extrinsic or introjected reasons did so regardless of whether they had any intrinsic or identified reasons, and despite not wanting to do the exam at all.

PIQ Variables

The average ratings for Divergent Thinking (4.08 ± 0.75) and Understanding (4.22 ± 0.67) fell between the "mostly true" and "very true" response categories, suggesting that the sample endorsed high levels of these learning outcomes. Average ratings of Preparation (3.52 ± 0.89) fell between the "somewhat true" and "mostly true" response categories, which indicates that the study sample endorsed a moderate level of thoroughness in using course material. The average ratings for Anxiety (2.35 ± 0.93) and Cursory Approach (2.15 ± 0.67) both fell between the "mostly not true" and "somewhat true" response categories, suggesting that, for the most part, the levels of anxiety and surface-level engagement were low to moderate among the study participants throughout the exam. These results are consisted with Hypothesis 4, which predicted moderate to high levels of deep learning (H4a) and low to moderate levels of surface learning (H4b). Results for anxiety are consisted with Hypothesis 2b, which predicted low to moderate levels of anxiety.

Correlational analysis revealed several statistically significant or otherwise non-trivial ($r > .10$) relationships among the PIQ subscales (see Table 7). The subscales of Divergent Thinking, Understanding, and Preparation were significantly strongly correlated with one another ($r = .40$ to .48), reflecting their shared focus on deep-level processing. Additionally, Divergent Thinking and Understanding were moderately negatively correlated with the Cursory Approach subscale that measured surface-level processing ($r = -.28$ and -.26), although these associations did not reach significance. This suggests that students who reported deeper comprehension of the material and greater engagement in higher-order thinking during the case-based exam were less likely to approach the exam with surface-level processing. These results provide preliminary convergent validity evidence for these PIQ subscales.

In terms of students' self-reports of anxiety and stress during the exam, the Anxiety subscale was significantly negatively associated with Understanding (r = -.50), and it also had weaker negative associations with Preparation and Divergent Thinking (r = -.31 and -.21), although these correlations did not reach significance. This suggests that students who were less thoroughly prepared, understood the material less well, and engaged in less higher-order thinking tended to feel more anxious and worried about doing poorly on the exam. Cursory Approach was unrelated to Anxiety or Preparation, suggesting surface-level processing could occur irrespective of students' anxiousness or preparation throughout the exam.

Correlations between IMMS and PIQ Variables

Pearson's correlations were computed to examine the relationships between participants' perceptions of the exam's motivational properties and perceived impact ratings of Anxiety, Divergent Thinking, Understanding, Cursory Approach, and Preparation (see Table 7). Total IMMS was significantly strongly and positively associated with Divergent Thinking and Understanding (r = .64 and .59). Broken down by subscale, these relationships were similarly strong for Attention, Relevance, and Satisfaction (r = .50 to .65), and moderate for Confidence (r = .35 to .41). Total IMMS, as well as Attention, Relevance, and Satisfaction were also strongly positively associated with Preparation (r = .30 to .35), whereas Confidence was not. Total IMMS was significantly strongly associated with lower Anxiety (r = -.38), with the strongest correlation coming from Relevance (r = -.42) and the weakest from Satisfaction (r = -.27). Taken together, these correlations indicate that students who perceived the Case-Based Exam as interesting, relevant, and enjoyable tended to engage in more deep level processing (greater understanding, divergent thinking, thorough preparation), which is consistent with Hypothesis 5a. Conversely, students who rated their anxiety levels high tended to perceive the exam as uninteresting,

irrelevant, and not enjoyable. Contrary to Hypothesis 5b, neither Total IMMS nor its subscales were reliably associated with Cursory Approach, indicating that this learning approach could occur irrespective of the exam's perceived motivational properties.

Correlations between IMMS and SRQ-A Variables

Pearson's and Spearman's (for Amotivation) correlations were computed to examine the relationships between participants' perceptions of the exam's motivational properties and students' Autonomous, Controlled and Amotivation regulation (see Table 7). Total IMMS was significantly strongly and positively associated with Autonomous regulation (r = .72). These relationships were similarly strong for Attention, Relevance, and Satisfaction (r = .72 to .81), and strong for Confidence (r = .30). Relevance and Confidence were moderately positively associated with Controlled regulation (r = .24 and .23), whereas Total IMMS, Attention and Satisfaction were not associated with Controlled regulation. Total IMMS was also significantly negatively associated with Amotivation (r_s = -.41), with the strongest correlation coming from Attention (r_s = -.47) and the weakest from Confidence (r_s = -.12). These correlations indicate that, when paired with students' motivation orientation, the interesting, relevant, and enjoyable elements of the exam tended to be more closely aligned with students' Autonomous orientation for completing the exam, which illustrates their shared emphasis on intrinsic and internalized value. In addition, the moderate associations between Controlled Regulation and Relevance and Confidence could suggest that students' extrinsic reasons for completing the exam may have tied into external gains in these stages of the assessment design (e.g., case-study was relevant to someone they know, confidence in achieving a certain grade). Regarding Amotivation, students who did not put time and effort into the exam tended to find the exam uninteresting, irrelevant, and not enjoyable.

Correlations between SRQ-A and PIQ Variables

Pearson's and Spearman's (for Amotivation) correlations were computed to examine the relationships between students' Autonomous, Controlled and Amotivation regulation and perceived impact ratings of Anxiety, Divergent Thinking, Understanding, Cursory Approach, and Preparation (see Table 7). Autonomous regulation was significantly strongly and positively associated with Divergent Thinking, Understanding, and Preparation (r = .42 to .56). In contrast, Autonomous regulation was strongly negatively associated with Anxiety and Cursory Approach (r = -.32 and -.30). Amotivation was strongly negatively associated with Understanding (r_s = -.55) and strongly negatively associated with Divergent Thinking and Preparation (r_s = -.33 and -.27). Amotivation was also positively moderately associated with Anxiety (r_s = .29). Consistent with Hypothesis 6, these associations suggest that students who put time and effort into the exam for intrinsic or identified reasons also tended to adopt a deep learning approach (H6a), and had less anxiety (H6d) and surface-level thinking (H6c). Furthermore, consistent with Hypothesis 8, students who did not have the motivation to put time and effort into the exam tended to engage in less deep-process thinking (H8a), and they experienced higher levels of anxiety (H8d). Contrary to Hypothesis 7 (a, b, and c), controlled regulation was not reliably associated with any of students' perceived impact ratings, indicating that students' deep versus surface approaches to learning could occur irrespective of the controlled motivation orientation.

Correlates of Exam Grades

Table 7 contains Spearman's rank correlations of all study variables with exam grades. Spearman's coefficients were interpreted due to violations of normality in the distribution of grades. Consistent with Hypothesis 5c, total IMMS was significantly strongly and positively associated with exam grades (r_s = .39), with the strongest correlation coming from Confidence

(r_s = .53), indicating that students who felt more confident in their mastery of course content and ability to complete the exam successfully tended to receive higher grades. In addition, Satisfaction and Attention had weaker positive associations with exam grades (r_s = .31 and .23), although these correlations did not reach statistical significance. Of the SRQ-A variables, Controlled motivation orientation was significantly positively associated with exam grades (r_s = .41), indicating that students who were more strongly motivated by external rewards (e.g., the pursuit of good grades, feeling good about oneself) tended to receive higher exam grades. This finding is only partially consistent with Hypothesis 7d, which predicted a weaker positive association between controlled motivation and exam grades. However, neither Autonomous motivation orientation nor Amotivation were reliably associated with exam grades, which contradicts Hypotheses 6b and 8b. Of the PIQ variables, only Divergent Thinking was positively correlated with exam grades (r_s = .24), though not statistically significantly. This is partially consistent with Hypothesis 9, which predicted that grades will be positively associated with deeper processing and comprehension because the exam was graded to reflect these outcomes.

Correlates of the Extra IMMS Variables

Table 9 contains Spearman's rho correlations of all study variables with the extra IMMS variables. Spearman's estimates were interpreted due to the ordinal nature of item scores. Regarding the motivational properties of the exam, Total IMMS significantly and positively correlated with students' perceptions of exam validity/fairness (r_s = .72) and challenge, (r_s = .51). These relationships were similarly strong with subscale items, with Satisfaction showing the strongest association with perceptions of valid/fair (r_s = .66), and Relevance showing the strongest association with perceptions of exam challenge (r_s = .56). Total IMMS and Attention, Relevance, and Satisfaction subscale scores were not associated with perceptions of exam

lengthiness, whereas Confidence was strongly negatively associated with lengthiness (r_s = -.38). Students' preference ratings for a traditional exam were strongly negatively associated with all IMMS variables (r_s = -.40 to -.55). Taken together, these associations suggest that students who found the exam to be interesting and relevant, also tended to perceive the exam as fair, appropriate, challenging, and preferred the take-home exam to a traditional exam.

Autonomous regulation was strongly positively associated with students' perceptions of Valid/Fair and Challenge (r_s = .69 and .44), and strongly negatively associated with preferences for a traditional exam (r_s = -.32). Like Autonomous regulation, Controlled regulation was positively moderately associated with Challenge, but positively moderately associated with preference for a traditional exam (r_s = .26 and .23). Amotivation was strongly negatively associated with Valid/Fair and Challenge perceptions (r_s = -.52 and -.34), and strongly and positively associated with preference for a traditional exam (r_s = .41). These Spearman associations suggest that students who put time and effort into the exam for intrinsic or identified reasons tended to perceive the exam as fair, appropriate, challenging, and preferred the take-home exam to a traditional exam. Students who were motivated by external or introjected reasons perceived the exam to be challenging but tended to indicate a preference for traditional-style exams. Students who were not motivated to put time/effort into the exam tended to report the exam was not fair/valid or challenging, and tended to prefer traditional-style exams. No motivation orientation dimension was associated with exam lengthiness, suggesting that regardless of how students rated their motivational reasons for completing the exam (internal or external or none), they could have still perceived the exam to be too long.

Students' perceptions of Valid/Fair were strongly and positively associated with Divergent Thinking and Understanding (r_s = .39 and .64) and moderately positively associated

with Preparation (r_s = .29). A similar trend was reflected in perceptions of Challenge, which was strongly positively associated with Understanding (r_s = .80) and strongly and moderately associated with Divergent Thinking and Preparation (r_s = .33 and .21). All three of these PIQ variables were strongly to moderately negatively associated with preferences for a traditional exam (r_s = -.35 to -.28). These findings suggest that students who perceived the exam as fair/valid and challenging tended to engage in deep-level learning during the exam. Students who reaped these benefits also reported a higher preference for the take-home exam.

A reverse trend was found for Anxiety and Cursory Approach. Perceptions of Valid/Fair was strongly negatively associated with Anxiety (r_s = -.45), and moderately negatively associated with Cursory Approach (r_s = -.22). Challenge was moderately negatively associated with Anxiety (r_s = -.26) but did not correlate with Cursory Approach. Anxiety and Cursory Approach were also moderately positively associated with preference for a traditional exam (r_s = .33 and .35). These findings suggest that students who perceived the exam as fair/valid and challenging tended to feel less anxious during the exam. Students who felt more anxious and adopted a surface-level approach tended to report a higher preference for a traditional exam rather than the take-home exam.

Cursory Approach was strongly positively associated with perceptions of lengthiness (r_s = .30), which indicates that to a small extent students who adopted a surface-level approach to the exam tended to rate the exam as taking a long time to complete. No other PIQ rating correlated with perceptions of lengthiness, suggesting that high or low perceptions of lengthiness could occur irrespective of students' Anxiety, Divergent Thinking, Understanding and Preparation.

Exam grade was strongly and moderately positively associated with Valid/Fair and Challenge perceptions (r_s = .36 and .21), but not associated with Length or preferences for a traditional exam.

Qualitative Analysis

Our qualitative research question sought to explore students' experiences with the Case-Based Exam expressed in their own words. A thematic analysis of students' feedback yielded three main themes that reveal the exam was generally well received, with many perceived learning benefits weighed against few technical drawbacks. *Prepared to Succeed* was the first theme identified in students' feedback. The students identified high instructor presence, in-course feedback, and the preparatory case-inquiry assignments as key to their success on the exam. In the second theme, *Better than Traditional Exam*, the students generally found the case-based exam experience to be more valuable when comparing it to a traditional memory-based exam. Lastly, the third theme, *Benefits Outweigh the Costs*, involved students' positive appraisals toward the exam as being effective at testing their higher-order cognitive skills and material comprehension, weighed against negative appraisals toward the technical aspects (e.g., time limit, layout) of the exam.

Prepared to Succeed

A theme frequently mentioned by the participants related to how prepared they felt going into the exam: "*All required materials to succeed are provided in the course, so the ability to succeed falls back on the students [sic] personal ability to learn and retain necessary information*" (Participant 8). Overall, the students recounted the in-course case-inquiry assignments as being important to their success. For instance, one participant thought the exam "*was pretty easy to do with the preparation from the case inquiries*" (Participant 26), with

another participant adding the exam was "*very similar to the case inquiries so it was very fair*" (Participant 36). This suggests that the students appreciated being familiarized with the case-based exercises throughout the course, as they were taken as good practice building up to the final exam.

The alignment between the course and exam objectives, and the feedback on case-inquiry assignments in particular was found to be valuable to the students' success on the exam, as two students wrote:

> *"I genuinely loved the idea for the Take-Home Exam. Especially since throughout the entire course we were preparing and focusing on topics that we would need a full well rounded understanding of each topic. But, if you did well on the Case inquiry assignments or if you didn't you were able to get feedback that overall would have helped with writing the take home exam. You felt prepared for it and was always able to reach out for clarification."* (Participant 1)

> *"I appreciate that what we work on throughout the semester is actually tied in to what we will be writing about. Instructor feedback throughout the course and feedback from the case inquiries was extremely helpful in how I ended up writing in my exam. I found it faster to do than a normal take home exam because of how much guidance and understanding I had of the materials I could use to fit the criteria of the rubric."* (Participant 23)

Receiving feedback beforehand from the practice case-inquiry assignments allowed the students time to adjust and make changes prior to the exam, and having samples and these materials available during the open-book exam helped make the exam process clear. For these

reasons, the students praised both the positive and negative instructor feedback as key for their preparation process. Instructor presence was also mentioned as an important factor for students' preparation, even during the final examination period, with one student writing: "*our professor was also always willing to help with questions that came up about the exam*" (Participant 19).

Despite these positive perceptions, some students had their reservations. In reference to the Take-Home Exam, one student shared "*it was easy but we should be given the rubric and more instructions of the expectations in the exam*" (Participant 28). These negative recounts might indicate that there could have been more communication between the instructor and the students regarding the expectations of the exam. Some students may have been unfamiliar with the exam's alternative format despite having case-inquiry assignments throughout the course in order to prepare for the exam. Thus, a potential implication for students who did not do the practice case inquiries is that they would be at a disadvantage on this exam:

> *"Very stressful, I could've done more to prepare so I would have a better understanding of what to expect" (Participant 27)*

Better Than Traditional Exam

Participants frequently juxtaposed their experience with the Case-Based Exam to their past experiences with traditional memory-based tests. Overall, participants found the Case-Based Exam more valuable for their learning experience. In line with Blooms' taxonomy of higher educational objectives, the students acknowledged that standard closed-book exams "*focus too much on rote memorization and do not give students that opportunity to truly think critically or present their knowledge to the best of their abilities*" (Participant 7). Given that the Take-Home Exam emphasized testing higher-order cognitive skills in addition to concept definitions, students were able to showcase the depth of their knowledge acquired throughout the course in

their write-up. Another participant added memory benefits on to the critical thinking benefits, saying "*I retained more from doing the case study then [sic] I would have studying for a traditional exam*" (Participant 35). Being assessed on higher-order cognition is more effective for long-term retention than a multiple-choice exam. In a well summarized quote, one student wrote:

> *"If I were to rate the exam compared to a multiple-choice exam, I would have to say the take-home exam likely reinforced the material covered throughout the course. Research is much more interesting than picking the best right answer for demonstrating retention of lecture/textbook and seminar information. Take-home exam likely took slightly more time than having to memorize content and [it] focuses on understanding."* (Participant 5)

Despite the Take-Home Exam being more time-consuming than a traditional exam, Participant 7 felt it "*was also significantly less stressful than a standard examination*". The preparation aspects highlighted in the previous theme, *Prepared to Succeed*, may have contributed to this more relaxed feeling. That is, even though the Take-Home Exam was lengthy, the students felt adequately prepared going into the exam and thus felt less stress because the majority knew what was expected of them. Another student found the take-home format to alleviate some of their test anxiety, writing:

> *"I am a big fan of the aspect of a Take-Home Exam. I believe that having the ability to complete the exam in my own household allows be [*sic*] to focus and think clearer on the exam. It also prevents the stress of having other students around while I write, which could in turn affect my abilities to remain focus [*sic*] and less stressed during the exam. If more courses were to implement the notion of a Take-Home Exam, I believe the exam grade of students would increase overall."* (Participant 11)

However, the Case-Based Exam was not for everyone, and some students found this format more frustrating than a traditional exam. As one student in our sample explained:

> *"I would have preferred to take an actual exam. As much as I was able to apply my learning I struggled to put all of my knowledge into the case as it did not encompass everything I learned through the course."* (Participant 25)

Benefits Outweigh the Costs

Throughout their feedback, the students often used positive emotion words to express how much they "enjoyed" or "loved" writing the Case-Based Exam, particularly in relation to being able to exercise their higher-order cognitive skills. One student remarked the exam sparked their interest and enjoyment:

> *"The take-home exam was an interesting way to assess the final state of comprehension of course material. I enjoyed the opportunity to go into more depth explaining and applying the concepts learned throughout the course. This allowed me to explain my thinking and provide evidence and rationale to support my answers."* (Participant 4)

In addition to the reported higher-order cognition benefits, another student highlighted the actual content of the case-study as being interesting and educational:

> *"I thought the final exam in psych 2300 was an effective way to evaluate the combination of skills we developed throughout the course. It was highly enjoyable as well as thought provoking. The article was thorough and provided a good basis for our exam. I enjoyed learning about the potential causes, treatments, and external/internal factors of the author's experience with anxiety".* (Participant 19)

An instructional design challenge when creating an effective Case-Based Exam is finding an appropriate and good-quality case study that is not only interesting, but also accurately

illustrates core course concepts, provokes critical thought, and fits within the criteria of the rubric. The above quote shows that, in addition to testing higher-order cognitions, providing students with interesting content to work with is equally important to keep them engaged in the task. Several participants also noted how working on the case study helped them "*think deeper*" (Participant 30), "*consolidate knowledge*" (Participant 3), and "*connect everything that was learned in the course*" (Participant 16). Another participant added: "*I found that the take home exam connected the dots for me in terms of the course material. To take what we had learned and then apply it to a study, opened up my learning experience*" (Participant 29).

The participants also gave input about the technical aspects of the exam (e.g., time limit, layout, page count), with a mix of negative and positive appraisals. Time and length of the exam were noted to be longer than expected by some participants. For example, one participant said the exam experience was "*good but long… was burned out by the time I finished it*" (Participant 31). Another participant concurred, saying they experienced some complications due to the length of the exam: "*It was hard to fit in getting the paper done tho [sic] and trying to make sure I studied enough for my other classes*" (Participant 26). In contrast, others found the timeframe reasonable, saying "*we were given ample time to work on the final paper*" (Participant 32). Another participant appreciated the timing flexibility, saying "*there were multiple days we could choose from to complete it [the exam] which relived a lot of stress as i [sic] could do it when I felt prepared*" (Participant 24).

Another source of mixed feedback came from the 10-page length requirement. Some participants thought it was too restrictive: "*I found it difficult to stick to the 10 page limit however fully understand the reasoning for this limit*" (Participant 12). Other participants found the page requirement to be excessive:

"While it was very long and I'm like 90% sure I did not meet the required length it was a very good assignment and there was nothing on it that surprised me, like stated the only thing that was hard for me was the amount of pages that We [sic] had to reach."

(Participant 13)

These quotes illustrate that despite some negative appraisals toward the exam, the students still believed it was a good assessment. In terms of exam difficulty, the students generally described the exam as "*challenging but quite manageable*" (Participant 5), which is an optimal level of difficulty for sustained motivation. Indeed, while acknowledging the challenge of being tested on higher-order cognition, the students found the Case-Based Exam made them more interested in the subject: "*the take-home exam was interesting to work on and help expand and peak [sic] my interest in the class, even though so what [sic] of a challenge*" (Participant 37). Based on this feedback, we can conclude that students' perceived benefits related to higher-order thinking skill development outweighed their negative appraisals toward the technical aspects of the exam (e.g., time limit, layout, page count).

Study 2 Discussion

To our knowledge, this was the first empirical study that applied case-based learning (CBL) pedagogy to a university exam and analyzed it using Keller's (1987) ARCS framework. The ARCS framework is traditionally applied to instructional design, whereas we applied it to our case-based summative exam. The intention was to use CBL and ARCS principles to create a motivating assessment design and then relate it to students' motivation orientations, emotional responses, and perceived learning outcomes, including deeper process thinking. In addition, we explored students' experiences with and opinions of the case-based exam through thematic analysis of their open-ended responses, which yielded three main themes: preparation to succeed,

students' preference toward the case-based exam, and the benefits outweighing the costs. These qualitative results are integrated with the quantitative findings throughout the discussion.

Motivational Properties of the Case-Based Exam

In the ARCS model, Keller (1987) suggested that students are motivated to engage with instructional materials when they perceive them to capture attention (evokes curiosity and arousal), to be relevant to personal interests (importance of the material tied to professional goals or their life; helping to sustain attention), enable confidence in their abilities (high perceived persistence and competence for task completion), and lead to satisfaction (feelings of overall enjoyment after the instructional design has ended; Li & Keller, 2018). In line with previous studies demonstrating higher motivational ARCS properties of CBL designs relative to traditional designs (Gholami et al., 2021; Yoo & Park, 2015), we predicted that students' evaluations of the ARCS dimensions in the case-based exam would be elevated at the medium to high range (hypothesis H1). Our findings supported this hypothesis. Based on Huang and Hew's (2016) range of motivation levels guidelines, students' average ratings of the IMMS dimensions of Attention, Relevance, and Confidence all fell into the top "High" level of motivation, while Satisfaction average rating fell into the "Upper Medium" level.

The high Attention levels indicate the case-based exam materials captured students' initial interest. The case study was presented as a personal narrative from a magazine article, rather than a dry academic text, and it was also available as an audio podcast. Moreover, the high ratings on the Relevance dimension speak to students being familiar with the topic or finding the case study relevant to their interests. The case study presented to the students involved an individual adapting to their experiences living with anxiety, which is one of the most common mental health conditions. As one participant stated in their open-ended response, they felt the case study used

provided a good basis for the exam, allowing them to directly apply and solidify their learning about the causes and treatments of anxiety. This case study was perceived to be highly appropriate and beneficial to the participants' learning, and high ratings on the Attention and Relevance dimensions speak to this.

The high ratings on the Confidence dimension indicate that students felt well prepared for the case-based exam. Our participants had opportunities to complete several practice case-based exercises prior to the exam, so they knew what was expected of them, this could have factored into their confidence during the exam. In their qualitative comments, many participants highlighted how extensive instructor feedback, extended time frame, and preparatory case-inquiry assignments were crucial to students' sense of confidence and success in the case-based exam. This aligns with existing literature on alternative take-home exams, emphasizing the importance of instructor support and structured preparatory assignments in helping prepare students for exams (Slack & Priestley, 2023; Tam, 2022). Prior CBL research has also noted that unfamiliar instructor expectations, insufficient time, and unstructured CBL formats heightened students' stress and dissatisfaction with CBL methods (Amiri & Heidari, 2014; Harman Hodges, 2005; Rhodes et al., 2020). However, preparatory efforts prior to the CBL activities, as well as scaffolded (gradual) CBL implementation helped mitigate these potential setbacks (Baeten et al., 2012, 2013a, 2013b). This implies that students benefit from building a base of knowledge before engaging in complex case-based activities, and our findings extend this conclusion to a case-based summative exam.

The upper-medium ratings of Satisfaction are consistent with our hypothesis H2, which predicted that students would generally hold favourable perceptions of the case-based exam. Perceptions of satisfaction could be attributed to several factors. For one, the majority of the

participants (over 80%) perceived the case-based exam to be a valid test of their learning and a fair assessment of their knowledge, and these perceptions were strongly positively correlated with satisfaction ratings. Moreover, 80% stated that they preferred the case-based exam to a traditional memory-based test. Participants' strong preference for the case-based exam was also expressed in their qualitative comments, where they generally found the case-based exam experience to be more valuable, and even more effective for retention, compared to traditional exams. Specifically, they perceived the chief benefit of the case-based exam to stimulate their higher-order thinking skills, considering it more worthwhile despite greater cognitive challenge. This argument is evidenced in participants' use of positive emotions words, such as "enjoyment" and "interesting", when referring to the case study as well as the challenge in their open-ended responses. Participants also appreciated the flexibility associated with the take-home exam format, explaining that the absence of stringent time limits alleviated their anxiety. However, some participants expressed frustration about page length, unclear instructions, and cognitive load related to the process, which may explain why satisfaction ratings were slightly lower than the other ARCS dimensions. Overall, participants perceived the benefits of the case-based exam as outweighing the costs.

Participants' qualitative comments, on how the case-based exam encouraged them to think deeper about course material and retain it better, were consistent with the high levels of deep processing and low-to-moderate levels of surface processing endorsed in the present sample. These results support our hypothesis H4, which was based on previous research showing that CBL designs promote less surface and more deep learning when implemented appropriately (Baeten et al., 2013b; Harman et al., 2015). For instance, participants in Harman et al. (2015) recognized the value of case-based approaches for improving their analysis, elaboration, and

application skills. Similarly, Baeten et al. (2012, 2013b) found that students in the gradual (scaffolded) CBL condition appreciated case-based learning for contextualizing theory, adopted deeper approaches to learning, and outperformed students in the lecture-only or CBL-only conditions. Mayo (2002; 2004) found that students who completed CBL assignments in conjunction with lectures found them to be helpful for their learning and outperformed students who only had traditional lecture instruction, suggesting that CBL methods helped them comprehend the material more deeply.

In a related hypothesis, H5, we predicted that total motivational scores (as measured by the IMMS) would correlate positively with students' self-reports of deeper process thinking about course material (H5a), negatively with surface-level thinking (H5b), and positively with exam performance (H5c). These predictions were based on previous research showing that ARCS-enhanced instructional designs promoted greater higher-order thinking (e.g., problem-solving) and academic achievement compared to traditional instructional designs (Gholami et al., 2021; Goksu & Islam Bolat, 2021; Yoo & Park, 2015). Consistent with past research and our hypothesis H5a, we found that students with higher IMMS scores tended to report more deep-process thinking during the exam, including greater comprehension, more critical and divergent thinking, and more thorough use of preparation materials. Although the IMMS scores were negatively associated with surface-level thinking as predicted, these associations were weak and non-significant, therefore hypothesis H5b was not supported. The association between students' overall IMMS scores and exam grades was found to be moderate and positive, with Confidence dimension being the strongest predictor of exam performance. This aligns with hypothesis H5c and with findings from Goksu and Islam Bolat (2021), who found a medium overall effect of the ARCS-enhanced designs on academic achievement among undergraduate students.

Motivation Orientations towards the Case-Based Exam

Aside from the ARCS framework, we also applied the SDT framework to explore students' motivation orientations towards the case-based exam. Based on previous research demonstrating increased autonomous motivation associated with CBL designs (Baeten et al., 2013a; Wijnia et al., 2024), we predicted moderate to high average levels of autonomous motivation and low to moderate average levels of amotivation towards the case-based exam, with no specific predictions for controlled motivation. Consistent with this hypothesis, our sample endorsed autonomous and controlled regulation, with participants' scores falling between "mostly true" and "very true" response categories. The average rating for Amotivation was very low, falling in "not at all true" and "mostly not true" rating categories. These findings are consistent with one of the clusters of motivational profiles identified by Ratelle et al. (2007) and Vansteenkiste et al. (2009). Ratelle et al.'s (2007) most prevalent motivation orientation profile included a combination of high levels of autonomous and controlled motivation, with low levels of amotivation. This mirrors the "high quantity" motivation orientation profile identified by Vansteenkiste et al. (2009), where both motivation orientations can exist simultaneously.

With regards to learning approaches, our hypotheses H6a and H6c proposed that autonomous regulation would positively correlate with deep process thinking and negatively correlate with surface-level processing. Conversely, we expected that amotivation would have the opposite pattern of associations (hypotheses H8a and H8c). Lastly, we predicted that controlled regulation would show weak positive associations with both learning approaches (hypotheses H7a and H7c). These hypotheses were informed by an extensive body of research linking motivation orientations to various learning outcomes (Howard et al., 2021). Our hypotheses H6a and H6c were both supported: students who were motivated by autonomous

reasons engaged in deeper processing and less surface-level thinking. Hypothesis H8a was supported but not H8c: students who lacked any motivation engaged in less deep-process thinking, but not necessarily in more surface-level thinking. Hypotheses H7a and H7c were not supported, as controlled regulation was not associated with either of the perceived learning outcomes.

These findings are mostly consistent with meta-analytic evidence by Howard et al. (2021), who found that adaptive outcomes such as mastery goals, effort, engagement, and enjoyment were moderately negatively associated with amotivation, unrelated or weakly positively related with controlled subtypes of motivation, and positively strongly associated with autonomous subtypes of motivation. The results for autonomous regulation in our sample appeared to mimic the "Autonomously Motivated Deep-Strategic Thinkers" cluster identified by Baeten and colleagues (2012), characterized by high levels of autonomous motivation and deep approaches to learning, as well as low levels of controlled regulation and surface approaches to learning. Similarly in our study, autonomous regulation was positively associated with divergent thinking, deeper understanding, and more thorough use of preparation materials, and negatively associated with surface-level processing. However, we did not replicate Baeten et al.'s (2012) second cluster of "Little Motivated and Less Pronounced Deep Strategic Thinkers", characterized by lower scores on autonomous motivation and deep thinking approach, but higher scores on controlled motivation and surface approach to learning. In our sample, controlled regulation was not associated with these patterns of outcomes. Amotivation, however, was more closely associated with Baeten et al.'s second cluster.

One aspect that our study added to these associations, that was not extensively discussed by Baeten et al. (2012), was the levels of anxiety associated with motivation orientation. We

found that exam-related anxiety correlated negatively with autonomous regulation (support for hypothesis H6d), positively with amotivation (support for hypothesis H8d), but was unrelated with controlled regulation (contrary to hypothesis H7d). These findings are partially consistent with Howard et al.'s (2021) meta-analysis, where maladaptive outcomes such as anxiety were moderately positively associated with amotivation, weakly positively associated with controlled forms of motivation, but unrelated or weakly negatively associated with autonomous forms of motivation. Our findings contradicted Vansteenkiste et al. (2009), who found that motivational profiles characterized by high levels of controlled regulation, even in the presence of high autonomous motivation, experienced more test anxiety than profiles characterized by low levels of controlled regulation, even when autonomous regulation was also low. Taken together, these findings indicate that as students' lack of motivation increases, anxiety becomes more likely.

With respect to exam performance, Howard et al.'s (2021) meta-analysis found that academic performance was moderately negatively correlated with amotivation, strongly positively associated with autonomous forms of regulation, and unrelated or weakly positively associated with controlled forms of regulation. Accordingly, we formulated our hypotheses H6b, H7b, and H8b based on these past findings. Surprisingly, we only found controlled regulation was moderately positively associated with exam grades, which we attributed to how students valued external rewards (e.g., the pursuit of good grades, feeling good about oneself) in shaping this association. Neither autonomous regulation nor amotivation were significantly associated with exam grades in the present study, which contradicts prior research (Baeten et al., 2013a; Howard et al., 2021).

The finding that autonomous regulation was associated with deep process thinking but not grades, whereas controlled orientation was associated with grades but not with deep process

thinking, is noteworthy. These differential findings highlight a clear distinction between the correlates of autonomous and controlled motivation orientations. Students with more autonomous regulation reported deeper levels of understanding and engagement in divergent thinking, however their exam grades did not reflect this learning approach. On the other hand, students with more controlled motivation tended to achieve higher grades, but this did not translate to deeper understanding or divergent thinking. In addition, the relationship between exam grades and deep-level processing was weak and non-significant in the present study, contradicting our hypothesis H9, which proposed that exam grades would be positively associated with deep-level processing.

Howard et al. (2021) found that students with mastery approaches to learning tend to associate with higher motivation internalization, meaning that students with intrinsic reasons to complete a task (i.e., autonomous regulation) tend to focus more on mastering material (i.e., deeper approaches to learning). They also found a positive association between performance approaches and both autonomous and controlled motivation types, suggesting that students can focus on performing regardless of whether their motivation is internal or external. These complementary findings provide evidence that the more internalized the motivation, the deeper the learning gains, preparation, and confidence, but not necessarily how well students will perform. In fact, a meta-analysis by Richardson et al. (2012) found that strategic learning approach was more strongly correlated with university grades than pure deep learning approach. This may be because deep learners who are motivated by intrinsic curiosity may pay less attention to assessment rubrics, compared to strategic learners who direct their efforts towards fulfilling the rubric and maximizing their grade.

In our data we also observed associations between students' ratings of the motivational properties of the case-based exam and their motivation orientations towards the exam. Students who perceived the case-based exam to be highly engaging in any of the ARCS aspects, but especially in attention, relevance, and satisfaction, tended to have a stronger autonomous orientation. In contrast, controlled orientation was only weakly positively associated with relevance and confidence dimensions, and unrelated with attention and satisfaction. Moreover, amotivation was found to be moderately negatively associated with all ARCS properties except confidence. Our findings indicate that the interesting, relevant, and enjoyable elements of the exam were more closely aligned with an autonomous orientation for completing the exam, which illustrates their shared emphasis on intrinsic joy and identified values. The associations between controlled regulation and relevance and confidence may indicate that students' extrinsic reasons for completing the exam were tied to elements such as relevance of the course to their educational goals and confidence in achieving a certain grade. Regarding amotivation, students who did not invest time and effort into the exam tended to find the exam uninteresting, irrelevant, and unenjoyable.

Methodological Considerations and Future Directions

This study had several methodological strengths. First, the exam we developed and used in this study was constructed the ARCS framework and we used IMMS to formally evaluate its motivational properties; applying ARCS and IMMS has been rarely formally been done before for a case-based exam. Moreover, we used mixed methods to evaluate participants' perceived learning outcomes. Given the newness associated with applying ARCS and IMMS to an exam, providing students with a medium to reflect open-endedly was warranted, as it is possible some areas or variables that were relevant to the study may have been overlooked in planning due to

the lack of research. Finally, we used actual exam grades as opposed to self-reported grades, the latter being a common practice in past research. Using participants' actual grades mitigated potential over-estimates of their exam or course grades.

Several limitations can be identified in this research. First, this was a retrospective correlational study where data were collected approximately one month after students completed the case-based exam. While this may have helped mitigate problems related to recency bias, we did not capture participants' immediate reactions to the case-based exam. Given that this research is descriptive and correlational in nature, in none of the analyses performed are we able to draw causal inferences. Though, we can highlight the strength and magnitude at which all relevant variables are related.

Second, our sample was small and not representative of the cohort of students who completed the case-based exam in this course. Specifically, students with exam grades in the A and B range were over-represented, and students with exam grades in the C and D range were under-represented in the sample. This means our conclusions were based on a high-achieving sample and may not be generalizable to the typical classroom population. Based on prior research, we would expect a lower-achieving sample to have lower levels of autonomous and controlled motivation with moderate levels of amotivation (Ratelle et al., 2007). Thus, our findings with regard to learning outcomes are not new for this demographic of students. For instance, other researchers have found CBL was more effective at shaping learning gains for students with stronger academic backgrounds (Rhodes et al., 2020). A larger and more varied sample of students would have also helped to increase the power to detect weak-to-moderate correlation coefficients between test variables.

Another limitation is specific to our measure of deep and surface learning approaches. The Perceived Impact Questionnaire we used was not a validated scale, and subscales developed from exploratory principal component analysis were based on our insufficient sample size. This could mean the subscales may not reliably capture the underlying dimensions from the scale items by including potential noise, as well as being underpowered. A larger sample would be needed to better validate this scale. In line with this, the Perceived Impact Questionnaire is based on students' perceptions of learning outcomes and not actual learning outcomes. This could be potentially problematic as research can show discrepancies between how students think versus how they behave (Deslauriers et al., 2019).

Future directions in designing case-based exams should focus on the issues of academic dishonesty. From an educator's perspective, maintaining academic integrity is necessary for ensuring students complete take-home exams honestly (Bengtsson et al., 2019). For this exam, several precautions were utilized to prevent academic dishonesty: (1) students were required to contextualize their answers using the assigned case and to cite course-specific materials to support their answers; (2) no marks were given for answers that had inappropriate or no source citations, or that were quoted verbatim or worded too closely to the sources; (3) academic integrity expectations were explicitly stated on the exam; and (4) exams were submitted electronically and scanned through a software that flags duplicate or substantially overlapping submissions. It should be noted, however, that the present study was conducted before generative artificial intelligence applications became available. Therefore, measures taken to prevent academic misconduct in the current study may no longer be sufficient in the current educational context. With more frequent use of generative artificial intelligence, it is becoming more difficult for educators to develop assessments that deter academic dishonesty. As evidenced in this study,

participants were strongly motivated by external reasons such as high grades, even in the presence of autonomous reasons such as internalized learning values. We believe this area warrants further investigation.

Another area worth investigation is for education researchers to further consider the distinction between effort and performance (Keller, 1979). Academic performance, which is considered the gold standard outcome variable in educational psychology research, is only an indirect measure of learning. It may be more worthwhile to shift focus to measuring effort and skill gain, which is a more direct measure of learning. It is important to differentiate between students who give more effort and take away more learning from the course than those who perform better academically. Most of the studies reviewed have tended to link motivation with academic performance, which may be more applicable to controlled/externally regulated students and those with a strategic learning approach. Effort, on the other hand, is a better indication of motivation according to Keller (1979), and may be more applicable to the autonomous/ intrinsically oriented students and those with a deep learning approach. Further research should investigate this distinction.

General Conclusion

The present thesis examined undergraduate Psychology students' perceptions of summative exams, focusing on their attributions of exam purpose and effectiveness, learning approaches, emotional responses, and motivational impacts. Through two studies, this research identified exam features associated with positive student experiences and engagement. Study 1 highlighted the importance students placed on the assessment of meaningful learning outcomes that went beyond mere memorization of facts, and how their experiences with traditional exams often deviated from these ideal objectives. Students' perceptions of exam fairness and

effectiveness hinged on clearly communicated expectations and the alignment between the exam and the intended learning outcomes. Students' attributions about test anxiety and emotional responses highlighted the role of preparation and examination environment. Study 2 demonstrated how features like real-life relevance and scaffolding in a case-based take-home exam were associated with greater perceived fairness and effectiveness and deeper cognitive and intrinsic engagement. These findings reinforce the importance of integrating student-centered approaches and motivational elements into assessment design. Given the high stakes of university exams for students' career, financial, social, and personal goals, instructors should focus on emphasizing the relevance of assessments and clearly communicating their purpose and expectations. These insights advocate for student-centered assessment designs that balance pedagogical objectives with practical constraints to foster fair, engaging, and meaningful learning opportunities for students when possible.

Reflexivity Statement

A lot of my struggles with learning in the public education became apparent in my middle school years. It was here where I quickly learned the magnitude at which impressions and impression management, and the perceptions of others was important to the developing brain. It was here where I realized the fun from elementary school was over and that things were beginning to get serious with more emphasis on academic performance and achievement. It's where I began to see the flaws of the education system and how teaching and learning became different. While at the time I obviously did not see these flaws and instead just attributed them to advanced instruction taking place beyond what I was prepared for, retrospectivity, I now view things very differently.

When I explain to my friends and family the topic I am researching, they usually cringe as soon as I mention Exams. “No, but, the idea is that we can make exams that are actually fun and enjoyable to take,” is often met with more looks of cringe. From my observations, students, and even the general public, do not enjoy taking tests. Going into this project I agreed with many of these negative stances. My testing history is not good. My negative stance was shaped early in Grades 3 and 6, where I failed Ontario’s EQAO math tests. I then failed the Ontario Secondary School Literacy Test in Grade 10 and was forced to take a literacy course in summer school in order to meet the high school diploma requirement. It was also common for me to fail other, smaller tests in elementary and secondary school. It was my creativity during in-class assignments that kept my mediocre grades afloat. But every time a test came up, my grades always took a hit for the worse.

My negative perceptions toward exams and the self-fulfilling prophecy that I do not test well followed me into university. These were unfortunately reinforced in my first year, where my grades again kept taking hits after tests. In two examples, I failed my Media Studies exam where I saw a 10 percent grade decrease and my Introduction to Psychology course, where I saw a nine percent grade decrease after those exams. First year courses also predominately use multiple-choice tests, and the bulk of my grades from these years were determined by these tests. Eventually, my progress on exams began to improve as the format of exams changed. In third-, and fourth-year classes, the emphasis on memorization became less pronounced, and questions that reflected deeper thinking and understanding of course material became more apparent. It wasn’t until my fourth-year honours thesis where I started to build an appreciation for summative assessments that tackled these learning outcomes. The honours thesis is where I begin to see the fruits of my labour; a tangible final product where enclosed was a big critical thinking exercise

completed by me. From my perspective, this is a more fulfilling and encapsulating experience than common memory-based tests that I was exposed to all throughout my schooling.

Through my honours thesis topic itself, I also learned the realities of developing knowledge and why a healthy dose of skepticism is good, how instructors can improve students' interests in completing coursework, and how this information can be used to leverage deep thinking about course material. Since I found the honours thesis experience so fulfilling after years of performing poorly on tests, I was curious as to why summative critical thinking exercises were not more common in instruction. On one end, evaluating these types of exercises requires a lot of time and resources, which makes its widespread use not feasible. On the other end, I am under the impression that the purpose of the education system is to train critical thinkers. This discrepancy led to my research search for alternative exam formats that tackle critical thinking learning outcomes, and was the basis for my search into how students think about university exams. These areas culminated into this master's thesis topic, where we looked all three of these areas (how things are constructed, how to engage, and how to promote deeper thinking) in summative assessments.

It is also important to acknowledge that I graduated with a Bachelor of Arts Honours Degree in Psychology, having taken numerous social science classes in psychology and sociology. During my master's, I held an inaugural graduate student researcher position in Trent's Centre for Teaching and Learning, along with various graduate teaching and research assistantship positions. I was also a student lab demonstrator for Trent's Introductory to Psychology course. Recently, and up until recently have acquired a part-time teaching load in Fleming College's School of General Arts and Science. These experiences have provided me

with both a theoretical understanding of adult learning and the practical application of teaching adults. The expertise I bring into this research project also comes from this experience.

It is possible that my history and academic experience influenced my approach to the research process. The focus area of interview questions was generated based on the hot topics in assessment design, though, how I interpreted the responses from these may be subject to individual interpretation. Reflecting on my own teaching values, I believe that genuine learning is difficult, receptivity is essential, and high expectations must be communicated. This position may mean that I come from a place having high expectations of other students. So, if a participant were to respond to an interview question something related to any of these three areas, my expressions and potential to ask follow-up questions may have increased. Several measures were put in place to mitigate bias, including the use of two research coders described in the Study 1 and 2 methods sections. The faculty supervisor on this project was the author of the case-based exam in Study 2, and thus had a vetted interest in understanding students' experiences to improve the exam experience for students, while still maintaining exam quality and integrity. Their mentorship came from these goals, though the findings described here are majorly constructed in the context of my own beliefs about assessment and instructional design, which were formed even before the analyses were completed.

The results found in this thesis strengthened my views that other students – like myself in my own undergrad – want to be tested critically but scaffolded effectively. My biases have followed me throughout my early academic and professional career in the higher-education sector. That is not to say my biases are inherently bad; they come from an education student researcher who was taught the value of becoming a critical thinker. I feel this makes me better apt at actively working toward finding solutions to adapt to the new wave of challenges in

Education – of which there will be many. We hope that other educators will find the insights from this study helpful in their own practice.

References

Acosta-Gonzaga, E., & Ramirez-Arellano, A. (2022). Scaffolding Matters? Investigating Its Role in Motivation, Engagement and Learning Achievements in Higher Education. *Sustainability, 14*(20), 13419-. https://doi.org/10.3390/su142013419

Almeqdad, Q. I., Alodat, A. M., Alquraan, M. F., Mohaidat, M. A., & Al-Makhzoomy, A. K. (2023). The effectiveness of universal design for learning: A systematic review of the literature and meta-analysis. *Cogent Education, 10*(1). https://doi.org/10.1080/2331186X.2023.2218191

Amiri Farahani, L., & Heidari, T. (2014). Effects of the case-based instruction method on the experience of learning. *Journal of Biological Education, 48*(1), 40–45. https://doi.org/10.1080/00219266.2013.788539

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*. New York: Addison Wesley Longman.

Andrade, H. G. (1999). *Student Self-Assessment: At the Intersection of Metacognition and Authentic Assessment*.

Archbald, D. A., & Newmann, F. M. (1988). *Beyond standardized testing: assessing authentic academic achievement in the secondary school*. National Association of Secondary School Principals.

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment and Evaluation in Higher Education, 39*(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

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*(2), 205–234. https://doi.org/10.1007/s10648-017-9406-6

Atwa, S., Gauci-Mansour, V. J., Thomson, R., & Hegazi, I. (2019). Team-based and case-based learning: a hybrid pedagogy model enhancing students' academic performance and experiences at first-year tertiary level. *Australian Educational Researcher, 46*(1), 93–112. https://doi.org/10.1007/s13384-018-0282-y

Baeten, M., Dochy, F., & Struyven, K. (2012). Using students' motivational and learning profiles in investigating their perceptions and achievement in case-based and lecture-based learning environments. *Educational Studies, 38*(5), 491–506. https://doi.org/10.1080/03055698.2011.643113

Baeten, M., Dochy, F., & Struyven, K. (2013a). The effects of different learning environments on students' motivation for learning and their achievement. *British journal of educational Psychology, 83*(3), 484-501.

Baeten, M., Dochy, F., & Struyven, K. (2013b). Enhancing students' approaches to learning: the added value of gradually implementing case-based learning. *European journal of psychology of education, 28*, 315-336.

Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education, 74*, 49-64.

Bengtsson, Lars. (2019). Take-home exams in higher education: A systematic review. *Education Sciences, 9*(267), 1-16.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York, NY: David McKay

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

Brown, G. T. L. (2011). Self-regulation of assessment beliefs and attitudes: a review of the Students' Conceptions of Assessment inventory. *Educational Psychology (Dorchester-on-Thames), 31*(6), 731–748. https://doi.org/10.1080/01443410.2011.599836

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice, 15*(1), 3–17. https://doi.org/10.1080/09695940701876003

Buckley, A. (2023). Are we answering the question that has been set? Exploring the gap between research and practice around examinations in higher education. *Studies in Higher Education*, 1-17.

Bureau, J. S., Howard, J. L., Chong, J. X., & Guay, F. (2022). Pathways to student motivation: A meta-analysis of antecedents of autonomous and controlled motivations. *Review of Educational Research, 92*(1), 46-72.

Cahill-Ripley, A. (2015). Innovative methods of assessment in law: The value of open-book exams as a catalyst for improving teaching and learning in the law school. *The Law Teacher, 49*(2), 206-218.

Cai, X., Li, Z., Zhang, J., Peng, M., Yang, S., Tian, X., … Yan, F. (2022). Effects of ARCS model-based motivational teaching strategies in community nursing: A mixed-methods

intervention study. *Nurse Education Today, 119*, 105583–105583. https://doi.org/10.1016/j.nedt.2022.105583

Chang, S.-C., & Hwang, G.-J. (2018). Impacts of an augmented reality-based flipped learning guiding approach on students' scientific project performance and perceptions. *Computers and Education, 125*, 226–239. https://doi.org/10.1016/j.compedu.2018.06.007

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (Second Edition.). SAGE.

Cypress, B. (2018). Qualitative Research Methods: A Phenomenological Focus. *Dimensions of Critical Care Nursing, 37*(6), 302–309. https://doi.org/10.1097/DCC.0000000000000322

Dale, E. (1969). Audiovisual methods in teaching (3rd ed.). New York: Dryden Press.

Darabi Bazvand, A., & Rasooli, A. (2022). Students' experiences of fairness in summative assessment: A study in a higher education context. *Studies in Educational Evaluation, 72*, 101118-. https://doi.org/10.1016/j.stueduc.2021.101118

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry, 11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Deslauriers, L., McCarty, L., Miller, K., Callaghan, K., & Kestin, G. (2019). Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. *Proceedings of the National Academy of Sciences - PNAS, 116*(39), 19251–19257. https://doi.org/10.1073/pnas.1821936116

Dinçer, S. (2020). The effects of materials based on arcs model on motivation: A meta-analysis. *Ilköğretim Online, 19*(2), 1016–1042. https://doi.org/10.17051/ilkonline.2020.695847

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public interest, 14*(1), 4-58.

Durak, H. I., Caliskan, S. A., Bor, S., & Van Der Vleuten, C. (2007). Use of case-based exams as an instructional teaching tool to teach clinical reasoning. *Medical teacher, 29*(6), e170-e174.

Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino Jr, A. R., Boulet, J. R., & Eva, K. (2016). Comparing open-book and closed-book examinations: a systematic review. *Academic medicine, 91*(4), 583-599.

Durrani, U. K., Al Naymat, G., Ayoubi, R. M., Kamal, M. M., & Hussain, H. (2022). Gamified flipped classroom versus traditional classroom learning: Which approach is more efficient in business education? *The International Journal of Management Education, 20*(1), 100595-. https://doi.org/10.1016/j.ijme.2021.100595

Fang, X., Ng, D. T. K., Leung, J. K. L., & Xu, H. (2023). The applications of the ARCS model in instructional design, theoretical framework, and measurement tool: a systematic review of empirical studies. *Interactive Learning Environments*, 1-28

Flores, M. A., Veiga Simão, A. M., Barros, A., & Pereira, D. (2015). Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. *Studies in Higher Education (Dorchester-on-Thames), 40*(9), 1523–1534. https://doi.org/10.1080/03075079.2014.881348

Fong, C. J., Gonzales, C., Cox, C. H-T., & Shinn, H. B. (2023). Academic help-seeking and achievement of postsecondary students: A meta-analytic investigation. *Journal of Educational Psychology, 115*(1), 1-21.

Fortun, J., Morales, A. C., & Tempest, H. G. (2017). Introduction and evaluation of case-based learning in the first foundational course of an undergraduate medical curriculum. *Journal of Biological Education, 51*(3), 295-304.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*(4), 331–362. https://doi.org/10.1002/job.322

Gardner, H. (1992). *Assessment in context: The alternative to standardized testing*. In B. R. Gifford, & M. C. O'Connor (Eds.), Changing assessment: Alternative view of aptitude, achievement and instruction (pp. 77-119). Boston: Kluwer Academic.

George, D., & Mallery, P. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference 17.0 Update*. 10th Edition, Pearson, Boston

Gerritsen-van Leeuwenkamp, K. J., Joosten-ten Brinke, D., & Kester, L. (2018). Developing questionnaires to measure students' expectations and perceptions of assessment quality. *Cogent Education, 5*(1), 1464425-. https://doi.org/10.1080/2331186X.2018.1464425

Gerritsen-van Leeuwenkamp, K. J., Joosten-Ten Brinke, D., & Kester, L. (2019). Students' perceptions of assessment quality related to their learning approaches and learning outcomes. *Studies in Educational Evaluation, 63*, 72-82.

Gholami, M., Changaee, F., Karami, K., Shahsavaripour, Z., Veiskaramian, A., & Birjandi, M. (2021). Effects of multiepisode case-based learning (CBL) on problem-solving ability and learning motivation of nursing students in an emergency care course. *Journal of Professional Nursing, 37*(3), 612–619. https://doi.org/10.1016/j.profnurs.2021.02.010

Ghosh, S., & Bowles, M. (2020). Challenges and implications in achieving content validity of an authentic assessment task designed to assess seafarer's leadership and managerial skills.

WMU Journal of Maritime Affairs, 19(3), 373–391. https://doi.org/10.1007/s13437-020-00209-9

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74-78.

Glazer, N. (2014). Formative plus Summative Assessment in Large Undergraduate Courses: Why Both? *Interna*tional *Journal of Teaching and Learning in Higher Education, 26*(2), 276-.

Goksu, I., & Islam Bolat, Y. (2021). Does the ARCS motivational model affect students' achievement and motivation? A meta-analysis. *Review of Education, 9*(1), 27-52.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative Assessment and Writing: A Meta-Analysis. *The Elementary School Journal, 115*(4), 523–547. https://doi.org/10.1086/681947

Guay, F. (2022). Applying Self-Determination Theory to Education: Regulations Types, Psychological Needs, and Autonomy Supporting Behaviors. *Canadian Journal of School Psychology, 37*(1), 75–92. https://doi.org/10.1177/08295735211055355

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Educational Technology Research and Development, 52*(3), 67–86. https://doi.org/10.1007/BF02504676

Gulikers, J., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction, 18*, 172–186.

Hagström, L., & Scheja, M. (2014). Using meta-reflection to improve learning and throughput: redesigning assessment procedures in a political science course on power. *Assessment & Evaluation in Higher Education, 39*(2), 242-252.

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

Harman, T., Bertrand, B., Greer, A., Pettus, A., Jennings, J., Wall-Bassett, E., & Babatunde, O. T. (2015). Case-based learning facilitates critical thinking in undergraduate nutrition education: students describe the big picture. *Journal of the Academy of Nutrition and Dietetics, 115*(3), 378-388.

Haynie III, W. J. (2003). Effects of Take-Home Tests and Study Questions on Retention Learning in Technology Education. *Journal of technology education, 14*(2), 6-18.

Henderson, M., Chung, J., Awdry, R., Mundy, M., Bryant, M., Ashford, C., & Ryan, K. (2023). Factors associated with online examination cheating. *Assessment and Evaluation in Higher Education, 48*(7), 980–994. https://doi.org/10.1080/02602938.2022.2144802

Herrington, J., & Herrington, A. (1998). Authentic Assessment and Multimedia: how university students respond to a model of authentic assessment. *Higher Education Research and Development, 17*(3), 305–322. https://doi.org/10.1080/0729436980170304

Hodges, L. C. (2005). From problem-based learning to interrupted lecture: Using case-based teaching in different class formats. *Biochemistry and Molecular Biology Education, 33*(2), 101-104.

Horne, A., Yuen, J. J., Beveridge, T. S., & McLean, S. (2022). Grade-focused interactions in higher education: has the pursuit for good grades replaced learning? *Advances in Physiology Education, 46*(4), 752–762. https://doi.org/10.1152/ADVAN.00021.2022

Howard, J. L., Bureau, J. S., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science, 16*(6), 1300-1323.

Huang, B., & Hew, K. F. (2016). Measuring Learners' Motivation Level in Massive Open Online Courses. *International Journal of Information and Education Technology, 6*(10), 759–764. https://doi.org/10.7763/IJIET.2016.V6.788

Humphries, B., & Clark, D. (2021). An examination of student preference for traditional didactic or chunking teaching strategies in an online learning environment. *Research in Learning Technology, 29*, 1–12. https://doi.org/10.25304/rlt.v29.2405

Iannone, P., & Simpson, A. (2017). University students' perceptions of summative assessment: The role of context. *Journal of Further and Higher Education, 41*(6), 785–801. https://doi.org/10.1080/0309877X.2016.1177172

Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education, 59*(2), 243–258. https://doi.org/10.1007/s10734-009-9245-6

Isa, N, K., Nordin, N, A., Saari, E, M., Isa, N, J., & Yunos, M, Y. (2023) Student Motivation in Learning through the Use of 21st Century Learning Activities. *Educational Administration: Theory and Practice, 28*(2), 222-230. https://doi.org/10.17762/kuey.v29i2.613

Jacobs, A. D. (2021). Utilizing take-home examinations in upper-level analytical lecture courses in the wake of the COVID-19 pandemic. *Journal of Chemical Education, 98*(2), 689-693.

Jennings, P. A., Frank, J. L., Snowberg, K. E., Coccia, M. A., & Greenberg, M. T. (2013). Improving classroom learning environments by Cultivating Awareness and Resilience in Education (CARE): Results of a randomized controlled trial. *School Psychology Quarterly, 28*(4), 374–390. https://doi.org/10.1037/spq0000035

Johanns, B., Dinkens, A., & Moore, J. (2017). A systematic review comparing open-book and closed-book examinations: Evaluating effects on development of critical thinking skills. *Nurse Educ Pract, 27*, 89−94.

Johnson, C. M., Green, K. A., Galbraith, B. J., & Anelli, C. M. (2015). Assessing and refining group take-home exams as authentic, effective learning experiences. *Research and Teaching, 44*(5), 61-71.

Juan, Y. K., & Chao, T. W. (2015). Game-based learning for green building education. *Sustainability, 7*(5), 5592–5608. https://doi.org/10.3390/su7055592

Keefer, K., & Bradley, H. (June 2020). *Evaluating motivational properties of inquiry-based assignments in sophomore psychology courses*. Poster presented at the annual convention of the Canadian Psychological Association. Virtual Event.

Keefer, K.V., Bodrug, T., & Girardi, L. (2021). *Bringing inquiry-based learning into online discussions: Student experiences and perceived outcomes*. Poster presented at the Canadian Psychological Association convention (Virtual).

Keller, J. M. (1979). Motivation and Instructional Design: A Theoretical Perspective. *Journal of Instructional Development, 2*(4), 26–34. https://doi.org/10.1007/bf02904345

Keller, J. M. (1987). Development and Use of the ARCS Model of Instructional Design. *Journal of Instructional Development, 10*(3), 2–10. https://doi.org/10.1007/BF02905780

Keller, J. M. (2010). Motivational Design for Learning and Performance The ARCS Model Approach (1st ed. 2010.). Springer US. https://doi.org/10.1007/978-1-4419-1250-3

Kibble, J. D. (2017). Best practices in summative assessment. *Advances in physiology education, 41*(1), 110-119.

Kim, A. S. N. & Shakory, S. (2017). Early, but not intermediate, evaluative feedback predicts cumulative exam scores in large lecture-style post-secondary education classrooms. *Scholarship of Teaching and Learning, 3*(2), 141-150.

Kreber, C., Klampfleitner, M., McCune, V., Bayne, S., & Knottenbelt, M. (2007). What do you mean by "authentic"?: A comparative review of the literature on conceptions of authenticity in teaching. *Adult Education Quarterly, 58*(1), 22–43. https://doi.org/10.1177/0741713607305939

Latif, R. (2014). Impact of case-based lectures on students' performance in vascular physiology module. *Advances in Physiology Education, 38*(3), 268–272. https://doi.org/10.1152/advan.00103.2013.

Latimier, A., Riegert, A., Peyre, H., Ly, S. T., Casati, R., & Ramus, F. (2019). Does pre-testing promote better retention than post-testing? *NPJ Science of Learning, 4*(1), 1–7. https://doi.org/10.1038/s41539-019-0053-1

Li, K., & Keller, J. M. (2018). Use of the ARCS model in education: A literature review. *Computers and Education, 122*, 54–62. https://doi.org/10.1016/j.compedu.2018.03.019

Litchfield, B. C., & Dempsey, J. V. (2015). Authentic Assessment of Knowledge, Skills, and Attitudes. *New Directions for Teaching and Learning, 2015*(142), 65–80. https://doi.org/10.1002/tl.20130

Lynam, S., & Cachia, M. (2018). Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education, 43*(2), 223-234.

Ma, L., & Lee, C. S. (2021). Evaluating the effectiveness of blended learning using the ARCS model. *Journal of Computer Assisted Learning, 37*(5), 1397–1408. https://doi.org/10.1111/jcal.12579

Marsh, R. (1984). A comparison of take-home versus in-class exams. *The Journal of Educational Research, 78*(2), 111-113.

Martínez-Huamán, E. L., Añazco, C. Q., Quispe, S. C., & Gutiérrez-Pérez. B. N. (2023). Teaching with Chunking in Synchronous Classes: The Influence on University Students' Intrinsic Motivation. *International Journal of Learning, Teaching and Educational Research, 22*(2), 377-391. https://doi.org/10.26803/ijlter.22.2.21

Mayo, J. 2002. Case-based instruction: A technique for increasing conceptual application in introductory psychology. *Journal of Constructivist Psychology 15*: 65–74.

Mayo, J. 2004. Using case-based instruction to bridge the gap between theory and practice in psychology of adjustment. *Journal of Constructivist Psychology 17*: 137–46.

Meyers, N. M., & Nulty, D. D. (2009). How to use (five) curriculum design principles to align authentic learning environments, assessment, students' approaches to thinking and learning outcomes. *Assessment and Evaluation in Higher Education, 34*(5), 565–577. https://doi.org/10.1080/02602930802226502

Moore, R., & Jensen, P. A. (2007). Do open-book exams impede long-term learning in introductory biology courses?. *Journal of College Science Teaching, 36*(7), 46.

Murillo, F. J., & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation, 53*, 10–16. https://doi.org/10.1016/j.stueduc.2017.01.001

Murillo, F. J., & Hidalgo, N. (2020). Fair student assessment: A phenomenographic study on teachers' conceptions. *Studies in Educational Evaluation, 65*, 100860-. https://doi.org/10.1016/j.stueduc.2020.100860

Nisbet, I. (2019). Fairness takes centre stage. *Assessment in Education Principles Policy and Practice, 26*(1), 111–117. https://doi.org/10.1080/0969594X.2017.1358151.

Nisbet, I., & Shaw, D. S. (2019). Fair assessment viewed through the lenses of measurement theory. *Assessment in Education Principles Policy and Practice, 26*(5), 612–629. https://doi.org/10.1080/0969594X.2019.1586643.

Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., ... & Roberts, T. (2011). Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical teacher, 33*(3), 206-214.

Olmos-Vega, F. M., Stalmeijer, R. E., Varpio, L., & Kahlke, R. (2023). A practical guide to reflexivity in qualitative research: AMEE Guide No. 149. *Medical Teacher, 45*(3), 241–251. https://doi.org/10.1080/0142159X.2022.2057287

Ontario Universities Council on Quality Assurance. (2010). *Quality Assurance Framework*. https://oucqa.ca/framework/introduction/

Pereira, D., Cadime, I., Brown, G., & Flores, M. A. (2022). How do undergraduates perceive the use of assessment? A study in higher education. *European Journal of Higher Education, 12*(1), 1-17.

Permzadian, V., & Cho, K. W. (2023). Does the format of an assessment (closed book or open book) affect learning? A systematic review of the literature. *Teaching in Higher Education*, 1-18.

Polack, C. W., & Miller, R. R. (2022). Testing Improves Performance as Well as Assesses Learning: A Review of the Testing Effect With Implications for Models of Learning. *Journal of Experimental Psychology. Animal Learning and Cognition, 48*(3), 222–241. https://doi.org/10.1037/xan0000323

Polat, M. (2020). Investigating the Backwash Effect of Higher Education Exam (YGS) on University Students' Attitudes. *International Journal of Psychology and Educational Studies, 7*(3), 152-. https://doi.org/10.17220/ijpes.2020.03.014

Rasooli, A., Zandi, H., & DeLuca, C. (2018). Re-conceptualizing classroom assessment fairness: A systematic meta-ethnography of assessment literature and beyond. *Studies in Educational Evaluation, 56*, 164–181. https://doi.org/10.1016/j. stueduc.2017.12.008

Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented analysis. Journal of educational psychology, 99(4), 734.

Rhodes, A., Wilson, A., & Rozell, T. (2020). Value of case-based learning within stem courses: Is it the method or is it the student? *CBE Life Sciences Education, 19*(3), 1–11. https://doi.org/10.1187/cbe.19-10-0200

Rich, J. D. (2011). An experimental study of differences in study habits and long-term retention rates between take-home and in-class examinations. *International journal of university teaching and faculty development, 2*(2), 121.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin, 138*(2), 353.

Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction, 49*, 142-156.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57*, 749–761.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology, 61*, Article 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Sabnis, S. V., & Wolgemuth, J. R. (2023). Reflexive disclosure in qualitative research publications in school psychology (2006–2021). *Psychology in the Schools, 60*(10), 3738–3753. https://doi.org/10.1002/pits.22965

Salehi, M., & Gholampour, S. (2021). Cheating on exams: Investigating Reasons, Attitudes, and the Role of Demographic Variables. *SAGE Open, 11*(2), 215824402110041-. https://doi.org/10.1177/21582440211004156

Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23*(4), 349–371. https://doi.org/10.1016/S0191-491X(97)86215-3

Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology, 109*(8), 1049–1066. https://doi.org/10.1037/edu0000190

Segers, M., Gijbels, D., & Thurlings, M. (2008). The relationship between students' perceptions of portfolio assessment practice and their approaches to learning. *Educational Studies, 34*, 35–44.

Simpson, J. A., & Weiner, E. S. C. (1989). *The Oxford English Dictionary*. (2nd ed. IX / prepared by J.A. Simpson and E.S.C. Weiner.). Oxford: Clarendon Press.

Slack, H. R., & Priestley, M. (2023). Online learning and assessment during the Covid-19 pandemic: exploring the impact on undergraduate student well-being. Assessment and *Evaluation in Higher Education, 48*(3), 333–349. https://doi.org/10.1080/02602938.2022.2076804

Spiegel, T., & Nivette, A. (2023). The relative impact of in-class closed-book versus take-home open-book examination type on academic performance, student knowledge retention and wellbeing. *Assessment and Evaluation in Higher Education, 48*(1), 27–40. https://doi.org/10.1080/02602938.2021.2016607

Steingut, R. R., Patall, E. A., & Trimble, S. S. (2017). The effect of rationale provision on motivation and performance outcomes: A meta-analysis. *Motivation Science, 3*(1), 19-50.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & evaluation in higher education, 30*(4), 325-341.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Tam, A. C. F. (2022). Students' perceptions of and learning practices in online timed take-home examinations during Covid-19. *Assessment & Evaluation in Higher Education, 47*(3), 477–492. https://doi.org/10.1080/02602938.2021.1928599

Theophilides, C., & Koutselini, M. (2000). Study behavior in the closed-book and the open-book examination: A comparative analysis. *Educational Research and Evaluation, 6*(4), 379-393.

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation, 43*, 55–69. https://doi.org/10.1016/j.stueduc.2013.12.003.

Tozoglu, D., Tozoglu, M. D., Gurses, A., & Dogar, C. (2004). The students' perceptions: Essay versus multiple-choice type exams. *Journal of Baltic Science Education, 2*(6), 52-59.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: the quality of motivation matters. *Journal of educational psychology, 101*(3), 671.

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment and Evaluation in Higher Education, 43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Way, K. A., Burrell, L., D'Allura, L., & Ashford-Rowe, K. (2021). Empirical investigation of authentic assessment theory: An application in online courses using mimetic simulation created in university learning management ecosystems. *Assessment and Evaluation in Higher Education, 46*(1), 17–35. https://doi.org/10.1080/02602938.2020.1740647

Wiggins, G. (1991). The case for authentic assessment. Practical Assessment, *Research & Evaluation, 2*(2).

Wijnia, L., Noordzij, G., Arends, L. R., Rikers, R. M. J. P., & Loyens, S. M. M. (2024). The Effects of Problem-Based, Project-Based, and Case-Based Learning on Students' Motivation: a Meta-Analysis. *Educational Psychology Review, 36*(1), 29-. https://doi.org/10.1007/s10648-024-09864-3

Worrell, F. (2016). *Commentary on perspectives in fair assessment*. In N. J. Dorans, L. Linda, & L. L. Cook (Eds.). Fairness in educational assessment and measurement (pp. 283–294). New York, NY: Routledge.

Yoo, M.-S., & Park, H.-R. (2015). Effects of case-based learning on communication skills, problem-solving ability, and learning motivation in nursing students. *Nursing & Health Sciences, 17*(2), 166–172. https://doi.org/10.1111/nhs.12151

Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. *The Journal of Educational Research, 80*(6), 352-358.

Appendices

Appendix A: Tables

Table 1

Participant Demographic Information (Study 1)

Participant	Major	Level of Study	Duration of Interview
P1	Psychology & Forensic Science	Third Year	78:58
P2	Nursing	Graduated (Four Years)	50:44
P3	Computer Science & Psychology	Third Year	45:33
P4	Psychology	Third Year	65:24
P5	Psychology	Third Year	50:06
P6	Biology & Psychology	Third Year	46:11
P7	Psychology	Third Year	58:27
P8	Biochemistry and Molecular Biology	Third Year	47:27
P9	Psychology	Third Year	41:07
P10	Psychology	Third Year	55:02

Note. Participants' level of education was determined through the number of course credits they said they had obtained at the time of the interview.

Table 2

Thematic Analysis Results (Study 1)

Theme	Subtheme
1. Purpose of Exams	Exams as Should Be
	Exams as Practiced
	Purpose tied to Program Goals
2. Exam Fairness and Effectiveness	Alignment with Learning Objectives and Expectations
	Effectiveness and Fairness in Exam Format
3. Affective Responses	Preparation Leading to Confidence
	Test Anxiety
	Grade Anxiety

Table 3

Representative Quotes for Theme 1: Purpose of Exams (Study 1)

Subtheme	Representative Quotes
Exams As Should Be	"[exams are] *supposed to test your comprehension. It's just sometimes it always depends on the class, but sometimes it just tests what you have memorized rather than what you understand."* (P8)
	"I find it nice when there's one question in there that's like, 'what did you gain from this course?' Because sometimes you really do learn things that, maybe not be as academic, but maybe more applicable to your life. For example… learning about a mental disorder that affects you personally or your family personally. It's obviously an academic gain, but also a personal gain as well." (P4)
	"*look like prioritizing learning over people's ability to retrieve information at a particular time… exploring different non-traditional methods of testing knowledge and making sure that people are actually learning.*" (P6)
Exams as Practiced	*"I think the on-paper focus of an exam is to display all of the knowledge that you've learned through the course in one spot. I feel like that is the bottleneck idea of what the final exam should be.*" (P4)
	"*Sometimes it's more of just, you know, recall like spitting facts… like, you know, when was this? Blah, blah blah. Where is this?*" (P10)
Purpose Tied to Program Goals	*"[Exams are] supposed to be the building block. Because obviously when you do a final exam to end the course, it doesn't stop there. Usually it's something that will lead into something else.*" (P7)
	"I know it's because of the lack of resources, and of course there's only one professor and a limited handful of TAs. So yeah, I'm well aware of like the situation. I know it's a lot, but I understand that too." (P7)
	"There's something to be said about testing things like how you handle information under pressure… under anxiety situations, what can you actually recall or what can you do? And again … it's kind of hitting back to that whole practical job level thing" … you have to make these snap decisions, or you have to do your job immediately. Don't think, just do, sort of thing. And I think the exams help with that as well." (P3)

Table 4

Representative Quotes for Theme 2: Exam Fairness and Effectiveness (Study 1)

Subtheme	Representative Quotes
Alignment with Learning Objectives and Expectations	
	"*As long as what you would said matches up with what shows up in the exam, I feel it's a fair exam. Because at the end of the day I'm a student, I'm expected to know and know the content. I'm supposed to study this content. That's why I'm here. So as long as you set your expectations for me that I can actually obtain them, and that when I go to the exam, I see what you would ask of me... I'm fine" (P7).*
	"*A lot of these guest speakers were over zoom… I guess leading up there were quite a few hiccups… stuff that the guest lecturers had talked about wasn't really on [the exam]"* (P1).
	"worrying that you didn't say something that you should that they wanted to know because I think you remember the final exam, it was like try to give as much as you can to show that you know the content". (P2)
	"as a student it's pretty clear when an exam is fair. Like say it's fair, but you do poorly, or unfair and you did poorly. Like it's pretty clear which, if it's fair or not, even if you did well or not" (P5).
	"*there are times when I'm writing an exam and I can kind of tell that if I studied better, I'd be able to do better on the test. I recognize that it's a fair exam. The questions aren't out of pocket… it's very much things that we have learned. So I'm able to recognize at that moment that, OK, this is me not studying properly" (P6).*
	"That's what's frustrating about it, cause they take a hard topic, you learn about it, you spend a lot of time to write about it perfectly, because they said it's gonna [sic] *be a long answer. And it's just not on there at all. And there is a question about something else. Would you spend less time studying because that question, which you studied for was supposed to be 15 points" (P8).*

"you're trying to figure out what to do in the question when you've never been presented a question even if you know how to do the question if it was worded differently. But it just adds an additional troubleshoot that is unnecessary and it's ineffective" (P4).

"I think that my professors have been pretty fair on, you know, what's on and what's not. I think some of them are kind of smarter in that they just give a blanket statement. So obviously it's not necessarily just a 'know everything'. It's like focus on XYZ" (P1).

"tests are way harder than what we learned in the class, and we just didn't have enough questions to, like, practice to. Like just five or like questions in the back of the chapter, I felt it wasn't enough". (P9)

"I think it's also kind of reflective on the professor and how well they taught. Because obviously if they're not very effective at teaching then their average won't be that high".

"[they] used test questions that students had submitted. It was for an assignment… if you saw your question on the exam you're like, 'oh, perfect'. Check it off. You obviously know the answer. Whereas I know my questions weren't on it. I didn't love that whole experience" (P1).

"*What makes an exam effective? Effective at what? Achieving its goal. What's its goal? What makes an exam effective? What is the goal of this exam we're talking about?*" (P2).

"I suppose the question needs to be justified. You know, why is the question being asked? Yeah, and If it's not really justified, then it shouldn't be on the exam" (P5).

"when I know it starts getting unfair is when I'm like I don't know the answer to this. I chose not to know the answer to this because I feel like this is not important to me in any way shape or form" (P2).

"If I felt good about it after… If I come out of an exam and I don't feel good about it. I have a tendency to blame it on the prof. Which is probably not fair, but I mean, a lot of the… I can't help but do that… I think that if I come out of an exam, I don't feel good about it, I just assumed that it was unfair if I studied everything and like felt like I understood everything and I was still sent with curveballs, I immediately think it was unfair." (P4)

Exam Performance is Not a Factor	*"as a student it's pretty clear when an exam is fair. Like say it's fair, but you do poorly, or unfair and you did poorly. Like it's pretty clear which, if it's fair or not, even if you did well or not" (P5).*
	"*there are times when I'm writing an exam and I can kind of tell that if I studied better, I'd be able to do better on the test. I recognize that it's a fair exam. The questions aren't out of pocket… it's very much things that we have learned. So I'm able to recognize at that moment that, OK, this is me not studying properly" (P6).*
	"If I felt good about it after… If I come out of an exam and I don't feel good about it. I have a tendency to blame it on the prof. Which is probably not fair, but I mean, a lot of the… I can't help but do that… I think that if I come out of an exam, I don't feel good about it, I just assumed that it was unfair if I studied everything and like felt like I understood everything and I was still sent with curveballs, I immediately think it was unfair." (P4)
Effectiveness and Fairness in Format	
Mixed Format	"*I feel like I can get the most from them [mixed format exams] or I can show the most from them.*" (P1)
	"*It [final reflect questions at the end of mixed-format exams] makes you really think about everything of what you learn, not just little bits and pieces like kind of everything throughout your 12 weeks all intertwines into your final thing.*" (P10)
Student Autonomy	"*Ideally [exams] should just be a way to reflect on the learning experience... You should be allowed to communicate that in your own way.*" (P2)
	"*I'll learn a lot about certain things, but I don't know what to do with that information if I'm not given opportunities to apply it… but I guess that's also my job as a student to figure out, OK, why is this important? Like, why am I learning this?*" (P4)
Recall Aids	"*It makes those other questions a little bit easier to answer because you're already kind of on a roll. So when the questions sort of start easier but less intense, I find that really effective*" (P4)

	"*Buzzwords are very triggering, so it could really help the students remember and I don't think an exam should aim towards not providing buzzwords because… why are you trying to block the recall process?"* (P6)
Universal Design for Learning	"*I have this one professor who uses what [they] calls a universal design, which I thought was a beautiful idea. None of the [student accessibility services] students actually have to leave the room… We all got a certain amount of time and that amount of time was enough where a student who may need extra time should still be able to answer the questions*." (P2)

Table 5

Representative Quotes for Theme 3: Affective Responses (Study 1)

Subtheme	Representative Quotes
Preparation Leading to Confidence	"*The best [exams] were probably the ones that I was just the most prepared for. Where I went in to just sit down in the gym and was able to go clean through the whole thing without having to stop to second guess… I knew the information. It was all good. I just had to put it on paper." (P3)*
	"*I do everything I can to avoid feeling [anxious]. I've never gone into an exam unprepared, and I've never gone into an exam without going over every single concept and making sure that I knew the important stuff or the main themes of every lecture, sort of thing." (P4)*
	"I had a class I was really prepared for and we also were allowed a cheat sheet… I had written the whole course material on this paper, open my final exam, I started flipping the pages and realized that it looked like I didn't know anything… It's frustrating. It was really sad because I expected to get a really good grade in this class." (P8)
Test Anxiety	"*I'll be totally prepared for an exam. I'll know everything … but I'm thinking about the person shaking their leg beside me, [or] somebody opening the door over there, [or] I can hear the clock ticking… those things completely take my attention away." (P4)*
	"*The door on the gym was super squeaky … And then also the person in the back corner of the gym decided to wear high heels to the exam and so when they were leaving it was just clicking. anyway" (P4)*.
	"Adding that two-hour/three-hour timer to everything just makes it that much more anxiety-inducing." (P3)
	"I think the difference is not so much the confidence, but, just kind of like whittling away at my perseverance that I can do this. 'I got this' just eventually becomes 'OK, yes, let's just get this over with.'" (P3)
	"*really enjoy my take-home exams... I sit at home, I write something, I can have a break whenever I want. It's not stressful. I'm not in a room with 300 other people sweating and being cold at the same time in the gym. The grad students are like hawks looking at you. Scary… it's also really cold in there." (P8)*
	"Worrying that you didn't say something that you should that they wanted to know because I think you remember the final exam, it was like try to give as much as you can to show that you know the content. And I'm like, oh, she has certain section she's looking for in this questions to demonstrate knowledge." (P2)
	"Sometimes because I'm anxious I tend to make a little checklist of little goals that I make throughout the exam… for example, long answer, sometimes I'll

	write point form on the actual paper of what topics I can talk about… I write down little points to remind myself…. If I go back I can continue writing." (P7) *"Really love writing in the same spot that I was taught or writing in the same environment that you were taught in. That really does help a lot." (P7)* *"When you're in the same room as when you learn it, it's easier. Your recall is a bit easier, however, that works" (P10).* *"If I sit in the front, I feel like I can kind of block out the other people, like hundreds of kids in the gym, if that makes sense. I can kind of try and zone out. But like, as soon as you walk in, there's the rows of desks kind of like a prison. That's what my friends and I kind of picture it. And there's, like, all these papers, and the teachers are just standing at the front, just staring at you, right? That setting definitely causes added stress" (P10).*
Grade Anxiety	*"I have one of those like the renewable scholarships. So if I get an 80 average for the year, I get $2500 for the next [year]. Which is a lot of money, especially when you're a broke [student]. So that is a really big incentive to me to bust my [butt] pretty much the whole year to get the 80 so it basically takes off a third of my tuition cost." (P10)* "*You can imagine how frustrating that is... you don't know what you're getting going into the exam. You can obviously speculate, but again you're waiting for marks... which is a big thing… if you get the [course work grades] before the exam, you can prepare better" (P7).* "*I know this usually stresses more people out than it's worth, but I always calculate what my grade is going into the exam and how much I need on the exam to get my desired grade… That helps me kind of calm down my thinking" (P10).*

Table 6

Descriptive Statistics for the Study Variables (Study 2)

	Valid *N*	*M*	*SD*	*z Skew*	*z Kurt*	*Min*	*Max*
IMMS Variables							
Attention	35	4.00	0.60	-0.83	-1.19	2.83	4.92
Relevance	35	4.00	0.57	-0.43	-0.54	2.67	4.89
Confidence	35	4.04	0.56	-2.22	0.49	2.67	5.00
Satisfaction	35	3.94	0.73	-0.20	-1.45	2.50	5.00
Total IMMS	35	4.00	0.54	-0.85	-1.01	2.92	4.83
PIQ Variables							
Anxiety	34	2.35	0.93	1.22	-0.39	1.00	4.67
Cursory Approach	34	2.15	0.67	0.61	-0.72	1.00	3.67
Divergent Thinking	34	4.08	0.75	-1.90	0.52	2.00	5.00
Understanding	34	4.22	0.67	-1.40	-0.46	2.60	5.00
Preparation	34	3.52	0.89	-0.57	-0.68	1.33	5.00
SRQ-A Variables							
Autonomous	34	4.00	0.71	0.09	-1.73	2.83	5.00
Controlled	34	4.02	0.58	-0.60	-0.71	2.71	5.00
Amotivation	34	1.21	0.47	6.35	8.28	1.00	3.00
Exam Grade	34	78.82	9.11	-3.59	4.06	48.00	92.00

Note. All scale scores are on a 1-5 rating scale; Exam Grade is on a 0-100 scale. **IMMS** = Instructional Materials Motivation Survey; **PIQ** = Perceived Impact Questionnaire; **SRQ-A** = Academic Self-Regulation Questionnaire.

Table 7

Pearson's (Below Diagonal) and Spearman's (Above Diagonal) Correlations among the Study Variables (Study 2)

		1	2	3	4	5	6	7	8	9	10	11
1.	**IMMS_A**	--										
2.	**IMMS_R**	.76**	--									
3.	**IMMS_C**	.56**	.57**	--								
4.	**IMMS_S**	.84**	.78**	.52*	--							
5.	**IMMS_TOT**	.91**	.89**	.75**	.92**	--						
6.	**PIQ_ANX**	-.33†	-.42*	-.30†	-.27†	-.38*	--					
7.	**PIQ_CUR**	-.22†	-.19	-.14	-.11	-.19	.04	--				
8.	**PIQ_DIV**	.56**	.56**	.41*	.62**	.64**	-.21†	-.28†	--			
9.	**PIQ_UND**	.50*	.65**	.35*	.53*	.59**	-.50*	-.26†	.48*	--		
10.	**PIQ_PREP**	.35*	.31†	.12	.30†	.32†	-.31†	-.15	.44*	.40*	--	
11.	**SRQ_AUT**	.72**	.76**	.30†	.81**	.77**	-.32†	-.30†	.56**	.56**	.42*	--
12.	**SRQ_CON**	-.05	.24†	.23†	.14	.16	-.07	.10	.12	.18	-.15	.15
13.	**SRQ_AMO**	-.55**	-.43*	-.30†	-.35*	-.47**	.19	.33†	-.40*	-.54**	-.38*	-.52**
14.	**Exam Grade**	.20†	.15	.55*	.27†	.33†	-.02	-.05	.20†	.01	-.09	.13

Note. N = 31 to 35 (pairwise). **IMMS** = Instructional Materials Motivation Survey; **IMMS_A** = Attention; **IMMS_R** = Relevance; **IMMS_C** = Confidence; **IMMS_S** = Satisfaction; **IMMS_TOT** = Total Score; **PIQ** = Perceived Impact Questionnaire; **PIQ_ANX** = Anxiety and Stress; **PIQ_CUR** = Cursory Approach; **PIQ_DIV** = Divergent Thinking; **PIQ_UND** = Understanding; **PIQ_PREP** = Preparation; **SRQ** = Academic Self-Regulation Questionnaire; **SRQ_AUT** = Autonomous Motivation; **SRQ_CON** = Controlled Motivation; **SRQ_AMO** = Amotivation. * $p < .05$. ** $p < .01$. † Non-trivial effect size (> .20).

Table 8

Descriptive Statistics and Inter-Correlations of the Extra IMMS Items (Study 2)

Abbreviated Item	*M*	*SD*	Frequency (*n*) 1 Not True	2 Slightly	3 Moderately	Mo
#37. Valid test of learning	4.43	0.92	0	2	4	
#41. Fair assessment of knowledge	4.40	0.95	0	2	5	
#38. Appropriate for learning objectives	4.34	0.87	0	1	6	
#43. Demonstrate strongest knowledge	4.14	1.09	1	2	6	
#39. Felt sufficiently challenged	4.00	1.14	2	1	7	1
#42. Took longer than anticipated	2.77	1.33	9	5	9	
#40. Prefer traditional memory-based test	1.83	1.36	24	2	2	
Spearman's Correlations	#37	#41	#38	#43	#39	#
#37. Valid test of learning	--					
#41. Fair assessment of knowledge	.75**	--				
#38. Appropriate for learning objectives	.77**	.69**	--			
#43. Demonstrate strongest knowledge	.80**	.73**	.79**	--		
#39. Felt sufficiently challenged	.54**	.54**	.64**	.42*	--	
#42. Took longer than anticipated	.01	-.20†	.10	.04	-.09	-
#40. Prefer traditional memory-based test	-.47†	-.45**	-.44†	-.49**	-.23	.0

Note. $N = 35$. Items are listed in order from most to least endorsed. **IMMS** = Instructional Materials Motivation Survey.

* $p < .05$. ** $p < .01$. † Non-trivial effect size (> .20)

Table 9

Spearman's Correlations of the Extra IMMS Variables with the Study Variables (Study 2)

	Valid/Fair	Challenge	Lengthy	Traditional
IMMS Attention	.64**	.36*	.05	-.51**
IMMS Relevance	.65**	.56**	-.09	-.44**
IMMS Confidence	.46**	.37*	-.38*	-.40**
IMMS Satisfaction	.66**	.49**	.10	-.46*
IMMS Total	.72**	.51**	-.05	-.55**
PIQ Anxiety	-.45**	-.26†	.08	.33†
PIQ Cursory Approach	-.22†	-.14	.30†	.35*
PIQ Divergent Thinking	.39*	.33†	-.06	-.35*
PIQ Understanding	.64**	.80**	-.07	-.28†
PIQ Preparation	.29†	.21†	.07	-.28†
SRQ-A Autonomous	.69**	.44*	.06	-.32†
SRQ-A Controlled	.17	.26†	.15	.23†
SRQ-A Amotivation	-.52**	-.34*	.07	.41*
Exam Grade	.36*	.21†	.03	.06

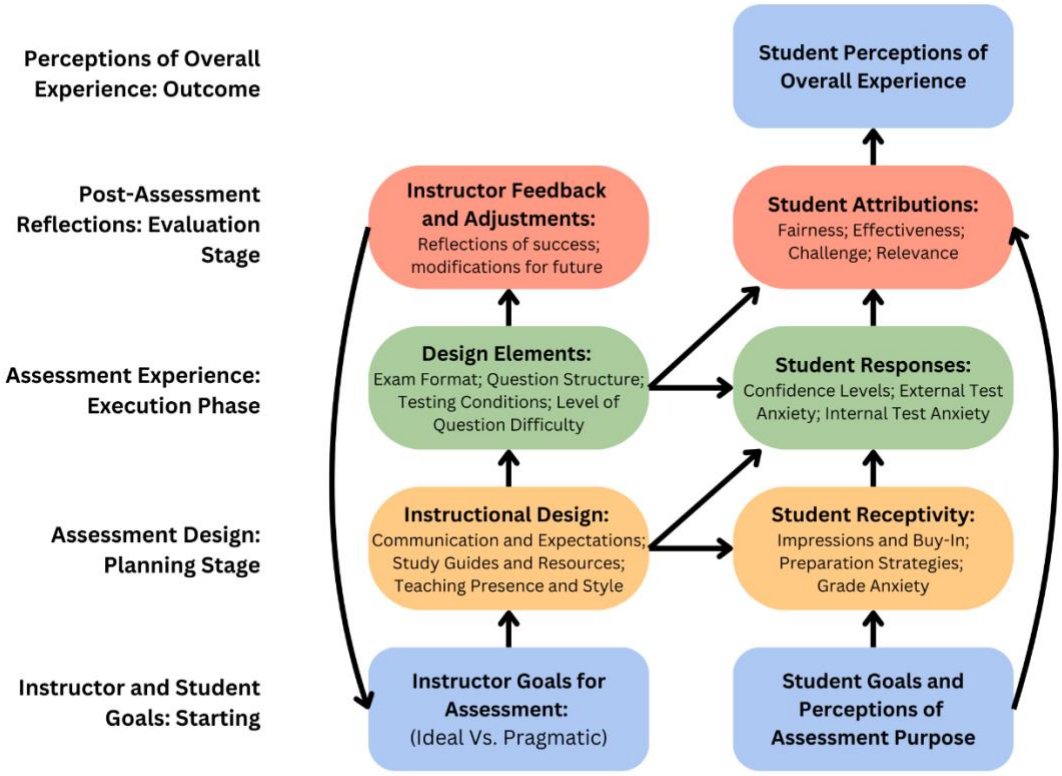
Note. N = 31 to 35 (pairwise). **IMMS** = Instructional Materials Motivation Survey; **PIQ** = Perceived Impact Questionnaire; **SRQ-A** = Academic Self-Regulation Questionnaire; **Valid/Fair** = Composite of items #37 (Valid test of learning), #41 (Fair assessment of knowledge), #38 (Appropriate for learning objectives), and #43 (Demonstrate strongest knowledge); **Challenge** = item #39 (Felt sufficiently challenged); **Lengthy** = item #42 (Took longer than anticipated); **Traditional** = item #40 (Prefer traditional memory-based test).

* $p < .05$. ** $p < .01$. † Non-trivial effect size (> .20)

Appendix B: Figures

Figure 1

Sources of Students' Perceptions of Exams: A Thematic Map (Study 1)



Note. The exam perceptions model outlines the assessment process from the students' perspective, divided into planning, execution, and evaluation stages. In the planning stage, instructors begin the assessment design and instructional design process, often balancing educational goals with practical constraints, while students approach assessments with their own expectations. During execution, the alignment between exam design and student preparation influences confidence and anxiety levels. In the evaluation stage, students reflect on the fairness and effectiveness of the assessment, with positive perceptions stemming from alignment with their goals, while misalignments can lead to feelings of unfairness and disengagement.

Appendix C: Study 1 Research Materials

LETTER OF INFORMATION

Project Title: Authentic Assessment: Phase II Interviews (REB #28469)

You are invited to participate in a study titled "Authentic Assessment: Phase II Interviews", conducted by Thomas Bodrug and Dr. Kateryna Keefer. The aim of this research is to understand students' perspectives, attitudes, and experiences with exams at university.

The information you provide will be used to build a richer understanding of the student perspective on exams, and to inform the design of exam assessments in university. The results of this study will be of value to instructors interested in improving the educational experience for their students. Participating in this study will also give you an opportunity to express your thoughts and opinions about taking exams at university. This is also an educational opportunity to experience some of the methods used in conducting educational research.

Student Principal Investigator:
Thomas Bodrug *
Graduate Student, Psychology M.Sc. Program
thomasbodrug@trentu.ca

Faculty Supervisor:
Dr. Kateryna Keefer
Associate Professor
katerynakeefer@trentu.ca

* Please direct all questions regarding this research and your participation to Thomas Bodrug, who is managing the data collection activities for this project.

PROCEDURES

If you agree to participate by clicking "Yes" at the bottom of this form, you will be sent an email inviting you to take part in a one-on-one Zoom interview with the Student Principal Investigator (Thomas Bodrug). The interview will be scheduled at a time that is convenient for you and will take approximately 60 minutes of your time. The purpose of the interview would be to chat about your opinions of university exams in general and to reflect on some of your previous exam experiences. The interview will be recorded, but the recording will then be deleted after an anonymized text transcript of the interview is created.

To show our appreciation for your participation, you will receive a $20 Amazon.ca eGift certificate. The certificate will be sent to the email address supplied at the bottom of this form within two weeks following the interview.

RISKS, PRIVACY, AND CONFIDENTIALITY

Participating in the interview will not involve any risk beyond what you, as an individual, may expect in daily life. Please know that your participation in this interview is entirely voluntary. You can refuse to participate without any penalty or negative consequences. You have the right to stop the interview at any time, skip any questions you do not wish to answer, request your comments to be redacted from the interview transcript, or withdraw your interview from the study; a request of partial or full withdrawal from this study should be sent to the Student Principal Investigator via email, no later than 48 hours after the interview has ended. Partial or full withdrawal from the interview after it has started will not disqualify you from receiving the eGift certificate. Participants wishing to withdraw from the study will have the right to determine what will happen to the interview data provided.

The Faculty Supervisor, Dr. Kateryna Keefer, will not know the names of those who participate in this study, and will have no knowledge of your individual interview responses. All interview transcripts will be anonymized by the Student Principal Investigator prior to sharing the dataset with Dr. Keefer. Neither your refusal to participate nor the answers you provide as part of this research will have any impact on your current or future interactions with Dr. Keefer.

Every effort will be made to maintain confidentiality of the information you provide for this study, which will be stored on encrypted computer files. No individual outside the research team will have access to the identifying information you provide. Interview recordings will be temporarily saved in a password protected file to the researcher's computer. The recordings will be deleted as soon as each interview is transcribed for analysis. The interview transcripts will be anonymized, and all identifying information that could make it possible to identify you as a participant will be removed from the transcripts. Each participant's transcript will be referenced using a unique participant ID code, with no names or emails linked to the IDs. All presentations of the data (e.g., conferences, publications, theses) will be in the form of common themes and anonymized quotes, with participants identified in generic terms (e.g., "participant one").

Please note that Zoom is an externally hosted cloud-based service. When information is transmitted over the internet absolute privacy cannot be guaranteed. There is always a risk your responses may be intercepted by a third party (e.g., government agencies, hackers). Further, while Trent University researchers will not use IP addresses or other information which could link your participation to your computer or electronic devices without informing you, there is a small risk with any platform such as this of data that are collected on external servers falling outside the control of the research team.

This study has been reviewed and received ethics clearance by the Trent University Research Ethics Board. If you have questions regarding your rights as a research participant, contact:

Anna Kisiala
Certifications and Regulatory Compliance Officer
Office of Research, Trent University
(705) 748 1011 ext. 7896
annakisiala@trentu.ca

CONSENT TO PARTICIPATE IN RESEARCH

I have read the information provided above, and I have had all questions answered to my satisfaction. It has been made clear to me that participation in this study is completely voluntary, that I have the right to refuse to answer any question or withdraw my participation, and that every participant who starts the interview will receive a $20 eGift certificate as an honorarium.

I understand that if I participate in the interview, I will have my video and voice recorded. I am aware that the video and audio files will be destroyed after transcripts have been created. I understand that all information collected for this study will be stored on encrypted computer files, that my personal and identifiable information will be kept confidential and separate from the interview transcripts, and that data from the interview will be used in an anonymized and de-identified format in reports, presentations, and journal articles.

I am aware that if, at any point during the research process, I have questions or concerns, or I wish to have my interview data destroyed, I may contact Thomas Bodrug (thomasbodrug@trenut.ca). I have adequate information to decide my participation in this study.

Consent to participate:

- □ Yes – I consent to participate in a videorecorded interview via Zoom
- □ No – I do not consent to participate in an interview

Full Name (please print): _____

Trent University Email Address (please print): _____

PLEASE RETAIN A COPY OF THIS INFORMED CONSENT FORM FOR YOUR RECORDS

Interview Protocol

Project Title: Authentic Assessment: Student Experiences and Learning Outcomes

Researchers: Kateryna Keefer (Faculty PI), Thomas Bodrug (Student PI)

Research Questions:

1. In students' own words, what is a fair and valid method to assess their learning?
2. What types of exams have students been exposed to, how do they evaluate different exam types, and what criteria are relevant to students when they evaluate exams?
3. What drivers motivate students' preferences for certain exam formats?
4. To what extent do emotions play a role in how students interpret and respond to exams?
5. To what extent do the aspects students value and prioritize in exams represent the learning outcomes outlined in the Undergraduate Degree Level Expectations?
6. What should educators know about the student perspective when designing university exams?

Points to Cover in Introduction:

- Thank interviewee for their participation.
- Interviewer introduction.
- The aim of this project is to examine students' experiences, attitudes, and perceptions of high-stakes summative exams, and its impacts on a range of learning outcomes, including motivation and engagement, higher-order thinking, learning approaches, and academic performance.
- Before we start, I would like to discuss confidentiality: Any identifying information you provide here will not be attached to the analysis or shared findings of this interview. What you tell me will not be associated with any identifying information.
    - If you would like, you can edit your name on Zoom to be a pseudonym that I can refer to you as that pseudonym.
    - Zoom is an externally hosted cloud-based service. There is a small risk with any platform such as this that data collected on external servers could fall outside of the researcher.
- If there is a question that you are uncomfortable with, feel free to tell me and we will skip it. Just respond "Skip!" if you want to skip a question. If you become uncomfortable with the interview, you are free to end it at any time and will still receive your eGift Certificate.
- Audio files will be destroyed after transcripts have been created.
- I will ask that if after the interview you do feel like you need to redact a comment or withdraw from the study entirely, that you send me an email within 48 hours after the interview ends and I will remove those responses.
- Before we begin the interview, can you please confirm your consent to participate and that you understand the purpose of this interview?
- Is it okay if I record?

Warm Up's

Why did you choose to pursue a university degree?

Can you tell me about your program?
Prompt: Tell me about: specializations, how classes are structured, content, instructors, course size, memorable projects, connections made.

What has been your favourite course to take at Trent and why?

Key Questions

What do you think the purpose of a university exam is?

In your opinion, what *should* be the purpose of a university exam?

What kind of exam formats are you familiar with?
Prompt: You mentioned X, did you consider Y as another type of exam?
(traditional/non-traditional)

Given these multiple ways to assess a students' knowledge, what do you think makes an exam effective?

How do you usually feel going into an exam?
Prompt: How do you prepare for an exam? Tell me about the study strategies you use.
Prompt: What goals do you have going into an exam?

How do you know when an exam is a fair and valid assessment of your learning?

Tell me about your best exam experience.
Prompt: Why did this exam stand out as a good experience?

Tell me about your worst exam experience.
Prompt: Why did this exam stand out as a poor experience?

Closing

What is one kind of exam you wish you had but never had?

If you could design your own ideal exam for one of the courses you are taking, what would this look like?
Prompt: You answered this question using a traditional sit-down exam format – how would you answer this question keeping non-traditional exams in mind?

Are there any last-minute thoughts you want to share that you would like to share?

E-Mail and Outreach Scripts

EMAIL: INVITATION TO PARTICIPATE

Subject: Research Participation Opportunity

Hi [Name of Potential Participant]!
Last year you participated in a research survey titled "Authentic Assessment: Student Experiences and Learning Outcomes", conducted by Thomas Bodrug and Dr. Kateryna Keefer. The purpose of our research is to understand student experiences with various exam formats, including the Take-Home Exam that you completed in PSYC 2300H – Introduction to Abnormal Psychology course.
On the survey consent form you had indicated your interest in being contacted with further information about this research. I am reaching out now as we are currently doing the second phase of our project, which involves follow-up interviews.
We were wondering – would you be interested in participating in a 60-minute interview over Zoom? The purpose of the interview would be to chat about university exams in general and to reflect on some of your previous exam experiences.
The information collected in these interviews will be used to understand the student perspective on exams and to inform the design of exam assessments in university. The results of this study will be of value to instructors interested in improving the educational experience for their students. Participating in the interview will also give you the opportunity to express your thoughts and opinions about taking exams at university.
To show our appreciation for participating in this interview you will receive a $20 Amazon.ca eGift Certificate as an honorarium, which will be emailed to you within two weeks following the interview. To learn more about the study and express interest in participating in an interview, please follow this link:
[QUALTRICS LINK TO LETTER OF INFORMED CONSENT]
If you have any questions or concerns, please contact me at thomasbodrug@trentu.ca
Thomas Bodrug
Student Principal Investigator
Graduate Student, Psychology M.Sc. Program

Dr. Kateryna Keefer
Project Supervisor
Associate Professor, Department of Psychology

EMAIL: INTERVIEW SCHEDULING

Subject: Research Interview Scheduling

Hi ,

Thank you for indicating your interest in participating in the interview phase of our research study titled “Authentic Assessment: Student Experiences and Learning Outcomes”. This email is to schedule the Zoom interview, which will take approximately 60 minutes. Please follow this Doodle link to sign up for a date and time that would work for you:

[DOODLE LINK TO SCHEDULE TIMES]
If none of these times work, please reply to this email, and we can find an alternative time that would work for both of us. Once the date and time have been scheduled, I will follow up with the Zoom link.

Thank you!

Thomas Bodrug
Graduate Student, Psychology M.Sc. Program
thomasbodrug@trentu.ca

EMAIL: ZOOM LINK

Subject: Research Interview Zoom Link

Hi ,

This is to confirm our research interview for **[Date, Time].** Please follow this link to connect:

[ZOOM LINK]
When you connect, you will first be directed to the virtual 'waiting room', before being admitted into the meeting room. This is a precaution to protect the privacy of our virtual meeting space. I was hoping we could both have our video turned on during the interview, for a more personable conversation, but I will leave this choice up to you.

I look forward to chatting with you!

Thomas Bodrug
Graduate Student, Psychology M.Sc. Program
thomasbodrug@trentu.ca

EMAIL: REMINDER INVITATION

Subject: Spots are still open! Research interview sign-up

Hi, [name of potential participant]
Last year when you completed a survey on your experiences with the Take-Home Exam in PSYC 2300H, you had indicated your interest in being contacted about a follow-up interview. I am reaching out now as we are currently doing interviews about students' experiences with university exams. If you are interested in participating in the interview but have not done so yet, the sign-up is still open and can be accessed at the following link:
[QUALTRICS LINK TO LETTER OF INFORMED CONSENT]
You will receive a $20 Amazon.ca eGift Certificate as a token of our appreciation for your time. If you have any questions or concerns, please contact me at thomasbodrug@trentu.ca

Thomas Bodrug
Student Principal Investigator
Graduate Student, Psychology M.Sc. Program

EMAIL: HONORARIUM

Subject: Honorarium for Interview Participation

Hi ,

This email is to follow-up on your participation in the interview portion of our research study titled “Authentic Assessment: Student Experiences and Learning Outcomes”, conducted by Thomas Bodrug and Dr. Kateryna Keefer.

In gratitude of your participation, **your $20 Amazon.ca eGift certificate is appended at the bottom of this email.**

Thank you again for your participation! Should you have any additional questions regarding this research project, please do not hesitate to contact me.

Best regards,

Thomas Bodrug
Graduate Student, Psychology M.Sc. Program
thomasbodrug@trentu.ca

Appendix D: Code Manual (Study 1)

Code	Code Description
Emotions, Cognitions Around Exams	Includes any reference to how students feel about the exam process, including positive and negative emotions. Also includes attitudes around exams in general, such as what learning outcomes exams test and what features they prefer in exams.
Format	Includes any reference exam formats (e.g., traditional and nontraditional).
Improvement	Refers to participant's ideas of how exams can improve.
Preparation and Process	Explicit reference to how participants prepare for exams, including the process leading up to exams (e.g., instructional design, teacher support).
Purpose, Effectiveness, Fairness	Includes any reference to how participants thought about the purpose of exams, which exams and features they find fair and effective.

Appendix E: Study 2 Research Materials

LETTER OF INFORMATION

Project title: Authentic Assessment: Student Experiences and Learning Outcomes

You have been invited to participate in a research study titled "Authentic Assessment: Student Experiences and Learning Outcomes", conducted by Thomas Bodrug and Dr. Kateryna Keefer. The purpose of this study is to understand student experiences with the final Take-Home Exam in **PSYC 2300H Introduction to Abnormal Psychology**, taught by Dr. Keefer in Winter 2022.

The main goal of this research is to examine students' perceptions of the Take-Home Exam and its impacts on a range of learning outcomes, including motivation and engagement, higher-order thinking, learning approaches, and academic performance.

The information collected for this study and the data you provide through your participation will be used to improve the quality of exam assessments used in future courses. The results of this study will have valuable benefits to future students taking this course, as well as to other instructors interested in improving the educational experience for their students. Participating in this research will also provide you with an opportunity to express your thoughts and opinions about the Take-Home Exam, and to gain insight into how educational research is conducted.

Upon completion of this study, we would be happy to provide you with a summary of the findings, if interested. To express your interest, or should you have any additional questions regarding the study or your participation, please do not hesitate to contact Thomas Bodrug.

Student Principal Investigator:*	**Faculty Principal Investigator:**
Thomas Bodrug	Dr. Kateryna Keefer
Graduate Student, Psychology M.Sc.	Associate Professor, Teaching Intensive
thomasbodrug@trentu.ca	katerynakeefer@trentu.ca

* Please direct all questions regarding this research and your participation to Thomas Bodrug, who is managing the data collection activities for this project.

PROCEDURES

If you agree to participate in this study, you will be asked to complete a brief online survey, which will take approximately 30 minutes of your time. This survey will ask specific questions about your academic background, your general habits as a learner, as well as both specific and open-ended questions about your experiences with the Take-Home Exam in PSYC 2300H.

To address the research questions pertaining to academic performance, we would like to ask your permission to access some of your grades in PSYC 2300H. Specifically, we are interested in the

grades you received on the Take-Home Exam, the Case Inquiries leading up to the exam, and your own self-assessments on the case inquiries. If you grant us permission to access these grades from the course records, this information will be matched with your survey responses. If you prefer not to disclose your grades to us, only your survey responses will be used in the study.

To show our appreciation for your participation, the name of each participant will be entered into a draw to win one of six electronic gift certificates from Amazon.ca, four valued at $25 and two valued at $50. Draw winners will receive the gift certificate via the email address supplied on the Participant Consent Form at the bottom of this document after all data has been collected.

After the data for this study has been processed, we would like to invite some of the participants for a follow-up interview study. The objective of the follow-up study will be to better understand students' unique and varied experiences with the Take-Home Exam. If you grant us permission to contact you at a later date, we will send you a new invitation email with the Letter of Information and Consent Form outlining the details of the follow-up interview study. At that time, you will have an opportunity to sign up for the interview should you choose to do so.

RISKS, PRIVACY, AND CONFIDENTIALITY

The proposed research will not involve any risk beyond what you, as an individual, may expect in daily life. Your participation in this study is entirely voluntary. You can refuse to participate without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time. You may also refuse to answer any questions you do not want to answer and still remain in the study.

Your former course instructor, Dr. Kateryna Keefer, will know that you participated in this study, but will have no knowledge of your individual survey responses. All survey responses will be anonymized by the Student Principal Investigator prior to sharing the dataset with Dr. Keefer. Neither your refusal to participate nor the answers you provide as part of this research will have any impact on your current or future relationship with Dr. Kateryna Keefer, or with other researchers conducting this study.

Every effort will be made to maintain confidentiality of the information you provide for this study, which will be stored on encrypted computer files. All data collected as part of this research will be anonymized after it is entered into the database. Your name and email address will not be attached to the data. Instead, each participant's data will be referenced using a unique participant code, with the master code key kept secure and separate from the data. All presentations of the data (e.g., conferences, publications, theses) will be in the form of group averages/aggregates and anonymized quotes, and will never include individual scores or any personal information that will make it possible to identify you as a participant. At no time will any individual outside of the immediate research team have access to the personal information that is provided.

This study has been reviewed and received ethics clearance by the Trent University Research Ethics Board. If you have questions regarding your rights as a research participant, contact:

Jamie Muckle
Certifications and Regulatory Compliance Officer
Office of Research, Trent University
(705) 748 1011 ext. 7896
jmuckle@trentu.ca

CONSENT TO PARTICIPATE IN RESEARCH

I have read the information about the research study provided above, and I have had all questions answered to my satisfaction. It has been made clear to me that participation in this study is completely voluntary, that my personal identifying information will be kept confidential and separate from the research data, that research results will be presented in an anonymized and aggregated format, that every participant will be entered into a prize draw, and that this will not impact my current or future relationships with the researchers, Dr. Kateryna Keefer and Thomas Bodrug.

I understand that I will complete a survey, which will ask about my academic background, my general habits as a learner, and my experiences with the Take-Home Exam in PSYC 2300H. I understand that, with my permission, my survey responses will be matched to my grades in PSYC 2300H, but I can still participate in the survey even if I do not give access to my grades.

I am also aware that if I have any questions or concerns, or to indicate my desire to withdraw from the study, I may contact Thomas Bodrug. I have sufficient information to make a decision about participating in this study.

I agree to participate in the survey:

- □ Yes
- □ No

I grant the researchers permission to:
(you may select all, some, or none of the options)

- □ a) access my Take-Home Exam grade in PSYC 2300H
- □ b) access my Case Inquiry grades in PSYC 2300H
- □ c) access my case inquiry self-assessments in PSYC 2300H

I grant the researchers permission to email me with more information about a follow-up interview study:

- □ Yes
- □ No

First and Last Name (please print): _____

Trent University Email Address (please print): _____

Please direct all questions regarding this research and your participation to Thomas Bodrug.

Student Principal Investigator:
Thomas Bodrug
Graduate Student, Psychology M.Sc.

thomasbodrug@trentu.ca

Faculty Principal Investigator:
Dr. Kateryna Keefer
Associate Professor, Teaching Intensive
katerynakeefer@trentu.ca

PLEASE RETAIN A COPY OF THIS CONSENT FORM

DEMOGRAPHICS FORM

Please provide the following information about yourself:

1. Gender:

☐ Female
☐ Male
☐ Prefer to self-describe: _____
☐ Prefer not to identify

2. Age (years): _____

3. Academic program/major(s): _____

4. Cumulative grade point average achieved to date:

☐ below 50%
☐ 50 to 59%
☐ 60 to 69%
☐ 70 to 79%
☐ 80 to 89%
☐ 90% or higher
☐ Other (please specify): _____

5. Total number of course credits completed to date (excluding courses currently in progress):

☐ under 5 credits
☐ 5 to 9.5 credits
☐ 10 to 14.5 credits
☐ 15 to 19.5 credits
☐ 20 or more credits
☐ Other (please specify): _____

6. Number of course credits you were enrolled in **during the Winter 2022 term**:

☐ 0.5 credit
☐ 1.0 credit
☐ 1.5 credits
☐ 2.0 credits
☐ 2.5 credits
☐ 3 or more credits
☐ Other (please specify): _____

7. How heavy do you feel your academic workload was **during the Winter 2022 term**?

5 = Too heavy
4 = Somewhat heavy
3 = Just right
2 = Somewhat light
1 = Too light

8. Approximately how many hours per week did you work **during the Winter 2022 term**? (please put zero if you did not work): _____

9. Reasons for taking **PSYC 2300H in the Winter 2022** (please check all that apply):

☐ Required course for my program/major
☐ Elective course for my program/major
☐ This course worked with my schedule
☐ Personal interest
☐ Other (please specify): _____

10. What was your final grade in **PSYC 2300H**?

☐ below 50%
☐ 50 to 59%
☐ 60 to 69%
☐ 70 to 79%
☐ 80 to 89%
☐ 90% or higher
☐ Other (please specify): _____

SECTION A - TAKE-HOME

We are interested in your experiences and thoughts about the final **Take-Home Exam in PSYC 2300H**, which was done in the format of an **applied case study** based on the case article "Surviving Anxiety" (Stossel, 2014). Your honest and specific feedback means a lot to us – it will help enhance the educational experience for future students.

1. Have you done a case-study exam in any of your other courses? (aside from PSYC 2300H)

☐ Yes, in one other course
☐ Yes, in more than one other course
☐ No, PSYC 2300H was my first course with a case-based exam

2. How helpful were the following aspects in preparing you for the final Take-Home Exam in PSYC 2300H?

(please circle one option for each question)

	Did not use / No experience	Not at all	Not much	Somewhat	Mostly	Very much
1. Lectures	n/a	1	2	3	4	5
2. Seminars	n/a	1	2	3	4	5
3. Course textbook	n/a	1	2	3	4	5
4. Case Inquiries	n/a	1	2	3	4	5
5. Inquiry self-assessments	n/a	1	2	3	4	5
6. Unit Tests	n/a	1	2	3	4	5
7. Q&A opportunities	n/a	1	2	3	4	5
8. Instructor support	n/a	1	2	3	4	5
9. Studying with peers	n/a	1	2	3	4	5
10. Online resources	n/a	1	2	3	4	5

3. Please share your thoughts about the final Take-Home Exam in PSYC 2300H:

SECTION B - IMMS

Please think about each of the following statements in relation to the **Take-Home Exam in PSYC 2300H** and indicate how true it is for you. Give the answer that truly applies to you, and not what you would like to be true, or what you think others want to hear.

(please circle one number for each question)

	Not true	Slightly true	Moderately true	Mostly true	Very true
1. When I first looked at the Take-Home exam, I had the impression that it would be easy for me.	1	2	3	4	5
2. There was something interesting about this exam that got my attention from the beginning.	1	2	3	4	5
3. The Take-Home exam was more difficult to understand than I would have liked it to be.	1	2	3	4	5
4. After reading the instructions, I felt confident that I knew what I was supposed to demonstrate on this exam.	1	2	3	4	5
5. Completing the Take-Home exam gave me a satisfying feeling of accomplishment.	1	2	3	4	5
6. It was clear to me how the content of this exam was related to things I already knew.	1	2	3	4	5
7. The exam materials had so much information that it was hard to pick out and remember the important points.	1	2	3	4	5
8. The case-study material was eye-catching.	1	2	3	4	5
	Not true	Slightly true	Moderately true	Mostly true	Very true
9. The exam had stories, pictures, or examples that showed me how this material could be important to some people.	1	2	3	4	5
10. Completing the Take-Home exam successfully was important to me.	1	2	3	4	5
11. The quality of the writing in the case study helped to hold my attention.	1	2	3	4	5
12. The Take-Home exam was so abstract that it was hard to keep my attention on it.	1	2	3	4	5
13. As I worked on the Take-Home exam, I was confident that I could master the content.	1	2	3	4	5
14. I enjoyed the Take-Home exam so much that I would like to know more about this subject.	1	2	3	4	5
15. The exam materials looked dry and unappealing.	1	2	3	4	5
16. The content of the Take-Home exam was relevant to my interests.	1	2	3	4	5
17. The way the information was arranged in the exam materials helped keep my attention.	1	2	3	4	5

SECTION B (continued)

	Not true	Slightly true	Moderately true	Mostly true	Very true
18. There were explanations or examples of how people might use the knowledge covered on this exam.	1	2	3	4	5
19. The Take-Home exam questions were too difficult.	1	2	3	4	5
20. The Take-Home exam had things that stimulated my curiosity.	1	2	3	4	5
21. I really enjoyed doing the Take-Home exam.	1	2	3	4	5
22. The amount of repetition in this exam caused me to get bored sometimes.	1	2	3	4	5
23. The content and style of this exam conveyed the impression that its content was worth knowing.	1	2	3	4	5
	Not true	Slightly true	Moderately true	Mostly true	Very true
24. I learned some things that were surprising or unexpected when doing the Take-Home exam.	1	2	3	4	5
25. After working on this exam for a while, I was confident that I would be able to pass it.	1	2	3	4	5
26. The content covered on this exam was not relevant to my needs because I already knew most of it.	1	2	3	4	5
27. The wording of messages after the exam, or of other comments about the exam, helped me feel rewarded for my effort.	1	2	3	4	5
28. The variety of reading passages, questions, illustrations, etc., helped keep my attention on the Take-Home exam.	1	2	3	4	5
29. The style of writing in the case study was boring.	1	2	3	4	5
	Not true	Slightly true	Moderately true	Mostly true	Very true
30. I could relate the content of the Take-Home exam to things I have seen, done, or thought about in my own life.	1	2	3	4	5
31. There were so many words in the exam materials that it was irritating.	1	2	3	4	5
32. It felt good to successfully complete the Take-Home exam.	1	2	3	4	5
33. The content tested on this exam will be useful to me.	1	2	3	4	5
34. I could not really understand quite a bit of the material on the Take-Home exam.	1	2	3	4	5
35. The good organization of this exam helped me be more confident that I would master this material.	1	2	3	4	5
36. It was a pleasure to work on such a well-designed exam.	1	2	3	4	5

SECTION B (continued)

	Not true	Slightly true	Moderately true	Mostly true	Very true
37. The Take-Home exam was a valid way to test my learning in this course.	1	2	3	4	5
38. The exam content was appropriate for the learning objectives of this course.	1	2	3	4	5
39. I felt sufficiently challenged by the Take-Home exam.	1	2	3	4	5
40. I would have preferred the final exam to be a traditional memory-based test, instead of the take-home case study.	1	2	3	4	5
41. The Take-Home exam was a fair assessment of my knowledge.	1	2	3	4	5
42. It took me longer to complete this exam than anticipated.	1	2	3	4	5
43. The Take-Home exam allowed me to demonstrate my strongest knowledge and skills.	1	2	3	4	5

SECTION C - PERCEIVED IMPACT

The following questions relate to the ways in which the **Take-Home Exam in PSYC 2300H** may have impacted you. Different students respond in different ways to such an assessment, and we want to know how true each of these impacts is for you.

(please circle one number for each question)

The Take-Home Exam prompted me to...	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
1. Engage actively with the course material.	1	2	3	4	5
2. Relate course concepts to real-life examples.	1	2	3	4	5
3. Discuss the topic with friends or family outside school.	1	2	3	4	5
4. Skip the course material I didn't quite understand.	1	2	3	4	5
5. Reflect on my own experiences, beliefs, or actions.	1	2	3	4	5
6. Do extra research beyond the assigned material.	1	2	3	4	5
7. Think originally and outside the box.	1	2	3	4	5
8. Seek clarification for concepts I didn't quite understand.	1	2	3	4	5
9. Wish I never had to do such an exam.	1	2	3	4	5
10. Read the textbook chapters thoroughly.	1	2	3	4	5
11. Consider alternative perspectives and points of view.	1	2	3	4	5
12. Realize that this subject was probably not for me.	1	2	3	4	5
13. Evaluate my areas of strengths and weaknesses.	1	2	3	4	5
The Take-Home Exam prompted me to...	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
14. Draw connections between different course concepts.	1	2	3	4	5
15. Think critically about the course material.	1	2	3	4	5
16. Stress over whether I was doing the task correctly.	1	2	3	4	5
17. Remember the course material better.	1	2	3	4	5
18. Identify gaps in my understanding of the material.	1	2	3	4	5
19. Put in extra effort towards the task.	1	2	3	4	5
20. Adopt one perspective and defend that point of view.	1	2	3	4	5
21. Deepen my interest in the subject matter.	1	2	3	4	5
22. Develop better study habits.	1	2	3	4	5
23. Study only those things that were asked on the exam.	1	2	3	4	5
24. Improve the overall quality of my academic work.	1	2	3	4	5
25. Articulate my ideas clearly and effectively.	1	2	3	4	5
26. Understand the course material in depth.	1	2	3	4	5

SECTION C (continued)

The Take-Home Exam prompted me to...	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
27. Think in new ways about people's behaviour.	1	2	3	4	5
28. Worry that I might do poorly on the exam.	1	2	3	4	5
29. Fulfill the exam requirements with minimal effort.	1	2	3	4	5
30. State some things without fully understanding them.	1	2	3	4	5
31. Apply what I was learning to my everyday life.	1	2	3	4	5
32. Feel uneasy and anxious.	1	2	3	4	5

SECTION D - MOTIVATION

The following questions relate to the reasons you may have had for putting time/effort into the **Take-Home Exam in PSYC 2300H**. Different students have different reasons for doing well on such an assessment, and we want to know how true each of these reasons is for you.

(please circle one number for each question)

I put time/effort into the Take-Home Exam...	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
1. Because the exam material was interesting to me.	1	2	3	4	5
2. Because I wanted the professor to think well of me as a student.	1	2	3	4	5
3. Because I wanted to get a good grade.	1	2	3	4	5
4. Because I found the exam material to be useful for me.	1	2	3	4	5
5. Because I wanted to make a good impression on my family or friends.	1	2	3	4	5
6. Because I enjoyed doing the exam.	1	2	3	4	5
7. Because otherwise I would risk losing grades.	1	2	3	4	5
8. Because it is important to me to apply myself in my studies.	1	2	3	4	5
I put time/effort into the Take-Home Exam...	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
9. Because the exam was required for the course.	1	2	3	4	5
10. Because I wanted to learn the course material.	1	2	3	4	5
11. Because the exam was fun to do.	1	2	3	4	5
12. Because otherwise I would feel bad about myself.	1	2	3	4	5
13. Because it made me feel proud of myself.	1	2	3	4	5
14. I don't know, I didn't really see the point to the exam.	1	2	3	4	5
15. I didn't, because I felt the exam was a waste of my time.	1	2	3	4	5
16. I didn't, because I didn't think the exam was worth the effort.	1	2	3	4	5
17. I didn't, because of competing priorities from other coursework.	1	2	3	4	5

SECTION E – BELIEFS ABOUT INTELLIGENCE

The following questions relate to your **ideas about intelligence**. There are no right or wrong answers. We are just interested in your views. Please indicate the extent to which you agree or disagree with the following statements.

(please circle one number for each question)

	Strongly disagree	Disagree	Mostly disagree	Mostly agree	Agree	Strongly agree
1. My intelligence is something about me that I personally can't change very much.	1	2	3	4	5	6
2. With enough time and effort I think I could significantly improve my intelligence level.	1	2	3	4	5	6

SECTION F – MAI

The following questions relate to **you as a learner**. Please rate each statement on how true it is of you in general when you are in the role of a learner (student, attending classes, etc.).

(please circle one number for each question)

	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
1. I ask myself periodically if I am meeting my goals.	1	2	3	4	5
2. I consider several alternatives to a problem before I answer.	1	2	3	4	5
3. I try to use strategies that have worked in the past.	1	2	3	4	5
4. I pace myself while learning in order to have enough time.	1	2	3	4	5
5. I understand my intellectual strengths and weaknesses.	1	2	3	4	5
6. I think about what I really need to learn before I begin a task.	1	2	3	4	5
7. I know how well I did once I finish a test.	1	2	3	4	5
8. I set specific goals before I begin a task.	1	2	3	4	5
9. I slow down when I encounter important information.	1	2	3	4	5
10. I know what kind of information is most important to learn.	1	2	3	4	5
11. I ask myself if I have considered all options when solving a problem.	1	2	3	4	5
	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
12. I am good at organizing information.	1	2	3	4	5
13. I consciously focus my attention on important information.	1	2	3	4	5
14. I have a specific purpose for each strategy I use.	1	2	3	4	5
15. I learn best when I know something about the topic.	1	2	3	4	5
16. I know what the teacher expects me to learn.	1	2	3	4	5
17. I am good at remembering information.	1	2	3	4	5
18. I use different learning strategies depending on the situation.	1	2	3	4	5
19. I ask myself if there was an easier way to do things after I finish a task.	1	2	3	4	5
20. I have control over how well I learn.	1	2	3	4	5
21. I periodically review to help me understand important relationships.	1	2	3	4	5

SECTION F (continued)

	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
22. I ask myself questions about the material before I begin.	1	2	3	4	5
23. I think of several ways to solve a problem and choose the best one.	1	2	3	4	5
24. I summarize what I've learned after I finish.	1	2	3	4	5
25. I ask others for help when I don't understand something.	1	2	3	4	5
26. I can motivate myself to learn when I need to.	1	2	3	4	5
27. I am aware of what strategies I use when I study.	1	2	3	4	5
28. I find myself analyzing the usefulness of strategies while I study.	1	2	3	4	5
29. I use my intellectual strengths to compensate for my weaknesses.	1	2	3	4	5
30. I focus on the meaning and significance of new information.	1	2	3	4	5
31. I create my own examples to make information more meaningful.	1	2	3	4	5
32. I am a good judge of how well I understand something.	1	2	3	4	5
	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
33. I find myself using helpful learning strategies automatically.	1	2	3	4	5
34. I find myself pausing regularly to check my comprehension.	1	2	3	4	5
35. I know when each strategy I use will be most effective.	1	2	3	4	5
36. I ask myself how well I accomplish my goals once I'm finished.	1	2	3	4	5
37. I draw pictures or diagrams to help me understand while learning.	1	2	3	4	5
38. I ask myself if I have considered all options after I solve a problem.	1	2	3	4	5
39. I try to translate new information into my own words.	1	2	3	4	5
40. I change strategies when I fail to understand.	1	2	3	4	5
41. I use the organizational structure of the text to help me learn.	1	2	3	4	5
42. I read instructions carefully before I begin a task.	1	2	3	4	5
43. I ask myself if what I'm reading is related to what I already know.	1	2	3	4	5

SECTION F (continued)

	Not at all true	Mostly not true	Somewhat true	Mostly true	Very true
44. I reevaluate my assumptions when I get confused.	1	2	3	4	5
45. I organize my time to best accomplish my goals.	1	2	3	4	5
46. I learn more when I am interested in the topic.	1	2	3	4	5
47. I try to break studying down into smaller steps.	1	2	3	4	5
48. I focus on overall meaning rather than specifics.	1	2	3	4	5
49. I ask myself questions about how well I am doing while I am learning something new.	1	2	3	4	5
50. I ask myself if I learned as much as I could have once I finish a task.	1	2	3	4	5
51. I stop and go back over new information that is not clear.	1	2	3	4	5
52. I stop and re-read when I get confused.	1	2	3	4	5

THANK YOU FOR YOUR PARTICIPATION!

Email and Outreach Scripts

EMAIL INVITATION

Subject: PSYC 2300H Take-Home Exam: Looking for your feedback

Hello!

You are invited to participate in a research survey about the Take-Home Exam that you completed as part of your PSYC 2300H - Introduction to Abnormal Psychology course in Winter 2022. The purpose of this research study, which I am conducting with Dr. Kateryna Keefer, is to understand students' experiences with the Take-Home Exam.

We would really appreciate your feedback about the Take-Home Exam. The information collected for this study will be used to improve the quality of exam assessments used in future courses. The results of this study will have valuable benefits to future students taking this course, as well as to other instructors interested in improving the educational experience for their students. This will also give you the opportunity to express your thoughts and opinions about the exam.

Your former course instructor, Dr. Kateryna Keefer, will know that you participated in this study, but will have no knowledge of your individual survey responses. I will anonymize all survey responses prior to transferring the dataset to Dr. Keefer.

To show our appreciation for your participation, the name of each participant will be entered into a draw to win one of six electronic gift certificates from Amazon.ca, four valued at $25 and two valued at $50. Draw winners will receive the gift certificate via the email address supplied on the Participant Consent Form after all data has been collected.

To learn more about the study and to participate, please follow this link:
[SURVEY LINK]

If you have any questions or concerns, please contact me at thomasbodrug@trentu.ca

Thank you for your time!

Thomas Bodrug
Student Principal Investigator
Graduate Student, Psychology M.Sc. Program

EMAIL REMINDER

Subject: PSYC 2300H Take-Home Exam: Last chance to win the draw

Hello!

Earlier this summer you were invited to participate in a research survey about the Take-Home Exam that you completed as part of your PSYC 2300H - Introduction to Abnormal Psychology course in Winter 2022. If you are interested in providing feedback about the Take-Home Exam but have not done so yet, the survey is still open and can be accessed at the following link: [SURVEY LINK]

If you complete the survey by July 31, 2022, your name will be entered into a draw to win one of six electronic gift certificates from Amazon.ca (four valued at $25 and two valued at $50), as a token of our appreciation for your time.

If you have any questions or concerns, please contact me at thomasbodrug@trentu.ca

Thomas Bodrug
Student Principal Investigator
Graduate Student, Psychology M.Sc. Program

EMAIL DRAW WINNER NOTIFICATION

Subject: Draw prize for participating in research study

Dear ,

This email is to follow-up on your participation in the research study titled "Authentic Assessment: Student Experiences and Learning Outcomes", conducted by Thomas Bodrug and Dr. Kateryna Keefer. You completed the study questionnaires earlier this summer.

In gratitude for your participation, your name was entered into a draw to win one of six electronic gift certificates from Amazon.ca, four valued at $25 and two valued at $50. I am happy to inform you that you were randomly selected as one of the prize winners.

Your gift certificate is appended at the bottom of this email.

Thank you again for your participation. Should you have any additional questions regarding this research project or the prize draw, please do not hesitate to contact us.

Best regards,

Thomas Bodrug
Graduate Student, Psychology M.Sc. Program
thomasbodrug@trentu.ca

Dr. Kateryna Keefer
Associate Professor, Teaching Intensive
Department of Psychology
katerynakeefer@trentu.ca

Appendix F: Code Manual (Study 2)

Code	Code Description
Preparation	Includes any reference to preparation materials prior to the exam (CBL activities, studying, etc.).
Resources/ Resource Presence	Includes any reference to resources utilized during the exam (textbook, instructor presence, Q and A, etc.).
Perceived Utility	Refers to how the students believed the exam was useful or beneficial to their learning, as well as the factors that may have influenced their learning experience (perceived takeaways) - Broken down into learning outcomes and technical aspects of the exam.
Comparison	Explicit reference to assessment/exam alternatives (e.g., traditional exams, m/c tests), in comparison to the case-based exam.

Appendix G: Principal Components Analysis of the PIQ Items (Study 2)

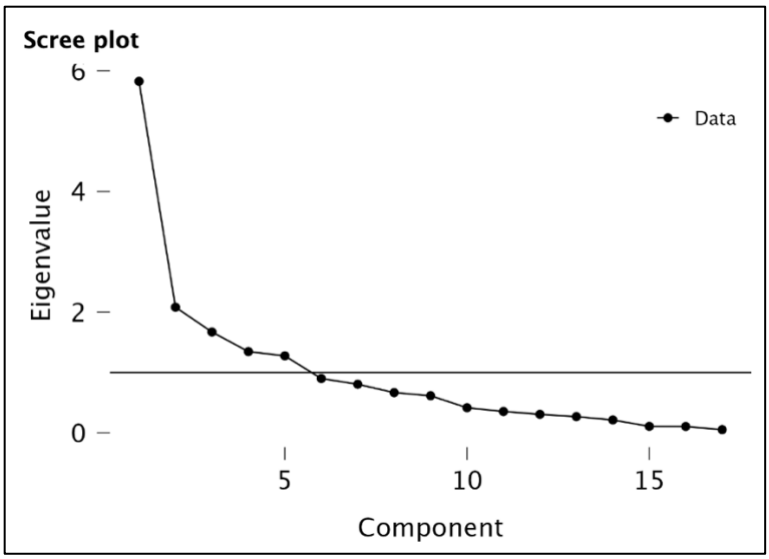
Based on initial content review of the 32 PIQ items, five items were eliminated from the pool because they had too much conceptual overlap with the SRQ-A measure of motivation: #1 ("Engage actively with the course material"), #9 ("Wish I never had to do such an exam"), #12 ("Realize that this subject was probably not for me"), #19 ("Put in extra effort towards the task"), and #21 ("Deepen my interest in the subject matter").

Using the remaining 27 items as the starting pool, a series of iterative Principal Components Analyses (PCAs) were conducted to identify groups of items that could be combined into subscales. At each iterative step, the Scree plot and Eigenvalues were examined to determine the number of components to extract, and varimax-rotated solutions were examined for item loadings, cross-loadings, and interpretability of the extracted components. Items that did not load substantively (.40 or higher) on any component were removed from subsequent iterations. Items that cross-loaded onto multiple components without a clear marker loading were also removed from subsequent iterations. Only components with three or more marker loadings were retained.

As a result of this iterative process, the pool was reduced to 17 items that loaded onto five interpretable components, each of which explained at least 10% of variance, for a cumulative 71.8% of variance. The item composition of the five resulting PIQ subscales is presented in Table B1, followed by the PCA outputs.

Subscales of the Perceived Impact Questionnaire (Study 1)

Subscale	Items
Understanding	8. Seek clarification for concepts I didn't quite understand. 15. Think critically about the course material. 18. Identify gaps in my understanding of the material. 26. Understand the course material in depth. 27. Think in new ways about people's behaviour.
Divergent Thinking	5. Reflect on my own experiences, beliefs, or actions. 7. Think originally and outside the box. 11. Consider alternative perspectives and points of view.
Anxiety	16. Stress over whether I was doing the task correctly. 28. Worry that I might do poorly on the exam. 32. Feel uneasy and anxious.
Preparation	6. Do extra research beyond the assigned material. 10. Read the textbook chapters thoroughly. 23. Study only those things that were asked on the exam. (R)
Cursory Approach	20. Adopt one perspective and defend that point of view. 29. Fulfill the exam requirements with minimal effort. 30. State some things without fully understanding them.



Component Characteristics

	Unrotated solution			Rotated solution		
	Eigenvalue	Proportion var.	Cumulative	SumSq. Loadings	Proportion var.	Cumulative
Component 1	5.826	0.343	0.343	3.698	0.218	0.218
Component 2	2.082	0.122	0.465	2.341	0.138	0.355
Component 3	1.671	0.098	0.563	2.299	0.135	0.491
Component 4	1.346	0.079	0.643	2.067	0.122	0.612
Component 5	1.274	0.075	0.718	1.793	0.105	0.718

Component Loadings

	PC1	PC2	PC3	PC4	PC5	Uniqueness
PI_5		0.914				0.149
PI_6		0.505		0.566		0.337
PI_7	0.567	0.586				0.281
PI_8	0.839					0.173
PI_10				0.796		0.248
PI_11		0.718				0.280
PI_15	0.781					0.269
PI_16			0.671			0.467
PI_18	0.795					0.255
PI_20					0.609	0.524
PI_23				−0.760		0.302
PI_26	0.724					0.238
PI_27	0.690					0.361
PI_28			0.827			0.245
PI_29					0.824	0.211
PI_30					0.605	0.281
PI_32			0.766			0.182

Note. Applied rotation method is varimax.