

University of Alberta

Shifting Summative Final Examination Policies and Teacher Assessment of Student
Achievement

by

Darren Phelps

A capping exercise submitted to the Faculty of Education
In partial fulfillment of the requirements for the degree of

Master of Education
in
Educational Administration and Leadership

Faculty of Education

Edmonton, Alberta

August 11, 2022

Abstract

The purpose of this study was to better understand how continually changing summative final examination policies are impacting teachers' ability to assess the academic achievement of their students. Four experienced secondary level teachers familiar with the implementation of high stakes summative examinations served as data sources in this basic qualitative study. Data was collected through semi-structured interviews and analyzed using the constant comparative method. Two categories of data emerged, one emphasizing teachercentric variables and the other student learner attributes (SLA's). Shifting examination policies had a toppling effect on the SLA's of students due to a decrease in the influence high stakes examinations had on their academic value judgements and a resultant decrease in levels of student self-efficacy. Changing policies led to inconsistencies in the deployment of formative and summative assessment strategies by teachers. They had had an influence on teachers' ability to assess the academic achievement of students by affecting their ability to accurately plan for the instructional pacing of their classes which in turn changed how teachers approached the delivery of curriculum objectives. These factors required teachers to hastily adjust their assessment plans creating further inconsistencies among educators. Recommendations for the future involve the improvement of student motivation and self-efficacy through increasing levels of teacher developmental leadership behaviours. Districts are also encouraged to adopt longer-term policy approaches to the implementation of summative final examinations while school leaders should engage in a renewed focus towards building assessment cultures in their schools.

Table of Contents

Shifting Summative Final Examination Policies and Teacher Assessment of Student	
Achievement	4
Researcher Beliefs	5
Review of the Literature	6
Methods.....	7
Design.....	8
Respondent Group.....	9
Data Collection.....	11
Trustworthiness	12
Limitations	13
Data Analysis	14
Findings and Discussion	15
Student Learner Attributes	16
Perceptions of Purpose for Final Examinations	20
Informing Teacher Understanding of Achievement.....	22
Ability to Assess Student Achievement	24
Summary	26
Conclusions and Recommendations	28
Recommendations	31
References	34
Appendix A.....	41
Appendix B	43
Appendix C.....	44
Appendix D.....	46

Shifting Summative Final Examination Policies and Teacher Assessment of Student Achievement

During the past two years, schools have been forced to adjust their end of term summative final examination assessment regimes in response to the evolving COVID-19 educational landscape. These adjustments have taken several forms within the Province of Alberta and more specifically the Peace River School Division, such as: developing teacher created examinations to replace cancelled departmental examinations, decreasing examination weightings, cancelling all end of term summative final examinations outright, and most recently, not allowing for some examinations to count towards a grade while allowing others to continue to do so. As of March 2022, a coherent long-term strategy for the implementation of an assessment regime involving summative final examinations has yet to be developed neither provincially nor at the district level. Consequently, teachers are still expected to develop and execute their courses with the anticipation that a final examination will be used as a measure of student academic achievement and in determining a significant proportion of their students' overall grades.

The ongoing uncertainty of whether teachers will be able to use final examinations as key assessment metrics may be causing some to question traditional assessment regimes all while considering new approaches to assess student academic achievement. Further, the ongoing educational disruptiveness caused by COVID-19 may be affording an opportunity for teachers to explore new ideas relating to assessment without the fear of reprisal typically derived from examination-based accountability.

The purpose of this study was to better understand how continually changing summative final examination policies are impacting teachers' ability to assess the academic achievement of their students. The study is guided by the following questions:

1. What are teacher's perceptions of the purpose for administering summative final examinations?
2. In what ways do summative final examinations inform teacher understanding of student academic achievement?
3. Given the cancellation of summative examinations, what impacts do teachers perceive on their abilities to assess student achievement?

Researcher Beliefs

Inconsistent assessment practices have impacted me as a practicing classroom teacher. When considering these questions, my personal beliefs are that COVID-19 is acting as a significant driver of change with respect to many educational practices. As new waves of COVID continue to evolve, we must place greater emphasis on establishing long-term sustainable methods allowing teachers to assess the academic achievement of their students. Traditional assessment regimes culminating in high-stakes final examinations do not seem to be one such sustainable practice. With respect to the courses I currently teach, I am open to exploring new approaches towards end of term summative assessment largely due to a growing sense our final examinations are a vestige of the past intended to satisfy a variety of forces, many of which lie outside of the classroom setting. I am aware these sentiments may introduce bias and impact the confirmability of this qualitative study hence my rationale for stating them overtly here.

Review of the Literature

“Assessment in the context of education is a process of deciding, collecting and reasoning from evidence about learners’ knowledge and skills” (Harlen, 2005, p. 263). Tools to collect assessment evidence can take many forms yet high-stakes examinations have become a predominant component of many North American assessment regimes since the mid-20th century (Moses & Nanna, 2007). High-stakes examinations are those which impact the opportunities and options available to learners such as their ability to progress to a higher grade level, the receipt of credentials, or access to post-secondary domains (Cairns, 2021).

Researchers cite many potential advantages when using high-stakes examinations to assess academic achievement. Lemann (1999) suggests the original intent of these examinations was to promote equality of opportunity by offering a uniform and objective means of assessing student knowledge and ability. Nichols et al. (2005) indicate high-stakes examinations work to motivate students and educators alike. Their deployment results in a collective effort towards improved academic achievement, largely due to the drawbacks associated with failure. Testing advocates also suggest properly developed examinations are aligned with curriculum and can therefore inform both educators and educational stakeholders as to whether students are learning the material they are supposed to know (Agrey, 2004). Further, high-stakes examinations embody considerable administrative utility for post-secondary institutions seeking to make candidate selections (Moses & Nanna, 2007). Success on these examinations, therefore, is used as a predictor of future success with respect to both intelligence and graduation potential (Thacker & Williams, 1974).

Opponents of high-stakes summative examinations have expressed a variety of concerns. Cairns (2021) cited their inflexibility for use as an assessment tool during times of disruption

such as those experienced during COVID-19. Others provide warnings about the power testing regimes have over “what we learn and how we learn” (Klenowski, 2012, p. 178). Suggesting their ability to narrow curriculum, constrain teacher practice, and disproportionately influence decisions relating to instructional pedagogy, an effect known as *examination washback* (Spratt, 2005).

Equity concerns are cited by various authors. Froese-Germain (2001) states test driven instruction impacts poor students most. Drill and practice modes of teaching within these populations tend to increase in frequency with an effort to raise achievement scores. Time for this is frequently pulled from supplementary programs such as art, drama, and music. Programs those students cannot afford to engage in outside of the school setting. In some situations, low achieving students are placed in special education programs to avoid having their results reflected in school accountability metrics. These students may also be refused access to numerous educational opportunities through efforts to engineer student populations (Darling-Hammond, 1991). High-stakes examinations have been used to label schools as underperforming. This label encourages many high caliber educational practitioners to look for different work placements. Further perpetuating the learning gaps faced by many schools in low socioeconomic areas (Froese-Germain, 2001). High-stakes examinations have an impact on students and their future. Understanding how teachers are adjusting to inconsistencies in their deployment therefore has merit.

Methods

This basic qualitative study utilized an action oriented theoretical framework (Borton, 1970). Purposeful convenience sampling was used to target several teacher respondents for data collection via online interviews. Member checks followed each interview, and the combined

data were coded allowing for category construction and further analysis. Two broad categories of data emerged. One I considered to contain teachercentric variables and the other student focused variables specific to their learner attributes. These findings were then combined and modeled as described later in this section.

Design

Borton's (1970) model of reflection is the theoretical framework underpinning this study. The model provides three questions used to guide the reflective process in relation to events and experiences. It asks: *What? So What? and Now What?* The *What* stage combines the reflective process with the identification of experiences and their detailed descriptions (Smith et al., 2016). The *So What* stage tries to evaluate and analyze what has transpired to create deeper levels of understanding. Finally, the *Now What* acts as a synthesis of the previous stages, having the researcher consider next steps. It includes exploring alternative courses of action and future opportunities. Using Borton's framework is advantageous here due to its applicability across many disciplines and its relevance to the field of education. Further, it is orientated towards future action. This pilot study introduces research which may support decision making related to final examination policies. The framework's action orientation was a key consideration in its selection.

The study is categorized as a basic qualitative study (Merriam & Tisdell, 2016). This design is appropriate for educational research involving teacher reflection as it seeks to uncover the participants experiences. Further, it endeavours to understand the meaning a participant ascribes to those experiences. Kahlke (2014) discussed two sub-categories of basic qualitative study designs. These are referred to as qualitative description and interpretive description. Interpretive description, used here, draws on a candidate's experiences to help learn from them.

These learnings are then redeployed within the original practice setting (Thorne et al., 2004).

“Interpretive description focuses on developing research questions from practice and providing theoretically and methodologically sound evidence that can then be used in the practice setting” (Kahlke, 2014, p. 41). This context specific design is appropriate for disciplines closely linked to the settings of their practice (Hunt, 2009), education being one such example.

Respondent Group

Teachers formed the entirety of the respondent group as I was seeking to understand how teachers have interpreted and responded to recent experiences. Purposeful convenience sampling was used to identify participants as it allowed me to non-randomly select individuals who have experienced the phenomena of interest (Palinkas, et al. 2015). Participants required more than three years of teaching experience. This was an important variable in candidate selection as they needed the ability to draw upon their teaching experiences prior to the pandemic. Participants also needed to teach two or more courses at the secondary level to ensure their familiarity with high-stakes final examinations. Examinations were characterized as high-stakes when their weightings comprised at least 30% of the overall course mark.

Four teachers, meeting the selection criteria, were approached to participate in the study. They were between the ages of 33 and 58 years old and their teaching experience ranged from 9 to 33 years in the classroom. The bulk of these teaching experiences occurred while working in the Peace River School Division (PRSD). A rural district comprised of twenty schools operating in communities characterized primarily by the industries of agriculture and oil and gas operations. PRSD schools have relatively small student populations ranging between thirty and three hundred students. Also, two of the teacher interviewees taught internationally within Asian settings for a brief time. These teachers were intentionally contacted through an in-person

conversation instead of a generic introductory email. The initial face-to-face discussion was important as I was able to provide a brief overview of the study's aims and gauge their interest in participating. Teachers tend to be highly collaborative by nature and will often accept such requests from their colleagues, even when they are not generally interested in participating. As such, reading their body language during this initial interaction was just as informative as hearing their feedback. Incentives were not proffered to attract staff participation, an effort to ensure honesty and improve credibility of results (Shenton, 2004).

A letter of introduction was then sent to each candidate along with a participant consent form for them to sign (see Appendices A and B). Each of the candidates I approached were willing to participate and returned the signed consent forms. After conducting the interviews, during my analysis, I did not have a sense of data saturation (Merriam & Tisdell, 2016). To remedy this, I sought out an additional respondent using the aforementioned selection process and they too accepted my invitation.

Purposeful convenience sampling was deemed appropriate for candidate selection due to the short-time frame within which the study took place, my limited experience playing the role of researcher, and the lack of funding for the research itself. Suri (2011) critiqued convenience sampling, recommending it should be considered for use only as a final option. When employing this method "the nature of its use and associated caveats must be clearly described" (p. 72). In the spirit of this, it is important to note that each research participant was a member of the teaching staff at the same school where I am currently working. These longstanding professional relationships, and the resultant sense of faculty trust, were an asset to the data collection process, however. Teachers in low-trust schools are often guarded in their communication "often blocking or distorting communication to avoid confrontation with colleagues" (Tschannen-

Moran, 2009, p. 222). An outcome that would clearly hinder the information gleaned from participants.

Data Collection

Data were collected during a two-month period between February and March of 2022. Video recordings and digitally transcribed records of interviews were collected as data. These data were collected using a semi-structured interview format. Doing so allowed me to ask a set of common questions in each interview while affording the flexibility to respond to any new ideas being shared by participants (Merriam & Tisdell, 2016). Each interview was conducted in a virtual setting. This allowed me to deploy both transcription and audiovisual recording software with ease. The virtual setting gave the interviewees flexibility in choosing the location from which they would participate. Thus, helping to minimize external distractions and maximize their personal levels of comfort.

The interview schedule (see Appendix C) drew upon Castillo-Montoya's (2016) Interview Protocol Refinement Framework (IPR) which details the four phases in "developing a well-vetted interview protocol that can help a researcher obtain robust and detailed interview data necessary to address research questions" (p. 812). The interview questions were designed with consideration to the six categories of questions elucidated by Patton (2015). Prior to the questioning portion of the interview, participants were introduced to the nature of the study. They were informed of the opportunity to remove their comments from the record and their ability to terminate the interview at any time. Further, they were informed about the nature of a member check and to expect further communication after the interview process had concluded.

Upon the conclusion of each interview, I watched its recording while correcting any errors within the digital transcript. Reviewing each interview in a timely fashion was useful in

anchoring key ideas within my mind for future consideration. Doing so was also grounded in the notion that data collection and analysis are simultaneous processes (Merriam & Tisdell, 2016). The transcript was then coded which led to a draft summary of the interview. This represented my interpretation of key ideas. This was completed as part of the member checking process. The summary, information about the purpose, and the role required of the interviewee in the member checking process were explained in an email (see Appendix D). Few adjustments were suggested by participants. They were incorporated into the summary documents for each respective interview. In retrospect, I may have had greater levels of feedback had I engaged in Doyle's (2007) notion of participative member checking. This method affords participants choice in how to receive member check materials including hard copies, digital copies, and even audio copies. It also gives them an opportunity to request the researcher's presence during the review of these materials.

Trustworthiness

Guba's (1981) criteria for building trust in qualitative research were considered prior to the initiation of this pilot study and during its evolution. Ensuring the study measures what is intended, that its findings align with reality, is referred to as credibility in naturalistic studies. To support this, I sought to undertake an inductive approach to building meaning in the data. Sources of personal bias were explained early in the research process, an attempt to practice reflexivity (Merriam & Tisdell, 2016). Efforts were made to bracket these sources of bias to minimize their impacts on the analysis of data. Triangulation occurred, on a limited scale, where multiple sources of data were derived through interviews from four different teachers, each with their own perspectives.

The transferability of results is an important consideration for this study as it seeks to influence final examination policymaking. Therefore, boundaries of the study must be made clear to others prior to them making attempts at transference (Shenton, 2004). In support of this, I outlined the organization involved in the study, the parameters surrounding candidate selection, the number of participants involved, an overview of data collection methods, the number of data collection sessions, and the period during which the data were collected. This contextual information informs others seeking to derive meaning from the study, especially if planning to apply it within their own setting.

Limitations

Information relating to the second research sub-question was not explored at length by any of the interviewees. This may indicate an issue with the question design within the interview guide which led to the inability to extract sufficient data specific to the second research sub-question. It is also indicative of the skills possessed by me, a novice researcher, to extract desirable information during the interview process. This limitation was not remedied for two reasons. First, changing the questions in the interview guide raised concerns about consistency. An act, deployed after analyzing much of the data, that would clearly change the methods used in the study and negatively impact its credibility and dependability overall. Second, after sufficient data had been analyzed to realize this data gap existed, there was an insufficient amount of time available for the researcher to take steps towards filling it.

Member checks only occurred during the data collection and analysis process but did not occur upon completion of the study. Guba (1981) argued the credibility of naturalistic studies can be increased by completing these member checks when collecting data and “after casting it into final form” (p. 86). This secondary check was again not possible due to time constraints.

Despite changing final examination policies having had a perceived effect on teacher's ability to assess the academic achievement of their students it would be inappropriate to suggest the observations are entirely the result of shifting examination policies. These events have transpired during one of the most tumultuous moments in modern history and there are likely a host of other factors influencing these results.

Data Analysis

To start the process of coding, I considered Borton's (1970) reflection framework and both the principal research question and sub-questions. This helped to inform the coding process and is in alignment with the four-step analysis process suggested by Merriam and Tisdell (2016). Interviews were initially coded by hand, independent from one another, with keywords and comments written on the margins of the transcript. These codes helped to inform the summary generated as part of the member checking process. Adjustments were made upon receipt of participant feedback and the interview summaries were then coded. At this point, I attempted to use computer software to assist the analysis of data. I input the interview summaries into a word cloud generator in hopes of visualizing terms predominant in the text. Doing so proved to be an ineffective approach and I reverted to coding by hand once more. Using technology to support this portion of the analysis process would be helpful to the researcher. It could serve to emphasize key ideas from the text which in turn could work against potential researcher bias. The initial coding stage serves as a primary foray through the qualitative data. If the researcher is not thorough in considering all of the data in this phase, it could easily be omitted later during category construction. Technology applications could serve as a check and balance for the researcher to aid them in the extraction of data in an unbiased fashion.

Instead of using the word cloud, the keywords from each interview summary were placed into a separate document with a note about the context of their use. During this process, several keywords overlapped allowing for the creation of themes across interviews and for the data to be further distilled. Pseudonyms for each interviewee were assigned at this point to maintain their confidentiality. This constant comparative method of data analysis (Merriam & Tisdell, 2016) led to the process of category construction.

While attempting to distill the data, several concepts remained which did not seemingly appear to fit into a singular category. To help visualize the link between concepts I used mind mapping software. Themes from the summary data were input into the program then sorted into related groupings based on the contextual notes from each interview summary. Two broad categories emerged while engaging in this process. One emphasizing teachercentric variables and the other focusing on students, specifically their learner attributes. The student focused data were discussed in a manner suggesting a connection among ideas. These were considered, grouped, and arranged into a larger model.

The teachercentric variables did not readily combine into a singular interconnected model. In fact, the more I reviewed that portion of the data the more I became concerned attempting to do so would push me towards an oversimplification of the ideas or the intentional removal of certain concepts. Thus, resulting in dilution of the data and less trustworthy results overall.

Findings and Discussion

With the establishment of two broad categories of data including teachercentric variables and student learner attributes (SLA's) I began to seek connections to the research sub-questions. Each teacher discussed SLA's and perceived a shift after the cancellation of final examinations.

These appeared to be impacting the teacher's ability to assess academic achievement. For these reasons, the SLA's were given initial priority during the analysis process.

Student Learner Attributes

Teachers perceived a sharp decline in student motivation upon the cancellation of examinations. Sarah indicated "I've really noticed a decrease in student motivation." (Line 180). Leah connected the notion of motivation to student engagement "they're not as engaged and not as motivated" (Line 300). Leah also connected motivation to competitiveness "they lost that competitive nature... it gives internal motivation for kids to achieve" (Line 430). These views are somewhat misaligned with research investigating high stakes testing and their influence on student motivation. O'Neil et al. (1995) found the threat of evaluative consequences and an increase in competition among peers would not lead to increases in test motivation nor performance on large-scale assessments at the secondary level. These contradicting views may be explained by Baumert and Demmrich (2001) who posit "the personal value of a task is dependent not only on the motivation preferences of the individual, but also on the implications and consequences attached to it" (p. 442). Both the cancellation of final examinations or the adoption of a significant reduction in their weightings minimized the negative consequences resulting from student test performance which serves as a partial explanation for the perceived drop in student motivation, engagement, and competition among peers.

Michelle spoke of impacts on student work completion and preparedness for assessment tasks. She considered their influence on her ability to assess student achievement. "If they sit there, blankly staring for 72 minutes like what? I can't write it for them. I need them to produce something" (Line 444). Effective learners participate in self-regulated learning where they identify knowledge gaps and develop study strategies to address them. These gaps appear to be

magnified here, possibly indicative of a reduction in the deployment of useful study strategies such as retrieval practice and cumulative review (Sein et al., 2021). Iterative processes, such as these, require learners to recall information and commit them to their long-term memory (Karpicke & Blunt, 2011). These mentally taxing exercises must be deemed as worthy for the undertaking by students since their performance improves “when the difficulty of a task is proportionate to the value assigned to the task” (Pearce, 2018, p. 248). A reduced sense of value compared to the considerable effort required of students to prepare for summative examinations may explain some of Michelle’s observations.

Sarah built upon these ideas in her discussion about student retention of concepts and prerequisite understandings in future courses. “You really can see the ones who have internalized that knowledge and have that long-term retention” (Line 216). While also expressing concern for the future “I feel like we’re going to see a lot more where kids are not fully understanding the concepts” (Line 223). A lack of retention combined with gaps in student learning will have a compounding effect as students move onto subsequent grade levels. Teachers and learners alike will be expected to overcome the resultant difficulties for which there are no simple solutions. A strategy preventing this from occurring should be considered instead of one that seeks to mitigate harm after the fact.

Leah spoke to the link between final summative assessments and student independence “so I think with the summative piece, they’re on their own, they have to read it and produce” (Line 371). When working on formative assessment tasks, students can work cooperatively with their peers, with their teacher, and with a variety of tools to demonstrate their understanding of the learning objectives. Test based summative assessments prohibit access to these supports allowing teachers to garner an understanding of individual student abilities while working

independently. Michelle connected a lack of student independence to decreased levels of student confidence with increased perceived levels of stress. “It [in-class examinations] causes more panic than I've ever seen before... In the past, even if a student had high test anxiety, they still felt confident that they could produce something. Whereas this year, I have found...that type of confidence just hasn't been there” (Line 431). This observation can be explained when viewed through the lens of the Transactional Model of Stress and Coping (Lazarus & Folkman, 1984).

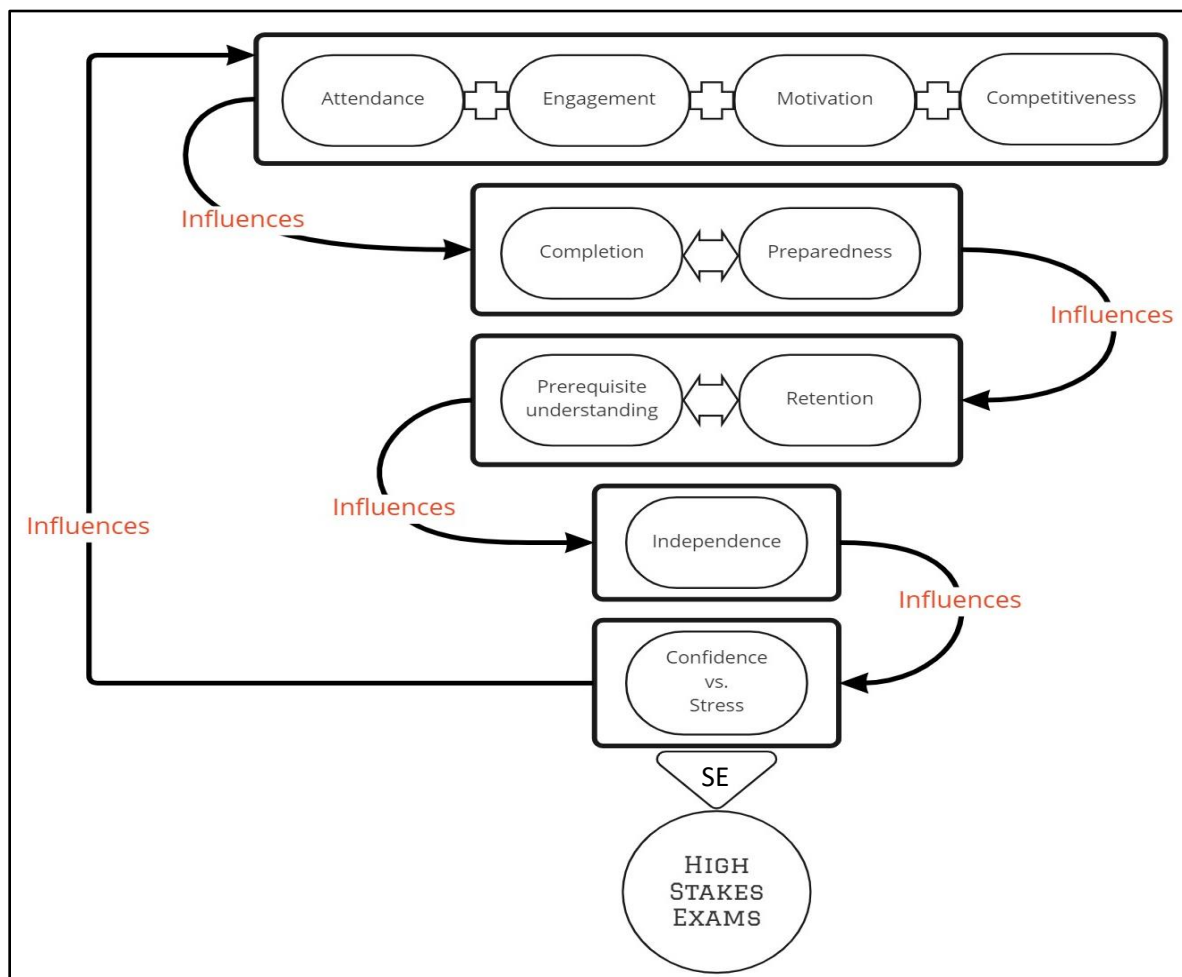
The model predicts students experience academic stress when they view academic events as threatening and interpret their ability to cope as deficient. Student self-efficacy (SE), the personal judgment in one's own ability to execute actions towards achieving goals (Bandura, 1977), influences how students evaluate their capacity to cope with these academic stressors (Meyer et al., 2022). More self-efficacious students may be more resilient when facing challenges and feel less stressed. This is partially due to SE's role in elevating one's sense of control when navigating academic situations and can therefore serve as a coping strategy. Burton and Campbell (2019) described a student's past performance as one of four sources of SE. Previous poor performance can lead to a sense of futility followed by a decrease in student motivation, perseverance during task completion, and a reduced sense of SE. When SE is diminished, student engagement also declines. Further, a student's confidence is grounded in their perceived sense of SE (Pollack & Lilly, 2008). Hence, a diminished sense of SE fosters declines in both student engagement and confidence levels.

Figure 1 was created to model my interpretations of the SLA data and their ability to feed into one another in an assumedly positive or negative fashion. At its base, lies high stakes examinations which teachers associated with many SLA's. It illustrates shifting examination policies had a toppling effect on the SLA's of many students. My findings suggest this was due

to a perceived decrease in the influence high stakes examinations had on academic value judgements being made by students and a resultant decrease in levels of student efficacy. Our education system is built around the notion of preparing students for final examinations and teachers condition their students to this fact. “I usually spend a lot of time teaching them how to properly write those examinations” (Sarah, Line 127). It is unsurprising to see students exhibit shifts in their learning attributes when a seemingly foundational aspect of their education has been altered.

Figure 1

Learner Attributes Impacting Teacher Assessment of Academic Achievement



Perceptions of Purpose for Final Examinations

Figure 2 outlines the themes derived from the coding process with respect to teacher's perceptions of purpose for administering final examinations. These are largely in alignment with the three major conceptions, or purposes, for assessment described by Brown et al. (2011). The first conception describes assessment as a mechanism to improve teaching practices and student learning. Several observations made by interviewees can be linked back to this notion of assessment as improvement. For instance, John stated the following with respect to diploma examinations: "It's the benchmark against which I measure my effectiveness... but there's a lot of variables that we can't control" (Line 46). Final examinations help teachers to reflect on their practice while serving as a partial indicator of their teaching effectiveness. Further, they push some teachers towards greater individual performance "I have generally regarded my performance on [diploma's] with a kind of zealousness that's bordering on monomania" (John, Line 39).

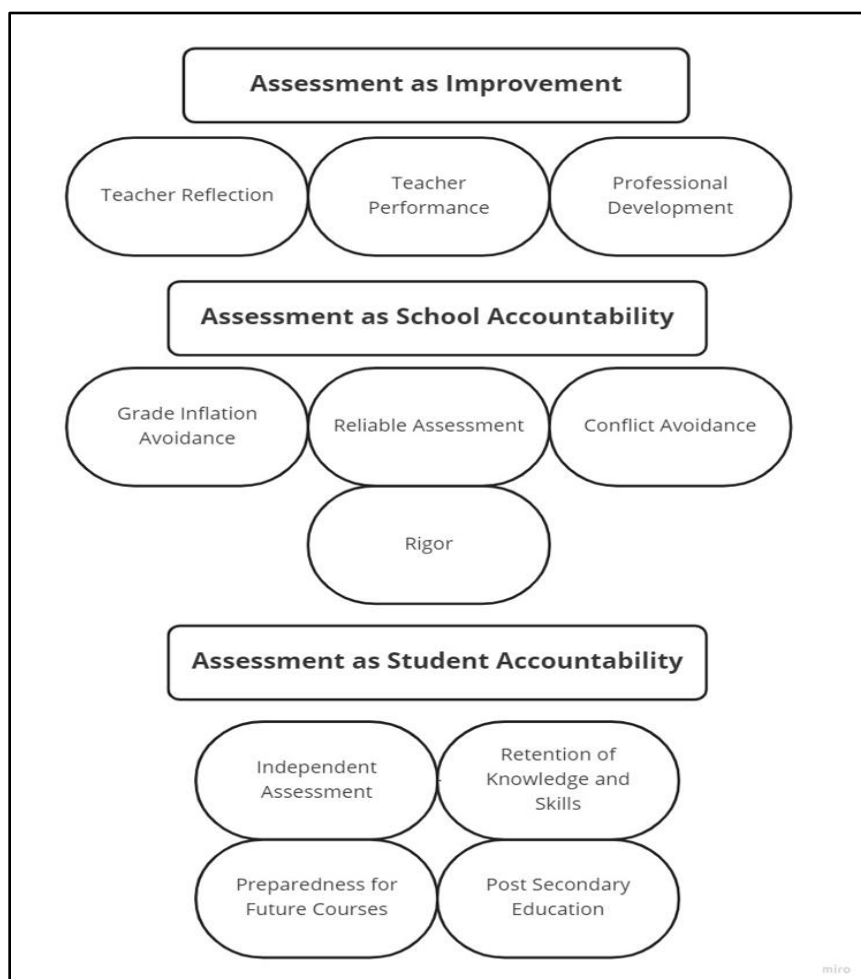
Additionally, final examination marking sessions act as valuable sources of professional development. They are vital for teachers wanting to improve upon their practice while working in the educational silo's commonplace in rural schools. Michelle articulated it this way "I really benefited from going to those provincial and local markings because when I was alone in the department, I didn't know if I was doing it right, I hoped" (Line 318).

Cizek (2001) detailed improvements in assessment quality as an unintended consequence resulting from the implementation of high-stakes examinations. He wrote "the quality of teacher made tests pales when compared with more rigorously developed, large-scale counterparts" (p. 25). John alluded to this idea suggesting provincially made examinations provide a reliable source of information for parents and teachers due to the steps taken in their development. This

can help to ground teachers in their practice and give them “something to stand behind” (John, Line 21). In turn, this may provide teachers with a greater sense of confidence, especially when grade related conflicts with parents and students arise. “It’s not just 30 kids from our school writing something made by me. I like that” (John, Line 68). John’s concern with conflict can be connected to the notion of teacher accountability where assessment results are used to publicly demonstrate teacher or school effectiveness. This links to the second major conception outlined by Brown et al. (2011) which describes assessment as a mechanism for school accountability.

Figure 2

Teachers Perception of Purpose for Administering Final Examinations



Final examinations play a role in sustaining high levels of academic rigor. Rigor involves course workload demands and expectations for course learning (Culver et al., 2021). John perceived a change here “I did not mark with quite the rigor that I have under normal circumstances” (Line 394). Whereas Michelle noted a change in students “the rigor and stamina that students previously had is not there” (Line 425). Further, teachers discussed variables playing into academic rigor, purporting the importance of a final examination. Sarah expressed concern, with their absence. Grades had “probably been inflated compared to where they normally would be” (Line 161). She also “struggled to know how well they retained the information” (Line 242). Leah indicated the need for assessment to target individuals “that they would be doing totally independently” (Line 553). While John felt his teacher created examinations, comprising the bulk of a student’s grade when examinations were cancelled, did not “have the same rigor” (Line 256) as their provincial counterparts. Long term, there is concern about a decrease in rigor impacting both student access to and success during post-secondary education. Final examinations can mitigate this by allowing students to “learn to handle stressful situations” and by helping “an individual learn how to prepare to perform” (Leah, Lines 455 and 457). Leah’s comments are reflective of the third major conception espoused by Brown et al. (2011) whereby assessment acts as a mechanism for student accountability. These assessments have high stakes consequences such as the attainment of graduation requirements or entry selection to higher levels of education, all for which the student is accountable.

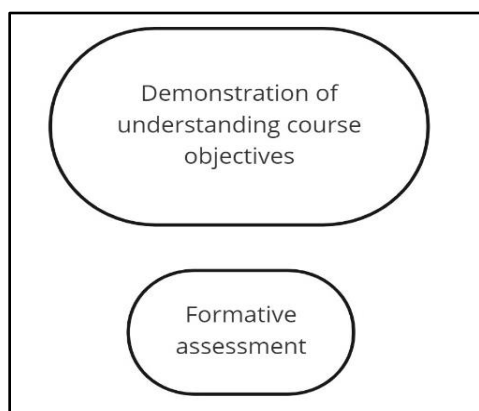
Informing Teacher Understanding of Achievement

Figure 3 illustrates limited information relating to the second research sub-question. Teachers were able to articulate some ways final examinations inform their understanding of

student academic achievement but on a limited scale. They inform “my ability to know exactly where they are, their understanding, and how much they retained by the end of the course” (Sarah, Line 251). Whereas John felt they “help to be able to see in the long term, how much have you retained of skills and knowledge over a five-month period” (Line 318). Despite the limited responses on this topic, there was consistent agreement among each interviewee that a final examination should occur at the end of a course. “I didn't like not having an exit examination.” (John, Line 317).

Figure 3

Summative Examinations Informing Teacher Understanding of Student Academic Achievement



There was considerable discussion about the role of formative assessment and its importance towards developing an understanding of student achievement. The implementation of formative assessment by teachers ranges on a continuum from little to substantial levels of integration (Marshall & Drummond, 2006). Leah and John described their role in implementing “a lot more formative assessment” (John, Line 155) in the absence of final examinations. There is general agreement among researchers suggesting increases in the deployment of formative assessment strategies can enhance learning and improve overall instruction (DeLuca et al., 2018). Citing the benefits of this practice, Leah described her intention to shift her future

assessment practices to include more assessments of a formative nature. “I honestly don’t think I would change that” (Line 583).

Pedagogical practices can be affected by teacher beliefs about the purpose of assessment (Brown, 2004). Formative assessment practices may be deemed as less important when there is a greater sense of pressure for summative requirements to be met (Black & Wiliam, 2018). They suggest these pressures may be derived from teacher concerns with respect to school and individual accountability. Sarah touted the need to deploy greater quantities of summative tasks while reducing those of a formative nature. “I’m marking more assignments than I typically would. To have really good evidence so that I know exactly where these kids are” (Line 149). A greater reliance on teacher generated summative assessment and the voluminous collection of learning evidence speaks to these accountability fears. With shifting examination policies, some teachers feel their own assessment practices will come under greater scrutiny and they must be in a position to prove the validity of their assessment judgments, hence increasing their emphasis on summative assessment practices. Some of the adjustments being made to teacher assessment practices resulting from shifting examination policies may be beneficial to students. Regardless, there appears to be a growing amount of inconsistency among practitioners which should be addressed in the future.

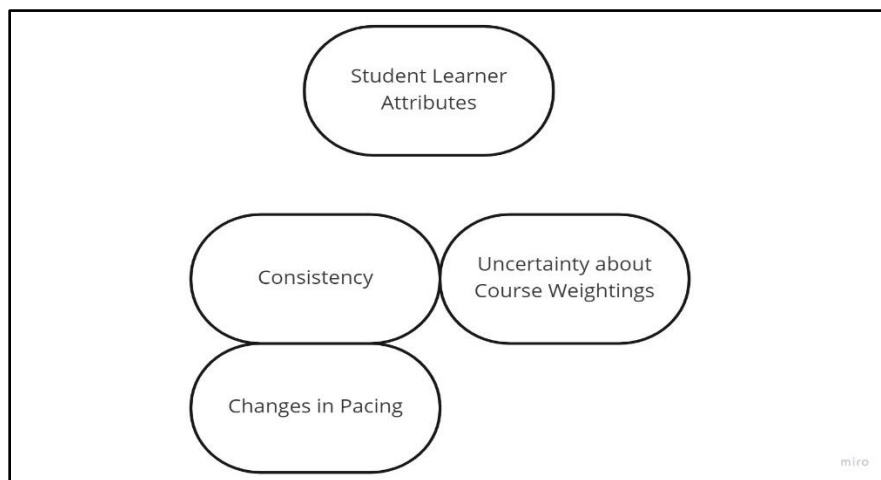
Ability to Assess Student Achievement

Figure 4 outlines how perceived changes in student learner attributes have impacted teacher’s ability to assess student achievement as was explored previously. Additionally, teachers discussed the benefits of maintaining consistent assessment policies and their influence on practice. Consistent assessment policies allow teachers to plan and design learning activities well in advance, followed by assessment and evaluation practices that generate relevant evidence

of student learning through a variety of methods while building towards required skills. These are important components of the professional expectations for Alberta teachers as outlined in the Teaching Quality Standard (Alberta Education, 2020). Michelle expressed concern about this “uncertainty” (Line 537) and how she could “have used that time [preparing for the examination] to better teach the curriculum” (Line 540). Consistency allows teachers to make better use of their professional judgments as they work to navigate the curriculum. It also affords them greater autonomy when making professional decisions, further enabling the maximal use of instructional time.

Figure 4

Impacts on Teacher’s Ability to Assess Student Achievement



The inconsistency introduced by changing examination policies also led teachers to change their instructional pacing. Sarah alluded to instructional benefits of cancelling examinations affording her more time to teach and assess outcomes. “I slowed down a little bit, I wasn’t pushing as hard” (Line 110). Whereas Leah alluded to pacing issues arising in future courses due to a lack of student retention “you spend so much time reviewing concepts” (Line 327). Michelle reinforced this idea suggesting final examinations are a useful metric for

analyzing student understanding of prerequisite concepts and this helps her to avoid “spending a large chunk of time re-teaching something” (Line 383).

Summary

Shifts in summative final examination policies, including their cancellation and changes to their weightings, appear to have had several impacts on teachers’ ability to assess the academic achievement of their students. Several interview responses were linked to a perceived transformation in student learner attributes largely considered detrimental to both student achievement and one’s ability to assess it. Student motivation, work completion, and task engagement all experienced declines, possibly due to a decreased sense of overall task value. Concerns about concept retention and student confidence levels were also cited. Combined, there was a perceived decrease in students’ sense of self-efficacy which, if left unchecked, could support additional declines in the student learner attributes noted in figure 1.

Teachers perceived the purpose of summative final examinations within the three conceptions described by Brown et al. (2011) including assessment as improvement, as school accountability, and as student accountability. Teachers indicated final examinations help them improve by encouraging them to reflect on and make adjustments to their practice. The examinations are also used by some teachers as benchmarks for which they measure their own professional effectiveness, pushing them towards greater levels of individual performance. Further, examination marking sessions serve as valuable assessment literacy themed professional development opportunities for teachers in rural schools.

Accountability themes were discussed by all interviewees. Provincially made final examinations were seen as more reliable and rigorous sources of assessment information when compared to their teacher developed counterparts. This reliability is useful for teachers wanting

to rationalize their own assessment specific professional judgements serving as evidence to mitigate conflicts with school stakeholders. Teachers also felt the examinations pushed students to be more accountable for their own learning by requiring students to demonstrate their understanding and retention of concepts while working independently. Additionally, they connected final examination performance to student accountability for meeting both graduation and post-secondary entrance requirements.

Interviewees provided somewhat limited information specific to how summative final examinations informed their understanding of student academic achievement. Each interviewee expressed a desire to continue using final examinations as part of their assessment regimes, stressing their importance for determining levels of student retention and understanding of curricular concepts. Each teacher discussed the adjustments they had made to the amount of formative and summative assessment being deployed in their classes resulting from shifts in examination policies. Some teachers touted the benefits of increasing the number of formative assessments used in their classes whereas others indicated the need to collect more summative assessment tasks serving as evidence of student learning. Greater reliance on this summative evidence may be indicative of teacher accountability fears.

Shifting examination policies appear to have created a lack of consistency in assessment practices among teachers which has influenced their ability to assess the academic achievement of students. It has affected their ability to accurately plan for the instructional pacing of their classes which in turn has changed how teachers approach the delivery of curriculum objectives. These factors have required teachers to hastily adjust their assessment plans creating further inconsistencies among educators.

Conclusions and Recommendations

This study sought to understand how changing summative final examination policies are impacting the ability of teachers to assess the academic achievement of students. The findings suggest these changing policies have had detrimental effects on a variety of student learner attributes such as motivation, engagement, and confidence. These have influenced work completion and student retention of concepts important to success in future courses. Students appear less able, or perhaps willing, to cope with academic stressors and work to achieve at levels seen prior to the COVID-19 pandemic. Combined, these changes in learner attributes have made it difficult for teachers to garner a true understanding of their student's academic achievement. One cannot accurately assess a student who is not regularly completing their work or not doing so to a degree reflective of their actual ability. Further, a perceived drop in students' sense of self-efficacy means less resilient students who will likely struggle to become suitably engaged in school, overcome deficits in their learning, and build the increased levels of academic confidence stemming from doing so. These results are troubling as finding ways in which to motivate and engage students in their learning is already an increasingly difficult challenge for division III and IV teachers. Policymakers are not doing any favours to students, or their teachers, by creating circumstances which further erode the attributes characteristic of successful learners.

Changing summative examination policies have influenced teachers' ability to meet the professional obligations required of them in the Teaching Quality Standard (Alberta Education, 2020). This is worth considering from a system level perspective as Alberta Education's (2021) student assessment policy requires education programs to provide for accurate assessment of

student achievement along with communication of said information to parents and other educational stakeholders.

Communication with stakeholders is integral to fostering effective relationships among the broader school community. Fostering these relationships is listed in a position of primary importance in all three professional practice standard documents outlining the competency requirements of teachers, school, and district leaders (2021). Relational trust is necessary between the school and its stakeholders for effective communication, cooperation, and educational improvements to occur. (Tschannen-Moran & Hoy, 2000). This trust is partially built through stakeholder confidence in schools with respect to elements of their reliability, openness, and honesty.

Peace River School Division has worked to openly communicate changes to its assessment policies throughout the COVID-19 pandemic and stakeholders have been generally understanding and tolerant of them. With many members of the public adopting a new mindset, one acknowledging COVID-19 will play a continual and potentially disruptive role in our daily lives, PRSD will be expected to adjust its policies to be more reflective of this thinking. Stakeholder trust is at considerable risk should teachers, their schools, and the district communicate assessment plans then suddenly change them once more. This could lead to the perception of poor planning and put into question the true reliability, openness, and honesty of PRSD. Distrust is established when accurate information is not being provided and commitments about future actions are not kept (Tschannen-Moran & Hoy, 2000). Educators should be weary of distrust as "once it is established, it has a strong tendency to be self-perpetuating" (p. 550). It leads stakeholders to view school actions with suspicion and impedes communication, a vital factor to the fostering of effective relationships. Therefore, consistent

assessment policies are not only integral to teachers for planning, curriculum delivery, and accuracy of their assessment regimes. They are vital to sustaining the trust of stakeholders within the broader educational organization.

It was somewhat unsurprising to hear teachers express the idea summative final examinations serve as mechanisms to improve their professional practice. Neither was the idea they serve as accountability tools for students and teachers alike. Of interest, however, were the various responses teachers took to engaging in classroom assessment during times of shifting examination policies. The variability in their responses to new educational policy speaks to their conceptions, or beliefs and attitudes, about assessment itself (Fives & Buehl, 2012). Teacher conceptions of assessment are influenced by the policy frameworks within which they operate which in turn influence the manifestation of their assessment practices (Brown et al, 2019). Some teachers adjusted their practice to safeguard against accountability pushback, knowing their assessment practices would likely face greater scrutiny in the absence of final examinations. Others adapted by implementing more opportunities for students to receive feedback exhibiting a greater focus on student learning objectives than fear of external reprisal. The primary purpose of assessment, to promote student learning and inform the teaching process (Peace River School Division, 2022), appears to have become muddled among teachers. The variability among teachers to shifting examination policies could work to undermine the collective dependability of a school's assessment regime. Therefore, teachers not only need to be confident in their ability to assess students, they also need to be continually working towards a more common conception of the purpose for assessment in their classes. Something worth considering by educational leaders, especially with the potential for government-imposed disruptions to summative final examinations resulting from COVID-19 in the future.

Recommendations

Three recommendations resulting from this study bear further consideration. Firstly, school leadership teams should consider various means of improving upon their students' levels of self-efficacy as those who have strong beliefs in themselves tend to persist in school and successfully achieve study outcomes (Zimmerman, 2000). Öqvist and Malmström (2018) provide an indirect pathway forward here. They suggest high levels of student self-efficacy can lead to increases in educational motivation while also highlighting the important role played by the developmental leadership behaviours expressed by teachers within their instructional settings. Öqvist and Malmström characterize these behaviours as those which are guiding, supportive, participation-, and performance-oriented. "When students are led by a teacher who adopts these four leadership behaviours, the chances for students to learn and advance their knowledge increase" (p. 169). Further, when students work with teachers possessing low degrees of developmental leadership, they risk losing motivation and may no longer muster the energy to learn. School leaders should bring to light the importance of teacher leadership behaviours and their connection to levels of student motivation and self-efficacy. They should facilitate opportunities to learn about, implement, and collectively reflect upon developmental leadership behaviours. By failing to do so "schools may prevent students from efficiently carrying out their educational tasks, achieving their goals and realising their full potentials" (Öqvist & Malmström, 2016, p. 376).

Provincially, we are not seeing a long-term plan to move away from the use of high stakes examinations. Doing so at the district level is therefore unrealistic. However, districts have considerable flexibility in their approach to assessment outside of these provincially mandated examinations (Alberta Education, 2021). Therefore, a second recommendation is for

school districts to adopt a longer-term policy approach to the implementation of summative final examinations within its division III and IV classes. This policy should enable the use of final examinations within teacher assessment regimes as they are deemed important for understanding the academic achievement of students and for teacher professional development. Further, this policy should clearly summarize district expectations and the parameters for implementation of summative final examinations expected of all schools. It should be developed with consistency in mind while considering potential future disruptions from COVID-19. This information, when communicated early in the academic year, will better enable teachers to perform their professional duties as outlined in the Teaching Quality Standard (Alberta Education, 2020), provide students with valuable learning targets, and work to build trust among stakeholders. Throughout this process, district leaders are encouraged to consider the words of Aitken and Aitken (2016):

Student learning and student wellbeing are the critical outcomes of the educational system, and well-founded assessment policies are the sole means that we have to measure how well we are doing in those two vital dimensions. Our assessment policies need to reflect the importance of learning and also *guide* learning in our schools. (p. 166)

Finally, a renewed emphasis should be placed upon building assessment cultures stressing the importance of sound assessment practices within schools. Encouraging teachers to add assessment themed goals as part of their annual growth plans can support this aim. School leaders can embed assessment topics linked to district policies into the school's annual professional development plans as well. With shifting summative examination policies there has been considerably more emphasis placed on teacher made evaluative judgements. Taking steps to improve the assessment literacy of teachers can further enable them to make defensible

decisions about high stakes tests and classroom-based assessments (Popham, 2009). The creation of localized, teacher led, professional development networks can support this aim (Tierney, 2006) while working to place greater emphasis on the primary purpose of assessment, the improvement of student learning. School and district leaders should ensure both structures and resources are in place to further enable the development of these professional development networks.

References

- Aitken, E. N., Aitken, A. J. (2016). Current policies surrounding assessment in Alberta: Future implications. In S. Scott, D. Scott, & C. Webber (Eds.), *Assessment in education: Implications for Leadership* (pp. 137-173). Springer. https://doi.org/10.1007/978-3-319-23398-7_7
- Alberta Education. (2020). *Leadership quality standard*. Alberta Queen's Printer. <https://bit.ly/3aLsewO>
- Alberta Education. (2021). *ECS to grade 12 guide to education*. Alberta Queen's Printer. <https://bit.ly/3bb0vG1>
- Agrey, L. (2004). The pressure cooker in education: Standardized assessment and high-stakes. *Canadian Social Studies*, 38(3), 1-12. <https://eric.ed.gov/?id=EJ1073917>
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychology Review*, 84(2), 191-215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education*, 16(3), 441-462. <https://doi.org/10.1007/BF03173192>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575. <https://doi.org/10.1080/0969594X.2018.1441807>
- Borton, T. (1970). *Reach, touch, and teach; student concerns and process education*. McGraw-Hill.

- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education Principles Policy and Practice*, 11(3), 301-318. <https://doi.org/10.1080/0969594042000304609>
- Brown, G. T. L., Gebril, A., & Michaelides, M. P. (2019). Teachers' conceptions of assessment: A global phenomenon or a global localism. *Frontiers in Education*, 4(16), 1-13. <https://doi.org/10.3389/feduc.2019.00016>
- Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, 27(1), 210-220. <https://doi.org/10.1016/j.tate.2010.08.003>
- Burton, L. M., & Campbell, B. D. (2019). Fostering self-efficacy and engagement: A case of alternative teaching and learning in high school mathematics. *Journal of Ethnographic & Qualitative Research*, 13(4), 272-288. <https://bit.ly/3zcrw1C>
- Cairns, R. (2021). Exams tested by COVID-19: An opportunity to rethink standardized senior secondary examinations. *Prospects: Comparative Journal of Curriculum, Learning, and Assessment*, 51(1-3), 331-345. <https://doi.org/10.1007/s11125-020-09515-9>
- Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *The Qualitative Report*, 21(5), 811-831. <https://doi.org/10.46743/2160-3715/2016.2337>
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27. <https://doi.org/10.1111/j.1745-3992.2001.tb00072.x>

- Culver, K. C., Braxton, J. M., Pascarella, E. T. (2021). What we talk about when we talk about rigor: Examining conceptions of academic rigor. *The Journal of Higher Education*, 92(7), 1140-1163. <https://doi.org/10.1080/00221546.2021.1920825>
- Darling-Hammond, L. (1991). The implications of testing policy for quality and equality. *Phi Delta Kappan*, 73(3), 220-225. <https://bit.ly/3rBfo94>
- DeLuca, C., Chapman-Chin, A. E. A., LaPointe-McEwan, D., & Klinger, D. A. (2018). Student perspectives on assessment for learning. *The Curriculum Journal*, 29(1), 77-94. <https://doi.org/10.1080/09585176.2017.1401550>
- Doyle, S. (2007). Member checking with older women: A framework for negotiating meaning. *Health Care for Women International*, 28(10), 888-908. <https://doi.org/10.1080/07399330701615325>
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, T. Urda, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 471–499). American Psychological Association. <https://doi.org/10.1037/13274-019>
- Froese-Germain, B. (2001). Standardized testing + high-stakes decisions = educational equity. *Interchange*, 32(2), 111-130. <https://doi.org/10.1023/A:1011985405392>
- Guba, E. (1981). ERIC/ECTJ Annual review paper: Criteria for assessing the trustworthiness of naturalistic inquiries. *Education Communication and Technology*, 29(2), 75-91. <https://bit.ly/3xw7tOC>

- Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20(3), 245-270. <https://doi.org/10.1080/02671520500193744>
- Hunt, M. R. (2009). Strengths and challenges in the use of interpretive description: Reflections arising from a study of the moral experience of health professionals in humanitarian work. *Qualitative Health Research*, 19(9), 1284–1292. <https://doi.org/10.1177/1049732309344612>
- Kahlke, R. M. (2014). Generic qualitative approaches: Pitfalls and benefits of methodological mixology. *International Journal of Qualitative Methods*, 13(1), 37-52. <https://doi.org/10.1177/160940691401300119>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772-775. <https://doi.org/10.1126/science.1199327>
- Klenowski, V. (2012). Raising the stakes: The challenges for teacher assessment. *The Australian Education Researcher*, 39(2), 173-192. <https://doi.org/10.1007/s13384-012-0057-9>
- Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer Pub.
- Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. Farrar, Straus, and Giroux.
- Marshall, B., & Drummond, J. M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, 21(2), 133-149. <https://doi.org/10.1080/02671520600615638>
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

- Meyer, N., Niemand, T., Davila, A., & Kraus, S. (2022). Biting the bullet: When self-efficacy mediates the stressful effects of COVID-19 beliefs. *PLOS ONE*, *17*(1), 1-16.
<https://doi.org/10.1371/journal.pone.0263022>
- Moses, M. S., & Nanna, M. J. (2007). The testing culture and the persistence of high stakes testing reforms. *Education and Culture*, *23*(1), 55-72. <https://bit.ly/3KiVQyN>
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, *14*(1), 1-172. <https://bit.ly/33OSsKX>
- O'Neil, H. F. Jr., Sugrue, B., & Baker, E. L. (1995). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, *3*(2), 135-157. https://doi.org/10.1207/s15326977ea0302_2
- Öqvist, A., & Malmström, M. (2016). Teachers' leadership: A maker or a breaker of students' educational motivation. *School Leadership & Management*, *36*(4), 365-380.
<https://doi.org/10.1080/13632434.2016.1247039>
- Öqvist, A., & Malmström, M. (2018). What motivates students? A study on the effects of teacher leadership and students' self-efficacy. *International Journal of Leadership in Education*, *21*(2), 155-175. <https://doi.org/10.1080/13603124.2017.1355480>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, *42*, 533-544. <https://doi.org/10.1007/s10488-013-0528-y>
- Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage.

- Peace River School Division. (2022, June). *Administrative Procedure 360: Assessment, evaluation, and communication of student learning*. <https://bit.ly/3Q944Md>
- Pearce, C. (2018). It pays to prepare: The value of low-stakes tutorial preparation exercises to student performance. *Australian Journal of Adult Learning*, 58(2), 246-265.
<https://bit.ly/3PuxlQM>
- Pollack, B. L., & Lilly, B. (2008). Gaining confidence and competence through experiential assignments: An exploration of student self-efficacy and spectrum of inquiry. *Marketing Education Review*, 18(2), 55-66. <https://doi.org/10.1080/10528008.2008.11489039>
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(4), 4-11. <https://doi.org/10.1080/00405840802577536>
- Sein, A. S., Dathatri, S., & Bates, T. A. (2021). Twelve tips on guiding preparation for both high-stakes exams and long-term learning. *Medical Teacher*, 43(5), 518-523.
<https://doi.org/10.1080/0142159X.2020.1828570>
- Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63-75. <https://doi.org/10.3233/EFI-2004-22201>
- Smith, S., James, A., Brogan, A., Adamson, E., & Gentleman, M. (2016). Reflections about experiences of compassionate care from award winning undergraduate nurses – What, so what ... now what? *Journal of Compassionate Health Care* 3(6), 1-11. <https://doi.org/10.1186/s40639-016-0023-x>
- Spratt, M. (2005). Washback and curriculum innovation: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
<https://doi.org/10.1191/1362168805lr152oa>

- Suri, H. (2011). Purposeful sampling in qualitative research synthesis. *Qualitative Research Journal*, 11(2), 63-75. <https://doi.org/10.3316/QRJ1102063>
- Thacker, A. J., & Williams, R. E. (1974). The relationship of the graduate record examination to grade point average and success in graduate school. *Educational and Psychological Measurement*, 34, 939-944. <https://doi.org/10.1177/001316447403400425>
- Thorne, S., Kirkham, S. R., & O'Flynn-Magee, K. (2004). The analytic challenge in interpretive description. *International Journal of Qualitative Methods*, 3(1), 1–11. <https://doi.org/10.1177/160940690400300101>
- Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education*, 13(3), 239-264. <https://doi.org/10.1080/09695940601035387>
- Tschannen-Moran, M., & Hoy, W. K. (2000). A multidisciplinary analysis of the nature, meaning, and measurement of trust. *Review of Educational Research*, 70(4), 547-593. <https://doi.org/10.3102/00346543070004547>
- Tschannen-Moran, M. (2009). Fostering teacher professionalism in schools: The role of leadership orientation and trust. *Educational Administration Quarterly*, 45(2), 217-247. <https://doi.org/10.1177/0013161X08330501>
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-81. <https://doi.org/10.1006/ceps.1999.1016>

Appendix A

Letter of Introduction



UNIVERSITY OF ALBERTA

Darren Phelps
Faculty of Graduate Studies

and Research
Edmonton, AB T6G 2R3

1-800-758-7136
dphelps@ualberta.ca

February 2, 2022

██████████
Fairview, AB T0H 1L0

Dear ██████████,

I am a graduate student in the Master of Education in Educational Policy Studies program at the University of Alberta. The purpose of this letter is for you to take part in a research assignment for my EDPS 509 Research Design and Data Analysis course. My assignment is intended to better understand how continually changing summative final exam policies are impacting teachers' ability to assess the academic achievement of their students. Your participation would involve a one hour recorded virtual interview session with myself. Following this, you would be asked to clarify the researchers' interpretations through a member check mechanism. Please note, your participation is voluntary; there will be no consequence to you should you decline to participate or decide to withdraw from participating.

To gather data for my research assignment, I will be engaging several participants in semi-structured interviews. The interview will be audio recorded and transcribed. Please note:

You may choose not to answer any question.

- You may opt out of this research assignment once responses have been submitted. To do so, please submit your request by email by March 28, 2022, and I will destroy all data.
- I will send you a transcription of the interview as well as a summary of the main points I understood you to make by email; you will have the opportunity to verify the accuracy of the transcription and my interpretation of it.

Should any concerns, complaints, or questions arise from your participation, you may contact me or my instructor, Dr. Jose da Costa (jdacosta@ualberta.ca).

All data will be handled in compliance with the University of Alberta Standards for the Protection of Human Research Participants:

- Participant names will not be revealed. To protect confidentiality, pseudonyms or numerical coding will be used in all written representations of the data.
- Hard copy data will be locked in my office filing cabinet. It will be destroyed upon completion of my graduate program.
- Digital data will be stored on my computer under a secure password-protected system and will be destroyed on my completion of my graduate program.
- Data will be used to complete my EDPS 509 course, my graduate program, and may be used in future presentations and publications in educational contexts.

Thank you for considering this invitation to participate in my research. If you wish to participate, please sign the attached consent form and return it to Darren Phelps either by hand or by email (dphelps@ualberta.ca) by February 3rd, 2022. I have included two copies of the consent form: one is to be signed by you and the other is for your own records.

The plan for this research has been reviewed for its adherence to ethical guidelines and approved by Research Ethics Board 1 at the University of Alberta. For questions regarding participant rights and ethical conduct of research, you can contact the Research Ethics Office at (780) 492-2615.

Should you wish a copy of my research findings, I would be pleased to provide one on your request.

Sincerely,

Darren Phelps

Graduate student in the Educational Administration and Leadership program
Faculty of Education, University of Alberta
780-835-5421
dphelps@ualberta.ca

University of Alberta Ethics ID# Pro00096710

Educational Policy Studies

7-104 Education North • University of Alberta • Edmonton • Canada • T6G 2G5

Telephone: (780) 492-5868 • Fax: (780) 492-2024

E-mail: Jose.da.Costa@ualberta.ca

Appendix B

Participant Consent Form



UNIVERSITY OF ALBERTA

EDPS 509 Research Assignment: Research Design and Data Analysis

Researcher: Darren Phelps

Date Range of Research: February 1, 2022 – April 30, 2022

I, _____ (name of participant), hereby consent to participate in the research, Research Design and Data Analysis.

I understand that my participation includes:

- Up to a one-hour semi-structured interview with Darren Phelps (researcher).
- The interview will take place using Google Meet.
- The interview will be recorded via Google Meet.
- Transcription software will be deployed to capture a written account of the interview.
- A member check will be conducted post-interview with the participant to ensure correct interpretations.

As per the Letter of Information, I understand that:

- My participation in this research is voluntary.
- I may withdraw from the research without penalty until March 28th, 2022.
- All information gathered will be treated confidentially.
- No identifying information will appear on written representations of the data: pseudonyms or numerical coding will be used to convey the data.
- The data will be used for the purposes of completion of the Master of Education in Educational Studies (MES) program and may be used in future presentations and publications in the educational context.
- The plan for this research has been reviewed for its adherence to ethical guidelines and approved by Research Ethics Board 1 at the University of Alberta. For questions regarding participant rights and ethical conduct of research, I can contact the Research Ethics Office at (780) 492-2615.

Signature of Participant

Date

University of Alberta Ethics ID# Pro00096710

Educational Policy Studies

7-104 Education North • University of Alberta • Edmonton • Canada • T6G 2G5

Telephone: (780) 492-5868 • Fax: (780) 492-2024

E-mail: Jose.da.Costa@ualberta.ca

Appendix C

Interview Guide

Script Prior to Interview

I would like to thank you once again for being willing to participate in the interview aspect of my study. As I have mentioned to you before, my study seeks to understand how changing summative final exam policies may be impacting teachers' ability to assess the academic achievement of their students. Our interview today will last up to one hour during which I will be asking you some questions relating to this topic. During the interview I encourage you to explore your thoughts and verbally share that thinking. Feel free to express these thoughts openly as your identity will remain confidential throughout this process.

At this point in time, you have been given a letter of introduction and a participation consent form. Have you had an opportunity to review and sign those documents as requested?

[wait for answer]

Great thank you. So just to reiterate I will be recording this discussion using Google Meet. Our discussion will also be transcribed, and copies of the transcription and recording will be saved in a secured cloud drive. After I have had some time to process the data from our interview today, I will reach out to you once more to ensure my interpretations of the information shared today are aligned with your own. This brief process is called a member check and helps to build trustworthiness and credibility of the collected data.

With this in mind, are you still ok with me recording (or not) our conversation today? ___Yes ___No

If yes: Thank you! Please let me know if at any point you want me to turn off the recorder or keep something you said off the record.

If no: Thank you for letting me know. I will only take notes of our conversation.

Before we begin the interview, do you have any questions? [*Discuss questions*]

If any questions (or other questions) arise at any point in this study, you can feel free to ask them at any time. I would be more than happy to answer your questions.

Question 1:

Tell me a bit about your teaching background?

Follow-up: So, you have taught courses with final exams including diploma exams for quite some time?

Question 2: Consider the typical assessment regime for one of your high school classes. What does that look like?

Follow-up: Is that how you commonly structure assessment in all your classes?

Question 3: Summative final exams comprise a large component of your assessment regime. Consider an alternate learning dimension where such assessments were not part of that regime. What would change within your classroom?

Question 4: In this alternate dimension, one without final summative assessments, what factors, external to your classroom, would be impacted?

Prompts = Program access? Accountability? Learning supports? Perception of teaching performance?

Make sure this is discussed – In what ways do summative final exams inform teacher understanding of student academic achievement?

Question 5: Has your assessment regime changed since the beginning of the pandemic, if so, how?

Question 6: Can you discuss the impacts, if any, this has had on your assessment of student academic achievement?

Conclusion: Before we conclude this interview, I would like to reiterate the purpose of my research. It is to better understand how continually changing summative final exam policies are impacting teachers' ability to assess the academic achievement of their students. Do you have any additional comments relating to this question that we have not had an opportunity to discuss?

Conclusion 2: I would like to take the opportunity to thank you for your time today. Your willingness to explore these ideas and to donate your valued time while doing so is greatly appreciated. I will be in touch, shortly, to engage in a member check where again I will seek your input to ensure my interpretations of this interview align with your own.

Do you have any final questions for me? (Answer them)


-Thanks again, Bye

Appendix D

Sample Member Check

Hello,

The **member checking** process is a means of creating trustworthiness in qualitative research data and any findings that are derived from them. In the table below, you will see a summary of ideas from our interview together. Please note that the summary represents my interpretations of what you said during the interview process. The information in the table below is not a verbatim transcription of the interview. With that in mind, could you please review the summary of ideas to ensure they are in alignment with your own thinking.

If you have any comments relating to the information presented below, please use the comment feature of Google Docs to express your thoughts on the pertinent sections. This can be accomplished by selecting the text you wish to comment on and clicking this icon: 

Your feedback will be considered during the analysis phase of my research. Therefore, I am very appreciative of your willingness to review the information. Please do not feel that you have to make changes if the ideas within the summary are aligned to your thinking.

Once you have had a chance to review the information and/or if you have any questions during the process, please do not hesitate to reach out in person or by email (phelpsd@prsd.ab.ca).

Thanks again,

Darren

Recall: Purpose of Study and Research Questions:

The purpose of this study is to better understand how continually changing summative final exam policies are impacting teachers' ability to assess the academic achievement of their students. The study is guided by the following questions:

1. What are teacher's perceptions of the purpose for administering summative final exams?
2. In what ways do summative final exams inform teacher understanding of student academic achievement?
3. Given the cancellation of summative examinations, what impacts do teachers perceive on their abilities to assess student achievement?

Interview Candidate #1 - summary:

Summary of Interview:

- Taught 30+ years.
- Senior high (10, 20, 30 level) with some Jr. High
- Used to diplomas as part of the evaluation process for gr. 12 courses.
 - Seen weightings of 50% and 30% on the diploma.
- Taught both full-time and part-time

- Has not taught gr. 9 courses when the PAT exam was around.
- Course assessment includes categories for assignments and quizzes.
- Exam marks comprise 60-70% of the course mark depending on the grade.
- Formative assessment gives students a chance to practice concepts.
 - Allows for preparedness on summative assessments.
- Large summative assessments (exams) occur at the end of each unit culminating with a final exam.
- Final exam helps the teacher to understand how much information has been retained by the students.
- Demonstrates how prepared to move into the next level in the course sequence (such as Math 10C to Math 20-1.)
- Students appear less confident in subsequent courses after not being exposed to a final exam in the previous course.
- Without an exam, one may think that students would learn for the sake of learning and growing knowledge.
 - This was not the case when given the chance.
 - Instead, many students saw not having a final exam as an excuse to not complete the in-class work.
 - Despite the in-class assessments comprising the totality of the student's grade.
 - Saw a decrease in the number of students doing the required work in the course.
- The final exam is a student motivator in that students know they will be assessed, independently, on the content of the course.
- Students are less engaged in their learning when a final exam is absent.
- Students do not seem to realize that despite not having a final exam, they will still need to know these prerequisite concepts for later courses.
 - Resultantly, prerequisite skills are lacking in subsequent courses.
 - As such it is to the students benefit to learn as much as possible and do well.
- There were times where little notice was given prior to removal of the final exam component.
- Prior to this, students knew to take good notes, there was a focus on exam preparation by both teacher and students.
- When weightings were reduced on the exam (10%) but the expectation was they were still writing one, students seemed to have a greater focus on their learning than when the exam was removed entirely.
 - Students knew they would be using the concepts they had learned again in the future.
- After the removal of the final exam requirement, other summative assessment supports for students had less effectiveness. For example: open-book testing did not seem to greatly benefit student achievement.
 - Likely due to retention of information which would allow them to meaningfully process the concepts as shown in the textbook.
 - Possibly due to a variety of variables present during the shifting COVID-19 educational landscape.
- When students have not retained their learning from previous courses, the amount of time reviewing concepts takes considerable time in subsequent courses. This may influence the learning timelines in those courses.
- Students who have not retained information demonstrate a lack of confidence in their abilities in subsequent courses.
- The summative final exam benefits the teacher. It shows if students can perform or understand

the concepts taught in the course.

- It also allows the teacher to assess a student's ability to navigate the learning objectives independently.
 - Have students independently achieved the objectives as they relate to the teacher's instruction?
- When students are on their own (in a final exam) they have to read and produce on their own.
- Since the pandemic began, students are much less independent in their learning.
 - They require much more attention from the teacher and are less able to navigate curricular concepts on their own.
- Not having final exams has contributed to this lack of learning independence.
 - Students are not preparing themselves for a final assessment.
 - Nor do they appear to know how to prepare for a final assessment.
 - This is leading to a decrease in student confidence.
 - It has become increasingly difficult to know what the student can do independently.
- Not having final exams has reduced a sense of student competitiveness.
 - There is a decrease in competitive nature among students.
 - Competitiveness can be very positive as it leads to intrinsic motivation towards working hard and achieving success in their courses.
- A lack of ability to prepare for summative exams will impact them upon moving into post-secondary.
 - How will they handle exams at the post secondary level when they haven't had success doing so in high school?
 - Or when they haven't written an exam with a significant weighting (impact) on their grade.
 - This will likely lead to considerable stress in their future.
- Preparing for exams helps students learn how to prepare for a variety of stressful situations. (E.g., job interview).
- Final assessments help an individual "prepare to perform."
- The interviewee expressed a strong desire to have a final exam at the end of the course, even if it was at a reduced weighting such as 10%.
 - A low weighting would still motivate students in higher academic streams. Not sure about those in a non-academic program.
- The final exam provides a variety of benefits to the teacher.
 - It allows for teacher reflection. It shows, to some degree, if the teacher has gotten their point across.
 - It also helps know if you are doing your job properly.
 - Encourages you to continually review your course materials and pedagogical practice.
 - The exams help you build confidence in your learners.
- If a final exam was not required, summative assessments could be completed in a more hands-on fashion.
 - There would still need to be an independently written summative assessment.
- Since the start of the pandemic, non-final exam assessment weightings have not changed.
- The interviewee has considerably increased their formative assessment practices.
 - This has been viewed as a generally positive improvement according to the interviewee.
 - Formative observation allows the teacher to focus their time on students who need additional support.

- There is a greater emphasis on teaching essential outcomes.
- There is a realization that not as much summative assessment is needed as was previously thought.
- The academic expectations placed on learners used to be higher. This has changed since the start of the pandemic.
- The increase in formative assessment has helped to build student confidence levels.
 - They seem more willing to ask questions.
 - Less of a sense that they are constantly being tested.
- There is a greater frequency in the need for the teacher to remind students to complete their work. Chasing or following-up is needed on a continual basis (lack of independence again)
- There has been a greater use of open-book tests in class. This has typically only benefited those who had a previous understanding of concepts largely due to the higher-level questioning that occurs on exams.
 - Open book is not an option on final exams.
- More opportunities to retry summative assessments have been proffered.
 - These rewrites usually focus on the concepts students did poorly on instead of redoing the exam in its entirety.
- When reflecting on student achievement pre-pandemic:
 - Students retained more information
 - Were more diligent, showed a stronger work ethic.
 - Had more obvious goals related to achievement.
 - Were more engaged in their learning.
 - More motivated to learn and achieve success in the course.
- Again, a lack of independent exam completion has led to poor results on the final exam.
- The interviewee felt that they still had a good understanding of student achievement levels in the course despite not writing an exam.
 - The exam helped show the level of student preparedness.
- Student stress levels have increased as they do not know how to prepare for an exam.
- The final exam helps students become more accountable.
- The final exam helps students retain information which can be recalled in later courses. This helps with their confidence as learners.