Mathematics Teachers' Practices and Perceptions of Excluding Students from Participating in Large-scale Assessments

Alex Yankey, University of Prince Edward Island, Canada
Tess Miller, University of Prince Edward Island, Canada

Abstract: This study examined large-scale assessment exclusion perceptions and practices from the perspective of Canadian math teachers. This study applied a descriptive research design involving an online survey to collect data during the height of the COVID-19 pandemic. Findings revealed that teachers' practices did not differ based on the division of education (e.g., primary/elementary or intermediate/senior) in which teachers taught. Findings also revealed that some teachers may have hesitated in recommending excluding students in areas permitted in international and national large-scale assessment guidelines. Further, a small percentage of teachers recommended excluding students in grey areas, such as students who missed a lot of school due to illness or have behaviour issues. It is possible that the small number of students recommended for being excluded from a large-scale assessment in these grey areas influences jurisdictions to exceed the 5% cap on exclusion rates. Providing teachers with short and detailed information on whether to recommend excluding a student from a large-scale assessment, or not, in these grey areas would be helpful.

Keywords: large-scale assessment, exclusion rates, social promotion

Introduction

The importance attached to participation rates in large-scale assessments (LSAs) has received more attention over recent years given large gains in some jurisdictions and the increasing number of non-participating students (Anders et al., 2021; Jerrim, 2021; Miller & Yan, 2022; Miller, et al., 2022). These LSA statistics, particularly from top-performing nations and provincial jurisdictions, have been a focus for policymakers, the general public, academics, and the broader education community (Anders et al., 2021). For example, policymakers use of LSA information has informed policy decision-making and LSA data is used to evaluate the impact of reform in education driven by policy changes (Clarke & Luna-Bazaldua, 2021).

The importance of including all students in LSAs is heightened given that the number of participating countries in the Programme for International Students Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) has increased over the past two decades by 46% and 42% respectively (Hernández-Torrano et al., 2021; CMEC, 2019; Lockheed, 2015). In Canada, the three territories (i.e., Northwest Territories, Nunavut, and Yukon) have not participated in the last four cycles of the Pan-Canadian Assessment Program (PCAP) (i.e., 2023, 2019, 2016, 2013) (PCAP, 2023, 2019, 2016, 2013, 2010, 2007). The Yukon had participated in the 2007 PCAP as well as on PCAP's predecessor assessment, the School Achievement Indicators Program (SAIP), which also included participation from the Northwest Territory (SAIP, 2004 – 1996). Nunavut has not participated in any LSAs. Hence, it is important to recognize that Canadian LSA results have not reflected Canada's northern student population. Despite the reluctance of the Canadian territories to join Canada's LSA program, LSAs have been adopted globally with significant technological advancements in how they are administered. Given that the interest in LSA data has continued to grow, participation in LSAs on an international scale has increased (Lockheed, 2015).

Notwithstanding the growth and use of LSA data in policy and program decision-making, the representativeness of LSA results has been questioned by researchers. This questioning of data reliability has raised a call for more consistent and accountable practice in determining which students should be included or excluded from LSAs (Anders et al., 2021; Jerrim, 2013, 2021; Spaull, 2019; Miller & Yan, 2022). Given the prominent role LSAs play, results should be reliable and have high face validity (Anders et al., 2021). Reliability of data is subsequently dependent on schools' and school boards' adherence to administration guidelines. Examples of such interactions include accurately identifying and deciding which students should be included or excluded from LSAs and what accommodations students can receive (Anders et al., 2021; Jerrim, 2013; Jerim & Sims, 2021; Miller et al., 2022; Spaull, 2019). Common factors influencing these decisions include eligibility criteria, exclusion rate criteria, and data representation.

The review of LSA scores and exclusions rates (Miller & Yan, 2022), followed by interviews with principals (Miller et al., 2022), has provided evidence of a selection bias. The next step in investigating a selection bias on LSAs involves surveying teachers' perceptions and practices in recommending excluding students from LSAs. Hence this

study is intended to contribute to the concerns of selection biasing on LSAs by examining the perceptions and practices of including and excluding students in LSAs from the perspective of mathematics teachers. The focus on mathematics teachers aligns with two previous studies that focused on mathematics teachers for the sole purpose of narrowing the focus of the study to manage the scope of the research project and in no way suggests that selection biasing is or is not occurring in other LSA domains. The two primary research questions posed in this study are:

1. What are teachers' general perceptions and practices towards LSAs?
2. To what extent do mathematics teachers' practices towards excluding or including students on LSAs differ in terms of the division in which they teach (i.e., primary, intermediate, senior)?

Theoretical Framework – Accountability

The theory of accountability underpins LSA. Accountability refers to monitoring and using student performance data to make judgments about schools, school systems, and teacher effectiveness (CMEC, 2019; Jerrim & Sims, 2021). Early observers of LSAs identified accountability as one of the most important frameworks for stakeholders to implement LSAs in various education contexts, be it Britain and Chile, where educational systems are highly centralized, or Canada and the U.S., where education is the responsibility of individual provinces or at the state level (Klinger et al., 2008; Nagy, 2000).

Although accountability in education can be conceptualized and operationalized in different ways (Gilbert, 2012), the key stakeholders within the school system such as teachers, principals, and school boards are held responsible for student learning (Brill et al., 2018). Accountability in the context of this study focuses on teachers and their role in or awareness of the role in deciding which students are included or excluded from LSAs. Although the decision to include or exclude a student from LSAs ultimately lies with the principal (Miller et al., 2022), the onus is on teachers to ensure that principals are informed of student progress and that LSAs are administered according to policy guidelines. Knowing that policy guidelines are vague and often lack consistency (Brzyska, 2018; Gamazo et al., 2019; LeRoy et al., 2019; Miller et al., 2022), there is likely to be variability in teachers' perceptions and practices about who should be included or excluded from a LSA. Subsequently, the practice of including and excluding students from LSAs is also likely to vary depending on one's interpretation of LSA guidelines (Anders et al., 2021; Spaull, 2019).

Literature Review

Given the limited number of studies on LSA exclusion rates in Canada and the absence of literature on mathematics teachers' perspectives, this review of literature focuses on general research about exclusion guidelines as well as international, national, and provincial exclusion rates over the past four assessment cycles and exclusion practices.

LSAs Exclusion Guidelines

Exclusion guidelines across international and national assessments such as PISA, TIMMS, and PCAP are similar. Jurisdictions are permitted to exclude some schools due to either logistical reasons or where there is an expectation that most pupils would not be eligible to participate (CMEC, 2016, 2019; LaRoche et al., 2015). Schools may also decide not to test some of the sampled students. However, the combined total of schools and within-school exclusion numbers must not exceed 5% of the target population (CMEC, 2016, 2019; LaRoche et al., 2015). This 5% ensures that any distortions in international, national, or provincial mean score would not exceed ±5 score points on the various LSAs scale.

PISA exclusion guidelines explicitly state that students should not be excluded solely because of poor academic performance or normal disciplinary problems (CMEC, 2019). It is also noted that students with dyslexia or other such learning *disabilities* should be accommodated in the test situation, if possible, rather than excluded.

Assessment guidelines, in terms of excluding students, have somewhat evolved over the past couple of decades to include a broader classification of disabilities (CMEC, 2016). Despite a broader classification of disabilities, various studies on LSA exclusion rates reported that guidelines were still too vague as shown in Table 1 showing descriptions of exclusion criteria on the international and national LSAs. Of the three assessments and accompanying criteria, the TIMSS exclusion criteria are the most detailed.

Combined with vague guidelines, the absence of monitoring the extent to which guidelines were implemented has led to a large variation in the exclusion of students (Anders et al., 2021; Spaull, 2019; Jerrim, 2013). Such vague exclusion guidelines made it challenging for administrators to accurately identify who should be assessed and who should not (Anders et al., 2021; Spaull, 2019; Miller et al., 2022). The practice of excluding students has also been fraught with misunderstanding due to an absence of clear guidelines and consistency from jurisdiction to jurisdiction (Brzyska, 2018; Gamazo et al., 2019; LeRoy et al., 2019). Again, issues around eligibility can lead to overestimating academic achievement and underestimating educational inequality (Spaull, 2019). It has also been reported that there is little alignment of international, national, and provincial exclusion guidelines (Miller & Yan, 2022).

In Miller et al. (2022)'s study interviewing principals about the practice of excluding students from LSAs, it was revealed that principals had a special role to play in deciding who to include and exclude in LSAs based on input from various sources, including provincial assessment guidelines and teachers. In particular, teachers offered unique student performance perspectives and should be further consulted for exclusion practices in LSAs. The findings revealed in Miller et al.'s (2022) paper suggest that excluding students from LSAs is open to teachers' and principals' interpretation and input.

Given all the vagueness surrounding the exclusion criteria, many researchers have called for greater inclusion of students with disabilities, greater transparency in practices related to excluding students from participating in LSAs, clearly articulated exclusion guidelines on LSAs, and alignment of exclusion guidelines between provincial/state, national, and international assessments (Anders et al., 2021; Jerrim, 2013; Miller et al., 2022).

PISA Exclusion Rates

On the 2012, 2015, and 2018 PISA assessments, most provinces in Canada have exceeded the allowable 5% PISA exclusion, as shown in Table 1 below[1]. Recall that the territories did not participate.

Table 1.1: Provincial and Canadian PISA student unweighted exclusion rates in percent*

Canada and Provinces	*2012*	*2015*	*2018*	*2021(22)*
Newfoundland and Labrador	*9.8*	*5.1*	*5.8*	*1.7*
Prince Edward Island	*8.3*	*14.7*	*7.0*	*1.9*
Nova Scotia	*7.7*	*7.6*	*7.6*	*2.5*
New Brunswick	*8.5*	*9.1*	*5.6*	*2.8*
Quebec	*4.2*	*2.5*	*3.7*	*1.8*
Ontario	*6.1*	*4.5*	*4.7*	*6.5*
Manitoba	*5.5*	*7.4*	*6.3*	*2.2*
Saskatchewan	*4.0*	*5.1*	*4.7*	*3.5*
Alberta	*4.5*	*6.0*	*5.1*	*2.6*
British Columbia	*5.4*	*10.3*	*6.6*	*4.9*

Note. *Unweighted exclusion rates are based on students selected to participate in the LSA but were excluded.
Sources: CMEC, 2019; O'Grady et al., 2016; Brochu et al., 2013)

In 2012, only four provinces, Quebec, Saskatchewan, and Alberta, achieved the 5% exclusion rate threshold. In 2015 there were even fewer provinces meeting the threshold provinces: Quebec and Saskatchewan. On the same 2015 PISA, the exclusion rate of Prince Edward Island and British Columbia was more than double the allowed rate. If a jurisdiction excludes many of its students from participating in the LSAs by deeming them ineligible or because they were absent on the day of the assessment, data will not be useful in making meaningful comparisons with other jurisdictions (Anders et al., 2021; Spaull, 2019).

In summary, current literature has indicated a continued interest in examining exclusion practices on LSAs through an accountability lens. Examining teachers' perspectives regarding exclusion practices on LSAs will add to this body of literature and ideally, contribute to reducing or eliminating potential selection biasing on LSAs.

[1] The 2021 PISA was postponed to 2022 due to COVID-19; hence, the scores are from 2022 rather than the original 2021 testing year.

Methodology

Data Collection

An online survey was created to measure teachers' practices and perceptions regarding including or excluding students on LSAs. The decision to collect responses via an online survey was in response to the COVID-19 pandemic measures that limited an initial plan of distributing surveys in person at a regional conference. A total of 20 mathematics teachers completed the survey.

The survey contained 45 items in total. The first seven items documented teachers' demographic information. Teachers' grade level taught (i.e., elementary, intermediate, or senior) was the grouping variable used to examine differences between the groups. The next 9 items examined teachers' LSA perceptions related to including or excluding students. The last section was the construct focusing on teachers' practices related to recommending including or excluding students on LSAs and contained 28 items. A 5-point rating scale anchored at each end with strongly disagree (1) and strongly agree (5) was used to collect participants' survey responses.

Data Analysis

Descriptive and inferential statistics were used to analyze the data. The descriptive statistics included frequency, percent, mean (*M*), and standard deviation (*SD*) for all survey items. Thus, the inferential analysis applied non-parametric statistics due to the small sample size. The non-parametric Mann-Whitney U test was used to see if there were any differences between the grouping variable, level of education taught, and the median score on the continuous variable measuring perceptions towards LSAs. The Mann-Whitney U test is used when comparing two groups which was the case in this study as there was only one intermediate teacher who was removed for this analysis.

Cronbach's alpha was used to determine the internal consistency of the 28-item practices scale. The alpha coefficient was 0.823. This scale was considered reliable since it was greater than the minimum coefficient of $\alpha = 0.7$ (Huck, 2012; Pallant, 2020).

Findings

Descriptive findings

Demographic Characteristics. The twenty participants from the province of Ontario (ON) had a range of teaching experience from 4 years to more than 20 years of teaching. Eleven participants were teaching mathematics at the elementary level, one at the intermediate level, and eight at the senior level. All participants were teaching mathematics at the time of the survey and had participated in a provincial LSA. Almost all participants were also familiar with the PISA and only 25% were familiar with PCAP while 31% were familiar with TIMMS.

General Perceptions Towards LSAs. The nine items in this section surveyed common perceptions about LSAs. Most participants (n = n14, 70%, *M* = 3.90, *SD* = 1.41) reported that teachers who teach the grade that is assessed feel more pressure to ensure students score well and that not all students should be assessed (n = 12, 60%, *M* = 3.90, *SD* = 1.11). Sixty percent (n = 12, *M* = 2.15, *SD* = 1.18) of teachers do not believe their LSA results reflected their students' abilities, and 12 (*M* = 2.10, *SD* = 0.97) of the teachers did not believe their LSA results were inflated. Further, most teachers (n = 11, 55%) were undecided as to whether the 5% exclusion cap was realistic, and 70% (n = 14) did not believe that more than 10% of students at their school had been excluded from participating in LSAs.

General Practices Related to LSAs. The majority (agreed or strongly agreed: n = 13, 65%, *M* = 3.8, *SD* = 1.28) of participants reported they encouraged students to show up on the day of the LSA. Although 7 (*M* =3.25, SD = 1.29) indicated they do not worry too much about students who do not show up on the day of the assessment, and 5 participants indicated they do worry when students do not show up for the LSA. This finding is aligned with responses indicating almost all teachers did not discourage students from attending on the day of LSA. Continuing with this line of inquiry, the majority of teachers also reported they refrained from providing incentives to encourage students to attend school on the day of the LSA, suggesting that attendance on LSAs is not of a concern. In terms of having received information about LSA exclusion guidelines and have read the guidelines, approximately 30% (n = 6) agreed

or strongly agreed to both items, while 40% (n = 8) indicated they had *not* read the guidelines (i.e. responded with strongly disagree or disagree, $M = 2.75$, $SD = 1.55$)).

Practices Related to Excluding Students in LSAs: Permissible Grounds to Exclude Students. When teachers were asked if they recommended excluding students with intellectual disabilities, socio-emotional difficulties, or a mental disorder, in each of these disabilities, only five teachers reported they would recommend excluding students with these disabilities. Upon examining these results further, most teachers were indecisive in their responses given that the majority responded in the middle of the 5-point scale, possibly because they have not had students with these disabilities in their classes.

When surveyed whether teachers would exclude a student who does not read or write the language of assessment, 40% (n = 8) of teachers (agreed or strongly agreed, $M = 3.30$, $SD = 1.42$) would exclude a student because of this difficulty. It is possible that more teachers did not recommend excluding a student because the student did not read or write the language of assessment, given that second language learners may be able to decipher what is being asked in a math assessment without being competent in the language of the LSA.

Inferential Findings

Mann Whitney U test was used to determine if there was a statistically significant difference in mean ranks between primary and senior teachers' responses to the 28-item practices scale. This test was not significant at the $p < 0.05$ level ($\text{MeanRank}_{\text{primary}} = 10.73$; $\text{Mean Rank}_{\text{senior}} = 7.53$).

Discussion and Conclusion

Research Question 1: What are teachers' general perceptions and practices towards LSAs?

The lens of teacher accountability presented in earlier studies (Brill et al., 2018; Jerrim & Sim, 2021; Klinger & Rogers, 2011), continues to resonate with teachers in this study given that 70% of teachers indicated that teachers who teach the grade that is assessed, feel more pressure to ensure students score well. There is also the perception that the 5% LSAs exclusion rate may not be a realistic benchmark for schools (30%, n = 6, $M = 3.25$, $SD = 1.02$). The 30% (n = 6) of teachers who did not support a 5% exclusion rate combined with 55% (n = 11) of teachers who responded in the middle of the 5-point scale is suggestive that ON teachers may not be supportive of this LSA guideline. This belief may have influenced ON's exclusion rates on the PISA, where ON's exclusion rates have been higher than the allowed 5% in three of the four past PISA cycles (CMEC, 2019; O'Grady et al., 2016; Brochu et al., 2013).

In contrast to a lack of agreement for the 5% exclusion cap, teachers were not overwhelmingly in favour of excluding students in the permissible areas such as intellectual disabilities (mental or emotional) (PISA, 2022). If teachers were not in favour of the 5% exclusion cap, then more teachers may have been in favour of excluding students with these learning challenges; however, only five teachers reported that they would exclude students because of intellectual, socio-emotional difficulties, or mental disorders. Students can also be excluded if they cannot speak/read the language of instruction (CMEC, 2022). Similarly, only eight teachers in this study indicated they would exclude a student from a LSA with this learning difficulty. In these areas where students can legitimately be excluded, teachers tended to be conservative in recommending that students be excluded. It is possible that schools only have a few students with severe disabilities in these areas, and subsequently, teachers have little experience in recognizing the need to exclude students with these disabilities, or they provided a neutral response to the survey question.

Further, it is possible that grey areas or areas not well articulated in LSA guidelines, may be influencing decisions to recommend excluding students from LSAs and subsequently, raising the 5% exclusion rate cap. Of the six grey areas surveyed (i.e., excluding a student because they are below the current grade (one grade below or two grades below), can only read a little bit of the language, behaviour issues, emotional issues, missing too much school, cannot sit for the duration of the assessment), the majority of teachers reported they were apt to exclude a student who is two grades below the grade being assessed. This finding is echoed in an earlier study by Miller and Yan (2022), who connected high exclusions with social promotion in that many students were being socially promoted to the next grade without achieving the learning outcomes of the current grade. If the same student is repeatedly socially promoted, it is feasible that the student would be operating a grade or more below their peers, and the teacher would subsequently

exclude the student from the LSA. The issues with social promotion in mathematics may be more problematic than in other disciplines, given that mathematics can be viewed as a scaffolded discipline. In this discipline, learning a concept is built on prior knowledge; therefore, it is feasible to question whether it is possible for a student to catch up when they have fallen far behind their peers given the restraints of teaching a diverse classroom. This practice of excluding students who are a grade or more below the level being assessed and exclusion of students in the other grey areas, is likely to increase the exclusion rate above the 5% cap and contribute to a selection bias.

It is likely that social promotion and the consequences and impact on LSA exclusion rates will stay, which has contributed to the high exclusion rates given that students are operating below grade level. Hence, finding a different mechanism to address this educational practice is important. Providing an option for schools to report socially promoted students who should have otherwise participated in the assessment but were excluded because they were working a grade or more below the grade level being assessed would capture a more accurate picture of students who are participating in the LSA and those who do not. Alternatively, the practice of social promotion could be withheld in disciplines like mathematics and possibly other courses that build new learning on prior learning given the challenges involved in catching up.

Last, it is important to report that although recommendations to exclude a student from a LSA practice did not always align with the LSA guidelines combined with previous studies acknowledging that exclusion rate guidelines were too vague (Anders et al., 2021; Spaull, 2019; Miller et al., 2021), only 20% of teachers indicated that schools should create their own guidelines. Hence, improved clarity in excluding students who fall into the grey areas would be beneficial. TIMSS's exclusion guidelines provide more guidelines, particularly for intellectual disabilities, which could be adopted by other LSAs (TIMSS, 2019).

In summary, the data suggests some misunderstanding regarding when a student can be excluded from a LSA. If only 55% (n = 11, M = 2.95, SD = 1.57) of teachers received information about exclusion rate guidelines and only 30% (n = 6, M = 2.75, SD = 1.55) of teachers reported reading the exclusion guidelines, it is reasonable to expect variation in teachers' exclusion rate recommendations. Given this finding, creating an infomercial that clarifies which students should be excluded from a LSA which should not is worth a financial investment.

Research Question 2: To what extent do mathematics teachers' practices towards excluding or including students on LSAs differ in terms of the division in which they teach (i.e., primary, intermediate, senior)?

Although we hypothesized that there would be a significant difference in teachers' responses to the perceptions scale based on the education level taught, there was not. It is possible that the small sample size influenced this finding or that teachers' responses simply do not differ based on the level of education taught.

Limitations

The COVID-19 pandemic greatly hindered the sample size of this study. Although, this study provides some insight into Ontario mathematics teachers' exclusion perceptions and practices, a larger and broader sample would confirm the findings presented in this paper. In addition, adding a response category for "I don't know" or "Have not experienced this" might help reduce the number of neutral responses in the middle category on the 5-point response scale. Adding the extra response category would require a much larger sample given that several records would not be included when calculating the total scale score due to the missing value(s).

REFERENCES

Anders, J., Has, S., Jerrim, J., Shure, N., & Zieger, L. (2021). Is Canada really an education superpower? The impact of non-participation on results from PISA 2015. *Educational Assessment, Evaluation and Accountability*, *33*(1), 229–249.

Brill, F., Greyson, H., Kuhn, L., & O'Donnell, S. (2018). What Impact Does Accountability Have on Curriculum, Standards and Engagement in Education? A Literature Review. *National Foundation for Educational Research.* The Mere, Upton Park, Slough https://files.eric.ed.gov/fulltext/ED590506.pdf

Brochu, P., Deussing, M. A., Houme, K., & Chuy, M. (2013). *Measuring up: Canadian results of the OECD PISA study: The performance of Canada's youth in mathematics, reading and science: 2012 first results for Canadians aged 15. Council of Ministers of Education, Canada*. http://www.cmec.ca/Publications/Lists/Publications/Attachments/318/PISA2012_CanadianReport_EN_Web.pdf

Brzyska, B. (2018). Trends in exclusion rates for students with special education needs within PISA. Oxford *Review of Education 44*(5), 633-650. DOI:10.1080/03054985.2018.1496907

CMEC. (2016). *Measuring up: Canadian results of the OECD PISA 2015 study.* Retrieved from https://www.cmec.ca/Publications/Lists/Publications/Attachments/389/PISA2015_CPS_EN.pdf

CMEC. (2019). Measuring up: Canadian Results of the OECD PISA 2018 Study Retrieved from https://www.cmec.ca/Publications/Lists/Publications/Attachments/396/PISA2018_PublicReport_EN.pdf

CMEC. (2022). Measuring up: Canadian Results of the OECD PISA 2022 Study. Retrieved from https://www.cmec.ca/Publications/Lists/Publications/Attachments/438/PISA-2022_Canadian_Report_EN.pdf

Clarke, M., & Luna-Bazaldua, D. (2021). *Primer on Large-Scale Assessments of Educational Achievement.* World Bank Publications. World Bank Publications. https://openknowledge.worldbank.org/entities/publication/618b21d1-e54e-5a94-a088-b7ed8965a990

OECD. (2013). Measuring up: Canadian results of the OECD PISA 2012 study. https://cmec.ca/Publications/Lists/Publications/Attachments/318/PISA2012_CanadianReport_EN_Web.pdf

OECD. (2016). Measuring up: Canadian results of the OECD PISA 2015 study. https://www.cmec.ca/Publications/Lists/Publications/Attachments/371/PISA2015_FL_EN.pdf

OECD. (2019). *Measuring up: Canadian results of the OECD PISA 2018 study*. https://www.cmec.ca/Publications/Lists/Publications/Attachments/396/PISA2018_PublicReport_EN.pdf

OECD. (2023). *Measuring up: Canadian Results of the OCED PISA 2022 study. https://www.cmec.ca/Publications/Lists/Publications/Attachments/438/PISA-2022_Canadian_Report_EN.pdf*

Gamazo, A., Martínez-Abad, F., Olmos-Migueláñez, S., & Rodríguez-Conde, M. J. (Eds.). (2019). Assessment of students with special education needs in PISA: Participation guidelines and data treatment. In Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'19). Association for Computing Machinery, 187–191. https://doi.org/10.1145/3362789.3362919

Gilbert, C. (2012). *Towards a self-improving system: the role of school accountability*. https://dera.ioe.ac.uk/14919/1/towards-a-self-improving-system-school-accountability-thinkpiece%5B1%5D.pdf

Hernández-Torrano, D., & Courtney, M. G. (2021). Modern international large-scale assessment in education: an integrative review and mapping of the literature. Large-scale Assessments in Education. *Large-Scale Assessments in Education*, *9*(1), 1–33. https://doi.org/10.1186/s40536-021-00109-1

Huck, S. W. (2012). *Reading statistics and research. Pearson Education.* (Fifth). Pearson Education.

Jerrim, J. (2021). PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK? *The Review of Education, 9*, e3270. https://bera-journals.onlinelibrary.wiley.com/doi/10.1002/rev3.3270

Jerrim, J., & Sims, S. (2021). School accountability and teacher stress: international evidence from the OECD TALIS study. *Educational Assessment, Evaluation and Accountability*, 1–28. https://link.springer.com/article/10.1007/s11092-021-09360-0

Jerrim, J. (2013). The reliability of trends over time in international education test scores: Is the performance of England's secondary school pupils really in relative decline? *Journal of Social Policy*, *42*(2), 259–279. https://johnjerrim.com/wp-content/uploads/2013/07/final_paper.pdf

Klinger, D., Deluca, C., & Miller, T. (2008). The evolving culture of large-scale assessment in Canada. Canadian Journal of Educational Administration and Policy, 76, 1–34. https://eric.ed.gov/?id=EJ807064

Klinger, D., & Rogers, T. (2011). Teachers' perceptions of large-scale assessment programs within low-stakes accountability frameworks. *International Journal of Testing, 11*(2), 122-143. https://www.researchgate.net/publication/239794802_Teachers'_Perceptions_of_Large-Scale_Assessment_Programs_Within_Low-Stakes_Accountability_Frameworks

LaRoche, S., Joncas, M., & Foy, P. (2015). *Sample design in TIMSS 2015. Methods and Procedures in TIMSS, 2016, 3-1*. https://isc.bc.edu/timss2019/methods/pdf/T19_MP_Ch3-sample-design.pdf

LeRoy, B. W., Samuel, P., DeLuca, M., & Evans, P. (2019). Students with special educational needs within PISA. Assessment in Education: Principles, Policy & Practice, 26(4), 386–396. https://doi.org/10.1080/0969594X.2017.1421523

Lockheed, M. (2015). *Why do countries participate in international large-scale assessments? The case of PISA. The Case of PISA (October 19, 2015). World Bank Policy Research Working Paper, (7447).* https://ssrn.com/abstract=2676416

Miller, T., FitzGerald, A. M., & Xu, H. (2022). The practice of excluding students from large-scale assessments: Interviews with principals. *Alberta Journal of Educational Research, 68*(2), 1-20. https://doi.org/10.11575/ajer.v68i2.70880

Miller, T., & Yan, Y. (2022). Factors influencing the accuracy of Canada's large-scale assessment data: Policies and practices of exclusion, absenteeism, and social promotion. *Educational Assessment, Evaluation, and Accountability, 35*(2). *https://link.springer.com/article/10.1007/s11092-022-09381-3*

Nagy, P. (2000). The three roles of assessment: Gatekeeping, accountability, and instructional diagnosis. Canadian Journal of Education 25(4), 262-279. https://www.jstor.org/stable/1585850

O'Grady, K., Deussing, M., Scerbina, T. Fung, K., & Muhe, N. (2016). *Measuring up: Canadian results of the OECD PISA study. 2015 first results for Canadians aged 15*. *https://www.cmec.ca/publications/lists/publications/attachments/365/pisa2015-cdnreport-en.pdf*

PCAP (2023). Overview. https://www.cmec.ca/754/PCAP_2023.html

PCAP (2019). PCAP 2019 Mathematics, Reading, and Science Highlights. https://www.cmec.ca/docs/pcap/pcap2019/PCAP2019-Highlights-EN.pdf

PCAP (2016). PCAP 2016 Reading, Mathematics, and Science Highlights. https://www.cmec.ca/docs/pcap/pcap2016/PCAP2016_Highlights_EN.pdf

PCAP (2013). PCAP 2013 Science, Reading, and Mathematics Highlights. https://www.cmec.ca/docs/pcap/pcap2013/PCAP-2013-Highlights-EN.pdf

PCAP (2010). PCAP 2010 Mathematics, Science, and Reading Highlights. https://www.cmec.ca/docs/pcap/pcap2013/PCAP-2013-Highlights-EN.pdf

PCAP (2007). PCAP 2007 Reading, Mathematics, and Science Highlights. https://www.cmec.ca/docs/pcap/pcap2007/PCAP2007-Highlights.en.pdf

PISA (2024). How does PISA shape education reform? https://www.oecd.org/pisa/aboutpisa/

Pallant, J. (2020). *Survival manual. A step-by-step guide to data analysis using SPSS, 7*.

Spaull, N. (2019). Who makes it into PISA? Understanding the impact of PISA sample eligibility using Turkey as a case study (PISA 2003–PISA 2012). *Assessment in Education: Principles, Policy & Practice*, *26*(4). https://doi.org/10.1080/0969594X.2018.1504742

TIMSS. (2019). TIMSS 2019 Canadian Results from the Trends in International Mathematics and Science Study. https://nces.ed.gov/timss/

ABOUT THE AUTHORS

Alex Yankey: Mr. Alex Yankey is a recent graduate of the Master of Education program at the University of Prince Edward Island. Alex has taught mathematics internationally and throughout Canada. His interests are in large-scale assessment of mathematics and mathematics pedagogy.

Tess Miller: Dr. Tess Miller is a Professor of Education at the University of Prince Edward Island. Her research focuses on large-scale assessment, mathematics education, and the use of artificial intelligence to support student learning in mathematics. She has co-authored on four mathematics textbooks for grades 9 to 12 and was a pedagogy and assessment consult on seven additional mathematics textbooks. Most recently, she has published three papers on mathematics anxiety and has several studies underway examining the use of artificial intelligence to supporting student learning in mathematics. Through her work, she aims to support and promote graduate students in conducting and publishing their research. Mr. Alex Yankey is one of her graduate students.