

ALBERTA'S STUDENT LEARNING ASSESSMENT PROGRAM

An Integrated Evaluation

Executive Summary

April 2017

David Slomp

*Board of Governors' Teaching Chair
University of Lethbridge*

Norbert Elliot

*Research Professor
University of South Florida*

Richelle Marynowski

*Tier II Board of Governors' Research Chair
University of Lethbridge*

Alex Rudniy

*Assistant Professor
Fairleigh Dickinson University*

EXECUTIVE SUMMARY

David Slomp, *University of Lethbridge*
Norbert Elliot, *University of South Florida*
Richelle Marynowski, *University of Lethbridge*
Alex Rudniy, *Fairleigh Dickinson University*

Core Finding

Alberta Education's shift toward the development of large-scale formative assessment programs, we believe, represents an exciting new direction for assessment in the province. Because assessment itself is a process made valuable when communities gather to advance student learning, the Student Learning Assessment (SLA), has been enriched by creating a culture of evidence used to design a valid, reliable, and fair assessment. Though there have been identified challenges with the development and implementation of the SLA program, these have been managed well, addressed competently, and resolved so that the program may continue to benefit all stakeholders.

The SLA program represents an innovative advance for large-scale assessment practice in Alberta. The SLA's digital platform holds the potential to advance all of Alberta Education's large-scale assessment programs. The multimodal capabilities of this platform will vastly expand the curriculum outcomes and construct facets currently being assessed by Alberta's Provincial Achievement Test (PAT) and Diploma Exam programs. Applied broadly as part of an expanding culture of evidence, these same features can reduce construct irrelevant variance, expand construct coverage, and usher in a new era of student performance data.

Understood to be the innovation it is, the SLA program holds the potential to serve as an incubator for the design of digital versions of PATs and Diploma Exams. If issues related to validity, reliability, and fairness continue to be addressed within a culture of evidence, opportunity advancement can be managed for all populations Alberta Education currently assesses.

The research team holds that this investment in Alberta's assessment future should be recognized for its innovative quality and its commitment to student learning, should continue, and should serve as a model for the development of other assessment programs. It is clear from our study that the SLA program holds great potential, and that challenges can be met through good will, informed pedagogy, technical expertise, and sustained commitment to student learning.

If the SLA program is to achieve its potential, it is important to recognize that students and teachers are first among equals of all stakeholder groups. As the design of the SLA moves forward, positive intended outcomes for these stakeholders should be enhanced, while negative unintended outcomes should be identified and mitigated.

Background

In 2013, the Government of Alberta announced that it was ending the Grade 3 PAT program and that beginning in the 2014/15 school year a new SLA program would take its place. The stated goal of the SLA program was to (a) improve student learning, (b) enhance instruction for students, and (c) assure Albertans that their education system was meeting the needs of students and achieving the outcomes of the *Ministerial Order on Student Learning*. In the 2016/17 school year, this third purpose was officially dropped from the SLA program; as a consequence, results from the SLA program will not be used as part of Alberta's Accountability Pillar system.

This shift from the Grade 3 PAT program to the SLA program represents an important change in the culture surrounding large-scale assessment practices in Alberta. The program represents a shift both from paper and pencil assessment to truly digital assessments and from summative assessment to formative assessment.

This new direction reflects a growing body of research into the power of formative assessment for fostering improvements in student growth and achievement. In their influential review of formative assessment literature, Black and Wiliam (1998) found that when properly designed and used, formative assessment provides the largest ever reported achievement gains for any form of educational intervention (p. 61).

Alberta Education has chosen to develop the SLA program through a series of pilots and it is "committed to remaining in pilot phase until the SLA program is deemed a valuable tool for teachers, parents, and students" (RFP No AE-16-05, p. 5). In conjunction with the 2016 pilot this study was commissioned to examine the following questions:

- ◆ *To what degree are the SLA results valid and reliable?*
- ◆ *Is the Grade 3 SLA deemed valuable by the 20 participating school authorities? Why or why not? What elements (e.g. beginning of the year assessment instrument, digital questions and performance tasks) of the Grade 3 SLA are deemed valuable?*
- ◆ *How are SLA implementation approaches and processes being used internally and among and between each of the participating school authorities, Alberta Education staff and/or the Alberta Regional Professional Development Consortia (ARPDC)?*
- ◆ *What were the successes and the opportunities for improvement relative to the 2016/17 Grade 3 SLA pilot?*
- ◆ *What are the lessons learned from the SLA pilot?*
- ◆ *How can the Grade 3 SLA be improved?*

Our study design was based on an Integrated Design and Appraisal Framework (Slomp, 2016) developed to guide the collection of evidence related to the validity, reliability, and fairness of an assessment program when determining its efficacy. Following this design, we collected and analyzed the following sets of data:

- ◆ We conducted a systematic analysis of literature on the literacy and numeracy constructs at the grade 3 level;
- ◆ We engaged in an expert analysis of SLA items;
- ◆ We collected and analyzed documentation on the SLA program provided by Alberta Education;

- ◆ We collected and analyzed archival data provided by Alberta Education and some participating school authorities.
- ◆ We interviewed Alberta Education and ARPDC personnel;
- ◆ We surveyed school authority personnel, school leaders and grade three teachers in participating school authorities; and
- ◆ We conducted construct remodeling interviews with grade three students.

Our full report provides data and analysis to support the responses to Alberta Education's questions. A summary response to these questions follows.

To what degree are the SLA results valid and reliable?

Key Finding: *Built on contemporary Item Response Theory models, the SLA produces evidence of a valid, reliable, and fair assessment. Specifically, the assessment serves its intended purpose of providing information about student ability in targeted constructs, is administered consistently over time and geographic region, and advances opportunity to learn for students, teachers, parents, administrators, and policy makers. Overall student populations, as well as subgroups of students, are well served by the assessment. Opportunities for further investigation arise when performance data is disaggregated by subgroups (Level 1-5, gender¹). The overall performance of the SLA will be enhanced as issues potentially related to construct irrelevant variance and construct underrepresentation are addressed.*

- ◆ The SLA performance data is being used for the purposes intended.
- ◆ The SLA program will benefit by construct expansion in terms of key literacy and numeracy construct facets, progression elements, and curriculum outcomes.
- ◆ Assessing grade three students reliably is a difficult undertaking because these students have limited test-taking experience and nascent literacy skills. Construct remodeling interviews with students revealed the current items will benefit from further development to minimize construct irrelevant variance and expand construct representation.
- ◆ Both the literacy and numeracy models perform well for overall student groups.
- ◆ Subgroup analysis by performance level (Level 1-5 as assigned by SLA theta-scores) raises questions that require further study. The strength of the literacy and numeracy models lessen when disaggregated by student level. Statistically significant correlations are largely absent among the items in English literacy performance at Levels 1, 2, 4, and 5. Statistically significant correlations are largely absent in numeracy performance at Levels 1, 4, and 5.
- ◆ Subgroup analysis by gender revealed a statistically significant difference between course grades and performance scores of male and female students in literacy. This pattern is common in literacy assessment. Coefficients of determination ($p < .001$) are present for both male and female students, with the model accounting substantively for variance.

¹ Datasets received by the research team contained student names but gender data were not included. To retrieve this information, we used R Language with its Gender package, implementing four lookup methods based on: (a) U.S. Social Security Administration baby name statistics, (b) U.S. Census data from the Integrated Public Use Microdata Series, (c) the Kantrowitz corpus of female and male names, and (d) Genderize.IO API based on data collected from user profiles from online social networks. This method should not be taken as a way to provide evidence about sub-group performance; rather, it should be taken as a proof of concept of information that may be gained through such analyses. No inference is intended for any individual student, and no conclusions should be drawn for any student.

- ◆ The numeracy model performs well for student subgroups.
- ◆ In numeracy, mixed patterns of performance difference between male and female students appear not to be related to numeracy ability of one subgroup: no statistically significant difference between the numeracy course grades of male and female students exists, yet performance differences do exist on the assessment. Coefficients of determination ($p < .001$) are present for both male and female students, with the model accounting substantively for variance.

Is the Grade 3 SLA deemed valuable by the 20 participating school authorities? Why or why not? What elements (e.g. beginning of the year assessment instrument, digital questions and performance tasks) of the Grade 3 SLA are deemed valuable?

Key Finding: *The SLA program is deemed valuable by the school authorities participating in the 2016 pilot. While the degree of appreciation varies by stakeholder group, especially noteworthy is stakeholder appreciation for the SLA's performance tasks and reporting capabilities.*

- ◆ Superintendents and principals value the SLA program more highly than teachers do. Currently, 44% of teachers (n=122) and 56% of principals (n=54) believe that the data generated by the SLA program is worth the time and resources required to collect. We believe that as the technical quality of the SLA program is enhanced, and as time to administer the program is addressed, higher levels of more broadly-based support for the program can be achieved
- ◆ Changes to the SLA program were valued by teachers, principals and superintendents as having improved the program. Changes that were most appreciated include:
 - Digital Assessments are marked within 24 hours of the assessment being completed.
 - A more flexible administration period was introduced in 2016.
 - The SLA can be re-administered throughout the year to assist in measuring growth.
 - Principals and teachers are permitted to exercise professional judgement to opt students out of the assessment if warranted.
- ◆ The SLA program is deemed valuable by the school authorities participating in the 2016 pilot. This belief in the SLA's value is held by a strong majority of jurisdiction leaders (90%), but this belief diminishes the closer to the classroom one is situated; 70% of principals (n=57), and 49% of teachers (n=126) believe that participating in the SLA program benefits their school or students. Similarly, 100% of jurisdiction leaders (n=10), 80% of principals (n=56) and 56% of teachers (n=122), report using information from the SLA program to design instructional interventions to improve literacy and numeracy.
- ◆ The SLAs are valued by jurisdiction leaders, school leaders, and teachers because they provide data early in the year that can be used to allocate resources, focus professional development activities, and plan targeted instruction.

- ◆ Performance tasks are valued because the collaborative marking sessions associated with them provide opportunities for professional development and collaboration. Teachers report using the performance tasks to discover weaknesses in literacy and numeracy that include students' ability to:
 - Attend to punctuation and conventions of language,
 - Brainstorm,
 - Develop supporting details,
 - Organize ideas,
 - Plan a writing task,
 - Structure sentences and paragraphs,
 - Solve word problems,
 - Employ mathematical reasoning.
- ◆ The reports generated by the SLA program are valued because jurisdiction leaders, principals, and teachers believe that they provide information that accurately reflects student literacy and numeracy abilities (84% of superintendents (n=9), 74% of principals (n=54), 70% of English Language Arts teachers (n=94), and 64% of Mathematics teachers (n=87) believe this). Similarly, 89% of jurisdiction leaders (n=9), 85% of principals (n=54), 78% of Mathematics teachers (n=87), and 63% of English Language Arts teachers (n=94) felt that information provided in the SLA reports is useful to them in their professional roles.

How are SLA implementation approaches and processes being used internally and among and between each of the participating school authorities, Alberta Education staff and/or the Alberta Regional Professional Development Consortia (ARPDC)?

Key Finding: *While an implementation plan is in place, site-embedded, sustained professional development regarding score interpretation and use is a goal for the next assessment cycle.*

- ◆ A strong majority of all superintendents, principals, and teachers agreed that supports from their school authority, school leaders, or teacher colleagues helped them to better administer the SLA program. Principals and Superintendents also valued the support they received from Alberta Education.
- ◆ Teachers rated support received from school leaders more highly than supports from any other source. While principals and superintendents rated supports from their peers more highly than supports received from any other source.
- ◆ Supports valued by these stakeholder groups include the following:
 - The SLA Client Services Help Desk,
 - Alberta Education conference calls,
 - The online FAQ and resource manual provided by Alberta Education,
 - Conversations, meetings, and PD sessions with colleagues,
 - Collaborative marking sessions,
 - Online rubrics, exemplars, and sample tasks,
 - Conversations with school leaders who understood the SLA program.
- ◆ Teachers require more time both for collaborative marking sessions and for professional development focused on interpreting and using SLA performance data. Future professional development should be site-based and sustained throughout the school year.

What were the successes and the opportunities for improvement relative to the 2016/17 Grade 3 SLA pilot?

Key Finding: *The SLA platform offers a significant advancement for the future of large-scale assessment in Alberta. Indeed, there is evidence that the SLA pilot can serve as an international model for large-scale, construct-rich, digitally-driven formative assessment.*

- ◆ The ability to utilize any combination of audio, video, and text to present students with an assessment task creates the potential to measure a broader range of curriculum outcomes and construct facets than is possible for traditional paper-based items. These same capabilities also enable test designers to minimize construct irrelevant variance by reducing the reliance on print-based literacy for all test items.
- ◆ The reporting platform embedded in the SLA program offers incredible functionality, enabling teachers to sort data by outcomes, difficulty levels, and literacy elements.
- ◆ There is clear alignment between the intended uses of the SLA program (to provide formative data to support teaching and learning) and the ways in which teachers, principals, and superintendents report using this data. Each stakeholder group reports using data from the SLA program in the following ways:
 - To foster collaborations across grade levels, schools, and school authorities.
 - To gain insight into individual student's strengths and weaknesses in literacy and numeracy.
 - To organize student groups for targeted activities.
 - To design lessons and activities targeted at individual or group learning needs.
 - To plan one-on-one interventions for students.
 - To develop district level professional learning plans.
 - To plan literacy and numeracy programs in schools.
 - To inform school development plans.
 - To guide PLC work.
 - To communicate with parents.
 - To establish learning goals with students.
 - To confirm or contextualize other assessment information or decisions.
- ◆ Superintendents, teachers and principals appreciate how the SLA design team has responded to feedback in each pilot phase of the program.

What are the lessons learned from the SLA pilot?

Key Finding: *The innovation of the SLA pilot has been accompanied by gains and challenges that are identifiable and therefore resolvable. The innovations accompanying the SLA have provided eight lessons; the challenges have provided nine lessons.*

Innovation Lessons:

- (1) The great innovation in the SLA program is its intent to serve as a formative assessment program for use on a large scale. This innovation will be further strengthened by creating a range of testlets, blueprinted to specific areas of need identified by teachers, to provide fine-grained information on areas of student strength and need for growth.
- (2) All stakeholder groups appreciate the ongoing innovation in the SLA program. Continued innovation in response to educators' concerns will strengthen its value for these stakeholders.
- (3) A majority of superintendents, principals, and teachers believe that the SLA program supports them in achieving core functions associated with formative assessment. While support for the SLA program is less strong among teachers than it is among school and jurisdiction leaders, continuing to develop the formative assessment capacity of the SLA program will enhance its support among these stakeholders.
- (4) Unlike summative assessments traditionally administered by Alberta Education, formative assessment programs will benefit by providing fine-grained information regarding the strengths and weakness of individual students. To achieve this aim, each test item needs to be rigorously evaluated to ensure it is measuring the skills it was designed to measure. Assessments also need to be blueprinted by modality and curriculum outcomes, literacy element, and construct facet.
- (5) The SLA's digital platform enables the design of complex multimodal items. When all test items are multimodal, however, the information value of student performance data is reduced because it becomes difficult to determine the effect that item modality exerts on student performance. Performance data, then, cannot inform teachers if students have strengths or weaknesses processing audio, visual, or linguistic text. As such, a broader view of the constructs associated with literacy and numeracy is in order.
- (6) The reports generated by the SLA program are rich and detailed. Nevertheless, they can be overwhelming for teachers. Site-embedded and sustained professional development support focused on how to interpret and use this data will benefit the next assessment cycle.
- (7) Supports for administering the SLA program provided by Alberta Education, school authorities, school leaders, and colleagues were valued by superintendents, principals and teachers.
- (8) Principals are a linchpin within the SLA program. Principals tended to bridge the gap between superintendent and teacher perspectives regarding the value and utility of the SLA program. Channeling site-embedded and sustained PD programming through this stakeholder group should be considered.

Challenge Lessons:

- (1) Students at the grade three level are a challenging population to reliably assess. Given their limited test writing experience and their nascent literacy skills—as well as their inexperience with test taking in digital environments—students at this grade level require a high level of support when responding to test items.
- (2) For the next assessment cycle, items should be studied carefully and, where necessary, redesigned to provide better guidance and support for students.
- (3) Teachers express concern regarding the quality of information received from the SLA program. They also indicated that the time required to administer the SLA program outweighs the value of the information provided by the program. This is an issue that needs to be addressed.
- (4) Literacy and numeracy are large, complex constructs that are difficult to measure well. The SLA assessments will benefit by further investigation of factors that may potentially underrepresent given constructs. The SLA will benefit by working with concurrent innovation regarding literacy and numeracy constructs, literacy and numeracy progressions, and English language arts and Mathematics curriculum. The SLA program will benefit by enhanced emphasis on measuring metacognitive dimensions of literacy and numeracy as these provide the foundation for students' long-term growth and development as literate and numerate beings.
- (5) There remains a concern from jurisdiction leaders, principals and teachers that the SLA will one day be used for accountability purposes. Alberta Education should continue to address these concerns through communication that reaffirms Alberta Education's commitment to using the SLA program for formative assessment purposes only. Alberta Education can also address this concern by ensuring that data generated by the SLA program is protected from improper and invalid uses by external special interest groups.
- (6) Current PD support for the SLA program was deemed insufficient for the challenges of administering the SLA program, marking SLA performance tasks, and learning how to interpret and use SLA student performance data. Alberta Education should explore site-embedded, sustained PD models to support the administration of the SLA program and to support educators as they learn how to interpret and use student performance data.
- (7) Alberta Education's SLA design team will benefit from the development and communication of a common vision for the SLA program. Parallel to developing this vision, Alberta Education personnel will benefit from the development of a set of principles, and parameters, grounded in current assessment theory for making decisions regarding evidence centered design that will promote the sustainable evolution of this program.
- (8) Data regarding students' duration on each task and their number of visits to each task will offer insights into the intrapersonal domain, providing evidence of students' conscientiousness or on-task behavior. The present analysis demonstrates, for example, that there is disjuncture between duration, number of visits, and performance. In future assessment cycles, it will be useful to investigate what can be learned about students from duration and visit data. Used formatively, this data will provide valuable instructional information to teachers and assessment designers.
- (9) It will benefit SLA leaders to consider what can be learned from studying subgroup performance in literacy and numeracy in order to advance opportunity to learn.

How can the Grade 3 SLA be improved?

Key Finding: *Issues related both to the administration of the SLA digital assessment, and to the information value of the SLA should be addressed in the next phase of the SLA's implementation.*

Recommendation 1:

The SLA program should continue and should be given the resources necessary to pursue continued innovation focused on advancing student learning through formative assessment.

Recommendation 2:

Alberta Education should reaffirm its commitment to using SLA performance data only for formative assessment purposes.

Recommendation 3:

The SLA design team and those who manage this team should develop a common vision for the future of the SLA program. This vision should be anchored in the current *Standards for Educational and Psychological Measurement* (AERA, APA, NCME, 2014) and in contemporary validation models derived from these standards.

Recommendation 4:

Alberta Education should commit to an extended program of research focused on the design and appraisal of digital test items.

Recommendation 5:

Alberta Education should commit to a program of research that focuses on the issue of fairness as it relates to issues of validity and reliability. This program of research should examine performance by subgroup in both literacy and numeracy, to understand (a) why models are substantially weaker for some performance groups, (b) to identify ways to strengthen the assessment models through item refinement, and (c) to draw lessons from high performing groups that could be applied to enhancing the performance of other subgroups.

Recommendation 6:

Alberta Education should commit to funding site-embedded and sustained professional development of teachers. This PD should focus both on collaborative marking, and on interpreting and using student performance data.

Recommendation 7:

In future assessment cycles, the SLA program should attend closely to enhancing construct validity and reducing construct irrelevant variance by committing to the following actions:

- (1) Develop and communicate clear construct models for literacy and numeracy that can inform all future design decisions.
- (2) Blueprint all SLA items to construct facets, literacy and numeracy progressions, and curriculum outcomes.
- (3) Design monomodal items (print, linguistic, or oral text only) in addition to current multimodal items and then blueprint items by modality.
- (4) Design practice items that teach students how to process and complete digital SLA items.
- (5) Revise current SLA items to expand construct coverage and reduce irrelevant variance.

Recommendation 8:

Embrace a truly formative assessment design by creating multiple parallel-form testlets yielding similar evidence of information related to fairness, validity, and reliability, mapped to areas of information need identified by teachers. Enable teachers to assess students multiple times each year using different validated testlet forms to measure the same outcomes and curriculum facets. Consider permitting teachers to choose which of these testlets they will administer to their students. In a truly formative assessment, not all students need to complete an identical set of items.

Recommendation 9:

Create an assessment interface that significantly reduces the time it takes for students to access the SLA.

Recommendation 10:

Create jurisdiction and school level reports that have the same functionality as classroom-level reports.

Recommendation 11:

Create a resource allocation network in which school authorities will leverage resources to provide teachers sufficient time for site embedded, sustained PD so that they can engage both in collaborative marking sessions and in training on how to interpret and use SLA student performance data.

✎ Acknowledgements ✎

A study of this size and complexity could not be completed without a small army of research assistants. The authors of this study would like to acknowledge the following individuals for their outstanding work and their commitment to this project:

Doug Checkley, University of Lethbridge

Julie Corrigan, University of Ottawa

Jenna Lowe, University of Lethbridge

Raphaël Gani, University of Ottawa

Lisa Prawdzik, University of Lethbridge

Kaeli New, University of Lethbridge

Calvin deHaan, University of Lethbridge

Brenna Quigley, University of Ottawa

Spencer Pinnock, University of Lethbridge

Len Sproule, University of Lethbridge

Josh Markle, University of Lethbridge

Landry Forand, University of Lethbridge

We would also like to thank the **18 school authorities** that participated in this study. Your collegial spirit, and your investment of time, energy, and resources made this study possible.

Thank you to the many **central office personnel** who helped facilitate requests for access or information.

Thank you to **all the superintendents, principals and teachers** who took the time to complete a detailed survey.

Thank you also to the **principals and teachers at the nine schools** who invited our research team in to talk with their students.

Finally, our profound appreciation to the **98 students** who took the time to help us 'test the test'.

Your insights are invaluable.
You are the reason this research matters.