

The reality of assessing 'authentic' electronic portfolios: Can electronic portfolios serve as a form of standardized assessment to measure literacy and self-regulated learning at the elementary level?

L'évaluation d'e-portfolio «authentiques» : Les e-portfolios peuvent-ils servir d'évaluation standardisée pour mesurer la littératie et l'apprentissage autorégulé à l'élémentaire?

Eva Mary Bures, Bishop's University

Alexandra Barclay, Mount Saint Vincent University

Philip C. Abrami, Concordia University

Elizabeth J. Meyer, California Polytechnic State University

Abstract

This study explores electronic portfolios and their potential to assess student literacy and self-regulated learning in elementary-aged children. Assessment tools were developed and include a holistic rubric that assigns a mark from 1 to 5 to self-regulated learning (SRL) and a mark to literacy, and an analytical rubric measuring multiple sub-scales of SRL and literacy. Participants in grades 4, 5 and 6 across two years created electronic portfolios, with $n=369$ volunteers. Some classes were excluded from statistical analyses in the first year due to low implementation and some individuals were excluded in both years, leaving $n=251$ included in analyses. All portfolios were coded by two coders, and the inter-rater reliability explored. During the first year Cohen's kappa ranged from 0.70 to 0.79 for literacy and SRL overall, but some sub-scales were unacceptably weak. The second year showed improvement in Cohen's kappa overall and especially for the sub-scales, reflecting improved implementation of the portfolios and use of the assessment tools. Validity was explored by comparing the relationship of portfolio scores to other measures, including the government scores on the open-response literacy questions for the Canadian Achievement Tests, fourth edition (CAT4s), the scores we assigned to the CAT4s using our assessment tools, and scores on the Student Learning Strategies Questionnaire (SLSQ) measuring SRL. The portfolio literacy scores correlated ($p<0.01$) to scores we assigned the CAT4s using our assessment tools, and to government pre-CAT4 scores, but the self-regulatory learning scores did not correlate to our measure of student's self-regulation. The results suggest that electronic portfolio assessment is time-consuming and difficult due to the range of varying

evidence within even a single individual's portfolio and that it may not be fair to do across diverse classrooms unless there are shared guidelines or tasks.

Résumé

Cette étude explore les e-portfolios et leur potentiel pour l'évaluation de la littératie et de l'apprentissage autorégulé chez les enfants de l'école primaire. Des outils d'évaluation ont été élaborés et comprennent un barème général qui attribue une note de 1 à 5 pour l'apprentissage autorégulé (AAR), une note pour la littératie et un barème d'analyse permettant de mesurer plusieurs sous-échelles de l'AAR et de l'alphabétisation. Les participants en 4^e, 5^e et 6^e années ont créé sur une période de deux ans des e-portfolios, avec $n = 369$ bénévoles. Certaines classes ont été exclues des analyses statistiques dans la première année en raison d'une faible mise en œuvre et certaines personnes ont été exclues dans les deux années, réduisant les analyses à $n = 251$. Tous les e-portfolios ont été codés par deux codeurs et la fiabilité entre les évaluateurs a été explorée. Au cours de la première année, le coefficient kappa de Cohen variait globalement de 0,70 à 0,79 pour la littératie et l'AAR, mais certaines sous-échelles étaient trop faibles. À la deuxième année, il y a eu une amélioration dans le coefficient kappa de Cohen en général et en particulier pour les sous-échelles, reflétant une amélioration de la mise en œuvre des e-portfolios et de l'utilisation des outils d'évaluation. La validité a été évaluée en comparant la relation entre les résultats des e-portfolios et d'autres mesures, y compris les résultats du gouvernement sur les questions de littératie à réponses ouvertes du Canadian Achievement Tests (version 4), les résultats que nous avons assignés au CAT-4 à l'aide de nos outils d'évaluation et les résultats du Student Learning Strategies Questionnaire (SLSQ) mesurant l'AAR. Les résultats de littératie du e-portfolio sont en corrélation ($p < 0,01$) avec les résultats que nous avons attribués aux CAT-4 à l'aide de nos outils d'évaluation et avec les résultats du pré-CAT du gouvernement, mais les résultats de l'AAR ne sont pas en corrélation avec notre mesure de l'autorégulation des étudiants. Les résultats suggèrent d'une part que l'évaluation par e-portfolio prend beaucoup de temps et s'avère difficile en raison de l'éventail de données variables dans le e-portfolio même d'une seule personne et, d'autre part, qu'elle ne peut pas être faite de façon appropriée dans plusieurs classes à la fois, sauf s'il existe des lignes directrices ou des tâches communes.

Introduction

Traditional pedagogical approaches are coming under increasing criticism in part due to alarming attrition rates in Canada coupled with low literacy and numeracy skills in most Westernized countries (OECD, 2010; Knighton, Brochu, & Gluszynski, 2010). Interest is growing in alternative ways to instruct and assess students. One exciting pedagogical innovation is electronic portfolio software. Electronic portfolios may provide an alternative way to support and measure learners' literacy and self-regulatory skills.

Electronic portfolios enable students to demonstrate both more traditional literacy skills as well as 'new literacy' skills as students can go beyond text to create visual, auditory and multimedia artifacts. They may support students' use of self-regulated learning (SRL) skills. For example, students can post draft versions, reflect upon them, and post a final version. They can also make connections between their work, their goals and strategies for learning.

Electronic portfolios offer a new more authentic approach for educators to assess their students. They offer diverse ways of viewing and presenting progress and achievement, simplify the process of creating a long-term collection of academic work (Sharples, Taylor, & Vavoula, 2007), and allow multiple stakeholders to view a student's showcased productions. They place students at the core of the assessment process, evaluating their own work: as students build a collection of artifacts, they choose pieces for assessment, reflecting upon their work and the reasons for inclusion. For all these reasons, electronic portfolios may be more authentic than traditional means of assessment (Love & Cooper, 2004) and may be a more accurate reflection of a student's overall achievements at university (Chambers & Wickermsham, 2007) as well as at K-12 (Barrett, 2007). In a climate of increasing standardized testing of students, they offer an alternative approach that allows for more contextualized learning and differentiation.

Assessment of electronic portfolios is more complex than typical standardized assessment. Are scores and assessment data from electronic portfolios consistent, meaningful, and trustworthy? Can we assess portfolios across diverse classrooms as a form of standardized assessment? This research will focus on assessing elementary children's literacy and self-regulatory skills through electronic portfolios.

Literature Review

Literacy Development and Electronic Portfolios

Language and literacy development is important for individual, social and academic growth. Literacy is a process of constructing meaning through language. Literacy also has a social role that extends beyond encoding and decoding language (Lankshear & Knobel, 2006) that involves how it is culturally 'situated' within the daily lives of students (Lave & Wenger, 1991). Literacy increases when students are motivated and able to express their experiences through language particularly when the context is important to them (Pintrich, 1993).

There is growing evidence that electronic portfolios can improve literacy (i.e. Abrami, Venkatesh, Meyer, & Wade, 2013; Meyer, Abrami, Wade, Aslan, & Deault, 2010). They can allow students to engage in more authentic literacy practices (Abrami & Barrett, 2005; Yancey, 2004). They allow students to display multiple forms of literacy: they require students to engage with text, audio and visual learning, a group of tools that when used in combination are changing the nature of communication and meaning making (Jewitt, Kress, Ogborn, & Tsatsarelis, 2001). Allowing learners to represent their understandings in multiple ways including voice and audio recordings better reflects our daily lives where balancing media is increasingly common and required.

Electronic portfolios can support a more authentic process of writing, as students go through a process of revising draft versions, reflecting upon them, and finally publishing or finalizing work. Publishing in the portfolio enables students to witness the growth of their writing skills through the process of writing multiple revisions leading to a final shared public version (Hill, Song, & West, 2009; Yancey, 2004). Engaging students in more authentic literate practices makes them more likely to set goals and apply strategies reflectively (Barrett, 2007).

Self-regulated Learning and Electronic Portfolios

Research on the long-term impact of using electronic portfolios on SRL in its infancy but growing evidence suggests they can increase students' ability to self-regulate their learning (Wade & Abrami, 2005; Meyer et al., 2010; Zellers & Mudrey, 2007). Students generally have insufficient self-regulatory skills (Schunk & Zimmerman, 2006). To develop effective self-regulated learning strategies "students need to be involved in complex meaningful tasks, choosing the products and processes that will be evaluated, modifying tasks and assessment criteria to attain an optimal challenge, obtaining support from peers, and evaluating their own work" (Perry, 1998, p. 716). Electronic portfolios can provide such an environment.

In using electronic portfolios effectively, the student sets goals, posts work, and reflects upon his/her work, while going through a process of planning, monitoring, and regulating, the three key processes of meta-cognitive self-regulation (Azevedo, Moos, Johnson & Chauncey, 2010; Zimmerman, 2000). During the reflection phase, the student may contemplate how the work generally links to his/her goals and use of strategies, and can involve the creation of setting new learning outcomes for future tasks, central to SRL theory (Bandura, 1993). When electronic portfolios are done ineffectively, they can be a mere collection of work; it is the process of reflection that makes them a tool for life-long learning and professional development (Barrett, 2007; Foote & Vermette, 2001).

Assessment and Portfolios

Evidence suggests that electronic portfolios support the development of literacy and SRL; they may also provide an alternative more authentic means of evaluation than traditional paper-and-pen tests. Opportunities for contextualized learning require opportunities for contextualized assessment. Authentic assessment includes assessment practices that hold value beyond school and which may better predict real-world abilities (Newmann, Brandt, & Wiggins, 1998; Reeves, 2000; Reeves, Herrington, Oliver, & Woo, 2004). Within electronic portfolios students demonstrate achievement within the context of multiple relevant variables. The process of creating an EP is inter-disciplinary and unique, the nature of the tool situating the learning in context.

Standardized assessment traditionally allows us to measure student achievement across a range of different teaching and learning contexts, creating comparable scores with objective marking approaches with high inter-rater reliability. Norm-based testing is commonly used as a measure for knowledge acquisition, but it is criticized for not taking into account cultural cognition; there is no guarantee that standardized tests will reflect what the student has learned, or what real abilities they can demonstrate in context (Herman, Gearhart, & Aschbacher, 1996). Hence standardized testing may have low consequential validity (Hickey & Zuiker, 2012). The construct validity of standardized measures of literacy has been particularly questioned as literacy is often situated in context and within culture (Lave & Wenger, 1991) making it harder to judge effectively out of context.

The shift toward student-centered assessment practices has led to the development of standardized 'alternative' assessment approaches. For example, the Canadian Achievement Test (CAT) has introduced open-ended constructed response items to measure literacy, complementing the multiple-choice format of the core CAT. In Quebec, the government has

introduced end-of-cycle tests at the end of elementary school which are more akin to a unit plan than to a test: the time constraints are not consistent across classes and peers collaborate at points during the process. These approaches involve more subjective marking as compared to the traditional standardized test, but may provide more meaningful results. Can electronic portfolios serve as a final product of student achievement, one which is more in keeping with real life where appropriate responses do not include a, b, c, d, or none of the above (Gardner, 1990)?

Large-Scale Assessment of Portfolios

Large-scale portfolio assessments have run into difficulties with reported varying inter-rater reliability and inadequate correlations with other achievement measures, even in contexts where the tasks are set (Koretz, 1998). Gearhart and Herman (1998) analyzed portfolios across several classrooms and the teachers' instructional practices. They found that often there was not the appropriate evidence in the portfolios to judge the intended competencies; when the evidence was present it was time-consuming to find and interpret. Furthermore, they found it challenging to compare portfolios across classrooms, noting the evidence of varying levels of help from the teacher; in some contexts it could be difficult to decipher whether it was the student or the teacher's understanding being represented. Stecher (1998) also found it difficult to assess portfolios across multiple classes for large-scale standardized assessment; he argued it is difficult to align scores from a range of contexts to curricular goals. On the other hand, he reported relative success with teachers using portfolios for assessment purposes within their own classrooms, with the caveat that it is very time-consuming.

It may be that such results reflect issues with implementation; electronic portfolios are still in their infancy and many instructors struggle with the challenges associated with innovations (Love & Cooper, 2004). Zellers and Mudrey (2007) had mixed results when studying professors' implementation of electronic portfolios to support reflection/metacognition. Meyer, Abrami, Wade, and Scherzer (2011) documented challenges for teachers integrating portfolios into the curriculum to support literacy at the elementary levels. Improved implementation increases the likelihood of establishing valid and reliable assessment approaches.

Establishing Validity and Reliability of Portfolio Assessment

Authentic assessments do not generally follow a traditional assessment format. Wiggins (1990) underscores that authentic assessment redefines validity: "Test validity" should depend in part upon whether the test simulates real-world "tests of ability," not just its fit with the curriculum or its correlation with other test results (Wiggins, 1990; Broadfoot & Black, 2004). One approach to validity is through rich description, which allows the researcher/practitioner to provide so many details that the reader can use his/her own judgment (Geertz, 1973). Another approach is to try to triangulate the results from one assessment to another (i.e. scores on electronic portfolios should correlate to other measures of literacy) to create a sense of trustworthiness.

The aim of the current study is to create tools for assessing electronic portfolios, and to explore their usability, reliability and validity across a range of elementary classrooms and triangulated against a set of literacy assessment (CAT4s). However, this is done within a socio-cultural framework of learning and teaching with the understanding that student-centered assessment is the ultimate aim. Hence, we tread carefully amongst an inter-disciplinary minefield. Is it even possible to correlate measures of authentic assessment with less authentic measures? Must our

results from ‘objective’ evaluation correlate to those from contextual tasks undertaken in real life classrooms or can we blend them into a richer picture? The central question behind the research is: Can we develop assessment tools that provide consistent results such that electronic portfolios from diverse classrooms could be assessed, offering an alternative form of standardized testing?

How to Assess Electronic Portfolios

The literature revealed a variety of approaches to assess portfolios. The literature heavily recommends teacher/student portfolio conferences, generally used in formative assessment. In addition, many rubrics are available for scoring portfolios. Their use is controversial: “There is concern that we are losing the ‘stories’ in electronic portfolios in favour of the skills checklists” (Barrett, 2007, p. 444; Hickey & Zuiker, 2012; Stiggins, 2004). Creating rubrics for assessment of electronic portfolios runs the risk of becoming as rigid as a standardized test, failing to acknowledge and grasp a student’s individuality and achievements (Yancey, 2004). Furthermore, inter-rater and validity are rarely explored. Still, rubrics with clear criteria can help align scores across raters. Carliner (2005) suggests that sample rubrics will be needed to support evaluators using electronic portfolios especially those created within a socio-constructivist approach.

Different types of rubrics exist. Analytical rubrics break down the targeted skills into several sub-skills, generating several scores that are then combined. A variant is to use an analytical rubric but to allow the teacher to weigh the various parts of the process depending on the student. Different weights could be determined collaboratively with students in teacher-student conferences in order to determine the final score. In lieu of analytical rubrics, others develop holistic rubrics where each portfolio is assigned a level such as ‘excellent’ or ‘poor’ based on the various criteria considered simultaneously. Holistic rubrics can also be used in conjunction with analytical rubrics. Student-teacher conferences and rubrics are the most reported forms of portfolio assessment.

Assessment	Criteria
Writing	Ideas & Details Sentences & Organization Word Choice Conventions Purpose Creativity Perceptions
Self-regulated learning	Goals Strategies Reflection

Figure 1: Criteria to assess literacy and SRL through electronic portfolios

Beyond what type of assessment tool to use is what to actually assess -- some approaches emphasize the process of creating the portfolio and focus on student’s self-regulatory processes involved in planning, strategy use, and/or reflection, whereas others emphasize the quality of the content and use that to assess student’s subject-matter skills such as their mathematical reasoning or literacy. Research suggests that measuring SRL is challenging: Schraw (2010) suggests drawing on a variety of sources and interpretations of evidence of SRL given that the relationship between students’ self-reports and self-regulation is not established (Zimmerman, 2008) and multiple perspectives exist as to how to conceive of SRL (Azevedo, et al., 2010; Winne, 2010). Electronic portfolios provide a way to measure SRL as an event or process rather than as a trait. In this study we will explore how to assess both literacy and SRL through electronic portfolios.

Writing: Holistic Judgment	
Evaluate the writing skills students demonstrate through the pieces in their portfolio using these criteria: Ideas and details, Voice, Organization and Sentences, Conventions, Purpose and Meaning, Creativity and Imagination, and Perceptions	
Please circle a holistic mark of 1, 2, 3, 4 or 5, evaluating the category as a whole where:	
Category	Description
5 is Extending	Writing supports ideas show evidence of thoughtful understanding of producing, extending and enhancing meaning and information with thorough explanations, details and reflection.
4 is Achieving	Writing and ideas are focused understandings that explain and demonstrate meaning and information achieving some good explanations, details and reflection
3 is Developing	Writing and ideas are developing and in progress with meanings and information that is in the process of gaining a more complete understanding and accurate method of self expression.
2 is Beginning	Writing and ideas demonstrate some vague meanings and information that shows a superficial or vague understanding of information.
1 is Experimenting	Writing is inconsistent, incomplete or very confused demonstrating the need for much more attention to details, explanations, ideas and accuracy.
Comments:	

Figure 2: Holistic Assessment of Writing (HAW)

For the purposes of this study, assessment tools were designed by a team of researchers and practitioners including teacher educators and educational technologists with feedback from literacy and assessment consultants from Quebec, Alberta and Manitoba. For both literacy and

SRL we developed a holistic rubric and an analytical rubric based on the same criteria. The completed tools include both a holistic analysis of writing (HAW) and self-regulated learning (HASRL) and an analytical rubric with 7 writing sub-scales (ARW) and 3 self-regulated learning sub-scales (ARSRL).

After establishing the criteria, we developed a holistic assessment of writing (Figure 2). The analytical rubric draws on the same criteria but breaks the assessment into separate components to evaluate, namely: ideas, voice, sentences and organization, conventions, purpose, creativity, and perception. Figure 3 shows the descriptors for level 5 ('extending' or 'excellent').

Writing	Indicators for Extending- 5
Ideas & Details	The details show evidence of careful attention with elements selected to enhance the communication of central ideas that thoughtfully and thoroughly explore meaning and content by producing and extending the information to the reader.
Voice	The writer's voice is consistent, compelling and engaging while respecting the intended purpose and audience.
Organization and Sentences	Written messages and ideas are thoroughly and thoughtfully crafted with close attention to the intended purpose and audience illustrated through very well written sentences and organized, effective paragraphing that conveys a very clear message to the reader.
Conventions	Capitalization, punctuation and spelling is thorough with excellent attention and adherence to editing and revision that enhances and extends communication with the reader.
Purpose & Meaning	Use of language, dialogue and descriptive word choice is very appropriate for the intended purpose and/or audience with careful attention paid to crafting writing and an understanding of purpose and meaning is clearly conveyed to the reader.
Creativity and imagination	Explanations and interpretations demonstrate original ideas with value that enhance and extend the writer's imaginative ideas painting a clear image for the reader.
Perceptions	Writing shows a carefully crafted, thoughtful and meaningful point of view that clearly and consistently expresses personal understanding, thoughts, feelings and perceptions of the task, subject content and world beyond.
Comments: Total Score: /35	

Figure 3: Descriptors for Level 5 ('Extending' or Excellent) of the Analytical Rubric for Writing

A similar approach was taken to produce analytical and holistic rubrics to measure SRL.

Research Questions

The research questions are:

- 1) Do the developed assessment tools have inter-rater reliability?
- 2) Do the scores from the tools correlate to scores from other measures of literacy and SRL?
- 3) What challenges in assessing portfolios across a range of classrooms emerge?

Research Methods

Design

Data was collected in elementary classrooms across several Canadian provinces, with matching control groups. This classroom-based research draws on qualitative and quantitative approaches. Tools were developed to assess the portfolios across several classrooms, exploring their inter-reliability and validity. Each portfolio was double-coded and the inter-rater reliability explored. Validity is explored looking at the relationship of portfolio scores to other measures, including the government scores on the CAT4s, the scores we assigned the CAT4s, and self-regulated learning scores generated from a questionnaire.

Participants

Participants are volunteer students in grades 4, 5 and 6 across three Canadian provinces whose parents have signed consent forms, and 16 teachers. Portfolios were created across two years. All teachers received at least a half-day of training on the use of ePEARL from research centre staff and follow-up support including lesson plans and job aids, an online discussion forum (in the form of a moderated wiki), as well as in-class observations and model lessons during the school year. In addition, multimedia scaffolding and support for teachers and students are embedded in the tool. All participants created portfolios for class purposes using our educational software tool called ePEARL.

During the first year, a range of implementation was observed within 7 experimental classes ($n=149$). Our analyses drew on 3 classes where both SRL and literacy were assessable, $n=53$. In the second year the portfolios were better integrated so all experimental classes ($n=9$) were included in the analysis ($n=220$ students); some individual portfolios were excluded, leaving $n=198$ portfolios in analyses.

Electronic Portfolio Encouraging Active Reflective Learning (ePEARL)

All portfolios are made using ePEARL, a tool designed at the Centre for the Study of Learning and Performance (CSLP) in collaboration with our partner LEARN. ePEARL is bilingual (English-French), web-based and student-centered EP software that is designed to support the phases of self-regulation. ePEARL contains four developmentally-appropriate levels for use in early elementary (Level 1), late elementary (Level 2), and secondary schools (Level 3) as well as in higher education and beyond (Level 4) (Abrami, Bures, Idan, Meyer, Venkatesh, & Wade,

2013). To view a sample ePEARL environment, log on to <http://grover.concordia.ca/ePEARL/promo/en/index.php>

ePEARL includes features designed to lead students through a process of self-regulation: they can set goals, create new work via a text editor and/or audio recorder or link to work created elsewhere, reflect on work, revise work and save multiple versions and/or goals and/or strategies. They can also share work, obtaining feedback from teachers, peers and parents. They can choose work to represent their achievements or growth. ePEARL is intended for use in all school subjects; we are currently trialing a version for use by the Royal Conservatory of Music, called iSCORE, as part of piano studio teaching (Upitis, Abrami, Brook, Troop, & Varela, 2012).

ePEARL allows students to collect their work over time, and to represent their understanding in a variety of ways (including text, audio recordings, and photographs). Figure 4 shows different options for creating an artifact in ePEARL.

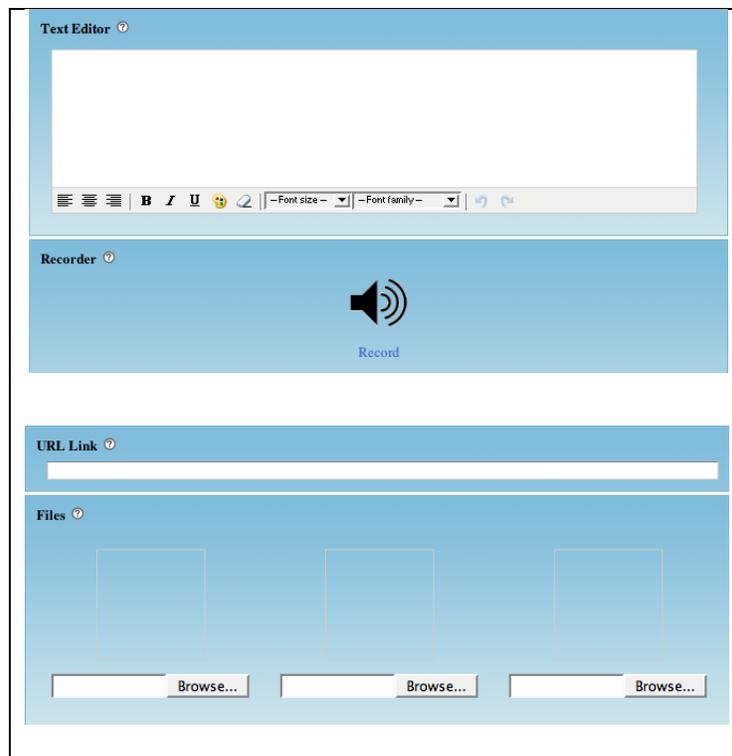


Figure 4: Text, Audio recording, URL link and file functions in ePEARL Level 2

ePEARL scaffolds the student through setting goals and strategies, creating artifacts, and reflecting upon them.

Data Sources

The data sources are student electronic portfolios to measure SRL and writing; CAT4 open-constructed responses to measure literacy; and the Student Learning Strategies Questionnaire (SLSQ) to measure SRL.

The primary data source are the electronic portfolios. These were double-coded using the assessment tools we created to assess writing and SRL, namely the Holistic Assessment of Self-Regulated Learning (HASRL), the Analytical Rubric for SRL (ARSRL), the Holistic Assessment of Writing (HAW) and the Analytical Rubric for Writing (HAW). This generates both holistic and analytical scores for writing and SRL.

The constructed response subtest of the fourth edition of the Canadian Achievement Tests (the CAT4) was also administered to students in both the fall and the spring (Canadian Achievement Tests, fourth edition, 2008). Those administered in the fall will be referred to as the pre-CATs and those in the spring will be referred to as the post-CATs. The CAT4 assesses both response to text (ideas, support) and writing (content, content management) using a rubric applied to two tasks, the first a response to text and the second a writing task. The Canadian Test Centre conducted all scoring as part of their norming study. The constructed response subtest depends on student narrative responses to prompts as opposed to the multiple-choice format of the main tests of the CAT4, which also measures student literacy. For the first task, multiple texts were used within each class at both pre-test and post-test, ranging from a Calvin & Hobbes comic strip to a several-page article about computers and ‘bots.’ For the second task multiple story prompts were used in each class at both pre-test and post-test but no student responded to the same prompt twice. We chose this form of measuring literacy achievement because it was compatible with notions of authentic assessment, even though it meant we generated a less detailed analysis of student learning than using the closed ended version of the CAT4. The reliability coefficients (KR-20) for CAT4 subtests range between 0.85 and 0.95, depending on the level and subtest. In the previous version (CAT3), test validity was established by showing that grade levels that were known to have different levels of achievement did indeed have different mean scores on the same test.

The Student Learning Strategies Questionnaire (SLSQ) (Abrami & Aslan, 2007) was administered to students near the beginning and end of the school year. It contains several open-ended and numerous Likert scale items to measure students’ perception of their ability to employ SRL strategies including their ability to set learning goals, observe and correct their performance and reflect on the learning outcome. The SLSQ contains six scales, namely, goal setting, strategy planning, self observation, self-instruction, feedback from adults, and self-evaluation. Students were also asked at the end of the SLSQ a series of open-ended questions about their experiences with ePEARL. These questions included items such as, “I like using ePEARL in my class because...” and “I did not like using ePEARL in my class because...” as well as “What I liked most about using ePEARL is...” and “What I liked least about ePEARL is...”

Analyses

Questionnaire data were analyzed by item, to obtain a fine-grained analysis of specific changes in self-regulation that occurred as a result of ePEARL use.

To assess writing through the electronic portfolios, the Holistic Assessment of Writing (HAW) and the Analytical Rubric for Writing (ARW) were used. Using the HAW, the team assigned an initial score on a scale from 1 to 5 where 1 is experimenting and 5 is extending. Using the ARW, the team used seven sub-scales (ideas, voice, sentences and organization, conventions, purpose, creativity, and perception), each ranging from 1 to 5 (Figure 2). Similarly, to assess SRL in the portfolios, the Holistic Assessment of Self-Regulated Learning (HASRL) and the Analytical

Rubric for SRL (ARSRL) were applied. The ARSRL contains 3 sub-scales (goals, strategies, and reflection, with reflection weighted twice), generating a score from 4 to 20). The HASRL results in a single score from 1 to 5. The research assistants were trained carefully with explanations of the tools developed, and exemplars of the different levels made available on a website. The training included the team of assistants and the first author assessing first 10, then 16, and finally 25 portfolios, and comparing results at each stage. All portfolios were double-coded, and discrepant cases were discussed.

The results of the Canadian Achievement Test’s constructed response reading and writing activities were sent to the Canadian Test Centre for evaluation. They assigned final scores to all the students that were then mailed to us for inclusion in our data set. We also coded the CAT4s ourselves using our own assessment tools developed for measuring literacy in the electronic portfolios, the HAW and the ARW, and applying them to the CAT open-constructed tasks, as a form of triangulation.

Results and Discussion

Applying the Assessment Tools

To provide a sense of how the tools were applied to the electronic portfolios, we will look at two cases. Figure 5 identifies an artifact posted by Scott illustrating a “Beginning” level on our assessment tool in both literacy and self-regulatory learning (SRL). The text or content box displays some writing but notice the grammatical errors and organization issues with the sentence structure. The reflection is also at a lower level as it offers a very basic explanation rather than a critical evaluation of their role in the task or the task itself.

The screenshot shows a digital portfolio interface. At the top, the word "Content" is written in bold. Below it is a light purple rectangular box labeled "Text". Inside this box, the word "Repone" is centered. Below "Repone" is a paragraph of text with several grammatical errors: "The picture is very good what they are doing. The langeag is very good. there was lots of quontons. The format is they are troing botleds of water in the recling. Yes I do drink from bottled of water. I drink bottled of water because it good yes but it's very bad because the lable has 1 that mens it bad to drink. Sometimes I drink bottled water. The features are in class. My opinion is they are doing a grat job. The conneste is they finsh." In the bottom right corner of the text box, there is a small blue link labeled "top". Below the text box, the word "Reflections" is written in bold. Underneath is a light blue rectangular box. Inside this box, the text reads: "Reflections Updated 03/30/09" followed by "My reflections is good."

Figure 5: Example Artifact and Reflection from Scott’s Portfolio Demonstrating Low-Level, “Beginning” Level of Literacy and Reflection

Figure 5 provides an example of low-level reflection within the interface with vague responses lacking in detail and complexity. The reflection box can be used to illustrate active thinking about the student’s strengths and weaknesses or ways to improve for the next task. Reflections such as Scott’s are too vague to indicate whether the student actively thought about the task

retrospectively. Figure 6 illustrates how a beginning goal or strategy may look. Students who use basic explanatory statements such as ‘to do well’ are providing very vague details and lower level self-regulatory skills.



Figure 6: Example of Scott’s Task Goals and Strategies, Demonstrating Low-Level, “Beginning” Level of SRL

Chloe scored at a high level in literacy and SRL. In contrast to the lower-level literacy students, Chloe’s text (Figure 7) illustrates focus, editing and consistent sentence organization.

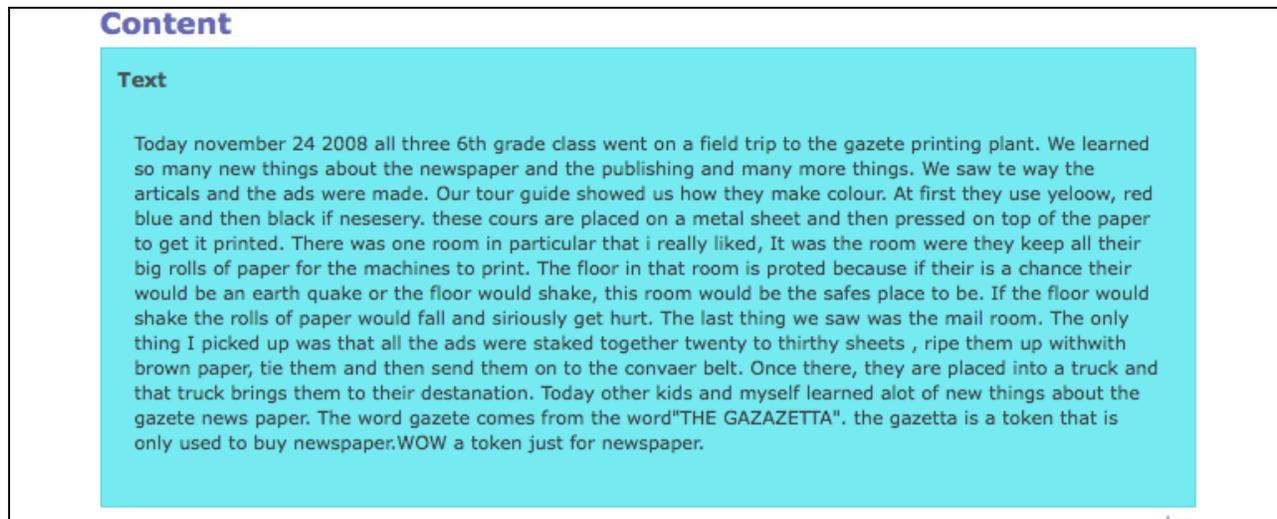


Figure 7: Chloe demonstrates high-level or “Achieving” score for literacy

In Figure Eight, Chloe offers some goals and strategies to complete the task. The goals show a dedication by the student to work towards overcoming their weakness for a busy schedule. In many ways, self-regulated learning exhibits stronger skills when the student is able to perceive their strengths and weaknesses and works towards a greater goal. She also describes a need to express herself clearly and she ends it by suggesting her willingness to find her voice. While there are a few spelling mistakes in her actual post, the significance of SRL is to develop thinking that is evaluative and shows progress for future work. What seems endemic in many of the portfolios is that goals and strategies are often scripted by the teacher to aid in the facilitation of the task(s). Hearing the voice of the student for these two aspects of SRL is more unusual.

Goals

Task Goals Updated 10/22/08

At home, I am a very busy perso so one of my goal is to advance my province research when I ahve enough time. My second goal is that I want to write good detailed sentences which makes it interesting and not a piece of junk. My last goal is to write the text into my own words which is sometimes hard so it proves that I had worked hard and it is worth to real!

Strategies Updated 10/22/08

- to take my time
- to read a lot of books about the two provinces
- to go on the computer to find answers to my questions

Figure 8: Chloe demonstrates high-level SRL skills or “Achieving” score

Measuring Literacy and SRL in the Electronic Portfolios: Inter-rater reliability

Each electronic portfolio was coded by at least two coders, either research assistants in teacher education or the first author, a teacher educator. Inter-rater reliability was adequate overall; looking at the holistic scores and the composite scores generated from the rubrics we found a range of Cohen’s kappa scores in the high 70’s for literacy and SRL (see Table 1).

Table 1: Inter-rater reliability of assessment tools

	Literacy Holistic	Literacy Rubric	SRL Holistic	SRL Rubric
Cohen’s kappa Year one	0.70	0.79	0.76	0.78
Cohen’s kappa Year two	0.77	0.83	0.73	0.76

The inter-rater reliability associated with the SRL sub-scales are noticeably lower, especially with strategies and reflection, but this improves in Year Two (Table 2).

Table 2: Inter-rater reliability of SRL sub-scales

	SRL – goals	SRL -- strategies	SRL – reflection
Cohen’s kappa Year one	0.68	0.45	0.46
Cohen’s kappa Year two	0.644	0.644	0.82

The literacy sub-scales display problems with two categories, ‘voice’ and ‘organization and sentences,’ which improved considerably in the second year.

Table 3: Inter-rater reliability of literacy sub-scales

	Ideas & Details	Voice	Organization & Sentences	Conventions	Purpose	Creativity	Perception
Cohen's kappa Year one	0.74	0.52	0.45	0.71	.72	.71	.70
Cohen's kappa Year two	0.724	0.691	0.75	0.70	0.67	0.76	0.70

Validity of Literacy Assessment

Do our assessment tools relate to other measures of literacy? The holistic scores assigned to the electronic portfolios for writing correlated ($p < 0.01$) to the holistic scores we assigned the post-CAT4s ($r = 0.48$). The holistic writing scores correlated to government-assigned pre-CATs ($r = 0.381$).

The rubric portfolio scores for literacy correlated to the holistic and rubric scores we assigned the CATs ($r = .603$ and $.479$ respectively). The rubric portfolio scores correlated to government pre-CAT scores ($r = .550$), but not the post-CATs.

Surprisingly, the portfolio scores for literacy generated from the portfolios correlated to the scores the government assigned to the CAT4s the students wrote earlier and not later. One interpretation is that our portfolio scores measure literacy as demonstrated through a range of pieces posted throughout the year, not just their final achievement; it would be more likely that if we coded only the 'best works' or 'presentation' pieces in the portfolio that a score would correlate to their measured 'literacy' at the end of the year. Notwithstanding, scores we assigned the post CATs did correlate to the scores we assigned to portfolios for literacy.

Validity of our Assessment of SRL

Do our assessment scores relate to other measures of SRL? Neither the holistic nor the rubric portfolio scores for SRL correlated to the SLSQ questionnaire results.

On the other hand, the scores for SRL in the portfolios correlated to various measures of literacy. The holistic scores correlated to the holistic scores we gave the post CAT4s ($r = .425$). Both holistic and rubric portfolio scores for SRL correlated to the pre-CAT4 government scores ($r = 0.465$ and $r = 0.593$ respectively).

Challenges in Assessing Electronic Portfolios and CAT4s

We faced many challenges in coding electronic portfolios. We found it challenging to deal with the inconsistency of student work in the portfolios from hastily done to carefully revised and crafted final work. There was a wide variability in quality even within one student's portfolio.

The wide range of varying evidence and the sheer amount of evidence in the portfolio made it difficult to capture the student's achievement in a score.

Furthermore, the wide diversity of work and teacher approaches in each class made it even harder to form a judgment. Coders found it very hard to judge across different classrooms; familiarity with a class seemed important for fair coding. For example, some teachers helped students with strategies and having coded a few, one would see that these were 'parroted' strategies. Coding without a sense of the classroom context in which ePEARL was being used proved challenging as teachers used the software in different ways, providing a varying level of support for both the content and SRL involved in effectively using the software. Coders felt it was particularly difficult to fairly judge SRL across the different classrooms through the portfolios.

It was much more efficient to generate a score for literacy based on the CAT4, which had set prompts and tasks. CAT4 scores were based on two reading and writing open responses, whereas portfolio literacy artifacts were a presentation of multiple pieces of work across multiple subjects. The limited amount of work in the CAT4 make it more efficient to assign a score, but we often felt the scores were not as accurate in terms of assessing a student's writing abilities as they were based on such scanty evidence. Furthermore, the evidence itself was varying (similar to the portfolios). There were several versions of each of the two tasks, which made up the open-constructed portion of the literacy component of the CAT4s. The tasks varied greatly and this called into question how the difficulty of the task and type of literacy piece affected the rubric scores. Inconsistent scores were often a result of two very different literacy tasks. Not only is the evidence limited, but the score is highly influenced by a student's ability to understand both of the tasks posed. The scores may better represent a student's self-regulated learning than his/her literacy. Both these non-traditional approaches to standardized assessment posed challenges.

Electronic portfolios were time-consuming and challenging to assess for literacy and SRL, but we were able to establish adequate inter-rater reliability of our assessment tools and support for the validity of our measure of literacy.

Conclusions

This research involved assessing student achievement in literacy and SRL through electronic portfolios in sixteen classrooms. We measured SRL through electronic portfolios and through a questionnaire; we measured literacy through electronic portfolios and through the open-constructed responses rather than the multiple-choice items of the CAT4. To interpret this evidence, we drew on the government's scores as well as our own, triangulating two ways to score the same source.

Although student portfolios provide rich evidence of self-regulated learning processes such as planning and strategy use, they pose a challenge to interpret as evidenced in this study. Inter-rater reliability for our measures of literacy and SRL within electronic portfolios is adequate; the validity of our literacy measure is somewhat supported by our results, but our SRL measures will need to be reconsidered.

Regarding SRL, concerns in the literature about the analytical rubric approach seemed to play out in this study regarding self-regulation: some students showed strong SRL abilities, but did

not write goals or strategies into ePEARL, and so their analytical rubric scores were not a good reflection of their SRL. Our analytical rubric for SRL may better measure how well the student used ePEARL than his/her self-regulatory abilities. Furthermore, sometimes the teacher provided support and some students just parroted strategies; other times, the students seemed to be left anchorless. This compromised the validity of our SRL scores.

Another challenge in measuring SRL through portfolios is that some students will choose to use strategies ‘off-line’ (Schraw, 2010) rather than within the portfolio. Our two measures of SRL, one a ‘self-report’ (a questionnaire) and the other a learning artifact (the portfolio), did not correlate, common with measures of SRL (Schraw). The relationship between students’ self-reports and self-regulation is not established (Zimmerman, 2008) and multiple perspectives exist as to how to conceive of SRL i.e. SRL as a state and an aptitude (Winne, 2010). Perhaps the portfolios and the self-reports represent complementary perspectives on a student’s self-regulated learning. Still, given the totality of our results, we feel a need to reconsider our approach to measuring SRL in electronic portfolios.

Coding of the electronic portfolios proved to be challenging. Our findings suggest it is essential when judging portfolios to discuss discrepant cases for more consistent, fair scoring even within the same class. In assigning scores, an overriding concern was inconsistent student work within each portfolio. The quality of portfolios varied greatly even within one portfolio, making it very complicated to judge and assess effectively. What score does one assign a ‘mixed bag’ such as the portfolio? Comparing results across different contexts of instruction, as large-scale standardized assessment necessitates, is even more challenging. The results of this study suggest it may not be fair to assess electronic portfolios across diverse classrooms unless there are some uniform aspects to the processes and/or products. Further research is needed with different approaches to measuring SRL and literacy.

We found it difficult to measure literacy and self-regulated learning through the electronic portfolios, but at the same time it provided rich evidence of student skills and learning. Less traditional means of ‘standardized assessment’ beg the question of how to fairly assess them. The constructed responses of the CATs and electronic portfolios both represent attempts at more authentic ways to assess student writing, but the gain in richness of evidence comes with challenges such as subjective and resource-intensive scoring (Herman, Gearhart, & Baker, 1993; Newmann, Brandt, & Wiggins, 1998). With the portfolios, time on task and tasks given vary widely within the different classrooms and there is a wide range of evidence within even an individual student’s portfolio, making assessment challenging, but the very fact that the portfolio prompts are diverse and that students choose their own unique pieces is why this type of assessment compels us. Double-coding the portfolios and discussing discrepant cases, as does the Ministère de l’Éducation du Québec with its grade 6 open literacy tests, helps. It would further be of interest to explore results if the teachers (and students) were aware of a common assessment approach, as suggested in Gearhart and Herman (1998); we do not want to constrict teachers to a particular portfolio assignment, but knowing that a general assessment approach would be applied could help create more alignment without creating uniform ‘standardized’ portfolios.

Acknowledgements:

The first author would like to acknowledge financial support for this research from the Fonds de Recherche du Québec sur la Société et la Culture and the Senate Research Committee of Bishop's University.

References

- Abrami, P. C. & Barrett, H. (2005). Directions for research and development on electronic portfolios. *Canadian Journal of Learning and Technology*, 31(3), 1-15. Available: <http://www.cjlt.ca/content/vol31.3/abrami.html>
- Abrami, P. C., Bures, E. M., Idan, E., Meyer, E., & Venkatesh, V. (2013). Electronic Portfolio Encouraging Active and Reflective Learning (ePEARL). In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 503-515). New York: Springer Science.
- Abrami, P. C., Venkatesh, V., Meyer, E., & Wade, A. (2013). Using electronic portfolios to foster literacy and self-regulated learning skills in elementary students. *Journal of Educational Psychology*, 105(4), 1188-1209.
- Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. (2010). Measuring cognitive and meta-cognitive regulatory processes during hypermedia: Learning and challenges. *Educational Psychologist*, 45(4), 210-223.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, 28, 117-148.
- Barrett, H. (2007). Research electronic portfolios and learner engagement: The REFLECT Initiative. *Journal of Adolescent & Adult Literacy*, 50(6), 436-449.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in education: Principles, policy and practice*, 11(1), 7-26.
- Bures, E., Abrami, P. & Bentley, C. (2007). Assessing electronic portfolios -- Now that we have them, what can we do with them? In G. Richards (Ed.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007 in Quebec* (pp. 7030-7038). Chesapeake, VA: AACE.
- Carliner, S. (2005.) Commentary: Assessing the current status of electronic portfolios. *Canadian Journal of Learning and Technology*, 31(3), 121-132.
- Chambers, S. M. & Wickersham, L. E. (2007). The electronic portfolios journey: A year later. *Education*, 127(3), 351-360.
- Foote, C. J. & Vermette, P. J. (2001). Teaching portfolio 101: Implementing the teaching portfolio in introductory courses. *Journal of Instructional Psychology*, 28(1), 31-38.
- Gardner, H. (1990). *On leadership*. New York: The Free Press.

- Gearhart, M. & Herman, J. L. (1998). Portfolio assessment: Whose work is it? Issues in the use of classroom assignments for accountability. *Educational Assessment*, 5, 41-56.
- Geertz, C. (1973). *The Interpretation of cultures: Selected essays*. New York: Basic Books.
- Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, 1(3), 201-224.
- Herman, J. L., Gearhart, M., & Aschbacher, P. R. (1996). Portfolios for classroom assessment: Design and implementation issues. In R. Calfee & P. Perfumo (Eds.), *Writing portfolios in the classroom: Policy and practice, promise and peril* (pp. 27-59). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hickey, D. T. & Zuiker, S. J. (2012). Multi-level assessment for discourse, understanding, and achievement in innovative learning contexts. *The Journal of the Learning Sciences*, 22(4), 1-65.
- Hill, J., Song, L., & West, R. (2009). Social learning theory and web-based learning environments: A review of research and discussion of implications. *The American Journal of Distance Education*, 23, 88-103.
- Jewitt, C., Kress, G., Ogborn, J., & Tsatsarelis, C., (2001). Exploring learning through visual, actional and linguistic communication: The multimodal environment of a science classroom. *Educational Review*, 53(1), 5-18.
- Koretz, D. (1998). Large-scale portfolio assessments in the US: Evidence pertaining to the quality of measurement. *Assessment in Education*, 5(3), 309-334.
- Knighton, T., Brochu, P., & Gluszynski, T. (2010). *Measuring up: Canadian results of the OECD PISA study - The performance of Canada's youth in reading, mathematics and science; 2009 first results for Canadians aged 15* (Catalogue No. 81-590-X). Ottawa: Human Resources and Skills Development Canada, Council of Ministers of Education, Canada and Statistics Canada.
- Lankshear, C. & Knobel, M. (2006). *New literacies: Everyday practices and classroom learning*. New York: McGraw-Hill.
- Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Love, T., & Cooper, T. (2004). Designing online information systems for portfolio-based assessment: Design criteria and heuristics. *Journal of Information Technology Education*, 3, 65-80.
- Meyer, E., Abrami, P. C., Wade, A., Aslan, O. & Deault, L. (2010). Improving literacy and metacognition with electronic portfolios: Teaching and learning with ePEARL. *Computers & Education*, 55(1), 84-91. doi: [10.1016/j.compedu.2009.12.005](https://doi.org/10.1016/j.compedu.2009.12.005)

- Meyer, E., Abrami, P. C., Wade, A., & Scherzer, R. (2011). Electronic portfolios in the classroom: Factors impacting teachers integration of new technologies and new pedagogies. *Technology, Pedagogy and Education*, 20(2), 191-207. doi: [10.1080/1475939X.2011.588415](https://doi.org/10.1080/1475939X.2011.588415)
- Newmann, F., Brandt, R., & Wiggins, G. (1998). An exchange of views on "Semantics, psychometrics, and assessment reform: A close look at 'Authentic' assessments." *Educational Researcher*, 27(6), 19-21.
- OECD. (2010). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)*. doi: 10.1787/9789264091450-en
- Perry, N. E. (1998). Young children's self-regulated learning and contexts that support it. *Journal of Educational Psychology*, 90, 715–729.
- Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95(4), 667– 686.
- Reeves, T. (2000). Alternative assessment approaches for online learning environments in higher education. *Journal of Educational Computing Research*, 23(1), 101-111.
- Reeves, T., Herrington, J., Oliver, R. & Woo, Y. (2004). Designing authentic activities in web-based courses. *Journal of Computing in Higher Education*, 16(1), 3-29.
- Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist*, 34(4), 258-266.
- Schunk, D. H., & Zimmerman, B. J. (2006). Competence and control beliefs: Distinguishing the means and the ends. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 349-367). Mahwah, NJ: Lawrence Erlbaum Associates.
- Sharples, M., Taylor, J. & Vavoula, G. (2007). A theory of learning for the mobile age. In R. Andrews & C. Haythornthwaite (Eds.), *The Sage handbook of elearning research* (pp. 221-247). London: Sage.
- Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessment. *Assessment in Education*, 5(3), 335-351.
- Stiggins, R. J. (2004). New assessment belief for a new school mission. *Phi Delta Kappan*, 86(1), 22-27.
- Upitis, R., Abrami, P. C., Brook, J., Troop, M., & Varela, W. (2012). Learning to play a musical instrument with a digital portfolio tool. *Journal of Instructional Pedagogy*, 9, 1-15.
- Wade, A., & Abrami, P. C. (2005). An electronic portfolio for learning. *Canadian Journal of Learning and Technology*, 3(3).
- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2). Retrieved from <http://PAREonline.net/getvn.asp?v=2&n=2>

- Winne, P.H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist*, 45(4), 267-276.
- Yancey, K. (2004). *Teaching literature as reflective practice*. Urbana, IL: National Council of Teachers of English.
- Zellers, M. & Mudrey, R. (2007). Electronic portfolios and metacognition: A phenomenological examination of the implementation of electronic portfolios from the instructors' perspective. *International Journal of Instructional Media*, 34(4), 419-430.
- Zimmerman, B. (2000). Attaining self-regulation: A social-cognitive perspective. In M. Boekaerts & P. R. Pintrich (Eds.), *Handbook of self-regulation* (pp. 13-39). New York: Academic Press.
- Zimmerman, B. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166–183.

Authors

Eva Mary Bures, PhD is an Associate Professor at the School of Education, Bishop's University, Lennoxville, Quebec and a member of the Centre for the Study of Learning and Performance, Montreal. A teacher educator, she focuses on supporting innovative assessment and instructional approaches through technology at all levels. Email: ebures@ubishops.ca.

Alexandra Barclay, BA has recently completed her Master's thesis at Mount Saint Vincent University. She is an educational consultant focused on new literacies and assessment practices in school-aged children. Email: alibarclay@hotmail.com.

Philip C. Abrami, PhD is a Professor, Director and Research Chair at the Centre for the Study of Learning and Performance, Concordia University, Montreal, Quebec with expertise on integrating technology across the curriculum especially to support literacy and self-regulated learning. His work ranges from meta-analyses to software development. Email: abrami@education.concordia.ca.

Elizabeth J. Meyer, PhD is an Assistant Professor in the School of Education at California Polytechnic State University in San Luis Obispo, California with expertise in media and electronic portfolios, and a strong focus on bullying. Email: ejmeyer@calpoly.edu.



This work is licensed under a Creative Commons Attribution 3.0 License.