ASSESSING LEARNING SKILLS AND WORK HABITS IN ONTARIO SECONDARY SCHOOLS

By

Stefan Merchant

Dissertation submitted to the Graduate Program in Education in conformity with the requirements for the degree of Doctor of Philosophy

Queen’s University
Kingston, Ontario, Canada
December 2018

Copyright © Stefan Merchant, 2018
ABSTRACT

Teachers in all Canadian provinces, and in many jurisdictions around the world, are required to assess elements of student performance other than academic achievement. In Ontario, these assessments are reported in a section of the report card called “Learning Skills and Work Habits” (LSWH). This dissertation examined Ontario secondary teachers’ assessment practice in relation to the LSWH. Empirical data were obtained through interviews, report cards, and an online survey to provide a better understanding of how teachers define, assess, and grade the LSWH. Relationships between LSWH grades and achievement grades, along with gender differences in LSWH grades were also examined. The first study used semi-structured interviews to establish a baseline understanding of teachers’ LSWH assessment practices, and to inform the development of the survey reported in the third study. The second study used Grade 9 and 12 report card grades from two different school districts to establish the dimensionality of the LSWH grades, to find correlations between LSWH grades and achievement, and to examine gender differences in LSWH grades. The third study focused on a single LSWH (self-regulation) and used interview data, report card data from a third district, and survey data to better understand how Ontario secondary teachers conceptualize and operationalize self-regulation. Findings from the three studies indicate a broad range of assessment practices are present in Ontario secondary classrooms, and that teachers make holistic judgements about students’ LSWH and use those judgements to inform their grading decisions. Findings from this research address a critical gap in the classroom assessment literature and are helpful to school systems seeking to improve teachers’ classroom assessment practices.
ACKNOWLEDGEMENTS

The completion of a Ph.D. is a prodigious undertaking that requires prodigious support. If I were to list every person who had been helpful to me in completing this work, this section would be longer than the actual dissertation, so I shall limit myself to a few individuals, but wish to recognize the many friends, neighbours, and colleagues who were an integral part of my success. Special thanks go to my supervisor Dr. Don Klinger for the guidance, support and whisky provided during every phase of my Ph.D. process. Thanks also to Dr. Nancy Hutchinson for her encouragement, patience and demonstration of how to conduct rigorous qualitative research. Dr. John Kirby pushed me to think deeper and more critically about my work, providing me with feedback that allowed me to grow as a scholar. Rani Cahyani not only moved to Canada with me from Indonesia so I could embark upon my Ph.D., but also spent countless hours looking after our children. Anyone who has both children and a Ph.D. knows what a valuable contribution this is. Similarly, I would like to express my gratitude to my mother, who was always ready to support me with childcare, an encouraging word, or a sympathetic ear. Finally, I would like to express my thanks to the hundreds of teachers who participated in this research. Revealing one’s practice to a researcher is an act that requires courage, and I admire you all for being so open, honest and helpful.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Study context</td>
<td>3</td>
</tr>
<tr>
<td>What are learning skills and work habits?</td>
<td>5</td>
</tr>
<tr>
<td>Organization of the dissertation</td>
<td>7</td>
</tr>
<tr>
<td>CHAPTER 2: LITERATURE REVIEW</td>
<td>9</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>10</td>
</tr>
<tr>
<td>Defining self-regulation</td>
<td>10</td>
</tr>
<tr>
<td>Assessment of self-regulation</td>
<td>12</td>
</tr>
<tr>
<td>Self-regulation and academic achievement</td>
<td>19</td>
</tr>
<tr>
<td>Teachers’ grading practices</td>
<td>22</td>
</tr>
<tr>
<td>CHAPTER 3: ASSESSING LEARNING SKILLS AND WORK HABITS: CAPTURING THE COMPLEXITY</td>
<td>31</td>
</tr>
<tr>
<td>Abstract</td>
<td>31</td>
</tr>
<tr>
<td>Prior research on teacher grading</td>
<td>34</td>
</tr>
<tr>
<td>The Ontario assessment context</td>
<td>37</td>
</tr>
<tr>
<td>Description of the study</td>
<td>38</td>
</tr>
<tr>
<td>Findings</td>
<td>42</td>
</tr>
<tr>
<td>General themes surrounding the research questions</td>
<td>47</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>Discussion</td>
<td>69</td>
</tr>
<tr>
<td>Limitations and further research</td>
<td>72</td>
</tr>
<tr>
<td>Summary</td>
<td>73</td>
</tr>
</tbody>
</table>

**CHAPTER 4: ASSESSING 21ST CENTURY SKILLS: WHAT DO REPORT CARD DATA TELL US?**

<table>
<thead>
<tr>
<th>Abstract</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>76</td>
</tr>
<tr>
<td>The Ontario context</td>
<td>79</td>
</tr>
<tr>
<td>Description of the study</td>
<td>84</td>
</tr>
<tr>
<td>Results</td>
<td>87</td>
</tr>
<tr>
<td>Discussion</td>
<td>94</td>
</tr>
<tr>
<td>Implications and limitations</td>
<td>99</td>
</tr>
<tr>
<td>Conclusion</td>
<td>101</td>
</tr>
</tbody>
</table>

**CHAPTER 5: HOW DO SECONDARY TEACHERS ASSESS SELF-REGULATION? EVIDENCE FROM ONTARIO**

<table>
<thead>
<tr>
<th>Abstract</th>
<th>103</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is self-regulation and how is it assessed?</td>
<td>106</td>
</tr>
<tr>
<td>Method</td>
<td>114</td>
</tr>
<tr>
<td>Results</td>
<td>117</td>
</tr>
<tr>
<td>Definition and dimensionality of self-regulation</td>
<td>117</td>
</tr>
<tr>
<td>Assessing self-regulation</td>
<td>122</td>
</tr>
<tr>
<td>Discussion</td>
<td>128</td>
</tr>
<tr>
<td>Summary</td>
<td>132</td>
</tr>
</tbody>
</table>

**CHAPTER 6: GENERAL DISCUSSION**

|                                                  | 135 |
LIST OF TABLES

CHAPTER 4
Table 4.1  Factor loadings for grades 9 and 12 in district 1  87
Table 4.2  Pearson correlation coefficients between academic achievement and LSWH total score.  90
Table 4.3  Linear model of predictors of standardized mathematics examination scores.  91
Table 4.4  Gender differences in LSWH and academic achievement – district 1 only.  92
Table 4.5  Gender differences in LSWH and academic achievement – district 2 only.  93
Table 4.6  Gender differences in LSWH and academic achievement for Grade 9 physical education – district 1 only.  94

CHAPTER 5
Table 5.1  Proportion of respondents who associated each behaviour with the individual LSWH  124
LIST OF FIGURES

CHAPTER 4

Figure 4.1  Histogram of total LSWH scores for grade 9 students in district 1.  89

Figure 4.2  Scatterplot of total LSWH scores vs. EQAO score for grade 9 mathematics students in district 2.  108
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADT</td>
<td>Academic Diligence Task</td>
</tr>
<tr>
<td>CFA</td>
<td>Confirmatory Factor Analysis</td>
</tr>
<tr>
<td>CFI</td>
<td>Comparative Fit Index</td>
</tr>
<tr>
<td>COPES</td>
<td>Conditions, Operations, Products, Evaluations, and Standards</td>
</tr>
<tr>
<td>EFA</td>
<td>Exploratory Factor Analysis</td>
</tr>
<tr>
<td>ELL</td>
<td>English Language Learner</td>
</tr>
<tr>
<td>GPA</td>
<td>Grade Point Average</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence Quotient</td>
</tr>
<tr>
<td>KIPP</td>
<td>Knowledge is Power Program</td>
</tr>
<tr>
<td>KMO</td>
<td>Kaiser-Meyer-Olkin measure of sampling adequacy</td>
</tr>
<tr>
<td>LSWH</td>
<td>Learning Skills and Work Habits</td>
</tr>
<tr>
<td>MSLQ</td>
<td>Motivated Strategies for Learning Questionnaire</td>
</tr>
<tr>
<td>PSD</td>
<td>Personal and Social Development</td>
</tr>
<tr>
<td>RMSEA</td>
<td>Root Mean Square Error of Approximation</td>
</tr>
<tr>
<td>RMSR</td>
<td>Root Mean Square Residual</td>
</tr>
<tr>
<td>RSSLR</td>
<td>Rating Student Self-Regulated Learning</td>
</tr>
<tr>
<td>SRL</td>
<td>Self-Regulated Learning</td>
</tr>
<tr>
<td>SRLIS</td>
<td>Self-Regulated Learning Interview Schedule</td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

Canadian public schools are expected to develop a broad range of knowledge, skills, and competencies in students. In addition to subject area knowledge and skills, many school systems aim to help students develop transferable skills that will help them succeed in school and contribute positively to Canadian society. Saskatchewan (2010), for example, includes “Understand, value and care for others” and “Use moral reasoning processes” as part of its broader aims for education. British Columbia (2001) developed a set of performance standards for “social responsibility” and Ontario (2010) has created a set of “Living Skills” expectations that are “designed to help students develop a positive sense of self, use coping and management skills, monitor their own progress, develop and maintain healthy relationships, and use critical and creative thinking processes as they set goals, make decisions, and solve problems” (p. 12). Outside of Ministries of Education, groups such as the Conference Board of Canada, Human Resources and Skills Development Canada, and People for Education are also calling for schools to develop these types of skills, as these skills are associated with positive economic and social outcomes such as higher income and lower drug use (Heatherton & Wagner, 2011; Heckman & Kautz, 2012).

Canadian schools are not just wanting to develop these skills in students, they are also wanting to assess them (Merchant, Klinger & Love, in press). While the exact skills vary from province to province, all provinces require teachers to assess elements of student performance beyond subject area achievement. These assessments are reported upon using different terms such as “competencies,” “learner profile,” and “social and behavioural indicators.” The term used in Ontario is learning skills and work habits (LSWH). While there appears to be broad agreement that schools should assess and report upon these skills, the methods teachers use to do
so is unknown. There is a large body of research examining teachers’ assessment and grading practices; however, this research focuses almost exclusively on assessment and grading of subject area knowledge (e.g., Brookhart et al., 2016; Nash & McMillan, 2000; Randall & Engelhard, 2010). To date, there has been minimal research on how teachers assess student skills such as the LSWH, and little of this research has taken place in the Canadian context. This dissertation aims to address this gap and provide baseline, foundational knowledge on how Ontario secondary teachers assess students’ LSWH.

All Ontario K-12 teachers must assess a set of six LSWH: collaboration, independent work, initiative, organization, responsibility, and self-regulation. When reporting student performance on the LSWH, teachers use a 4-point scale with scale points of “Needs improvement,” “Satisfactory,” “Good,” and “Excellent.” Because of the lack of research on how teachers assess LSWH, it is unknown what behaviours teachers are using to assess each LSWH, how these assessments are conducted, and how they make grading decisions. Thus we do not know the constructs these grades represent, nor how to interpret the grades. The immediate goal of my research is to illuminate not only teachers’ LSWH assessment practices, but also to better understand how we may interpret and use this grading information, and, ultimately, improve this aspect of teachers’ practice. The major research questions guiding this study are:

1. How are Ontario secondary teachers defining each of the six LSWH? What student behaviours do they use to assess them?
2. To what extent are the LSWH assessed independently of each other?
3. How strongly are the LSWH associated with academic achievement?
4. How do teachers’ conceptions of the LSWH align with established definitions of self-regulation?
Multiple methods of data collection were employed to obtain the data to answer these questions. These methods included semi-structured interviews with teachers, quantitative analyses of report card data, and an online survey.

Study Context

My research occurred in the province of Ontario and was situated at the secondary level. Ontario provided an ideal context in which to conduct this research as all K-12 teachers are required to assess the same six LSWH and have been doing so since 2010. Thus Ontario teachers have had time to refine and stabilize their LSWH assessment and grading processes. Further, schools and school districts have had time to devise and enact their own policies with respect to LSWH assessment. Lastly, the Ontario Ministry of Education (2010) provides guidance to teachers on how to assess the LSWH through a comprehensive assessment guide called Growing Success. This guide identifies sample student behaviours teachers may use for assessing each LSWH. According to the assessment guide, the sample behaviours provided “are intended to assist but not restrict teachers” (p. 10, emphasis in the original) in how they assess the LSWH.

This comment indicates that Ontario teachers have a high degree of control over their LSWH assessment and grading practice. In fact, Ontario teachers enjoy a high degree of professional autonomy in their overall assessment practice. Growing Success states that, “determining a report card grade will involve teachers’ professional judgement and interpretation” (p. 39) and includes specific examples of teacher autonomy such as teacher control over weighting of summative assessments, and the option to deduct marks for late assignments. Further, secondary teachers in Ontario have few standardized examinations for which to prepare their students. Ontario secondary students write two standardized examinations; a Grade 9 mathematics examination, and a Grade 10 literacy examination that is a minimum
competency examination intended to ensure that Ontario graduates have a minimum level of literacy. Neither of these examinations impact the teacher awarded grades—although a Grade 9 mathematics teacher may include the Grade 9 mathematics examination score, or a portion of it, into the final grade if they wish.

The result of this professional autonomy, and the lack of high-stake, large scale testing, is that classroom assessment of achievement can have high stakes, as teachers’ grades are the primary determinant of admissions for tertiary education. In contrast, the LSWH grades appear to have little, if any, impact on subsequent student opportunities. There is no evidence these grades are used for admission into post secondary programs, and minimal anecdotal evidence they are used for internal purposes within schools. While Ontario secondary teachers’ achievement grades can be high stakes, the LSWH grades appear to be not just low stakes, but no stakes. This is an important contextual element as the low-stakes nature of LSWH grades may lead to teachers not valuing this portion of their assessment practice, and to stakeholders (e.g., parents, school administrators) not holding them accountable for weak or inconsistent assessment and grading practices.

Another important contextual element is that most Ontario secondary teachers receive little training on how to assess LSWH prior to becoming teachers. While it is difficult to determine exactly what assessment and grading training teacher candidates receive across different B.Ed. programs in Ontario, it is clear much of the assessment training teacher candidates receive is embedded within curriculum courses and focuses on subject area achievement. Further, given that secondary school teachers’ undergraduate degrees are typically in the subject areas they teach, it is unlikely their prior undergraduate experiences provided them with the skills and knowledge required to be good assessors of LSWH. While it is possible that
teachers are acquiring these assessment skills through professional development, anecdotal evidence suggests this is happening on an inconsistent basis. As an example, one district from which I collected data for my research adopted the LSWH as a focus for professional development, offering single-day workshop training, and encouraging teachers to form collaborative inquiry groups. Another district presents a one-hour session on assessing LSWH as part of their new teacher induction program, but offered no further mandatory training, while one of the other districts I worked with appeared to offer no training or guidance on how to assess the LSWH.

What are Learning Skills and Work Habits?

The language surrounding learning skills and work habits can be confusing and there is little agreement on terminology (Duckworth & Schulze, 2009; Farrington et al., 2012). While learning skills is a commonly used term within the education profession, I could find no accepted definition of it in the research literature. In their meta-analysis on the effectiveness of learning skills interventions, Hattie, Biggs, and Purdie (1996) listed motivation, mnemonic skills, and self-regulation as skills that could be classified as learning skills; however, they did not define the term itself. Boyatzis and Kolb (1991) defined learning skills in their validation study of the “Learning Skills Profile,” a 72-item self-report assessment of learning skills. According to Boyatzis and Kolb, a learning skill is a “generic heuristic that enables mastery of a specific performance domain” (p. 280). Other authors also acknowledged that learning skills are generic in nature but emphasized that learning skills allow mastery of a wide range of domains (Costa & Kallick, 2000; Duckworth & Seligman, 2005; Dweck, Walton & Cohen, 2011; Rychen & Salganik, 2000; Tough, 2012).
For the purposes of my research a “learning skill” will be defined as a set of knowledge and abilities that facilitate learning. For example, a student who is proficient at “organization” would have knowledge of why it is beneficial to be organized, and of different strategies that can help organization. This student would then be able to implement those strategies in an effective and consistent manner. Learning skills can be developed through practice and are applied thoughtfully and intentionally. For example, metacognition would be classified as a learning skill because of its general applicability and mutability (Biggs, 1988; Kramarski & Mevarech, 2003). This is not to say that students who are metacognitive in one subject will be metacognitive in all subjects, only that metacognition is useful for learning across subjects—if it is actually used is a different matter. In contrast, a trait such as IQ would not be considered a learning skill. Although IQ underpins learning across different subject areas, there is evidence that it is a stable trait, relatively unaffected by training (Beaver et al., 2013). Learning skills should also be distinguished from characteristics such as humility or politeness. While a school system may decide it is worthwhile to develop and assess these characteristics in students, there is no evidence they are related to better learning, and so would not be considered learning skills.

Habits are different from skills in that they are not applied effortfully (Corno, 2004). Indeed, it takes conscious thought and effort to break a habit. Habits are “automatic response tendencies that are triggered by contextual cues” (Galla & Duckworth, 2015, p. 509). Galla and Duckworth (2015) argued that habits may be more beneficial than learning skills because they require less self-control, and so are more likely to be applied consistently. Habits form by repeating a behaviour in response to a cue. For example, before going to bed, checking one’s organizer to ensure that all homework assignments have been completed, and required materials
for the next school day are in your school bag. While creating this association may take effort at first, eventually it becomes automatic.

I take the view that the learning skills and work habits, as used by the Ontario Ministry of Education, refer to a common set of skills and behaviours that vary in terms of the effort required to implement them. This is to say we cannot classify some of the six Ontario learning skills and work habits as “skills” and others as “habits.” All of them are learning skills, but students may differ in how habitually these skills are applied. For some students, writing down homework in their diary may be an automatic response done almost thoughtlessly, while for others this same action may require conscious thought and discipline. Thus, the six LSWH refer to a set of skills and behaviours that are applicable to all students in all subject areas at all grade levels, but there may be substantial variance among students in how effective they are at demonstrating these skills and in how habitually these skills are applied across educational contexts. There is some support for this concept of the LSWH from other jurisdictions. For example, Manitoba and Prince Edward Island anchor the scale points of their learning skills grades with terms such as “consistently,” “sometimes,” and “rarely” (Manitoba Education and Advanced Learning, 2015; Prince Edward Island English Language School Board, 2015).

Organization of the dissertation

This dissertation is comprised of three related studies of Ontario teachers’ assessment of LSWH. Each study forms a separate chapter. These studies are preceded by a literature review which summarizes two distinct bodies of research. The first is teacher grading practices, and the studies are drawn from the literature on educational assessment. The second is the assessment of constructs related to the LSWH such as self-regulation, conscientiousness, and executive function. Studies in this portion of the literature review are drawn from both education and
The first of the three studies is presented in Chapter 3. This is a qualitative study which used semi-structured interviews with 26 teachers from three school districts. Participants encompassed a range of subject areas and years of experience. This study provided rich, baseline data used to inform the research design of the other two studies. Interview questions focused on how teachers define each of the LSWH, what behaviours they use to assess them, and how they distinguish among achievement levels.

The second study is presented in Chapter 4. This is a quantitative analysis of over 80,000 sets of Grade 9 and 12 grades from two districts. Descriptive statistics and t-tests were used to find differences in gender achievement on the LSWH, and multiple linear regression was used to determine how well LSWH grades serve as predictors of academic achievement. Factor analyses were also performed to examine to what extent the six LSWH are graded as separate constructs.

The third study is presented in Chapter 5. It used a mixed methods approach to investigate how Ontario secondary teachers define, assess, and report upon self-regulation. In addition to the interview and report card data already mentioned, data were collected using an online survey. Findings from the three data sources were used to form a deeper understanding of how teachers define self-regulation and what behaviours they use to assess it.

The final chapter of the dissertation synthesizes the findings from the three studies to inform broader conclusions that can be reached from these three studies. Common findings and important conflicts are examined, as are the limitations and implications of the work. The final chapter situates the three studies within the existing literature on teachers’ grading practices and demonstrates its unique contribution to this body of research.
CHAPTER 2: LITERATURE REVIEW

Research into teachers’ assessment and grading practices is an established field with a long history. Brookhart et al. (2016) noted that grades “were the focus of some of the earliest educational research” (p. 804) and that grading continues to be a fruitful area of educational research. One consistent finding from the last century of grading research is that teacher constructed grades are not pure measures of achievement, but a mix of achievement and non-achievement factors (e.g., Bowers, 2011; Miner, 1967; Sobel, 1936). Non-achievement factors included in grading decisions may include effort, participation, and improvement. In response to this, some assessment and grading experts recommended that teachers report upon achievement and non-achievement factors separately (e.g. Guskey & Bailey, 2010; Stiggins, Frisbie, & Griswold, 1989). Canadian school systems have wholeheartedly adopted this recommendation, with every province in Canada asking teachers to report non-achievement factors separately from subject area achievement (Merchant, Klinger & Love, in press).

There is an obvious logic to asking teachers to assess and report upon non-achievement factors separately from subject area achievement. There is the potential to obtain a more complete picture of student achievement by providing two distinct measures of student performance in the classroom. However, to date, there has been minimal research on how teachers assess and report upon non-achievement factors. My research aims to address this gap by investigating both how Ontario secondary teachers assess the six Learning Skills and Work Habits (LSWH), and what they are assessing. For these reasons, this review examines two distinct areas of educational research—student self-regulation and teacher grading practices.

The first part of the review focuses on self-regulation. Self-regulation was chosen because it is the only LSWH for which there exists an extensive literature, and the LSWH for
which the greatest number, and variety, of assessments exist. The review presented here examines differing concepts of self-regulation, how it is assessed, and its association with academic success. The second portion of the review focuses on teachers’ grading practices. Both areas of research are vast and conducting a thorough, systematic review would require time and resources beyond what is available. Further, the purpose of this review is not to review these areas exhaustively, but to provide the reader with a preliminary understanding of these two areas of research. This understanding will serve as a foundation for reading and interpreting the studies presented within this dissertation.

Self-Regulation

Central to my research is an understanding of teachers’ conceptions of LSWH. Unless it is known what teachers are assessing, it is of limited value to discuss the quality or validity of their assessments. A thorough review of all six LSWH and their associated constructs would be far too lengthy, so this review focuses on self-regulation. Self-regulation was chosen as a focus not only because of the large volume of literature on the topic, but also because preliminary interviews revealed that a substantial proportion of the teachers interviewed thought of self-regulation as an umbrella concept that encompasses the other LSWH. This view of self-regulation has some appeal, as there is evidence to support that self-regulation underpins, or is related to, other LSWH such as organization, responsibility, and independent work (Zimmerman, 2002).

Defining Self-Regulation

One definition of self-regulation is the “capability of controlling or directing one’s attention, thoughts, emotions, and actions” (McClelland & Cameron, 2012, p. 136). It is not difficult to imagine using the same set of words to define self-control or self-discipline, and in
fact it appears the three terms are often used to refer to the same construct (Duckworth & Schulze, 2009). Boekarts and Corno (2005) noted that “over the past two decades, researchers have struggled with the conceptualisation and operationalisation of self-regulatory capacity, coming to the conclusion that there is no simple and straightforward definition of the construct of SR [self-regulation]” (p. 200).

An early, influential model of self-regulation was put forward by Kopp (1982). Kopp described self-regulation as “the ability to comply with a request” (p. 199) and the ability to “generate socially approved behavior” (p. 199). Kopp’s paper demonstrates that self-regulation is a multi-dimensional construct involving compliance, self-control, and awareness of social norms. More recent research puts forward a different description of self-regulation, but still acknowledges its multi-dimensionality (Duckworth & Kern, 2011). Self-regulation involves more than just regulating cognitive processes. It encompasses a range of regulatory behaviours that work together to maximize an individual’s potential for achieving a goal (Duckworth & Carlson, 2013). Diamond (2013) described self-regulation as “processes that enable us to maintain optimal levels of emotional, motivational, and cognitive arousal” (p. 152). Boekarts (1997) linked self-regulation in school settings to cognition, motivation, volition, goal setting, effort, and prior knowledge, whereas Moffitt et al. (2011) linked self-regulation with conscientiousness. Nigg (2017) listed 19 different constructs related to self-regulation and noted that it was an incomplete list. With such a broad range of constructs involved in self-regulation, it is perhaps not surprising that a consistent, widely accepted definition eludes us.

One way of narrowing the definition of self-regulation is to consider it only in a classroom or academic setting. This is self-regulation in the service of learning. Some authors (e.g. Schunk, 2005) have equated self-regulation with self-regulated learning, but more
commonly self-regulated learning (SRL) is viewed as a cyclical process involving planning, action, and reflection (Winne & Hadwin, 1998; Zimmerman, 2013). One influential model of SRL is Winne’s (1997) Conditions, Operations, Products, Evaluations, and Standards (COPES) model. Key elements of the COPES model include goal setting, taking actions, and reflecting upon the effectiveness of those actions to inform future actions (Hadwin & Winne, 2012). Another influential model is Zimmerman and Campillo’s (2003) cyclical phase model of SRL. This model includes goal setting, metacognitive monitoring, self-evaluation, task strategies, and help-seeking. Both these models overlap substantially (if not completely) with the behaviours the Ontario Ministry of Education associates with self-regulation. As an example, Growing Success, the Ontario Ministry of Education (2010) assessment guide suggests teachers may assess self-regulation using student behaviours such as, “sets own individual goals and monitors progress towards achieving them” (p.11), and “assesses and reflects critically on own strengths, needs, and interests” (p. 11). These behaviours fit into all three models of SRL mentioned above. Given the overlap between the ministry’s description of self-regulation and SRL, this review includes works from both the self-regulation and the SRL literature.

Assessment of Self-Regulation

Part of the reason self-regulation is difficult to define precisely is that it is a complex phenomenon, especially in the context of learning in a classroom (Credé & Kuncel, 2008; Dinsmore, Alexander & Loughlin, 2008; Stroud, 2013). Perhaps as a result of this complexity, myriad assessments of self-regulation have been developed including; self-report inventories, observation, structured interviews, think-alouds, error detection tasks, online exercises, and teacher judgments (Galla et al., 2014; Stroud, 2013). It appears most measures are used only in one study or by a very limited number of researchers. Duckworth and Schulze (2009) meta-
analyzed 282 studies on self-control (a construct the authors equate to self-discipline and self-regulation) and identified 154 different measures across the studies. However, some measures, especially self-report measures, have enjoyed greater popularity (Winne & Perry, 2000). One popular self-report measure of self-regulation is the Motivated Strategies for Learning Questionnaire (Dinsmore, Alexander & Loughlin, 2008; Panadero, Jonsson, & Botella, 2017).

The Motivated Strategies for Learning Questionnaire (MSLQ) is a self-report instrument developed for college students to measure both their motivation and use of learning strategies (Pintrich, Smith, García & McKeachie, 1993). While not a tool specifically designed to measure self-regulation per se, it originates from a social cognition perspective and the questionnaire includes a 50-item learning strategies portion that aligns well with self-regulation. The MSLQ requires respondents to reflect upon their general learning behaviours and, therefore, measures aptitude and not maximum performance. Self-report measures are popular for assessing self-regulation because they can be useful for accessing internal thought processes and because they are economical and easy to implement (Winne & Perry, 2000). One advantage of the MSLQ’s popularity as an instrument is that its psychometric properties are well known. Unfortunately, the psychometric properties are not ideal, with Cronbach’s alpha ranging from $\alpha=0.52$ to $\alpha=0.80$ for the different learning strategies subscales (Pintrich, 1991). The test manual further states that goodness of fit indices “are not stellar” (p. 79) with the factor model. The relatively weak psychometric properties of the instrument may be due to respondents not being reliable reporters of their own learning strategies. Lapses in memory, distorted self-image, changing self-perception, and different ways of interpreting inventory items are all factors that can reduce the validity and reliability of self-report measures (Duckworth & Yeager, 2015; Hirschfeld, Thomas & McNatt, 2007).
Interviews are another means of assessing self-regulation. As with questionnaires, they involve self-report, but interviews allow researchers to explore a student’s thought processes in greater depth than questionnaire items. The most popular interview protocol appears to be the Self-Regulated Learning Interview Schedule (SRLIS). Zimmerman and Pons (1986) developed the SRLIS as a means of assessing self-regulation in high school students. The SRLIS was developed with an a priori structure of 14 categories of self-regulation strategies, although later work determined the interview items loaded onto 3 factors (Zimmerman & Martinez-Pons, 1988). The 14 categories were formulated based upon prior research. The initial study using the SRLIS had 80 Grade 10 students; 40 from a high achievement stream and 40 from other (lower achievement) streams. The SRLIS is essentially a count of self-regulation strategies and participants are awarded a point for each strategy they reported they used. The validation study results showed that 91% of the students were correctly classified as being in a normal or low achieving stream by the SRLIS. The authors noted that SRLIS had better predictive power than other variables they studied including SES and gender. They further found that consistent strategy use mattered more than which strategy was used. However, it could be argued that these results are not as impressive as they seem at first. Firstly, there was a substantial difference in achievement between the two groups. In standardized testing, the high achievement group scored 2.63 standard deviations higher in math achievement, and 1.75 standard deviations higher in English achievement than the lower achievement group. These are substantial differences in achievement. Given this, it is not clear if the SRLIS would be sensitive to comparisons between students with more similar levels of achievement. It is possible the SRLIS is only effective in finding large differences in self-regulation. Secondly, the participants were streamed into higher and lower achieving groups based partially upon their GPA. If their teachers included factors
such as effort and self-regulation into their grading, it seems likely that the two groups were “pre-screened” in effect, and that students in the low achieving group would have likely have been placed there at least partially as a result of their lower self-regulation.

These critiques of the SRLIS are not intended to demonstrate that the SRLIS is a profoundly flawed tool, rather they serve to illustrate the difficulty in validating any assessment of self-regulation. With no absolute standard to use, any comparison used for validation is open to criticism. One of the more common comparisons made is to teacher or parent report. Given that teachers have extended ongoing contact with students, it might seem they are ideally positioned to assess students’ self-regulation, but there have been concerns about teachers’ ability to be reliable raters of students’ self-regulation (Winne & Perry, 2000). Of particular concern are teachers’ abilities to separate self-regulation from other constructs such as achievement and motivation. This is a valid concern, given there is limited empirical evidence indicating teachers are able to distinguish between self-regulation and other constructs. Much of the research involving teacher assessments of self-regulation requires teachers to assess only self-regulation, and so it is not possible to determine to what extent teachers assess it independently of other constructs such as compliant behaviour (e.g., Boekaerts & Cascallar, 2006).

Zimmerman and Martinez-Pons (1988) used teacher reports as a comparison to help validate their SRLIS. Teacher ratings were done using the Rating Student Self-Regulated Learning Outcomes: A Teacher Scale (RSSRL). The RSSRL is a 12-item questionnaire that asks teachers to reflect on student self-regulation. Zimmerman and Martinez-Pons found that teachers assessed self-regulation as a unitary construct as the teacher ratings all loaded onto a single factor. Encouragingly, the teachers did not include confounding factors such as verbal
expressiveness in their ratings and the reliability was excellent (internal consistency = 0.95 using the Kuder-Richardson formula 20). However, these results must be interpreted with extreme caution as the sample size was only 3 teachers.

Interviews, self-report and retrospective teacher ratings of self-regulation all rely on memory. Memories can be unreliable and changing (e.g., Patihis et al., 2013). One way to avoid these problems is to ask participants to perform a specific task requiring self-regulation and then observe that task. The Academic Diligence Task (ADT) developed by Galla et al. (2014) is an example of this type of measure. The ADT asked participants (high school seniors) to solve a series of single digit subtraction problems using an online software tool that has built in distractors. The distractors were Tetris and YouTube videos. This type of assessment allows a researcher to see self-regulation in situ, but there remains a problem of generalizability. Can we know if the self-regulation (or lack thereof) demonstrated by participants is typical of their ability to self-regulate? In this particular assessment, it is difficult to see any motivation for participants to persist in the task. Given that the participants were high school seniors, what reasons would they have for persisting with single digit subtraction problems? One presumes that high school seniors are already expert at single digit subtraction, and so there is no benefit, to persisting in the task. We know that motivation impacts self-regulation (Baumeister & Vohs, 2007), so the absence of obvious motivation to complete the tasks may explain why the results of this study showed that while time spent on the subtraction problems significantly correlated to conscientiousness ($r = 0.09, p < 0.01$), the correlation to agreeableness ($r = 0.11, p < 0.01$) was slightly stronger. Given these weak correlations, it is difficult to determine if the ADT is a measure of diligence or a measure of compliance, or neither. This type of problem plagues
laboratory measures of self-regulation. There is always the question of how well
decontextualized measures of self-regulation apply to complex environments such as classrooms.

The measures of self-regulation described above give an indication of the difficulties of
assessing self-regulation with acceptable levels of reliability and validity. These measures have
fundamental flaws in how they measure self-regulation. Part of the difficulty may be poor
design, but clearly part, if not most, of the difficulty comes from the complex, context dependent
nature of self-regulation. Self-report inventories are quick, easy to score and tap into internal
thought processes that may not be evident when observing behaviours. However, these
inventories also have validity and reliability problems due to participants’ changing moods, lack
of self-knowledge, and differing interpretations of the items (Paulhus & Vazire, 2007). Interview
protocols allow respondents to expand upon their answers and explore a range of responses
beyond what was originally conceived by the researcher, but they are time consuming and still
rely upon the respondent’s memory. With their better ability to control the environment, task-
based assessments may provide more reliable data; however, there is a question as to how well
the results generalize, or how meaningful the results are in a school context. There are other
ways to assess self-regulation, such as think alouds and trace data (e.g. Greene, Robertson &
Costa, 2011; Hadwin, Nesbit, Jamieson-Noel, Code & Winne, 2007; Zimmerman, 2008), but
these methods also have weaknesses. Think-alouds place cognitive demands on the respondent,
which can lead to changes in the respondents’ thinking (Ericsson & Simon, 1998). Trace data
require a great deal of inferencing and there can be issues with reliability in interpreting the
meaning of the traces (Winne & Perry, 2000). This means that assessors need to be well trained
for trace data to be properly collected and interpreted (Winne, 2010). Further, some methods of
obtaining trace data, such as through online tools (e.g., Hadwin, Nesbit, Jamieson-Noel, Code, &
Winne, 2007), do not align well with how most classroom teachers conduct their assessments, although this could change as online tools become more integrated into classroom teaching and assessment. As Duckworth & Yeager (2015) remarked “no measure is perfect” (p. 239), and so it seems likely that good assessment of students’ self-regulation will require multiple measures, using multiple methods, over an extended period of time.

Teachers are in a unique and privileged position to do this. Teachers have sustained interaction with their students and so can offer a context rich perspective that is missing from assessments of a single event such as think aloud processes or task-based measures (Zimmerman & Pons, 1988). Teachers are also exposed to a range of self-regulatory styles and capabilities among their students and so have a sense of how a student’s self-regulation compares with norms of the current classroom and the teacher’s prior classrooms. Further, teachers can assess student self-regulation using a variety of tools. Nothing prevents teachers from using questionnaires or interview protocols with their students. These data could then be supplemented with day-to-day observations, student self-assessments, reflections, and journals.

The concerns raised by some researchers (e.g., Hoge, & Butcher, 1984; Winne & Perry, 2000) about teachers’ abilities to assess self-regulation are worthy of consideration but these concerns are counterbalanced by research that indicates that teachers have the potential to be competent assessors of self-regulation (Klug, Bruder, Kelava, Spiel & Schmitz, 2013; Zimmerman & Martinez-Pons, 1988). Given that self-regulation is context dependent, and changing as students develop, it makes sense that any single measure will be inadequate. Thus the ability for teachers to give multiple assessments of self-regulation over time may compensate for their lack of assessment expertise and allow them to be trustworthy reporters of student self-regulation.
Self-Regulation and Academic Achievement

There is overwhelming evidence linking self-regulation (and associated constructs) with academic achievement (Tangney, Baumeister, & Boone, 2004). The positive correlation between self-regulation and academic achievement holds true from kindergarten (e.g. Allan, Hume, Allan, Farrington & Lonigan, 2014; McLelland & Wanless, 2012; Sawyer et al., 2015) to higher education (Bail, Zhang & Tachiyama, 2008; Kitsantas, Winsler & Huie, 2008; Schapiro & Livingston, 2000). The link holds true not just in North America but also in Europe (Gestsdottir et al., 2014; Hubert, Guimard, Florin & Tracy, 2015), Asia (Wanless, McClelland, Acock, Chen & Chen, 2011) and in online environments (Cho, & Shen, 2013; Winters, Greene, & Costich, 2008). Summarizing this volume of evidence would require an extensive meta-analysis, and so this review will focus on three representative studies. These studies were selected because they were conducted with high school students, and because the results can be linked to research on teachers’ grading practices, including the current studies.

Zimmerman and Kitsantas (2014) conducted a study in Virginia that compared the links between self-regulation and self-discipline and academic achievement across 504 high school students. The self-regulation and self-discipline measures included self-report questionnaires and teacher observation (using formalized observation protocols). Academic achievement was measured using both GPA and a state-wide standardized test. The overall findings were that the composite measures of self-regulation and self-discipline were correlated to each other ($r = 0.54$, $p < 0.01$), and both self-regulation and self-discipline were correlated to academic achievement. Student GPA was most strongly correlated to the composite measure of self-regulation ($r = 0.81$, $p < 0.01$), but it was also highly correlated with teacher ratings of self-regulation ($r = 0.64$, $p < 0.01$). While one obvious explanation for these correlations would be that better self-regulation
leads to better learning, it is also possible that teachers include constructs such as self-regulation in their grading of academic achievement, leading to a strong correlation between GPA and teacher-reported self-regulation. The finding that teachers’ ratings of self-regulation had a lower correlation with the state-wide standardized test scores \((r = 0.33, p < 0.01)\) is consistent with this explanation.

Nota, Soresi and Zimmerman (2004) conducted a study of 81 high school students in Italy, 49 of whom provided follow-up data from university. The study used structured interviews to assess the types of self-regulated learning strategies students used, and how effectively they employed these strategies. One strategy in particular, “organizing and transforming” stood out as particularly effective. “Organizing and transforming” occurs when students rearrange or modify instructional materials to improve learning. This could be anything from creating an outline of a paper before writing the paper to rearranging terms in an equation to make the equation easier to understand and solve. This strategy was the only strategy that significantly predicted achievement in the three subjects included in the study (Italian, Mathematics, and Technical subjects), and accounted for 83% of the variance in Technical subject grades. The amount of variance accounted for was much lower in Mathematics (23%). When the self-regulated learning strategies were considered as a composite there were moderate to strong correlations with academic achievement as measured by final high school grades. These grades were a mix of teacher evaluations and standardized scores, although how the exact grades were determined was not described. The results showed correlation coefficients ranging from a low of \(r = 0.47\) between SRL and mathematics grade to a high of \(r = 0.91\) between SRL and grades in Technical subjects. The findings also showed that academic grades had lower correlations with SRL than did non-academic grades. This could be because teachers of non-academic subjects included
SRL to a greater extent in their grading decisions, consistent with the findings of Duncan and Noonan (2007), or it could be because SRL is of greater importance for success in non-academic subjects. Unfortunately, the data offered by Nota et al. do not allow us to determine which of these explanations is correct.

The study conducted by Duckworth, Quinn, and Tsukayama (2012) may be the most relevant to my research. The authors tested different structural equation models to examine the causal relationship between self-control and academic achievement. The first phase of their research involved 1364 Grade 9 students. Report card grades for mathematics, English, science, and social studies in Grades 8 and 9 were used as the outcome variable. Because these grades fit a one factor model, the average grade across the four subjects was used. Two predictor variables were included. The first was self-control and was measured using parent and teacher reports. The instrument was the Social Skills Rating System. This questionnaire lists a variety of behaviours and asks how often the child exhibits the behaviour. The other predictor variable was IQ as measured by the Wechsler Abbreviated Scale of Intelligence. This test was conducted when the students were in Grade 4. The finding from this portion of the study was that changes in report card grades from Grade 8 to Grade 9 were better predicted by self-control ($\beta = .20, p < .002$) than IQ ($\beta = .09, p = .044$), whereas changes in standardized test scores were predicted by IQ ($\beta = .29, p < .001$), but not self-control ($\beta = .01, p = .88$). The second portion of the study involved a different cohort of students (N = 510) in Grades 5 through 9. IQ was measured using Raven’s Progressive Matrices, and self-control was measured using the Impulsivity Scale for Children, which was completed by students, parents, and teachers. In addition, homework completion and classroom behaviour were measured using teacher report, in which teachers rated students on a 5-point scale. Confirmatory factor analysis showed homework completion and classroom
behaviour were best described using a one-factor model ($\chi^2(5) = 34.76, p < .001, \text{CFI} = .98$).

Once again, self-control was a better predictor of grades ($\beta = .22, p < .001$) than IQ ($\beta = .01, p = .60$), and IQ was a better predictor of standardized test scores ($\beta = .12, p < .001$) than self-control ($\beta = -.05, p = .44$). However, the association between self-control and grades was mediated by homework completion and classroom behaviour. Sobel’s test ($z = 2.31, p = .02$) confirmed the mediated model fit the data better than the non-mediated model.

Taken together, these three studies demonstrate self-regulation is positively associated with report card grades. However, the link between self-regulation and standardized test scores is much weaker and may be non-existent. Since the self-regulation measures used in all three studies include teacher report methods, it is impossible to draw a causal link between better self-regulation and better learning. The idea that better self-regulation leads to better learning is appealing, but knowing that teachers include factors such as homework completion, behaviour, and participation in their grading decisions (e.g., Cross & Frary, 1999; Duncan & Noonan, 2007; Russell & Austin, 2010) means that a reasonable alternative hypothesis would be that higher levels of self-regulation lead to better grades not because of better learning, but because of how teachers construct grades.

Teachers’ Grading Practices

How teachers assess and grade constructs such as self-regulation is unknown. There has been minimal research on classroom assessment of these constructs. Thus, the studies cited here focus on how teachers assess and grade academic achievement. This research has shown repeatedly that report card grades are not pure measures of achievement. Instead, teachers are known to deliberately or unconsciously include other factors such as effort, attendance, and behaviour when assigning grades (Brookhart, 1993; Cross & Frary, 1999; Russell & Austin,
One of the reasons that grades are not pure measures of achievement may be that teachers use grades to give a holistic sense of how that child performs the role of “student” (Allen, 2005; Taylor, 1964), which could include attention to other factors such as behaviour and effort. A teacher’s grade for a student may be further complicated by other factors such as the desire to minimize conflict, motivate a student, and maintain course enrolment (Bowers, 2011; Brookhart, 2004; Cizek, Fitzgerald & Rachor, 1995; Cross and Frary, 1999; Hunter, Mayenga & Gambell, 2006; Russell & Austin, 2010).

What information do teachers consider when determining report card grades? There is no single answer to this question, with research showing that teachers’ grading practices are highly variable and dependent upon subject, grade level, the individual teacher, and the individual student (Bowers, 2011; Brown, 2011; Howley, Kusimo & Parrott, 2000; Lekholm & Cliffordson, 2008). For instance, Duncan and Noonan (2007) found that mathematics and science teachers included LSWH (which they term “academic enabling behaviours”) to a lesser extent than teachers of other subjects such as English, social studies, and performing arts. As an extreme example of this, Russel and Austin (2010) found that attendance accounted for about 65% of the students’ secondary music grades. This is at odds with most studies on teachers’ grading that show that achievement is the dominant factor in determining students’ grades (Kelly, 2008; Randall & Engelhard, 2009). In comparing different high school subjects, it has been found that mathematics teachers incorporate achievement measures to the highest degree in their grading (Frary, Cross & Weber, 1993). Most teachers will also consider how close a student is to a grade boundary (Randall & Engelhard, 2010). While there is evidence that teachers try to communicate their grading criteria clearly (Brookhart, 1994), and teachers generally follow their stated criteria, it also appears that, in the case of atypical students, teachers’ grading practices become highly
inconsistent (Canal, Bonini, Micciolo & Tentori, 2012). Finally, we know that teachers incorporate LSWH to a greater extent with low achieving students, than they do with high achieving students (McMillan, 2001). In total, teacher grading practices are complex, idiosyncratic, and often at odds with practices recommended by experts in educational assessment (Cox, 2011). This suggests that a likely finding in this dissertation will be that great variance exists in how Ontario secondary teachers grade the LSWH.

The wide variety of factors that teachers consider in assigning a grade has made it difficult to ascertain the exact meaning of a teacher’s grade, and this has led some in the educational community to question the validity or reliability of teachers’ grades (Brookhart, 2004; Marzano, 2000). The fact that teachers do not grade according to the recommended practice of using only achievement data has been called, “disconcerting,” by Brookhart (1994, p. 289). Other researchers have made similar comments (e.g., Campbell, 2012). Students, parents, and the public at large should know what a grade represents and incorporating diverse factors such as LSWH into the grade in varying and inconsistent amounts makes this impossible.

Despite the documented flaws and problems with teachers’ grades, there are some positive aspects to teachers’ grades that must be noted. For instance, it appears that because teachers’ grades incorporate multiple facets of student performance such as effort and behaviour, they serve as better predictors of future success than achievement tests alone (Chamorro-Premuzic & Furnham, 2003; Heckman & Kautz, 2012). It is known that teachers’ grades are excellent predictors of high school completion (Allensworth & Easton, 2007) and serve as good predictors of college achievement and completion (Bowen, Chingos, & McPherson, 2009; Sparkman, Maulding, & Roberts 2012; Wolfe & Johnson 1995). Even the College Board, the
developers of the SAT, have shown that teacher determined grades are better predictors of first year college grades than the SAT (Camara & Echternacht, 2000; Kobrin et al., 2008).

How is it possible that the SAT, with its sophisticated development and implementation, is a poorer predictor of future performance than teachers’ grades? It has been hypothesized that this is because teachers’ grades include LSWH (Duckworth & Seligman, 2006). By including LSWH in their grading decisions, teachers may be providing a more holistic picture of student achievement—a picture that includes factors likely to lead to long-term success in academics and in life. For example, it is known that personality impacts achievement. Conscientiousness, one of the Big 5 personality traits, has been shown to correlate with both academic achievement and job success (Levin, 2012; Heckman & Kautz 2012; Salgado, Moscoso, & Berges; 2013). Conscientiousness is also related to LSWH such as responsibility, organization, and self-regulation (Bidjerano & Dai, 2007; Costa, McCrae, & Dye, 1991). If teachers are including LSWH in their grading decisions, they are including factors known to correlate with academic and job performance. This may explain why teachers’ grades serve as better predictors of future outcomes such as high school completion, college completion, income, and job retention than achievement tests alone.

The ability of grades to serve as useful predictors of other important outcomes appears to be leading to a shift in how grades are viewed by assessment and grading researchers (Brookhart et al., 2016). As an example, Brookhart (1993) reported that teachers do not follow recommended grading practices. Brookhart created an intervention to change teachers’ grading practices so they would be more closely aligned with recommended grading practices. She instructed teachers in educational measurement and found this training did not shift teachers’ thinking surrounding grades. Even after training, teachers continued to include non-achievement
constructs such as effort and remained more concerned with fairness and compassion than with validity and reliability. Data were analyzed using Messick’s (1989) validity framework—a framework rooted in a psychometric/measurement tradition. While Brookhart was sympathetic to teachers in the research, her foundations were rooted in the traditions of educational measurement and psychometrics and, as a result, she adopted a critical stance towards teachers because their classroom assessment and grading practices did not meet accepted quality standards developed by measurement specialists. More recently, Brookhart (2015) has suggested that teacher awarded grades, or what she terms “graded achievement” should be considered as a construct that is distinct from “tested achievement.” In this recent paper, Brookhart recognized that while teacher awarded grades are not pure measures of subject area knowledge, they have the potential to communicate important information about the student. “Graded Achievement” represents a complex, ill-defined construct that includes subject area achievement and other constructs such as citizenship, motivation, effort, and learning skills. Brookhart (2015) no longer used measurement theory or psychometrics to examine “grade achievement,” rather she adopted a historical perspective and noted that “recent research is more sanguine about the meaning of graded achievement” (p. 292).

While current grading research may be more sympathetic towards classroom teachers, this does not solve the very real problem of having a measure (teachers’ grades) that predicts future outcomes (e.g., school completion and college success) but not knowing what the measure represents. We know it is a mix of achievement and LSWH, but we do not know what LSWH are included, how they are assessed, or how they are incorporated into a student’s grade. One proposed solution to this problem is to ask teachers to grade and report upon LSWH separately.
from academic achievement (Brookhart, 2004; Guskey, 2006). This is a recommendation that has been adopted in many educational jurisdictions—including Ontario.

On the surface, adding a section to the report card allowing teachers to communicate student performance in a variety of categories seems an obvious solution to the problem of report card grades representing an inconsistent and unknown construct. However, this solution rests upon three untested assumptions. The first is that if teachers are given the opportunity to factor out LSWH from achievement, they will. A teacher may be offered a separate space on the report card to provide a grade for “organization” or “collaboration” that is independent from achievement, but this does not mean the teacher sees these LSWH as independent from achievement. It is still possible for a teacher to allow a student’s LSWH to impact the achievement grade they award to a student, just as it is possible for a teacher to consider achievement when awarding the LSWH grades. While there is much evidence that teachers include LSWH in their grading of academic achievement (e.g. Sun & Cheng, 2014), there is no research that investigates whether teachers include academic achievement when assigning LSWH grades. The assumption that teachers can, and will, separate achievement and LSWH when making grading decisions appears to be untested.

The second assumption is that teachers can devise valid and reliable assessments of LSWH. This assumption also appears to be untested, and it is very reasonable to question its veracity. The difficulties in assessing self-regulation were described earlier, but this is not the only LSWH that is difficult to assess. Effort, for example, is notoriously difficult to measure (Linn & Miller, 2005). Assessing participation appears to be just as problematic as some researchers have argued that teachers have “no concrete means of measuring students’ participation” (Miller, Klinger & Shulha, 2006, p. 2). This suggests that the assessment of
participation may be open to a broad range of practices. Kelly (2008) found that teachers awarded students with higher academic grades if they were engaged in the classroom, but only if they were “substantively” engaged. According to this study, a student who is “substantively” engaged completes all assignments and asks questions that are genuine and relevant (as opposed to procedural questions). These findings indicate that teachers value some forms of participation more than others, leading to the conclusion that the manner in which teachers conceptualize, assess, and grade participation is complex.

This is not to deny that some LSWH are easier to measure accurately and reliably (e.g., attendance, punctuality, completion rate of assignments), but many LSWH or components of LSWH are internal to the student (e.g., metacognition) and therefore difficult to measure in a classroom setting (Duckworth & Yeager, 2015; Lai, 2011). While there are many examples of researchers assessing metacognition or other constructs, the reality is that classroom teachers receive minimal training in assessment and have low levels of assessment literacy (Cizek, 1996; DeLuca & Bellara, 2013). Furthermore, whatever assessment training teachers have received, most, if not all of it, has likely focused on assessment of achievement. It is therefore reasonable to question the assumption that teachers are able to create reliable assessments that provide meaningful, actionable information about student performance on constructs such as effort, metacognition, and organization.

Finally, there is the assumption that teachers, parents, and students share a common conception of LSWH. It is known that teachers themselves cannot agree on what activities should be considered when grading academic achievement (Allen, 2005); therefore, it seems unlikely there will be common agreement among teachers as to what activities will serve as useful measures of different LSWH. If a teacher, student, and parent all have different concepts
of what constitutes “self-regulation,” and different ideas about what separates high levels of self-regulation from low levels of self-regulation, then grading self-regulation separately from achievement will add little clarity to the report card for students or parents.

Clearly, the extent to which teachers separate achievement and learning skills and work habits (LSWH) and the benefits of such separation, have not been adequately explored. Further, if such grading practices are to provide greater clarity of what each grade is communicating, teachers must be able to distinguish across the constructs being graded. Given that constructs such as self-regulation have varying definitions within the literature, it seems likely that when it comes to LSWH, there will be wide variation in how teachers define each of the six LSWH to be assessed. Further, given the lack of preparation teachers receive to assess LSWH, there is a question as to what extent teachers are capable of assessing LSWH in a manner that is valid and reliable, and whether they separate academic achievement from LSWH, even when given the opportunity to do so. These are critical gaps in teacher grading research that need to be addressed.

The three studies that comprise this dissertation address these gaps in different ways. Because there is so little empirical data on how teachers assess and report upon non-achievement factors such as the LSWH, a foundational study was needed to examine teachers’ practice. This was the rationale for the first study. The use of semi-structured interviews allowed for the collection of a rich data set to inform how Ontario secondary teachers defined the six LSWH, how they assessed them, and how they made grading decisions. This study addresses several gaps in the research such as how teachers distinguish among achievement levels for the LSWH, and what activities they design to assess LSWH. The results of this study served to inform the development and analysis of the other two studies.
The first study has a small sample size (N = 26) and it is likely this sample does not represent the general teaching population. This limitation is addressed by the second study which uses large-scale report card data (N = 78,032) from two school districts. This study serves to confirm prior research associating academic achievement with skills like the LSWH, and gender differences in these skills, but also adds a new dimension to the research by illuminating teachers’ ability to distinguish among different LSWH and revealing to what extent LSWH grades are norm-referenced. The third study is the first research I am aware of that explicitly addresses how Ontario secondary teachers operationalize self-regulation and compares teachers’ concepts of self-regulation with extant research, and Ontario educational policy. Taken together, these three studies provide a detailed picture of how Ontario secondary teachers assess and grade LSWH.
Abstract:

Grading and assessment research have traditionally focused on how teachers assess and report upon academic achievement. However, teachers in many school systems must also assess and report upon additional aspects of student performance. As an example, teachers in Ontario, Canada are required to assess and report upon a set of six learning skills and work habits. Semi-structured interviews with 26 secondary teachers from three school districts in Ontario were conducted to investigate how teachers undertake these assessments and make grading decisions. Results indicated wide variations in practice. Some teachers reported making no formal assessments of learning skills and work habits and recording no assessment information. Thus, they made quick grading decisions as they completed report cards. At the other end of the spectrum, teachers reported co-creating definitions of each learning skill and work habit with their students, using rubrics to assess the skills, and keeping detailed records of assessment information using online tools. However, even these teachers tended to trust their holistic judgment more than the assessment information they collected. Further data reveal the complexity of these assessments, and the challenges teachers face in making them.
The aim of this qualitative study is to better understand how Ontario secondary teachers assess and grade Learning Skills and Work Habits. It has long been known that report card grades are not pure measures of achievement (e.g., Parsons, 1959). Instead, teachers’ grades are reflective of complex, idiosyncratic constructs that include factors such as effort, attendance, and behaviour (Brookhart, 1993; Cross & Frary, 1999; Russell & Austin, 2010). Some researchers believe that teacher awarded grades are not pure measures of achievement, but a holistic judgment as to how well a child is performing the role of “student” (e.g., Allen, 2005; Bowers, 2009). Thus, a grade is a teacher’s overall judgment of a student’s classroom performance based upon diverse, and sometimes latent, factors such as achievement, behaviour, effort, and improvement. Grading decisions may be further impacted by factors external to a student’s performance such as the teacher wanting to minimize conflict, motivate a student, and maintain course enrolment (Bowers 2011; Brookhart 2004; Cizek, Fitzgerald, & Rachor, 1995; Cross & Frary 1999; Hunter, Mayenga, & Gambell 2006; Russell & Austin, 2010).

In an attempt to bring clarity to what report card grades mean, many jurisdictions, including Ontario, require teachers to separate achievement and non-achievement constructs when reporting on student learning (Merchant, Klinger, & Love, in press; Ontario Ministry of Education, 2010). Achievement in the subject area is to be based solely upon scores students have received on identified, summative assessments. Other constructs such as organization, responsibility, collaboration, and self-regulation are reported separately on the report card. In Ontario, this section of the report card is called “Learning Skills and Work Habits” (LSWH).

Allowing teachers to report on both achievement and non-achievement elements of student performance seems an obvious solution to the problem of report card grades representing an inconsistent and unknown construct. However, this solution rests upon three untested
assumptions. The first is that if teachers are given the opportunity to factor out LSWH from achievement, they will. Teachers may have separate sections of the report card to provide grades for “organization” or “collaboration,” but this does not mean that the teacher sees these LSWH as independent from achievement. It is still possible for teachers to consider LSWH when assigning achievement grades, just as it is possible for teachers to consider achievement when assigning LSWH grades.

The second assumption is that teachers have the skills to devise valid and reliable assessments of LSWH. This assumption also appears to be untested. Some assessment experts have suggested that because defining and assessing complex constructs such as the LSWH is so difficult, teachers should not undertake such assessments (Stiggins, Frisbie, & Griswold, 1989). As an example, it is known that teachers find effort difficult to measure (Linn & Miller, 2005). Assessing participation appears to be equally difficult, as teachers have “no concrete means of measuring students’ participation” (Miller, Klinger, & Shulha, 2006, p. 2). There is further evidence that how teachers assess and grade non-achievement constructs is inconsistent. Kelly (2008) found that teachers awarded students with higher academic grades if they were engaged in the classroom, but only if they were “substantively” engaged. According to this study, a student who is “substantively” engaged completes all assignments and asks questions that are genuine and relevant (as opposed to procedural questions). These findings indicate that teachers value some forms of participation more than others, leading to the conclusion that how teachers conceptualize, assess and grade participation is complex.

Clearly there are some non-achievement constructs that are easy to measure accurately and reliably (e.g., attendance, punctuality, completion rate of assignments), but many of these constructs, or components of them, are internal to the student (e.g., metacognition) and difficult
to measure in a classroom setting (Lai, 2011). It is known classroom teachers receive minimal training in assessment and have low levels of assessment literacy (Cizek, 1996; DeLuca & Bellara, 2013). Further, what assessment training teachers have received has likely focused on assessing subject area achievement. Given this, it is reasonable to question whether teachers can create high quality LSWH assessments that yield useful, actionable information about student performance on these skills.

Lastly, there is the assumption that teachers, parents, and students share a common conception of LSWH. It is known that teachers themselves cannot agree on what activities should be considered when grading academic achievement (Allen, 2005). Therefore, it seems unlikely the activities and behaviours different teachers will use as indicators of LSWH will be consistent. What is more likely is that teacher, student, and parent will all have different concepts of self-regulation (if parents and students even know what self-regulation is) and different ideas about what separates high levels of self-regulation from low levels of self-regulation. If different stakeholders do not share common understandings of what the LSWH are, and what distinguishes different levels of performance, then reporting LSWH grades separately from achievement is not likely to give additional, meaningful information to the student or parent.

Prior Research on Teacher Grading

What do report card grades tell us? There is no single answer to this question, as research into teachers’ grading practices shows that what information influences teachers’ grading decisions depends upon the subject, grade level, individual teacher, and individual student (Bowers, 2011; Brown 2011; Howley, Kusimo, & Parrott, 2000; Lekholm & Cliffordson 2008). In a Canadian context, Duncan and Noonan (2007) reported that mathematics and science teachers included LSWH (which they term “academic enabling behaviours”) to a lesser extent
than teachers of English, social studies, and performing arts. As an extreme example of this, Russell and Austin (2010) found that attendance accounted for about 65% of students’ secondary music grades. However, most studies on teacher grading have shown that achievement is the dominant factor in determining students’ grades (Kelly, 2008; Randall & Engelhard, 2009; Thorsen & Cliffordson, 2012). Like Duncan and Noonan (2007), Frary, Cross and Weber (1993) found that mathematics teachers incorporate achievement measures to the highest degree in their grading. Beyond subject area, we also know of other factors that impact grading decisions such as how close a student is to a grade boundary (Randall & Engelhard 2010), and whether the student is high achieving or low achieving (McMillan, 2001). While there is evidence that teachers try to communicate their grading criteria clearly (Brookhart 1994), and teachers generally follow their stated criteria, it also appears that, in the case of atypical students, teachers’ grading practices become highly inconsistent (Canal, Bonini, Micciolo, & Tentori, 2012). Encouragingly, it appears that teachers want the best for their students and use their grading to help achieve this (Nash & McMillan, 2000). However, it appears this approach also leads to grading practices that are complex, idiosyncratic, and at odds with practices recommended by experts in educational assessment (Cox, 2011).

Of particular relevance to this research is a qualitative study conducted by Nash and McMillan (2000). They interviewed 24 teachers (23 of whom were secondary teachers) about their grading practices. As is typical in the grading literature, the focus was on how teachers assign achievement grades. Nash and McMillan identified six themes in their interviews, only some of which are relevant to grading LSWH in Ontario secondary schools. The findings from this study provide some indication that decision-making processes surrounding LSWH grading may be distinct from grading achievement. For example, Nash and McMillan examined the
impact of standardized testing on teachers’ grading practice. However, no standardized tests of LSWH exist in Ontario, and so Ontario secondary teachers do not face this pressure. Indeed, even other pressures such as parents and students wanting high grades to allow entry into post-secondary programs are absent from the LSWH grading context. Nash and McMillan wrote about teachers “pulling for students” (p. 7) and using their grades to ensure that students pass courses, but LSWH grades have no impact on students’ pass/fail status in Ontario. The extremely low stakes attached to LSWH grades makes them a unique context.

There is a large volume of research focused on assessing constructs such as self-regulation, but very little of that research is connected to classroom assessment. The few studies that have been completed report that teachers struggle with assessing constructs beyond academic achievement. For instance, Miller, Klinger, and Shulha’s (2006) survey of Ontario mathematics teachers found that teachers conflated achievement and LSWH in their grading decisions, despite clear policy directions to report on them separately. These teachers also reported that they did not assess LSWH as frequently as they should, and that LSWH were difficult to assess. Further, parents and students did not appear to value the LSWH grades as they rarely asked about them. Nash and McMillan (2000) directly examined non-achievement factors in their interviews with teachers. The authors found that how teachers recognize and incorporate student effort into their grading was one of the most variable aspects of teachers’ grading practices.

As part of his doctoral dissertation, Ferrito (2015) examined the “Personal and Social Development” (PSD) grades on grade 4 report cards from a school in New Jersey. There were seven separate PSD grades, and all were reported using a 3-point scale. Exploratory factor analysis found the PSD grades to be unidimensional. This indicates either that teachers were not
assessing the seven PSD constructs separately, or the seven constructs were highly related. For example, three of the PSD constructs (uses keep calm, uses listening position, uses speaker power) related to instruction students had received on how to behave properly in the classroom. Another construct was “follows classroom directions and rules.” It is easy to see how a student who follows classroom directions and rules would also use strategies and instructions taught about how to behave in the classroom. Another possibility is that teachers use holistic impressions of the student to inform grading decisions of all seven PSD constructs. Unfortunately, it is not possible to determine why teachers’ grades of the PSD grades were unidimensional from his data.

In summary, prior grading research has focused on academic achievement, and this research has shown that teacher grades tend to be complex, idiosyncratic, and related to their philosophy of education (Nash & McMillan, 2000). What little research exists on grading LSWH indicates that teachers find these assessments difficult, and do not complete them as often, or in the same manner, as they complete their assessments of academic achievement.

The Ontario Assessment Context

The main policy document guiding assessment practice in Ontario schools is called Growing Success (Ontario Ministry of Education, 2010), and it outlines how teachers are expected to assess and report upon academic achievement and LSWH. The document states that achievement grades on secondary report cards are to reflect how well a student is meeting curricular expectations and should be expressed as a percentage. There are no curricular expectations, or grade level standards given for the LSWH. Instead, teachers are expected to use their professional judgment about what evidence to consider when making LSWH grading decisions. LSWH grades are reported on a 4-point scale that uses the terms: Needs Improvement,
Satisfactory, Good, and Excellent. Report cards are distributed after the end of the course; however, most districts in Ontario (including the ones from which the teachers in this study were sampled) send out an interim report card halfway through the course. This report card contains LSWH grades and comments, but no achievement grades. Thus parents and students receive information about LSWH performance midway through the course.

Curriculum documents offer teachers little additional guidance on how to assess LSWH. As an example, the Ontario Grade 9/10 science curriculum guide (Ontario Ministry of Education, 2008) contains seven pages of assessment policy and guidance for teachers. To help teachers understand how to assess subject area achievement, the curriculum guide provides teachers with assessment criteria, an achievement chart, descriptors of the different achievement levels, and prescriptive weightings of different assessment tasks. In contrast, a single paragraph addresses assessing the LSWH. This is a generic paragraph that appears in all the secondary curriculum guides. Most of the paragraph is description, with only one sentence offering direction to teachers, advising that LSWH “should not be considered in the determination of percentage grades” (Ontario Ministry of Education, 2008, p. 28).

Description of the Study

The purpose of this qualitative study was to develop a preliminary understanding of how Ontario secondary teachers define, assess, and report upon the six LSWH that are included on the provincial report card. The specific research questions were:

1. What processes do teachers report they are using to determine the LSWH grades?
2. How do teachers define the LSWH?
3. How do teachers distinguish among levels of performance on the LSWH?
The research employed semi-structured interviews for data collection. An interview protocol was developed and then refined based upon feedback from an expert in qualitative methodology (Appendix A). A total of 26 secondary teachers were interviewed in person. Twenty-five of the teachers taught in public schools within three school districts and one teacher taught in an independent school. After receiving ethical clearance from the university’s review board and the participating school districts, recruitment emails were sent to secondary principals, who then forwarded the notice to teachers. Nineteen of the teachers were recruited in this manner. The six remaining teachers working in public schools were recruited using snowball methods, in which a participant referred me to other teachers who might be interested in participating. Three of these teachers played in the same ice hockey league. Teachers in the sample taught a variety of subject areas including English, mathematics, science, technological education, music, and geography. Fifteen teachers were males and eleven were female. The number of years of experience reported by these teachers ranged from 2 to 35. The schools they taught in included rural, suburban, and urban secondary schools in Ontario.

Interviews were audio recorded and transcribed verbatim. The original intent had been to have teachers use their gradebook as an elicitation tool, but this proved problematic for several reasons. Firstly, some teachers did not keep written records to inform the awarding of LSWH grades, and so the gradebooks contained no relevant information. Secondly, many of the teachers who did record LSWH assessment data used online tools that were not accessible during the interview. As a result, 21 of the participants relied solely on memory to answer the interview questions. Transcripts were coded and analysed to find commonalities and variability in how these teachers reported that they defined and graded the six LSWH. The initial codes were based on the research questions, but as other important ideas emerged in the data, they were
recognized, and appropriate codes were developed and incorporated into the analysis. In total, 30 codes were used, including those based on the research questions and those emerging during the analysis. Coding took place in two stages. Firstly, entire transcripts were coded and relevant passages were placed into a spreadsheet next to the appropriate code. Transcripts were deliberately coded in a different order from the order the interviews took place. This was done to minimize threats to validity such as recency bias and salience of first impressions (Huberman & Miles, 1998). Secondly, the transcripts were re-analysed by selecting a code and re-rereading all of the transcripts to find sections relevant to that code. Adopting these two different means of coding was done to ensure that all relevant passages were identified for each code. After all transcripts had been coded, the method of constant comparison was used to combine codes to develop themes and write analytic memos for each research question (Birks, Chapman & Francis, 2008; Boeije, 2002; Center for Evaluation and Research, 2012).

To help me find commonalities within the data, I engaged in a process inspired by customer journey mapping (Temkin, 2010). Customer journey mapping is a technique used by businesses where they use customer data to create different customer archetypes to better understand how customers interact with their business. Each archetype represents a different type of customer. In the case of this research, the process of customer journey mapping was used to classify teachers into different categories of LSWH assessment practice. Classification into a group was based upon the amount of documentation the teacher collected to inform their LSWH grades. Group A includes teachers who collected no documentation, while group B contains teachers who collected some documentation. “Some” documentation is defined as any combination of lates, absences, missing or late assignments, and student self-assessments. If teachers recorded more information than this (e.g., recorded observational notes of student
behaviour), they were considered to have recorded substantial documentation and were put into group C. Once teachers were categorized, other data (e.g., demographic data) were used to create a composite picture of a “typical” teacher in each of the three categories.

Of the 26 teachers interviewed, eight were placed into group A, six into group B, and twelve into group C. There were differences in the demographics of the teachers in each group, although with such a small sample it is not possible to know if these differences are an artefact of the sample, or representative of a more general population. Of the eight teachers in group A, seven were male, and six had more than 15 years experience as teachers. Two teachers had three or fewer years of experience and a range of subject areas were represented. Four of the six Group B teachers were male, and again a range of subject areas were represented. One female and one male from this group had between five and ten years teaching experience, the rest had more than 15 years. Group C teachers had an even gender split and a range of subject areas. One person was in her second year as a teacher, and the rest had between five and 15 years experience. There were no teachers in this group with more than 15 years experience.

Amount of documentation was chosen as the distinguishing factor between the three groups because this aspect of LSWH practice contrasted starkly with the assessment of academic achievement. All teachers in the study reported that for academic achievement they adhere to district policies including: using their course outlines to communicate assessment policies, stating which assessments are summative at the beginning of the course, providing students with marking rubrics or schemes for summative assessments, and documenting achievement on summative assessments using gradebooks or software. Given the professional and legal expectations that teachers document students’ achievement on assessments of academic achievement, the finding that teachers’ documentation processes for assessments of LSWH were
variable provided the rationale for using amount of documentation as the criterion for
categorizing teachers into groups.

Findings

The findings are first reported using a narrative to create a portrait of a teacher that
represents each group. Portraits reflect a composite of the typical characteristics of the teachers
placed into that category, and not an actual person. To create a composite, the most common
characteristics of teachers within that category were amalgamated into a single character. Each
portrait is designed to give the reader an impression of how a teacher in each of the three
different groups might approach their practice. After the portraits, results are reported by
research question, and separated by group. This approach provides a straightforward way to
describe commonalities and differences among the groups with respect to the research questions.

Portrait A

Les is an English teacher with 27 years’ experience in the classroom. Over his career he
has taught English, Drama, History, Citizenship, Careers, and even took over a special education
classroom for a semester. Les takes pride in his teaching and is confident about his abilities in the
classroom. He has a reputation as a good teacher and the students generally like him. Over time,
he has become less concerned with adhering to school policy, especially policies he disagrees
with, but his general sense of professionalism means that his assessment and grading practices
are in line with the expectations of the school, the students, and the parents. Les’ course outline
clearly outlines which assignments are summative, and will count towards the final achievement
grade, and the grading rubric for each summative assessment is shared with the students. At
report card time, Les tallies up the students’ scores on their summative assessment, and


determines a final achievement grade, weighting each assignment as described on the course outline.

After entering a student’s academic grade, Les enters the LSWH grades. He does this from memory and a general impression of the student. While Les has recorded achievement on the summative assessments, there are no records or products informing his LSWH grades. This does not concern Les for two reasons. Firstly, he knows his students well, and believes he can make accurate judgments about their LSWH based upon his experiences with them in class. Secondly, Les knows that students, parents, and administrators never question his LSWH grades. Because these grades are not used by universities or colleges for admissions purposes, there are no consequences attached to the grades, and therefore students and parents do not care about them. Why would a teacher agonize over grades nobody uses or cares about? Thus, Les completes his LSWH grades very quickly, and his students tend to get the same grade for all six LSWH.

This is not to say that Les does not value the skills themselves. Organization, responsibility, and initiative are all important qualities for students to have—both in and out of school. However, when there are problems with students not submitting work or arriving to class late, he speaks to the student immediately and does not wait until report card time to deal with the issue. Les remembers his early teaching career when he could include things like effort, participation, and attendance into his achievement marks and misses those days. In his mind, the separation of LSWH from the achievement mark has contributed to students feeling free to submit work late, or to misbehave. The marks associated with good behaviour and participation were a great way of encouraging students to stay on task and work productively in the classroom.
Portrait B

Wanda is beginning to feel established in her school. A mid-career teacher, she has taught many subjects as a substitute teacher, but is now working as a humanities teacher. In her third year at the same school, Wanda is developing a reputation as a teacher who cares about students and is willing to coach or sponsor clubs. She pays close attention to school policies and procedures around assessment and reporting and ensures that her course outline includes all the information required. This means that not only are her summative assessments and marking rubrics given to the students at the beginning of the year, but also that she mentioned to the students that LSWH are part of the assessment and reporting process. She tells the students that LSWH are important not just at school but are vital life skills. Wanda does not have grading rubrics for the LSWH and is unclear on the distinctions between the performance levels. She uses her observations and judgment to decide when a student merits “excellent” or “good” on the report card. Wanda records some data about the students’ LSWH such as which assignments were submitted late, and if students are absent or late to class.

When completing the report cards, Wanda cannot remember what the six individual LSWH are, so she has the district policy on her desk to remind her what behaviours serve as indicators of each skill. These documents help her better understand how to distinguish between the skills, but because she has limited records of students’ LSWH, she tends to grade the students using her memory of their performance and behaviour in class. Wanda does not use student-self assessment, but she knows teachers who do, and thinks it is a good idea. Likewise, Wanda does not use rubrics to grade the LSWH, but sees how that could be valuable. However, she is quite busy, and creating LSWH rubrics is not high on her list of priorities.
Students, parents and administrators do not ask about the LSWH grades she gives, and this reduces her motivation to be more purposeful about her LSWH grading practices. In her mind she thinks this is a shame because she cares deeply about the LSWH grades her own children receive. In fact, she has noticed that parents who happen to be teachers are the ones who care most about the LSWH grades. She can appreciate why this is, as everyday she sees the connection between effort and achievement.

Portrait C

Viola teaches mathematics. She is a mid-career teacher, but so many of her colleagues have decades of experience that at times she feels like a beginner. Viola became a teacher because she wanted to inspire kids to reach their dreams, and for her the LSWH are a critical component of this. Not every kid wants to study mathematics, or even cares about the subject, and Viola recognizes this. However, no matter what you want to accomplish in life, organization, responsibility, initiative, and collaboration are useful. Thus, for Viola teaching is about developing these skills through mathematics. Long after her students have forgotten how to factor polynomials, they will remember the perseverance and discipline it takes to tackle difficult problems.

Viola’s classroom walls have inspirational posters, along with a large poster from the district that outlines what the LSWH are. Further, Viola has created her own mini-posters that describe behaviours associated with each LSWH. Each poster has a stem followed by bullet points. The poster for responsibility looks like:
During the semester Viola makes observations of the students’ LSWH and records them in a file on her computer. Viola uses group work to assess collaboration, but she has also developed a sense of which tasks are useful for assessing different LSWH. She uses seat work to assess independent work, and she has a research project she feels is a good opportunity to assess initiative. These tasks were designed to assess elements of the mathematics curriculum, but she also uses them to assess LSWH. Viola has also created rubrics to grade the LSWH. These rubrics are used on the summative assessments and are added to the bottom of the achievement rubric for that assessment. Supplementing her own observations are self-assessments the students complete two weeks before report cards are due. She compares the self-assessments to her own observations of the students and uses them as a gauge of her students’ ability to reflect on their learning.

At report card time, Viola considers all the data she has collected when making grading decisions, but there is no set method or weighting to the data. Rather, the data are considered in light of her experiences with the student and her holistic impressions. The grade she gives is the one she feels best reflects the data she has collected and her own general impression of the

---

**Responsible students:**

- Come to class on time
- Submit work on or before the due date
- Notify the teacher in advance if they will be absent or submit work late
- Make up missed work
student. Despite the effort Viola has put into her LSWH assessment practice, she still struggles to remember the definitions of all six LSWH, and so she uses the posters in her classroom to remind her what each LSWH represents as she is determining the grades.

The time and care Viola puts in to her LSWH assessments add substantially to her workload. This is a concern for her, but this concern is overridden by her desire to have a defensible rationale for her grades. Students, parents, and administrators demand the achievement grades are based on evidence, so why should the LSWH grades be any different? Viola has noticed that in the last few years parents and students seem to be more concerned about the LSWH grades, and, because these skills are critical to doing well in jobs or university, it is important to her that she be thoughtful in her assessment approach.

The teachers mentioned in each of these three portraits may not be real, but they provide an indication of the types of LSWH assessment and grading practices that Ontario secondary teachers engage in and allow the reader to imagine how these practices play out. To address the research questions more directly, general themes related to each question are presented below. Supporting quotations from teachers within each portrait group are offered as evidence.

General themes surrounding the research questions

RQ1. What processes are teachers using to determine the LSWH grades?

“So number one for me is just observation—what I see day to day.”

Group A: All eight teachers in group A reported using observation as their primary assessment tool for LSWH. This was not unique to this group, as all teachers in the sample used observation as their primary assessment tool. However, for members of group A it was their only assessment tool, as they did not have written records to rely on for grading decisions. Teachers in this group
typically justified the lack of documentation by stating that because of their frequent interaction with the students, they know their students well. As a business studies teacher with over 30 years of experience stated, “we see them every day for four plus months, four and a half months kind of thing. And you get a really good impression from that what they are, where they are at in terms of what the assessment is going to be.” An English teacher spoke about his experience allowing him to get a quick sense of which students will have weak LSWH, saying, “having taught three and a half decades, you know early on the cases that might be problematic.” A special education teacher who also had decades of experience described grading LSWH as, “more of a qualitative thing, it’s not a quantitative. And I don’t think many teachers in the profession think of it quantitatively. When we look at them, it’s obviously ‘Needs Improvement,’ ‘Satisfactory,’ ‘Good,’ or ‘Excellent’.”

The other argument teachers in group A used to support their practice of not collecting documentation was their belief that the LSWH grades were not of high importance. One of the two early career teachers in group A (there were no mid-career teachers in group A) said that her grading procedure was “I just go through as I’m doing the report card and click on the drop down” and that she adopted this quick, ad hoc process because a LSWH grade “doesn’t really matter.” An experienced English teacher described his grading process as, “Quick thought, serious student, tries hard. That’s all.” The short time spent grading LSWH was because the teacher puts “very little value” in the LSWH grades and finds them “not interesting.” Another experienced English teacher stated he used to record whether assignments were submitted late but had stopped the practice “because we are not supposed to take marks off for late assignments.” This same teacher described the LSWH as “all kind of hoo-haw,” and further
extrapolated that, “if you were to ask an administrator about these things. In their heart of hearts they might say some of it is hooey.”

Seven of the eight teachers in group A viewed academic achievement and performance on the LSWH as related in some way, but two of the teachers formalized the association by using the achievement grade to inform or even determine the LSWH grades. One of the experienced English teachers in group A stated, “if they’re in the 80s, it’s probably ‘Excellent’ across the board, if they are in the 70s it’s good, like the ‘G’ [Good], 60s satisfactory.” He later described his grading process as arising “from memory, impressions and looking at their marks.” A different teacher made a parallel statement,

I just think the synchronicity between the learning skills and the grade, I think they play out. You don’t see them as separate entities. If the student has got ‘Good’ all the way through, their mark is probably between 68 and 75 or so, if they got ‘Excellent’ all the way down through they are probably anywhere, somewhere 80 to 90 something as a mark. So, there is congruency that way.

Teachers in group A believed LSWH grades should reflect only events and behaviours from inside the classroom. This belief was shared by teachers in all three groups. Thus, aspects of student life such as participation in clubs or athletics, should not be counted towards the LSWH grades. This belief was not unanimous, but it was a strong majority of teachers who believed that LSWH grades should reflect only classroom behaviours. One of the three teachers who believed it was justifiable to include extra curricular work came from Group A. She was a music teacher, and, like a music teacher in group B, included participation in extracurricular bands or ensembles when assigning LSWH grades. There was also a teacher in group A who felt it was reasonable to include students’ personal or family circumstances. This teacher taught in an
alternative school and had many students who were older, employed, or came from difficult family circumstances. He felt that students who came from poor families or needed to supervise younger siblings should have those circumstances taken into consideration when assigning LSWH grades. The remaining five teachers in group A believed that including external events was poor LSWH assessment practice, although two recognized that events or interactions outside the classroom may colour their overall impression of the student.

**Group B:** Like group A, teachers in group B did not have specific activities or assessment tasks for the LSWH. Instead, the teachers in group B used observation as their primary assessment tool, but observation was sometimes supplemented with other data such as how many assignments were submitted late, or student self-assessments. This gave teachers in group B some assessment data to inform their grading decisions, but like group A, they used holistic judgment to arrive at the final LSWH grades. A mid-career civics teacher who described his LSWH data collection as “the odd homework check” described his grading process as,

> I have a mental picture of the person and I look at the criteria, and then I decide based on the criteria, what I’ve seen of that person, and it’s completely anecdotal, most of the time, except when it comes to homework checks and the amount of work that they have submitted.

A mid-career business teacher in group B summarized her assessment practice in three words, “Observation, conversation, continuous.” She then went on to clarify that assessing LSWH was about “having a conversation with them, walking around, sitting beside them, chatting with them.” She also asked, “If that student hasn’t given me a product, those conversations, and those observations, then how I am able to give them a mark?” This teacher recorded observations in a notebook, but noted that her records were inconsistent by declaring, “I am not good enough. You
can see the holistic part of me, where I won’t always do it.” When I asked a geography teacher if he documented his observations of students’ LSWH, he responded, “some stuff, and I am getting better at it, but definitely a lot of it is just I know the kid.”

Four of the six members of group B had used student self-assessment as part of their LSWH assessment process. Self-assessments were typically given shortly before report cards were due. Even though these teachers generally saw their students as “being pretty honest with their learning skills,” the self-assessments did not play a significant role in determining the LSWH grades. A mid-career mathematics teacher described the self-assessment process as, “just before midterm report… I had them do a self-evaluation and then I handed it back to them based on my feedback—where I thought they were.” The business studies teacher in group B described a similar process, saying,

Right before report cards, we’ve done it verbally more often, but right before report cards, ‘What would you give yourself?’ and an explanation. So, why are you here, and where do you want to be? You gave yourself an S, where do you want to be ultimately? And what will you do to get there? And why did you give yourself satisfactory? And then I also put mine on the same page, and then they get that sheet back.

A geography teacher described a similar use of self-assessment,

I kind of get them to give an assessment and see how they feel about it, and then I have a column for my own comments, so that I can take a look at what they have assessed themselves, and then I will tweak it based on how realistic their feedback is, and what I have actually seen as well.
Thus, for teachers in group B student self-assessment served not to inform grading decisions, but as a means of engaging the students in metacognitive thinking, and as a way of creating a dialogue about their LSWH.

Like group A, the teachers in group B believe the LSWH grades should reflect only behaviours or events that occurred within the classroom, although one music teacher was an exception. This music teacher also tied LSWH performance to achievement in the subject area. For her, practice was a vital component of the subject. Thus, students who practiced, received high LSWH grades, but also demonstrated superior performance skills, thus the achievement grade and LSWH grades were tightly linked.

**Group C:** The remaining twelve participants were placed into group C because they kept substantial documentation. They described this documentation as coming in the form of recorded observational notes, binder checks, rubrics, quantitative data such as number of lates to class, late assignments, and student self-assessments. All teachers in group C who kept observational notes stored them using online tools such as Google Classroom. Most of the teachers in group C taught in a single district, where LSWH had been a district focus for professional learning for the previous three years. Further, four of these teachers were involved in a collaborative inquiry project centered on how to assess LSWH. One of these teachers recounted:

> So, what I tried to do last year, and it worked, but I still need to do it deeper, like with deeper meaning, is make a chart with all their names on it, with big boxes beside it, where I can record if they have done their homework, if they have been to class on time every day, if they contribute to discussions, and just make little notes in it.
Group C teachers documented students’ LSWH because they saw it as part of their professional responsibilities, and to communicate to their students that LSWH are an important part of the assessment and grading process. A science teacher summarized this as

Right now, I can get away with just “Yeah, Gs, Es, Ss.” I can just wing it as I am filling [it] out in 30 seconds. I feel I owe the kids more than that, I feel my employer expects more than that. If it is going on the legal document there has got to be more to it.

Three of the teachers who kept written documentation about students’ LSWH recognized they tended to record negative events, and that this could skew the LSWH grades (This observation was also made by one teacher in group B). A computer science teacher noted that for his strong LSWH students, he often had very little data collected, and an English teacher reported that, “it’s not normally the good things they do that get written down.” These teachers reported recording events such as missed or late assignments, misbehaviour, and conflicts. The recording of negative events resulted from teachers paying greater attention to student behaviours that deviated from the expectations of the classroom, because “those are the types of things that stick out.”

Interestingly, teachers in group C reported that they too relied upon holistic judgment to determine the LSWH grades. An early career mathematics teacher who collected large amounts of LSWH assessment data admitted she used her “general knowledge of the student to support what I am assessing” instead of relying solely on her recorded data. An English teacher who was part of a collaborative inquiry group aimed at improving members’ LSWH assessment practices offered that while she collected LSWH throughout the semester, “you go through a process to come up with that final grade, it is not the computer spitting a number out at you.”
Some teachers in group C described trying to use an algorithmic approach to assign LSWH grades, but when the algorithmic grade conflicted with their holistic judgment, the holistic judgment was given preference. One such teacher reflected on this by saying that after she had calculated LSWH grades from her data, she would “sometimes go and change them, or even when I am typing them in—Nah, that’s not a ‘Good’ and then change it from that.” Another teacher had tried using an algorithmic approach, but then reverted to a holistic approach because “I just sort of realized my observations were far better.” Likewise, a mid-career English teacher talking about LSWH assessment data said, “Sometimes that gets pushed to the side, often I rely on my mind.” A chemistry teacher found that when he struggled to decide between two grades, that he used “what I remember in the class,” and then went on to acknowledge, “which isn’t the best, you shouldn’t just go off of memory.”

The only teacher interviewed who did not rely on holistic judgment to assign LSWH grades was team teaching a course, and needed a clear, transparent means of recording LSWH that could be used by both teachers. As a result, every student received a numerical grade on every LSWH on a weekly basis. These numbers were averaged to get a final LSWH grade to put on the report card.

All group C teachers, except one, incorporated student self-assessment into their practice. Like the teachers in group B, self-assessments were not used for grading decisions, but rather as a means of engaging students in metacognition and dialogue. An English teacher described self-assessment as “part of their learning cycle,” while a French teacher characterized it as “largely a reflective thing leading up to report cards.” An ESL teacher made a similar comment, saying the students “self-evaluate for all of the curriculum expectations we were looking for, and then they self-reflect about their learning strategies as well.” This teacher was the only one who felt the
students were usually not accurate in their self-assessments. A special education teacher described students as being “brazenly honest” in their self-assessments, and a male English teacher thought that in self-assessments, students “tend to be tougher on themselves than we are” but valued self-assessment because the students “tend to know more about something like organization.” One important comment about student self-assessment came from a mathematics teacher who worried that students, especially younger ones, do not understand the terms. She noted that “sometimes at the beginning of the semester they don’t even really know, especially in grade 9, they might not even know what self-regulation is.”

While there were clear differences in the groups, especially with regards to the amount of documentation collected, the use of student self-assessment, and how much these assessments are valued, there were also some commonalities across the groups. All 26 Ontario secondary teachers described assessment processes rooted in observation, and allowed holistic judgments of the students to influence, or determine, their grading decisions. Although teachers varied in the extent to which they documented their observations, there was consistency in their accounts of the importance of knowing the student. This was regardless of the academic subject they taught or their years of experience.

RQ2. How are teachers defining the LSWH?

“Some of them have crossover and so I decide—OK, this is going to be this, and this is going to be this.”

**Group A:** A theme common to all groups was difficulty in defining the six LSWH as separate constructs. The notion that the LSWH represent a unitary construct was shared by seven of the eight teachers in group A. One experienced English teacher viewed the six LSWH as a single construct representing compliance. Thus, his students were awarded the same grade across
all six LSWH “95% of the time.” A physical education/English teacher also saw compliance as an overarching construct for the six LSWH. A different English teacher felt his lack of clarity on what the LSWH represent led to his use of academic achievement to inform his LSWH grades. He also felt that his colleagues adopted a similar approach for similar reasons,

With these different categories, there’s some overlap I think in the minds of, at least there is in mine. And I think there is at least in the minds of the colleagues that I work with. And that’s one of the reasons I think that a lot of people, when they come to reporting time just go ‘OK, the kid has got an 80 something so it is all E.’

A business studies teacher nearing retirement said that “in terms of their learning skills, it’s not a specific. That’s not the way that I look at it. I look at it overall”, and another experienced business studies teacher saw the six LSWH as representing a construct related to how much he would like to hire this student to work in a business. This construct was related to effort, or, as he put it, “nose to the grindstone.” A beginning music teacher in this group described self-regulation as the overarching construct, saying, “that one [self-regulation] overlaps with almost all of them because if a student can’t self-regulate most of the other things suffer.”

The only teacher in group A who saw the six LSWH as representing different constructs was a beginning history teacher. He categorized the six LSWH as four different constructs. Organization and responsibility were described as, “I think of the responsibility one as being, you know, do you hand things in on time, being…but if they are organized, then they are going to do that, so they are paired together.” He further saw independent work and collaboration as two independent constructs, saying “and then there is the independent and the collaborative, and it is sort of like ‘OK, there is yin-yang’ I can sort of conceptualize it that way in my own head.”
He equated initiative with academic risk-taking but struggled with self-regulation. As he put it, “I don’t understand self-regulation. Not really. So, in terms of assessing, I find that one a bit trickier, because I don’t know exactly, I don’t know what I am looking for.”

**Group B:** The teachers in group B also struggled to distinguish among all six LSWH, but there were different approaches to how they managed this. A mid-career geography and civics teacher saw the LSWH as representing job skills and was able to articulate distinctions among the LSWH clearly. He did this by narrowing the constructs for some of the LSWH. As an example, he framed initiative as participation, by saying “I am also looking for initiative in the classroom, trying to participate in class discussion. It is just seeing who is continually trying to participate, trying to get others to participate.” He framed organization as “How organized is their binder? Do they have all their materials that they have been given?” and responsibility as “Do you show up?” and “Do you use your time wisely?”

A mid-career mathematics teacher saw the six LSWH as a measure of self-regulation, stating “self-regulation to me is almost the cumulation of them. It’s the summative of the other five”, but he was also able to distinguish between some of the LSWH. Initiative, for example, was seen as “participating in class, asking for help” whereas independent work was assessed using homework checks. This teacher was the source of the quote that opens this section and had made conscious decisions about which specific behaviours he would associate with the different learning skills.

A late career technical education teacher had public, posted definitions of all the LSWH in his classroom, but still struggled to distinguish among them. He described the six LSWH as being “all so close,” and further noted that “even my criteria sometimes it is overlapping.” He went on to say that when you read his definitions, “you see it and you think ‘well that’s sort of
like what you said there,’ but it is kind of like what you said there because there is not a lot of difference.” The one LSWH which he felt clearly stood out was collaboration, which he conceptualized as sharing and helping others.

An early career business studies teacher saw the LSWH as employability skills, explaining that when she worked in the corporate sector, “those skills were on my performance review.” Like the technical education teacher above, this teacher had posters of the LSWH definitions in her classroom, but the definitions she offered during the interview were slightly different. So, where the poster described ‘Organization’ as “devising and following a plan,” she described ‘Organization’ as “you know what is due, and when it’s due.” The poster for ‘Responsibility’ included “completes and submits class work, homework, and assignments according to agreed-upon timelines” but she defined it as “thinking about the people in your class.” The definitions offered by her during interview appear to be how she operationalizes the definitions shown on the poster, which came from district policy.

One interesting outlier was a civics teacher with a strong social justice orientation. This teacher had created his own LSWH which he termed “Challenging conventional authority”. This was a construct he valued deeply in his subject area and felt was important to report. The reporting of this was done in class to students, and via an interim report card that is sent out to parents halfway through the semester. With respect to this teacher-created LSWH, he stated, “nobody has ever said anything to me about it. I am surprised. I thought at some point they’d come and say to me—That’s not a learning skill. You can’t assess that.”

**Group C:** Like teachers in groups A and B, teachers in group C, struggled to define the LSWH as distinct constructs. As one teacher put it, “trying to remember all the categories, this is something I struggle with.” This struggle was despite all group C teachers having shared, or co-
created, definitions of the LSWH with their students. The primary cause for this difficulty was that teachers saw the six LSWH as overlapping constructs. An ESL teacher’s difficulty in distinguishing among the LSWH arose from the fact that “they do overlap a lot, unless you have that chart in front of you that says—‘homework falls under here’.” This teacher also admitted she could not remember all six LSWH, and so she relied on online tools to remind her what the LSWH are, and to record her LSWH assessment data. A mathematics teacher with seven years of experience found herself frequently consulting the policy document to remind her of the six LSWH as she found it tricky to “remember what’s what.” Later in the interview she described how she used the report card comment bank to help her remember the LSWH because “there is overlap,” and the definitions of the skills are “kind of fluid.” The difficulty in remembering the LSWH persisted despite this teacher co-creating definitions of the LSWH with her students and posting those definitions in her classroom.

While teachers in group C struggled as much as teachers in the other groups to separate the six LSWH as distinct constructs, they had more tools to help them. These tools included definitions they had co-created with their students, policy documents, and LSWH grading rubrics they had created. Thus, when teachers in group C were asked to define the LSWH, the struggle was usually accompanied by evidence that they had thought extensively about these skills. When I asked an English teacher what distinguishes organization from responsibility, she responded

I really focus on those two [organization and responsibility] a lot. And one has, if you look at the descriptors, one seems to have more to do with independently completing daily work and homework. One seems to be more the longer-term goals, but also if they were working with a group, that they were completing
their portion. I really do try to separate the two, but they do seem to go hand in hand.

A geography teacher distinguished between organization and responsibility in this way,

Organization—I use the notebook checks for that one. Responsibility, you’ve got a football game, and you know you are gone last period, but you come and hand something in to me, if it is due, ahead of time. So, you are being responsible for the work that you are going to miss.

This teacher was not alone in defining responsibility as meeting the obligations of the classroom. When teachers were able to articulate what responsibility meant to them, meeting obligations or expectations was the most common response.

The one LSWH that most teachers felt they could easily separate from the others was collaboration. As one teacher put it, “they all sort of correlate except for collaboration.” An art teacher offered more detail,

Collaboration, that one is the most unique, I think. Everything else, like self-regulation, you could argue that it relates to initiative; and if you are regulating yourself you might take more initiative; if you are organized you are doing well with independent work; if you are responsible, you are organized; they all sort of correlate except for collaboration.

Further evidence of teachers seeing collaboration as being distinct from the other 5 LSWH was that collaboration was the only LSWH for which teachers had specific activities that served as assessments. These activities were always group tasks or projects. Most teachers agreed that collaboration was easy to assess because it was very observable. An English teacher
expressed this as “you have them working in specific activities where they are working with a partner or a group, and they have to do something together, and they have to demonstrate their part.” Another English teacher often had students collaborating online using Google Classroom, and the analytics provided made assessing collaboration “pretty obvious.” Teachers in group C were usually able to clearly explain what they looked for when assessing collaboration. A humanities teacher saw good collaboration as “If they can work in groups that I have put them in, rather than just with their friends. And also, the kids that include, or are inclusive.” Another saw it as “participating when doing group work and sharing ideas and taking advice.”

The clear definitions of collaboration held by teachers meant they were able to give specific feedback to students to help them improve their collaborative skills. A teacher who was leading a collaborative inquiry project on assessing LSWH talked about how the board’s recent focus on collaboration had given both him and the students a clear, shared definition of collaboration. This meant that his feedback was more direct, and more actionable. He said,

Now I can actually look at them and say “You know what? Your group has been sitting there talking, you and your friend have been sitting up against the wall, laughing about the weekend. Well, that is not good collaboration.” And much in the same way I can go to that person who is trying to micro-manage the whole group and say “Are you showing good collaboration? I know you think you are because this group needs a leader but look at the language you are using. And is that language alienating them?”

This is not to say that all teachers found collaboration easy to assess. A computer science teacher found collaboration difficult to assess because “we don’t necessarily have it in programming.” This teacher assessed collaboration using helping, but recognized this was problematic, because
students who were weak in the course, could not help others, and so their collaboration grade would suffer. The other group C teacher who found collaboration difficult to assess believed that observing collaboration was difficult because much of what happens in the group is not observed. As she put it,

I don’t know necessarily how well they worked with their partner. Unless they have told me, unless there has been some kind of issue that they have told me about, or I have been able to observe while going around the room.

While peer assessment could provide additional information to teachers for assessing collaboration, this was not a tactic used by many teachers. In fact, only one of the teachers interviewed incorporated peer assessments into their LSWH assessment practice.

From the interviews it was clear that teachers in all three groups struggled to define the six LSWH, although there appeared to be a common belief that these skills could be conceived as generic job or life skills. Much of the difficulty in providing precise definitions of the LSWH arose from teachers seeing the six LSWH as overlapping constructs. There was evidence that teachers found collaboration easier to assess than the other five LSWH as collaboration skills were easily observed during group work. Further, collaboration was the only LSWH teachers assessed using specific assignments instead of general observation.

RQ3. How are teachers distinguishing between levels of performance?

“*It is just not doing things on a consistent basis, which can lower them, that excellent to a good.*”

**Group A:** In Ontario, the LSWH are graded on a 4-point scale, with ‘Excellent’ being the top point, followed by ‘Good,’ ‘Satisfactory,’ and ‘Needs Improvement.’ There was little
consistency in how group A teachers distinguished among these levels of performance. Two of the teachers had a binary view of the grades, in which if students were not ‘Excellent’ then they needed improvement. One of these teachers was a late career business studies teacher, and the other was a beginning history teacher. The business studies teacher reported he would tell students “if you turn in one thing late, you get ‘Needs Improvement’.” The teacher felt this approach was justified because he very strongly viewed the LSWH as job skills and felt students should be held to the same standard they would be in the workplace. In contrast, the history teacher adopted this binary view of the LSWH grades because he didn’t have a clear concept of what the four different grade levels represent. As he put it, “‘Good’ and ‘Satisfactory’ don’t mean anything to [students], those words don’t really work. They don’t know what that means. Those would be ‘Needs Improvement.’” He further went on to explain that needing improvement does not mean the situation is dire,

They are not an E yet, so what exactly that gradient is between all 4 is unclear.

But I know that I can’t put down ‘Needs improvement’ because then I might get parents saying like “they are like failing, they have got a 1 on a 4-point scale – instead of a 4”. And my bosses would be like “Really? Are they that bad at cooperating or collaborating?” No, they just need improvement.

Thus, this teacher awarded ‘Good’ when students needed to improve their LSWH, and then ensured the report card contained comments addressing the improvements needed. Many of the teachers in this group used comments to address issues with students’ LSWH, although it should be mentioned that four teachers came from a district where it was policy that report cards must contain at least one comment addressing LSWH.
The teachers with the most precise method of distinguishing among levels of performance for the LSWH were the two teachers in this group who used achievement grades to determine the LSWH grades. The remaining teachers had varying concepts of what distinguished the different LSWH grade levels. For one experienced business studies teacher it was consistency. He described the top three grade levels as,

‘Excellent’ to me means you never have a question. As a teacher. Like, we never question it, it’s just always done and done well. ‘Good’ means it can be inconsistent. Yeah, they were, it’s up and down, and ‘Satisfactory’ means I can live with it, but it could be better.

For another late career teacher, consistency was also mentioned, but he also talked about the outcomes of LSWH. For this teacher, what separated ‘Good’ from ‘Excellent’ was the level of engagement that students had with the course, was that students with excellent LSWH had “a deep understanding instead of a superficial understanding” of the material.

With the exception of the two teachers who took a binary approach to the LSWH grades, group A teachers viewed ‘Needs Improvement’ as indicating a serious problem. One teacher said “I don’t get down to ‘Needs Improvement’ unless I am letting the parents know that it’s not really satisfactory,” and another teacher expressed it as, “if I have N [Needs Improvement] across all the learning things, then it is showing below level 1.”

**Group B:** Five of the six teachers in group B used consistency as the discriminator between LSWH grade levels. One teacher linked the use of consistency to the Ontario assessment policy that grades should be related to typical performance,

I think it’s probably described in the document, it’s “consistently” does these things. So, there is an element of that in it. It’s not a snapshot on a given day, it’s
over the span of the course that … your behaviour consistently is that. If you get excellent, it should be, according to the descriptors we’re given, it should be a regular occurrence. It’s not like a one-off. Today is not “learning skills test day and you have to show me what you can do,” it’s every day.

A mathematics teacher in the same district described what separates ‘Good’ from ‘Excellent’ as “they maybe missed a few more days, they’re late, they are not always there with their stuff,” and a humanities teacher made a similar statement, saying that ‘Good’ students “may not consistently bring materials to class. So, leaving a textbook, not bringing writing utensils, that type of thing, on a consistent basis.” At the other end of the performance scale, a technology education teacher described ‘Needs Improvement’ as “are they never responsible, are they never organized, are they never collaborative?”

One of the group B teachers commented that the 4-point scale used to grade LSWH served as “a reference point” that “gives you some kind of framework,” but then went on to say he found it unlikely teachers were all grading to the same standard, offering that “the argument is there is a standard and it is going to be consistent, and I would argue that’s not really, because my view of what self-regulation is going to be is going to be very different than yours. So, what are we really measuring?” While the data presented here offer evidence that standards differ among teachers, the group B teachers also agreed that ‘Needs Improvement’ represented a serious problem. A geography teacher described ‘Needs Improvement’ as “really low,” and a civics teacher awarded this grade only “if they’ve given me nothing.” A technological education teacher noted that he had never given a student ‘Needs Improvement’ for all six LSWH as this would be below what “even a kid on the lowest part of the behavioural [scale]” would achieve.
Group B teachers did not speak much about using report card comments as a means of reporting or giving feedback about students’ LSWH. One teacher described himself as “liking to give lots of feedback” and this included feedback on students’ LSWH. This feedback was delivered orally, and in the moment, to address an immediate problem. The specific example he gave was research projects where he would observe students’ initiative in finding resources and information.

**Group C:** Like group B, most teachers in group C used consistency to distinguish among the levels of performance when grading LSWH. This was true for seven of the twelve teachers in this group. A science teacher who created an online spreadsheet to record his LSWH observations, had headings that read “always,” “usually,” and “sometimes.” These categories represented the different levels of LSWH performance. This teacher did not have a “never” category in his spreadsheet, but he did note that the grade of ‘Needs Improvement’ was reserved for the student who “doesn’t literally give me anything.” An English teacher described ‘Good’ LSWH as “doing everything you need to be doing, but not necessarily consistently,” and a history teacher made a similar comment, saying “‘Excellent’ is you are consistently doing something. ‘Good’ is… sometimes you are doing it, but there are times where you haven’t.”

The other factor that Group C teachers used to distinguish among levels of LSWH performance was their own expectations. Students who met or exceeded the teacher’s expectations were awarded ‘Excellent.’ An English Language Learners (ELL) teacher described ‘Excellent’ as “over the top,” whereas ‘Good’ was “they are not going above and beyond, but they are meeting the expectations.” A computer science teacher gave a similar account saying ‘Good’ is “they did what I asked, but then did nothing beyond that,” whereas ‘Excellent’ is awarded to students “that go—OK, I did what you asked, but then I also added this to it.”
geography teacher related ‘Good’ to the provincial academic expectations by making this comparison “‘Good’ is provincial standard, sort of 75. Does the work, might answer the question, but doesn’t add anything else to it.” At the bottom end of the scale, a humanities teacher described ‘Needs Improvement’ as “not being anywhere near the expectations.”

While teachers in group C articulated what separates the different levels of performance, they also reported difficulty in making those distinctions reliably. Which levels of performance were most difficult to distinguish varied with the teacher. One teacher found ‘Satisfactory’ and ‘Needs Improvement’ a difficult distinction because “it is harder to distinguish who never ever does and who sometimes does.” This contrasted with a computer science teacher who found ‘Excellent’ to be a difficult level to measure well as “it is easy to lose track of the ones that are always on task, because you don’t necessarily spend time with them.” The most common difficulty was distinguishing between ‘Satisfactory’ and ‘Good.’ One teacher explained the difficulty as “sometimes I find the satisfactory kids will sometimes be good, and the good kids will sometimes be satisfactory,” which does not clarify what separates the two levels. Another teacher also thought the ‘Satisfactory’ to ‘Good’ distinction was difficult because these levels were the least well defined. For her, ‘Needs Improvement’ was used only when “it is pretty clear there is a problem,” and ‘Excellent’ was reserved for standout students, which meant the middle two scale points were less well defined. An English teacher also struggled with defining ‘Satisfactory’ performance, asking “What does satisfactory mean?” and then adding she viewed ‘Satisfactory’ as “kind of heading towards needing to improve.” This view aligned with that of a geography teacher who saw ‘Satisfactory’ as “sort of bare minimum.”

About half of the teachers in group C remarked that the difficulty in defining the LSWH, and in distinguishing among levels of performance likely resulted in grading standards not being
consistent across teachers. An ELL teacher directly stated “there is no consistency between
teachers on how they approach it,” and this was likely because “there is no direction from admin
on how we should approach it.” As an example, she mentioned, “there has never been a staff
meeting where people have said—Here! This will help you with this.” After describing all the
LSWH data she collects, and how she uses it, a history teacher ended with the caveat, “I don’t
know if that is consistent with every teacher.” The French teacher in this group talked about how
she did not know if her assessment process was aligned with those of her peers.

I reflect on each student and look at the information that I put down, but I feel
like what if I am not doing what other people are doing? What other teachers are
doing? And so, it is one those things where I feel like it is not as consistent as it
could be.

The ELL teacher mentioned above is not the only one who remarked upon the lack of
administrative guidance on how to assess and grade LSWH. Only two of the 26 teachers in the
study had been given training in how to assess LSWH. While the four teachers who were part of
a collaborative inquiry group received support from their school administration in the form of
encouragement and release time, there was no direct support in terms of policy, advice, or
training. This was a finding consistent to all groups.

Teachers in all groups found it difficult to describe the distinctions among the levels of
performance, but it appears the extremes of the scale were easiest to articulate. For many
teachers, the ‘habit’ portion of learning skills and work habits is how they distinguished between
levels of performance. The consistency with which students displayed these skills was then
compared to the teacher’s expectations to arrive at a grade. ‘Good’ was the LSWH grade
associated with meeting, but not exceeding expectations, and ‘Needs Improvement’ was for
serious problems. ‘Excellent’ was for students who exceeded expectations, whereas ‘Satisfactory’ was not consistently described. The results are generally consistent across all three groups.

Discussion

It has been suggested that grading policies can be used as tools to reduce the variability in teachers’ grading practices (DeLuca, Braund, Valiquette & Cheng, 2017). While the research described here cannot demonstrate this, it does demonstrate that in the absence of well-defined policies, it is possible for extreme variations in practice to flourish. As an example, all teachers interviewed in this study claimed to determine academic achievement grades in accordance with district policies. This meant that achievement grades were based upon predefined summative assessments, with public, transparent marking schemes. Further, all the participating teachers recorded assessment information, and used that information to determine achievement grades. This contrasts starkly with the teachers’ assessment practices surrounding the LSWH, where a third of the participating teachers did not collect or record assessment data to inform their final grading decisions. Another quarter collected minimal data, meaning that less than half of teachers were able to base grading decisions upon recorded evidence. Even among the teachers who did collect and record LSWH assessment information, some preferred to use their holistic impressions and memory to determine the LSWH grades.

The variability in teacher grading practices extended to how they define the constructs. Construct definitions were vague and inconsistent, with teachers struggling to articulate what observable behaviours or products could be used to inform the assessment of LSWH. Teachers do receive some guidance in this area. The Ontario assessment guide, Growing Success (Ontario Ministry of Education, 2010), provides example behaviours that may be used as indicators of
each LSWH. However, these behaviours are presented as being “intended to assist but not restrict teachers” (p. 10, emphasis in the original) and teachers are further told the behaviours “will look different at the various grade levels.” (p. 10). These statements leave much flexibility for teachers to define the LSWH, and to determine what activities or behaviours should inform the grades.

The variability in how teachers defined the constructs arose for a variety of reasons. First, there were teachers who were either not aware of, or did not refer to, district and ministry advice surrounding LSWH assessment. These teachers graded students’ LSWH based upon holistic impressions only and made grading decisions as report cards were being completed. Other teachers created their own behavioural indicators of the LSWH, and these indicators varied depending on subject area. This was especially evident for teachers in non-academic subjects such as technical education or in special education settings. Finally, there were some teachers who tried to adhere to district and ministry descriptors of each LSWH. While these teachers used the descriptors to inform their grading decisions, they had not incorporated them into their concept of the LSWH. Even among the category C teachers, more than half admitted to not being able to remember what all 6 LSWH were, or what the behavioural indicators for each LSWH are. Thus, as teachers are making observations of students’ LSWH, those observations are not rooted in a conceptual framework that is based upon LSWH assessment advice from the ministry, but rather on their own latent definition of each LSWH.

A further distinction between grading achievement and grading LSWH comes in the form of the assessment activities. Apart from collaboration, teachers did not design assessment activities to measure the LSWH. No teacher spoke of designing activities to assess students’ initiative, or self-regulation. Rather, these skills were observed as students engaged in activities
designed to assess (or teach) subject area knowledge or skills. Referring to academic achievement grades, Willingham, Pollack and Lewis (2002) wrote that “A grade represents each teacher’s judgment as to how well a student has fulfilled the implicit local contract between teacher and student.” (p. 28). It appears this statement holds even more true for LSWH grades. In the absence of well-defined constructs, or specific assessment activities, all a teacher can do is make a general judgment as to how well the student is fulfilling their role. However, exactly what it means to fulfill the role of student is unclear, as different teachers had different opinions about what was important in the classroom. This means that interpreting the meaning behind LSWH grades can be challenging, as it is unknown as to how the teacher is defining the constructs.

Finally, it was clear that how teachers distinguish among levels of performance was different for LSWH grades than it was for achievement grades. Teachers struggled to articulate the differences between levels of LSWH performance. This difficulty may be related to the lack of concrete definitions of the LSWH—without good definitions, it is difficult to determine what separates one level of performance from another. It appears that many teachers internally use a 3-point scale with points of below expectations, meeting expectations, and exceeding expectations. This internal 3-point scale is then mapped onto the 4-point scale used on the report cards. As a consequence of this mapping, the grade of ‘Satisfactory’ is not well defined. This difficulty may reflect that no guidance is given to teachers on what separates the different levels of performance, and so they must make this determination based upon their own judgment and experience.

Clearly, the LSWH assessments as described here do not meet accepted standards for quality classroom assessment (Klinger et al., 2015). Construct definitions are vague and
inconsistent, there are few activities designed to measure the construct, and what distinguishes different levels of performance is not clear. However, it is unknown to what extent this is a problem. It is clearly problematic if the only information parents get about their child’s LSWH comes from the report card. However, evidence from these interview data suggests that in cases where LSWH need improvement, some teachers are leveraging other forms of communication such as email, parent-teacher interviews, and phone calls home. Further, several teachers commented that when problems are noticed with a student’s LSWH, these problems are addressed immediately—the teacher does not wait until report card time to report upon the problem. This immediate addressing of the problem indicates that while LSWH assessment information is used for formative purposes, report cards are not part of the formative assessment process. Finally, there is the possibility that holistic impressions formed over time are an effective way of assessing the LSWH. While such a process contravenes psychometric guidelines for assessment quality, it may more accurately reflect how such judgments are made in other contexts such as employment situations (Arvey & Murphy, 1998; Iqbal, Akbar, & Budhwar, 2015).

Limitations and Further Research

One important limitation in this research is the sampling method. Ideally, random sampling would be used to obtain a sample likely to represent the general teaching population. The convenience sampling method used, combined with the fact that teachers who volunteered to participate in the study were likely to have a high interest in assessing LSWH, likely means the number of teachers categorized into group C is higher than what would be obtained using a random sample. However, several of the teachers were recruited based upon personal contact, or through a teacher ice hockey league. The responses of these teachers may be more representative
of the general teaching population. Of the twelve teachers recruited in this manner, five were placed into group A, five into group B, and two into group C. Thus, while the sampling method likely resulted in the proportion of teachers categorized into group C being not representative of the general population, it is also likely the breadth of assessment processes and beliefs demonstrated by teachers in this sample is representative of the general teaching population.

It became clear through the interviews that teachers in all groups value the LSWH as constructs and want students to have strong LSWH. Thus, formative uses of LSWH assessment information are important to them. Because the LSWH grades are summative, this study did not investigate how Ontario teachers use their LSWH assessment data for formative purposes, nor how they use alternate methods of reporting such as parent-teacher interviews, to communicate students’ LSWH. This would be an area worthy of further study.

Summary

The results presented here demonstrate that teachers’ reports of LSWH grading practices are more variable than their reported grading practices for academic achievement. There appear to be a combination of factors at play. The LSWH constructs are more difficult to define and report upon than academic achievement, and there is much less guidance and training given to teachers in assessing and grading LSWH than there is in assessing and grading achievement. Further, LSWH grades are extremely low stakes. These grades are not used to inform promotion, university entrance or other educational decisions. Consequently, there is minimal pushback from parents, students, or administrators on these grades, so if a teacher makes an unjustified grading decision, it does not result in problems for them. The combination of poorly defined constructs and targeted assessment activities, along with poorly articulated distinctions between performance levels, mean these grades are difficult to interpret, and could not defensibly be used
for high stakes decisions. The current study suggests that if schools value general learning skills such as the LSWH, and are going to continue to ask teachers to assess these skills, clearer policies and procedures are required to guide teachers, and teachers need to be given training and professional development to help them improve and refine this aspect of their assessment practice.
CHAPTER 4: ASSESSING 21st CENTURY SKILLS: WHAT DO REPORT CARD DATA TELL US?

Abstract:

Many school systems are aiming to develop 21st Century skills in their students. Accordingly, teachers are increasingly being asked to assess and report upon such skills. However, how well teachers are able to undertake these assessments is not well studied. This study examines Grade 9 and 12 report card data from two districts in Ontario, Canada to determine to what extent different 21st Century skills are assessed independently of each other, and to what extent they are associated with teacher awarded academic achievement and achievement on a standardized Grade 9 mathematics examination. Results indicate that a set of six different skills (known in Ontario as “learning skills and work habits”) are assessed as a unitary construct. Grades on these skills have higher correlations with teacher awarded grades than with standardized test scores. Finally, gender differences in both academic achievement and achievement on the set of skills are investigated.
Introduction

Teachers are commonly expected to assess elements of student performance beyond academic achievement. These assessments include constructs such as effort, participation, and collaboration. Examples of such expectations for student evaluation can be found in many countries, including, but not limited to, Canada, Hong Kong, Northern Ireland, and Singapore (Hong Kong, N.D.; Merchant, Klinger & Love, in press; Northern Ireland, 2007; Singapore, 2014). Further, similar sets of expectations are included in international education systems such as the International Baccalaureate program, and the International Primary Curriculum (International Baccalaureate Organization, 2009; International Primary Curriculum, 2014). Educational organisations in the United States have also acknowledged the need for teachers to evaluate other educational outcomes beyond achievement. As an example, the Association for Supervision and Curriculum Development (ASCD) argued for a new learning compact, that focuses not just on academics, but also on developing other factors such as empathy, curiosity, creativity, self-discipline, and social competence (ASCD, 2007). More recently, the National Education Association (NEA, 2012) initiated a discussion about how to develop and assess critical thinking, communication, collaboration, and creativity (collectively known as the “4 C’s”) in American public schools. Thus, there appears to be widespread agreement that it is desirable for schools to develop and assess students’ educational outcomes that reflect skills beyond subject area achievement.

While there is broad consensus that classroom teachers should include the assessment of a broad set of learning skills and other competencies, there is no consensus on an appropriate umbrella term for such skills or of the components that such skills should encompass (Duckworth & Schulze, 2009). The lack of a common term and framework has resulted in
myriad descriptors appearing within educational jurisdictions and students’ report cards, such as: “citizenship”, “21st Century skills,” and “cross-curricular competencies.” For example, teachers in Knowledge is Power Program (KIPP) schools report on 24 different aspects of “character,” while teachers within the International Baccalaureate programme are expected to assess elements of the “learner profile” (International Baccalaureate Organization, 2009; KIPP, N.D.). In accordance with the common, current use of “21st century skills” across both educational researchers and practitioners, we will use this term throughout. Admittedly, the term is somewhat of a misnomer, as many of these skills are timeless (Silva, 2009). Nevertheless, it is a phrase that currently enjoys broad appeal and widespread use.

A strong rationale supports including 21st century skills as part of teachers’ assessment and evaluation of their students. Firstly, a large research base connects these skills with improved learning (e.g., Jacob, 2002; Farrington et al., 2012; Muenks, Wigfield, Yang & O’Neal, 2017; Zimmerman, 1990). Skills such as metacognition, self-regulation, and self-efficacy are significantly and positively correlated with learning (Ivcevic & Brackett, 2014; Kleitman & Costa, 2014; Zimmerman & Kitsantas, 2014; Zuffianò et al., 2013). In addition, economic research has found that 21st century skills are associated with better long-term outcomes such as higher employment income, relationship stability, health, lower criminality, and lower drug use (e.g., Almlund, Duckworth, Heckman & Kautz, 2011; Borghans, Duckworth, Heckman & Ter Weel, 2008). Finally, employers have identified 21st century skills as vital to workplace performance and are demanding that students develop these skills in schools (Casner-Lotto & Barrington, 2006; Conference Board of Canada, 2015; Levin, 2012).

There is a further argument that assessing and reporting separately on achievement and non-achievement factors yields a more complete picture of student performance. It is known that
teachers’ achievement grades are not pure measures of achievement (Cross & Frary, 1999). Instead, they reflect a mix of constructs that includes not only achievement, but factors such as effort, focus, and improvement (McMillan, 2001). Some classroom assessment experts (e.g., Guskey, Swan, & Jung, 2011) have argued that by reporting skills such as 21st century skills separately from achievement, we obtain a more complete picture of student performance in the classroom. However, this argument is predicated on the assumption that teachers can and do grade achievement and non-achievement factors separately. If, for example, the grades teachers award for 21st century skills are partially based upon achievement, then the additional information provided by these grades may be minimal.

While the rationale to assess and evaluate 21st century skills is potentially strong, there is an open question as to how well teachers are able to do this. Many of these skills are inconsistently defined in the research literature and precisely defining a measurable construct is critical if assessment and grading are to be reliable and result in valid interpretations of students’ skills (Bass, 2005). In addition, many teachers struggle with assessment and grading not just in North America (e.g., Cizek, 1996; DeLuca & Bellara, 2013), but in other countries as well (e.g., Alkharusi, Kazem & Al-Musawai, 2011). Further, there is evidence that assessment education in Canadian teacher preparation programs focuses on assessment of subject matter knowledge and skills, and not on more general skills such as collaboration or perseverance (Poth, 2012). Our own scan of assessment courses offered by teacher preparation programs in Ontario found only one program (out of 13) that offered a course specifically addressing how to assess constructs other than subject area achievement—and this course is an elective. These facts call into question whether teachers can effectively assess their students’ 21st century skills in their classrooms, and if teachers’ grading of 21st century skills yields useful, actionable information.
The Ontario Context

In Ontario, 21st century skills are labelled “Learning Skills and Work Habits” (LSWH). We will use the term LSWH when referring to the six skills assessed and graded in Ontario, and 21st century skills when referring to the broader set of skills and competencies that are separate from academic achievement. All Ontario K-12 teachers must assess, grade, and report upon a set of six LSWH. The six LSWH are: collaboration, initiative, independent work, organization, responsibility, and self-regulation, and are considered to be “an integral part of students’ learning” (Ontario Ministry of Education, 2010, p. 10). A 4-point scale is used for reporting performance on the LSWH, with teachers selecting from “Excellent”, “Good”, “Satisfactory”, and “Needs Improvement”. Teachers report on the LSWH halfway through a course, and at the end.

The Ontario Ministry of Education (2010) does not define these six skills. Rather, they provide teachers with examples of observable classroom behaviours that may serve as indicators of the skills. For example, behaviours associated with “organization” include, “devises and follows a plan and process for completing work and tasks,” and “establishes priorities and manages time to complete tasks and achieve goals” (p.10). While these descriptors are very reasonable for organization, they also very closely match the description of self-regulated learning found in the literature (Hadwin & Winne, 2012; Zimmerman, 2013). “Self-regulation” has associated behaviours of “assesses and reflects critically on own strengths, needs, and interests” and, “perseveres and makes an effort when responding to challenges” (Ontario Ministry of Education, 2010, p.10). These behaviours also match the definitions of grit and metacognition found in the research literature (Duckworth & Gross, 2014; Flavell, 1979).
The example behaviors described here illustrate the challenge in precisely defining and distinguishing learning skills such as “organization” or “initiative.” Adding to the challenge, these skills are context dependent—organization may look different in an English classroom than it does in a physical education classroom. Further, these skills are not independent (Diamond, 2013; Saskatchewan, 2010; Stecher & Hamilton, 2014). Organization requires self-regulation, and a sense of responsibility towards the group is necessary for effective collaboration. The complex, context dependent, and interrelated nature of these skills may explain why researchers and policy makers have struggled to provide consistent and distinct definitions of such 21st century skills (Duckworth & Schulze, 2009; Farrington et al., 2012). Of importance for our work, if those responsible for deeply understanding these skills struggle to define them precisely, a legitimate question arises as to how teachers are defining these skills. We know that imprecise definitions lead to poor construct validity and interpretations in assessment (Watson & Emeery, 2010). Thus, understanding how teachers define the LSWH is critical if we are to accurately interpret the meaning of the grades, and how they may be used for formative and summative purposes.

The difficulties of defining LSWH are just the first hurdle with respect to assessing LSWH. Even if teachers are devising precise definitions of the LSWH, they must also create good measurements of the LSWH. Measuring a construct becomes more difficult when the construct is multi-dimensional and varies depending on context. We can use one of the Ontario LSWH, self-regulation, as an example. Not only are there a variety of definitions of self-regulation (e.g., Baumeister & Vohs, 2007; Dinsmore, Alexander & Loughlin, 2008; Schunk, 2008; Zimmerman, 2000), but student self-regulation is known to vary depending on motivation (Pintrich & De Groot, 1990). Self-regulation has been measured using a broad variety of
instruments. These include: self-report questionnaires such as the Motivated Strategies for Learning Questionnaire (Dinsmore, Alexander & Loughlin, 2008), standardized observation protocols (Zimmerman & Kitsantas, 2014), interviews (Zimmerman & Martinez-Pons, 1988), specific tasks (Galla et al., 2014), think-alouds (Greene, Robertson & Costa, 2011), and trace data (Winne & Perry, 2000). These tools and methods have various advantages and disadvantages, but one common issue with these assessments of self-regulation in the research literature is that they tended to be conducted once, or over a short period of time. This is not reflective of the classroom environment.

Assessments occurring at a single point in time are not able to measure the “habit” portion of LSWH. A student who demonstrates strong LSWH on a single task may or may not consistently demonstrate such behaviors in the classroom. Accurate and meaningful assessment of the LSWH requires sustained interaction with the students. Hence it is reasonable to consider teachers as ideally positioned to make these assessments (Zimmerman & Martinez-Pons, 1988). Teachers have daily interactions with their students over an extended period of time and can therefore offer a context-rich perspective that is missing from assessments of a single event or at a single point in time such as a task-based measure, self-report questionnaire, or think aloud process. Teachers are also exposed to a range of self-regulatory styles and capabilities among their students, enabling them to have a sense of how a student’s self-regulation compares with norms of the current classroom and the teacher’s prior classrooms (Wigelsworth, Humphrey, Kalambouka & Lendrum, 2010). Finally, teachers are positioned to assess self-regulation using a variety of tools. Nothing prevents teachers from using questionnaires or interview protocols with their students. These types of data could then be supplemented with day-to-day observational
data and student self-assessments, such as reflections or journals, allowing for the collection of a rich data set from which to make a judgment about a student’s ability to self-regulate.

While the potential for teachers to be good assessors of LSWH may be high, certain realities cannot be ignored. For instance, many aspects of LSWH are internal to the student, and difficult to observe directly. The Ontario Ministry of Education (2010) describes one aspect of self-regulation as “reflects critically on own strengths” (p.11), but how does a teacher observe critical reflection on one’s own strengths? Student self-assessment may provide one useful tool, but the reality is that latent thought processes are not directly observable, and therefore difficult for classroom teachers to assess (Lai, 2011). Further, there is strong evidence that teachers’ grading practices are inconsistent and highly variable (Bowers, 2011; Brown 2011; Howley, Kusimo & Parrott, 2000; Lekholm & Cliffordson 2008). If teachers experience challenges with assessing subject area achievement, is it reasonable to expect them to be competent at assessing LSWH? Previous research has highlighted that teachers report struggling to assess simpler aspects of LSWH such as student effort and participation (Linn & Miller, 2005; Miller, Klinger & Shulha, 2006). For example, it is easy to conflate effort with achievement, and participation can take many forms, some of which are not observable.

One study that directly addressed how teachers assess non-achievement constructs was conducted by Ferrito (2015). He found that teachers struggle to assess different “Personal and Social Development” items independently. In his study of the report cards of 113 Grade 4 students in New Jersey, using exploratory factor analysis, Ferrito found that teachers’ ratings of seven such items were best described using a one-factor model. This model accounted for 75% of the variance in the ratings, and all constructs loaded at 0.84 and above onto the single factor. Examples of the “Personal and Social Development” constructs include, “Is able to follow
classroom directions,” “Is able to follow rules,” and “Is able to use Listening Position” (p. 74). A 
surface inspection of these constructs indicates they are likely all related to compliance, and so a 
one-dimensional factor structure is perhaps not surprising.

Ferrito’s results may also rise from a halo effect in teachers’ ratings of “Personal and 
Social Development.” In his discussion of the halo effect, Thorndike (1920) stated that “even a 
very capable foreman, employer, teacher, or department head is unable to treat an individual as a 
compound of separate qualities and to assign a magnitude to each of these in independence of the 
others” (p. 28). Thus, it is possible that while teachers may be required to assess different 21st 
century skills as independent constructs, they are incapable of doing so. There is some research 
to support this hypothesis. Babad, Inbar, and Rosenthal (1982) found that some physical 
education teachers’ ratings of students were impacted by irrelevant factors such as 
socioeconomic status and physical attractiveness. More recently, Duckworth and Yeager (2015) 
suggested that “Teachers’ ratings of students’ specific qualities can also be colored by their top-
down, global evaluations” (p. 241). The halo effect appears in many different rating contexts 
such as student ratings of instructors (Keeley, English, Irons, & Henslee, 2013), supervisor 
ratings of medical residents (McGill, Van der Vleuten, & Clarke, 2011), and ratings of students’ 
academic engagement (Briesch, Chafouleas, & Riley-Tillman, 2010). Given these findings, there 
remains an open question as to how well teachers can assess different 21st century skills as 
independent constructs.

Our own qualitative work in Ontario revealed that teachers struggled to articulate how the 
six LSWH were distinct—except for collaboration (Merchant, 2016). During interviews, teachers 
reported collaboration was the easiest LSWH to assess because it is visible in the classroom. 
Further, collaboration was the only LSWH for which teachers had specific assignments or tasks
that served as assessments. Based on these data, we concluded that Ontario secondary teachers hold a two-dimensional view of the LSWH, with collaboration forming one dimension, and the other five LSWH coalescing into a second dimension. This view is consistent with the perspective that 21st century skills can be divided into interpersonal and intrapersonal skills (Pellegrino & Hilton 2013; Stecher & Hamilton, 2014).

Description of the Study

The paucity of research on how teachers define, assess, and report upon 21st century skills is surprising as there have long been calls for research to be done in this area (e.g., McMillan & Workman, 1998; Stecher & Hamilton, 2014). Our research not only acknowledges the need for such research, but also more deeply explores the issue than found in previous research. In doing so, we hope to illuminate teacher practices in this important, but poorly understood area of classroom assessment. Our efforts to better understand how secondary teachers define, assess, and report upon the six LSWH were supported by a large set of secondary school report card data. Three specific research questions guided our work:

4. To what extent are the six LSWH assessed independently of each other?
5. To what extent are the six LSWH assessed independently of academic achievement?
6. What gender differences exist in patterns of grades on the LSWH?

Report card data were obtained from two school districts in Ontario, Canada. The data included final report card grades for all Grade 9 and 12 students within each district. District 1 raw data consisted of 57,230 sets of grades, but 982 of those were missing the LSWH component. We could see no obvious patterns as to which courses or types of students did not have LSWH grades inputted but noted for students where the LSWH grade was not present, the achievement grade was also frequently not present. As an example, for District 1 of the 982
excluded data sets, 775 were also missing the achievement grade. For the excluded data sets that
did have the achievement grade, the mean grade ($M = 71.17, SD = 21.10$) was not significantly
different than that of the final sample ($M = 71.99, SD = 16.85; t(30,816) = -0.70, p = 0.49$). There
was a significant difference in gender balance ($\chi^2(31,794) = 7.82, p < .01$) with fewer males
(48.8%) in the excluded data than in the included data (53.3%). Unfortunately, we have no way
of knowing why missing data is excluded, but because only 1.7% of the data were excluded we
feel confident our final sample for District 1 was representative of the population of that district.

District 2 raw data contained 26,024 sets of grades, but 1,004 of those were missing the
LSWH components, and so were not usable. Patterns to the missing data were slightly different
than for District 1. The gender balance was not significantly different ($\chi^2(25,394) = 3.59, p =
.06$), but the mean achievement grade of the excluded sample ($M = 70.26, SD = 16.57$) was
significantly lower than for the final sample ($M = 76.44, SD = 13.67; t(25,252) = -13.15, p <
.001$). In this case, the excluded data accounts for 3.9% of the total data, and again we feel this
number is small enough that our sample is likely representative of the population.

Note that in Ontario, high school students typically take six to eight courses per year, and
so the actual number of students included in the sample is lower than the number of sets of
grades. To comply with Canadian privacy laws, data were anonymized by the school districts so
that no identifying information with regards to student, teacher or school was available. This
prevented us from conducting analyses that examined effects at the teacher or school level. So,
while we know that District 1 has 15 secondary schools, and District 2 has 4 secondary schools,
we do not know which grades came from which school, or which class. All statistical analyses
were conducted using SPSS version 24.0, except for the confirmatory factory analysis, which
was conducted using the student edition of LISREL 9.2.
To answer RQ1, an exploratory factor analysis was conducted on the LSWH report card data from one school district. The data from the second school district were then analyzed using a confirmatory factor analysis to determine if the factor structure remained constant across districts. Based upon findings from our prior qualitative study (Merchant, 2016), we hypothesized that a two-dimensional factor structure would emerge. We expected collaboration to be a distinct factor, and that the other five LSWH would form a second factor.

RQ2 was answered using correlational analyses between LSWH grades, teacher awarded subject grades and scores on a standardized mathematics examination. Students in Ontario receive only two standardized tests during high school, and one of those tests is a minimum competency literacy test graded on a pass/fail basis, and therefore not suitable for correlational analyses. The other test is a Grade 9 mathematics examination. Hence the data used to answer this research question was restricted to Grade 9 mathematics only. Fisher’s r to Z transformation was used to test whether the correlation coefficients were significantly different. Based on Steiger’s work (1980), we felt this would provide the best test of significance. We hypothesized that LSWH grades would show a stronger correlation with teacher awarded grades than with standardized test scores. This hypothesis was based upon earlier findings that teachers include constructs such as effort, participation, and attendance in their achievement grades, whereas these constructs are absent from standardized testing results (Brookhart, 1993; Cross & Frary, 1999; Russell & Austin, 2010). Hierarchical multiple regression was used to examine whether LSWH grades could predict standardized mathematics examination scores beyond the teacher awarded grade.

RQ3 was answered by comparing means using t-tests and calculating effect sizes. A further analysis was conducted separately for Grade 9 physical education, as this is the only course in
which students are separated by gender. By exploring if gender differences in grades remain constant in single gender courses, it is possible to illuminate the extent to which LSWH grades are norm referenced. Based upon findings by Duckworth & Seligman (2006), we hypothesized that girls would be awarded higher LSWH grades than boys.

Results

The first RQ was answered using exploratory factor analysis for the LSWH ratings. Ratings were recoded to a numerical scale so that the top point of the scale (excellent) equated to a 4, and the bottom point (needs improvement) equated to 1. The first analysis focused on the Grade 9 report card data from District 1 only (n = 31,087). The Kaiser-Meyer-Olkin measure of sampling adequacy was used to determine if the data distribution was suitable for factor analysis. The results (KMO = 0.945) demonstrated that patterns of correlation were very compact, and therefore exploratory factor analysis was likely to yield interpretable results. EFA was conducted using a maximum likelihood algorithm within SPSS (ver. 24.0). A scree plot was used to determine the dimensionality of the data set. The EFA revealed a unidimensional factor structure with the single factor accounting for 82.1% of the variance. The factor loadings are provided in Table 1. The EFA was repeated for the Grade 12 data (n = 22,854) within the same district, and the results were nearly identical. The single factor accounted for 82.6% of the variance (see Table 1).

To determine if the factor structure was invariant across contexts, the district 1 data were reanalyzed separately for each course. For statistical purposes, only courses where the district
Collaboration was always the lowest loading LSWH, and the remaining five LSWH were tightly clustered in terms of factor loading, although the ordering was not identical across courses. The amount of variance accounted for by the single factor tended to be lowest in mathematics and science courses. As an example, Grade 12 biology exhibited the lowest amount of variance accounted for by the single factor, with the single factor accounting for 73.43% of the variance. At the other extreme, a single factor accounted for 89.23% of the variance in Grade 9 music. The EFA was further repeated separately for each gender. Once again, the factor structure remained consistent, with a single factor accounting for 80.17% of the variance for boys, and 82.79% of the variance for girls. Collaboration retained the lowest factor loading (0.83 for boys, and 0.86 for girls), and the other five LSWH were all above 0.90, except for boys’ responsibility, which had a factor loading of 0.88.

With the EFA giving such strong evidence for a one-dimensional factor structure, it is not surprising that the CFA confirmed that a one-factor model was appropriate for this data. The CFA was completed using the Grade 9 data from District 2, and the model fit parameters were generally very good. The single factor model yielded CFI = 0.99, and the standardized root mean square residual (SRMSR) was very low at 0.01. In contrast, the finding that RMSEA = 0.088, CI

<table>
<thead>
<tr>
<th>LSWH</th>
<th>Grade 9 Factor Loading</th>
<th>Grade 12 Factor Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration</td>
<td>.85</td>
<td>.84</td>
</tr>
<tr>
<td>Independent Work</td>
<td>.91</td>
<td>.91</td>
</tr>
<tr>
<td>Initiative</td>
<td>.92</td>
<td>.93</td>
</tr>
<tr>
<td>Organization</td>
<td>.90</td>
<td>.91</td>
</tr>
<tr>
<td>Responsibility</td>
<td>.93</td>
<td>.93</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>.92</td>
<td>.92</td>
</tr>
</tbody>
</table>
was surprising. At the suggestion of a reviewer, we repeated the analyses treating the LSWH grades as ordered data. This change made little difference to the numerical results, and no difference to the general conclusions. Treating the data as ordered, the one-factor CFA gave fit indices of $\text{CFI} = 0.999$, $\text{SRMSR} = 0.009$, and $\text{RMSEA} = 0.105$, CI [0.102, 0.109]. The high RMSEA values are not ideal, but there is evidence that models with few variables have an inflated RMSEA value (Kenny & McCoach, 2003), and so we do not take these values as evidence of a poorly fitting model. We would have liked to test other models with CFA, but the only different model with a theoretical justification was a 2-dimensional model with collaboration as the second dimension. Because collaboration has only a single measure (the collaboration grade), if we allow collaboration to freely correlate with the other dimension, this 2-dimensional model is mathematically identical to a single factor model. The correlation between collaboration and the other dimension will be identical to the factor loading of correlation in the single factor model. Thus, we gain no new information in testing this model. Instead, as an indicator of how well a two-dimensional model would fit the data, we used EFA and forced a two-factor model onto the data. The variance accounted for increased only by 0.79% in District 1, and 1.71% in District 2. Further, none of the LSWH loaded onto the second factor.

Another way of visualizing the data is through a histogram of the total LSWH scores. For each set of LSWH grades, a total score was calculated, by summing the six individual LSWH grades. There are spikes at total LSWH scores of 6, 12, 18, and 24 (Figure 1). The histogram further reveals a strong negative skew to the grade 9 data (skewness = -0.567 for District 1, and -0.888 for District 2). This skewness is due to a pronounced ceiling effect that is occurring with
the LSWH grades. When further investigated, it was found that 45% of students in District 1 and 53% of students in District 2 received the same grade for all 6 LSWH.

*Figure 1. Histogram of total LSWH scores for grade 9 students in district 1*

The total LSWH score was also used to address the second research question. Since teachers’ ratings of the six LSWH were so strongly unidimensional, and this unidimensionality extended across all subject areas, summing the six LSWH ratings created a good measure of student performance with respect to LSWH. Students’ total LSWH scores were then correlated to both the teacher awarded final grade in the course, and to the score the student received on a province-wide standardized mathematics examination. As the only applicable standardized test given to high school students in Ontario is for Grade 9 mathematics, the analyses were restricted to Grade 9 mathematics. Further, high school students in Ontario are streamed into either applied or academic mathematics courses. Applied mathematics is a lower level course intended for students who would struggle with the demands of the academic course. For the purpose of these analyses, the two courses were separated. Results from the two districts showed that the total
LSWH scores were more strongly correlated with teacher awarded grades, than with standardized test scores (see Table 2). Using Fisher’s r to z transformation, it was determined that for both districts and both streams, the differences in correlation coefficients were significant ($p < 0.001$). Correlations between the teacher awarded grade and the standardized math test score were high. For District 1 the value was $r = 0.76$ for academic math ($n = 2096$) and $r = 0.75$ for applied math ($n = 1250$). In District 2 the values were $r = 0.77$ and $r = 0.80$ for academic ($n = 497$) and applied ($n = 245$) math respectively.

Table 4.2. *Pearson correlation coefficients between academic achievement and LSWH total score*

<table>
<thead>
<tr>
<th></th>
<th>District 1</th>
<th>District 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher awarded final grade</td>
<td>0.79</td>
<td>0.78</td>
</tr>
<tr>
<td>Standardized mathematics examination score</td>
<td>0.52</td>
<td>0.53</td>
</tr>
<tr>
<td>z-value (Fisher’s r)</td>
<td>12.47**</td>
<td>15.02**</td>
</tr>
</tbody>
</table>

** $p < 0.001$

To further investigate the relationship between LSWH and achievement, a hierarchical multiple regression was conducted to determine if LSWH contributed to the prediction of the standardized mathematics examination score. Data from District 2 were used as they contained information about students’ attendance. Gender, lates, and absences were entered into the first step of the regression, the teacher awarded grade into the second step, and the LSWH total into the third step. The final model accounted for 63% of the variance in the standardized mathematics examination score (Table 3). However, the addition of the third step, added only
0.8% to the variance accounted for, meaning the total LSWH score accounted for a minimal amount of variance after the teacher awarded grade had been included in the model. It is interesting to note the regression coefficient for the LSWH was negative in the final model. This means that when controlling for teacher awarded grades, higher LSWH grades predicted lower scores on the standardized mathematics examination. Because of the negative regression coefficient, we followed the advice of Smith, Ager, and Williams (1992) and tested for suppression effects using semipartial correlations but found no evidence of such effects.

Table 4.3. Linear model of predictors of standardized mathematics examination scores

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>β</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>3.347</td>
<td>.078</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>.005</td>
<td>.005</td>
<td>.041</td>
<td>.236</td>
</tr>
<tr>
<td>Lates</td>
<td>-.042</td>
<td>.014</td>
<td>-1.08</td>
<td>.003</td>
</tr>
<tr>
<td>Absences</td>
<td>-.014</td>
<td>.003</td>
<td>-1.194</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.878</td>
<td>.088</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.006</td>
<td>.003</td>
<td>-1.043</td>
<td>.055</td>
</tr>
<tr>
<td>Lates</td>
<td>-.001</td>
<td>.009</td>
<td>-1.002</td>
<td>.936</td>
</tr>
<tr>
<td>Absences</td>
<td>-.001</td>
<td>.002</td>
<td>-1.007</td>
<td>.766</td>
</tr>
<tr>
<td>Teacher awarded grade</td>
<td>.035</td>
<td>.001</td>
<td>.790</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.886</td>
<td>.087</td>
<td>&lt; .001</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.002</td>
<td>.003</td>
<td>-1.019</td>
<td>.412</td>
</tr>
<tr>
<td>Lates</td>
<td>-.009</td>
<td>.009</td>
<td>-1.022</td>
<td>.339</td>
</tr>
<tr>
<td>Absences</td>
<td>-.002</td>
<td>.002</td>
<td>-1.023</td>
<td>.319</td>
</tr>
<tr>
<td>Teacher awarded grade</td>
<td>.039</td>
<td>.001</td>
<td>.890</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>LSWH</td>
<td>-.019</td>
<td>.005</td>
<td>-1.148</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Adjusted R² = .059 for Step 1; ΔR² = 0.560 for Step 2; ΔR² = 0.008 for Step 3.

Gender (1 = boys, 2 = girls)

While the LSWH factor structure remained constant for boys and girls, girls significantly outperformed boys on all 6 LSWH. The difference was lower for Grade 12 students than Grade 9 students, but still significant (see Tables 4 and 5). At both the Grade 9 and 12 levels, girls
received higher academic grades than boys across all courses. There was no gender difference in score on the Grade 9 standardized mathematics examination, even though girls received higher teacher grades in the course than boys \((p < 0.001)\). T-tests were used to find statistically significant differences between genders.

Table 4.4. Gender differences in LSWH and academic achievement – district 1 only

<table>
<thead>
<tr>
<th></th>
<th>Boys (N=16131)</th>
<th>Girls (N=14182)</th>
<th>p-value</th>
<th>Effect Size</th>
<th>Boys (N=11162)</th>
<th>Girls (N=11296)</th>
<th>p-value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Collaboration</strong></td>
<td>2.92 (0.01)*</td>
<td>3.26 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.36</td>
<td>3.02 (0.01)</td>
<td>3.29 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.28</td>
</tr>
<tr>
<td><strong>Independent Work</strong></td>
<td>2.76 (0.01)</td>
<td>3.19 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.42</td>
<td>2.84 (0.01)</td>
<td>3.20 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Initiative</strong></td>
<td>2.69 (0.01)</td>
<td>3.10 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.39</td>
<td>2.72 (0.01)</td>
<td>3.09 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>2.68 (0.01)</td>
<td>3.22 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.52</td>
<td>2.73 (0.01)</td>
<td>3.18 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.43</td>
</tr>
<tr>
<td><strong>Responsibility</strong></td>
<td>2.71 (0.01)</td>
<td>3.15 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.42</td>
<td>2.72 (0.01)</td>
<td>3.11 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Self-Regulation</strong></td>
<td>2.71 (0.01)</td>
<td>3.17 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.45</td>
<td>2.77 (0.01)</td>
<td>3.14 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>Teacher Awarded</strong></td>
<td>69.42 (0.13)</td>
<td>75.1 (0.14)</td>
<td>&lt; 0.001</td>
<td>0.34</td>
<td>68.98 (0.18)</td>
<td>74.85 (0.17)</td>
<td>&lt; 0.001</td>
<td>0.31</td>
</tr>
<tr>
<td>Grade – All courses**</td>
<td>66.27 (0.39)</td>
<td>69.38 (0.42)</td>
<td>&lt; 0.001</td>
<td>0.18</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Teacher Awarded</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade – Mathematics Only Standardized Mathematics Examination ***</td>
<td>3.06 (0.02)</td>
<td>3.03 (0.02)</td>
<td>= 0.34</td>
<td>0.03</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* Number in parentheses is the standard error of the mean.
** Teacher awarded grades are given as a percentage in Ontario
*** The standardized mathematics examination is scored on a scale from 0.0 to 4.9. \(N = 3489\)

Grade 9 physical education is the only course in which there exist separate gender classes. To examine if teachers are norm-referencing their LSWH grades with respect to their classroom norms, the LSWH results for Grade 9 physical education were analysed separately (see Table 6). If the gender gap disappeared, it would provide evidence that LSWH grades are norm referenced. For this course, the gap between girls’ and boys’ mean LSWH grade narrowed but did not disappear. Initiative was the only LSWH without a statistically significant difference
(\(p = 0.16\)) in scores between genders. All other LSWH show significant differences (\(p < 0.001\)) in mean grade between girls and boys, but effect sizes were smaller than when the LSWH grades were compared from all courses. Organization and self-regulation have the largest effect sizes.

Table 4.5. Gender differences in LSWH and academic achievement – district 2 only

<table>
<thead>
<tr>
<th></th>
<th>Grade 9</th>
<th></th>
<th>Grade 12</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys (N=6395)</td>
<td>Girls (N=6941)</td>
<td>(p)-value</td>
<td>Boys (N=6069)</td>
</tr>
<tr>
<td>Collaboration</td>
<td>3.18 (0.01)*</td>
<td>3.47 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.39</td>
</tr>
<tr>
<td>Independent Work</td>
<td>2.98 (0.01)</td>
<td>3.41 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.50</td>
</tr>
<tr>
<td>Initiative</td>
<td>2.93 (0.01)</td>
<td>3.33 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.45</td>
</tr>
<tr>
<td>Organization</td>
<td>2.92 (0.01)</td>
<td>3.41 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.54</td>
</tr>
<tr>
<td>Responsibility</td>
<td>2.92 (0.01)</td>
<td>3.35 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.48</td>
</tr>
<tr>
<td>Self-Regulation</td>
<td>2.95 (0.01)</td>
<td>3.37 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.48</td>
</tr>
<tr>
<td>Teacher Awarded Grade – All Courses</td>
<td>73.92 (0.13)</td>
<td>79.63 (0.12)</td>
<td>&lt; 0.001</td>
<td>0.43</td>
</tr>
<tr>
<td>Teacher Awarded Grade – Mathematics Only</td>
<td>70.40 (0.57)</td>
<td>74.08 (0.50)</td>
<td>&lt; 0.001</td>
<td>0.25</td>
</tr>
<tr>
<td>Standardized Mathematics Examination Score**</td>
<td>2.62 (0.02)</td>
<td>2.71 (0.02)</td>
<td>= 0.448</td>
<td>0.09</td>
</tr>
</tbody>
</table>

* Number in parentheses is the standard error of the mean.
** N = 1569 for the Teacher Awarded Grade – Mathematics Only and the Standardized Mathematics Examination Score

Discussion

The analyses of our data highlight several potential issues with teachers’ assessment of LSWH. The factor analysis results clearly indicate that these six LSWH grades represent a unidimensional construct, suggesting that teachers are not assessing the six LSWH as distinct. Admittedly, it is possible that the six LSWH are assessed independently of each other, but the
constructs themselves are so closely correlated, that the results appear unidimensional. A second possibility is that teachers are assessing some or all of the six LSWH as distinct constructs, but that teachers’ definitions of each construct are randomly varied, such that the entire set of grades appears unidimensional. However, given that approximately 80% of the variance in LSWH grades can be accounted for by one factor, and based on our earlier qualitative work, we believe the most likely explanation is that teachers’ assessments of the six LSWH are founded upon holistic impressions of their students formed over time. Interviews with Ontario teachers revealed that over half the participants could not name all six LSWH (Merchant, 2016). If teachers are not aware of what the six LSWH are, they cannot be assessing them as separate constructs. Holistic impressions would also explain why approximately half of students receive the same rating across all six LSWH.

Table 4.6. Gender differences in LSWH and academic achievement for Grade 9 physical education – district 1 only

<table>
<thead>
<tr>
<th></th>
<th>Grade 9 Physical Education</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Boys (N=1619)</td>
<td>Girls (N=1477)</td>
<td>p-value</td>
<td>Effect Size</td>
</tr>
<tr>
<td>Collaboration</td>
<td>3.29 (0.02)*</td>
<td>3.40 (0.02)</td>
<td>&lt; 0.001</td>
<td>0.14</td>
</tr>
<tr>
<td>Independent Work</td>
<td>3.11 (0.02)</td>
<td>3.26 (0.02)</td>
<td>&lt; 0.001</td>
<td>0.17</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.05 (0.02)</td>
<td>3.09 (0.02)</td>
<td>= 0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>Organization</td>
<td>2.99 (0.01)</td>
<td>3.20 (0.02)</td>
<td>&lt; 0.001</td>
<td>0.23</td>
</tr>
<tr>
<td>Responsibility</td>
<td>3.08 (0.02)</td>
<td>3.21 (0.01)</td>
<td>&lt; 0.001</td>
<td>0.15</td>
</tr>
<tr>
<td>Self-Regulation</td>
<td>3.10 (0.02)</td>
<td>3.31 (0.02)</td>
<td>&lt; 0.001</td>
<td>0.24</td>
</tr>
<tr>
<td>Teacher Awarded Grade</td>
<td>76.15 (0.33)</td>
<td>77.87 (0.38)</td>
<td>&lt; 0.001</td>
<td>0.12</td>
</tr>
</tbody>
</table>

* Number in parentheses is the standard error of the mean.
We justified testing whether collaboration emerged as a distinct factor because it is the only LSWH that is interpersonal, rather than intrapersonal, and because in prior research teachers reported they were able to separate collaboration as a separate construct, while they were more likely to group the other five LSWH into a single construct. The factor analyses provide weak support for this in the report card data, as collaboration always had the lowest loading within the single factor model, regardless of district, grade, course, or stream. However, when we forced a two-factor model onto the data, the variance accounted for increased by a minimal amount, and collaboration did not emerge as a separate factor. Thus, it appears that either collaboration is very highly correlated with the other five LSWH, or teachers are not grading it as a separate construct. Collaboration also had the highest mean score and lowest standard deviation of any of the LSWH, and the smaller variance likely reduced factor loadings.

Based on the assumption that strong 21st century skills lead to better learning, it was expected that LSWH grades would correlate positively with academic achievement and standardized test scores. A stronger correlation between LSWH and teacher awarded grades than with standardized test scores was also expected, based upon prior research demonstrating teacher awarded grades are not pure measures of achievement, but include subjective judgments, and other factors such as effort and participation (Allal, 2013; Hunter, Mayenga & Gambell, 2006; McMillan, 2001). While it is likely that strong LSWH are positively impacting students’ achievement grades, the strength of this association may be influenced by other factors. For example, the direction of influence between achievement and LSWH grades may be bidirectional. It is possible that teachers use achievement grades to inform their LSWH grades and vice-versa. Another possibility is that because achievement grades and LSWH grades are
both assigned by teachers, common method variance is creating strong correlations between the LSWH grades and achievement grades.

At first glance, the negative regression coefficient for LSWH and standardized mathematics examination scores appears surprising. However, this negative value is small, and appears after controlling for the teacher awarded grade. One possible explanation of the negative effect is that it represents attempts by teachers to compensate some students’ poor grades with an acknowledgement of positive effort. We tested this explanation by examining a scatter plot of the LSWH grades vs. the EQAO scores (Figure 2). If this explanation were correct, there should be a preponderance of data points with high LSWH grades and low EQAO scores.

*Figure 2. Scatterplot of total LSWH scores vs. EQAO score for grade 9 mathematics students in district 2.*

An examination of the Figure 2 shows this to be not the case. Instead, it can be seen that the data points with the largest number of people have LSWH total scores of 24 and EQAO
scores between 3 and 4. These data points fall outside the $p = 0.99$ confidence interval lines for the line of best fit, meaning these students scored lower on the EQAO mathematics examination than predicted by the line of best fit. Additionally, the figure shows that many students who received low LSWH grades did better on the EQAO mathematics examination than predicted. Thus the figure shows many high LSWH students scoring lower on the EQAO examination than predicted, and low LSWH students scoring higher than predicted. This is likely what is responsible for the negative association between LSWH and EQAO scores found in the multiple linear regression. However, it should be re-emphasized that while statistically significant, this coefficient is small. Therefore, the conclusion holds that LSWH scores do not serve as useful predictors of EQAO math achievement above and beyond the teacher awarded grades.

Prior research indicates that girls’ higher achievement in school is a result of them possessing better self-discipline than boys (Duckworth & Seligman, 2006). The results here show that secondary teachers in Ontario rated girls as having better LSWH than boys in all categories. This is true for both grade 9 and 12, but the difference is smaller in grade 12. Duckworth and Seligman (2006) found that better self-discipline accounted for girls’ achieving higher grades than boys in school. Like this study, Duckworth and Seligman used teacher report to measure self-control. However, their use of standardized questionnaires developed by researchers avoided the potential of common method variance accounting for strong correlations between self-control measures and achievement grades. In the present study, both the achievement grades and the LSWH grades were determined solely by the teacher, meaning that strong correlations between them may have been partially due to common method variance. This gender difference was true for both Grades 9 and 12, but the difference was smaller in grade 12. One potential explanation for this decrease in the gap would be that low achieving males are
more likely to drop out than other students (Stetser & Stillwell, 2014; Wang & Fredricks, 2014). It may also be a result of students in Grade 12 becoming more serious about their educational success because of a desire to maximize options for post-secondary education and employment.

Gender gaps in LSWH performance narrow, but do not disappear, in Grade 9 physical education (the only class that is single gender). This suggests that teachers are using a mix of norm- and criterion-referenced frameworks when grading LSWH. Alternatively, it is possible that Grade 9 students demonstrate different behaviours in single gender classrooms, or that boys have a higher interest in physical education than girls, and this is reflected in improved LSWH grades. While these other explanations are plausible, we contend that in the presence of other research that demonstrates teachers often interpret grading criteria in relation to the norms of their classroom (e.g., Sadler, 2009; Wyatt-Smith & Klenowski, 2013), it is reasonable to conclude that Ontario secondary teachers use a mix of norm- and criterion-referencing when grading LSWH.

Our findings highlight a broader issue with LSWH assessment policies in Ontario. While teachers are given strong guidance and support about how students’ academic abilities should progress as they move through the school system, there appears to be no guidance or direction about how students’ LSWH should develop over time. Ontario provincial curriculum documents and assessment policies do not provide teachers with criteria or standards for the LSWH. Thus, teachers are creating their own, internal, criteria and standards for the different levels of LSWH performance at each grade level. We do not know, for example, how teachers distinguish between “excellent” collaboration and “good” collaboration, nor whether those standards are different for Grade 12 students than for Grade 9 students. Clearly, more research needs to be
done to better understand how teachers define these constructs, what activities they use to assess them, and what student behaviours and characteristics influence the grades awarded.

Implications and Limitations

Stecher and Hamilton (2014) found that while teachers are interested in developing and assessing 21st century skills, they “do not have the resources to develop programs or assessments on their own” (p. 7). Our findings here provide further empirical evidence of the necessity to support teachers’ classroom assessment practices in relation to 21st century skills. Given their sustained interaction with students over time, and in a variety of contexts, teachers would appear to be well positioned to be quality assessors and reporters of students’ 21st century skills (Zimmerman & Martinez-Pons, 1988). However, our evidence strongly suggests that teachers’ grades of such skills are likely based more upon holistic impressions of the student than well-defined constructs. Consistent with this hypothesis is our finding that approximately half of students receive the same grade across all six LSWH. Such findings are not surprising given that very few teachers receive training or guidance on how to define, assess, and grade such skills. For example, in Ontario, there is no requirement for preservice teacher education to include courses or even a single lesson on how to assess the LSWH. This points to the need for teachers to be given training, and concrete examples of how to assess 21st century skills in a manner that is defensible and valid. School systems interested in having teachers assess and report upon 21st century skills need to give careful consideration as to what skills should be assessed, how they can be defined clearly for teachers, students, and parents, and what are appropriate standards for these skills at different grade levels. Further, teachers will need to have resources in place including; ongoing training, example assessment activities, and grade level standards.
Even with such training, it is not clear that teachers would assess the LSWH as independent constructs. Evidence from the rater training literature demonstrates inconsistent effects on how effective rater training is at reducing halo effects (e.g., Bernardin, 1978; Woehr & Huffcutt, 1994), and that the effects of rater training diminish over time (Ivancevich, 1979). One possible way to prevent halo effects from impacting teachers’ LSWH grades would be to stop asking teachers to assess multiple LSWH for the same student. It would be possible in secondary schools for different subjects to be responsible for assessing and reporting upon a single LSWH. Because students generally take more than six subjects during a school year, all six LSWH could be assessed and reported for most students. The province of Quebec adopts a similar approach where teachers may choose which two (of four) “cross-curricular competencies” they report upon (Quebec Ministry of Education, Leisure and Sport, n.d.).

Girls obtaining higher grades than boys is a well-known phenomenon that is documented across grade levels, races, and contexts (Voyer & Voyer, 2014). Our research finds that, in Ontario, this gender gap extends to non-achievement constructs such as the LSWH. This finding is consistent with the findings of Duckworth and Seligman (2006), although the results shown here demonstrate a smaller gender gap than they found. Our results add to the growing body of research indicating that boys have not only lower grades than girls, but also weaker 21st century skills. Further, our finding that gender differences in standardized mathematics scores are minimal is consistent with other data such as the Canadian PISA results (Brochu, Deussing, Houme & Chuy, 2013). The lack of significant gender differences in standardized mathematics examination scores suggests that boys and girls have equal mathematical ability, and so girls’ higher achievement on report cards is due to other factors. One possibility is that a gender bias exists in Ontario mathematics teachers’ assessment and grading practices. Another possibility is
teachers perceive girls to have better LSWH, and this leads to better grades as teachers consider the LSWH when assigning achievement grades. This explanation is consistent with prior research on teachers’ grading practices that demonstrates teachers include non-achievement factors in their achievement grades (e.g., Bowers, 2011; McMillan, 2001). Thus, educators with an interest in improving boys’ academic achievement at school may wish to focus on improving their 21st century skills. Unfortunately, the unidimensional nature of the LSWH grades reported here precludes us from suggesting which skills would be most useful to develop first.

The most severe limitations of the study are imposed by the nature of the data obtained. While the large number of data points improves statistical power, the limited amount of information collected for each set of grades constrained the analyses that could be performed. As an example, it would be interesting to test how variable the LSWH grades are across contexts. Are the biggest sources of variance the student, the teacher, or the school? Unfortunately, the anonymity of the data prevents us from answering these questions. It would also be interesting to look at standardized test scores for subjects other than mathematics, as mathematics appears to be the subject where non-achievement factors have the smallest impact on achievement grades (Bol, Stephenson, O’Connell, & Nunnery, 1998; Duncan & Noonan, 2007; Pilcher, 1994). Further, it would be helpful to complement the data with different ratings of the LSWH or related constructs such as conscientiousness or self-control. These additional measures could help determine to what extent the strong correlations between achievement and the LSWH grades are due to common method variance.

Conclusions

Ontario secondary teachers assess the 6 LSWH as a unitary construct, which accounts for approximately 80% of the variance in the LSWH. This indicates that LSWH grades likely do not
reflect separate performance levels on six distinct constructs, but a teacher’s overall impression of the student. Consistent with other research, we found LSWH grades are more strongly correlated to teacher-awarded grades than standardized test scores. This implies that teachers conflate achievement and non-achievement factors when grading, although other explanations, such as common method variance, exist. Gender gaps in LSWH achievement are significant, both statistically and practically, and it would be worthwhile to further examine the impact this has on gender gaps in academic achievement, and whether improving boys’ LSWH would help close this gap. Taking a broader perspective, these findings from Ontario imply that if school systems wish to incorporate the assessment and grading of 21st century skills at the classroom level, they need to provide teachers with appropriate supports including training, sample assessment activities, and grade level standards.
CHAPTER 5: HOW DO SECONDARY TEACHERS ASSESS SELF-REGULATION?
EVIDENCE FROM ONTARIO

Abstract:

Self-regulation is known to be positively associated with better academic, and life, outcomes. As a result, many school systems aim to develop self-regulation, or related constructs. Thus, teachers in many school systems are asked to assess and report upon self-regulation (or related constructs). How secondary teachers in Ontario, Canada accomplish this task was investigated using mixed methods research. Phase 1 involved semi-structured interviews with 26 secondary teachers. These interviews informed the development of an online survey. Finally, report card data was used to examine large scale trends in self-regulation grades. The overall findings are that Ontario secondary teachers vary in their definitions of self-regulation, but the most common definition is more closely aligned with complying with the teacher’s expectations than with maintaining focus on a goal. Report card data suggest that teachers struggle to assess self-regulation independently from other constructs such as initiative or collaboration.
Self-regulation has become a buzzword in education. As an example, a Google Scholar search on self-regulation AND education yielded over 30,000 hits over the last 5 years. The popularity of self-regulation as a construct to be studied is not without good reason. Besides an ongoing and large research base demonstrating self-regulation is related to academic achievement (Nuckles, Hubner & Renkl, 2012; Pintrich & De Groot, 1990; Zimmerman & Kitsantas, 2014), there is evidence that teachers can effectively teach self-regulation (Perry, Hutchinson & Thauberger, 2007). Given the mutable nature of self-regulation, and its positive associations with learning, it is perhaps not surprising that many school systems are making self-regulation a focus.

One systematic example is illustrated in the Canadian province of Ontario. Ontario’s secondary teachers are required to assess and grade a set of six learning skills and work habits (LSWH). The six LSWH are collaboration, independent work, initiative, organization, responsibility, and self-regulation. Thus, as Ontario secondary teachers assess self-regulation, they must also assess five other non-achievement constructs. All six LSWH are included on report cards and graded using a 4-point scale with scale points of excellent, good, satisfactory, and needs improvement. This current study explores how Ontario secondary teachers operationalize self-regulation. Self-regulation was selected as it is the LSWH with the strongest research base. The specific research questions guiding this study were:

1. How do Ontario secondary teachers define self-regulation? How well do their definitions align with those put forward by researchers and by the Ontario Ministry of Education?

While there have been many studies on classroom teachers’ assessment and grading practices (see Brookhart et al., 2016 for an overview), these studies have focused on assessment and grading of academic achievement. However, school systems across the world, including educational systems in Singapore, many schools in the United States, and the internationally renowned International Baccalaureate program, require teachers to assess elements of student performance beyond subject area achievement (Boise School District, 2017; International Baccalaureate, 2009; Singapore Ministry of Education, 2014). Canadian school systems are no exception, and teachers in all provinces must assess and report upon diverse student skills beyond subject area achievement (Merchant, Klinger & Love, 2018). Despite such assessments being a common part of teachers’ practices, there has been minimal research on how teachers assess and report upon these aspects of student learning. This study addresses this gap in the assessment and grading literature. Ontario is an ideal location for research on teachers’ assessment of self-regulation as it has a large population base from which to draw data, and these assessments have been required since 2010, so teachers have had time to refine and stabilize how they assess these skills, and specifically for this study, self-regulation.

There is a need to understand how teachers assess self-regulation, and what meaning can be derived from self-regulation grades on report cards if we are to use this information for formative or summative purposes. For example, if teachers do not have shared understandings of what is meant by self-regulation, it will be very challenging to have meaningful discussions or make policy decisions to improve students’ self-regulation skills. Similarly, we need to determine if these assessments and grades have sufficient rigour, validity, and reliability to be used for important decisions such as admission into special programs, or post-secondary opportunities. Further, it will be critical to understand the relationship between teachers’
conceptions of self-regulation and current definitions found in the research literature. Do teachers’ concepts of self-regulation align with current definitions found in the research literature? If teachers and researchers have differing concepts of self-regulation, then communication between the two groups (e.g., professional development or other knowledge mobilization efforts) will likely be impeded. A better understanding of how teachers define and assess self-regulation could potentially lead to more effective use of these assessments to improve students’ self-regulation, and to identify and inform areas of need for teachers’ future professional development.

What is self-regulation and how is it assessed?

There are many definitions of self-regulation. Burman, Green & Shanker (2015) claimed to find 447 different terms related to self-regulation in the research literature. Similarly, Boekarts and Corno (2005) noted that “over the past two decades, researchers have struggled with the conceptualisation and operationalisation of self-regulatory capacity, coming to the conclusion that there is no simple and straightforward definition of the construct of SR [self-regulation]” (p. 200). Early definitions of self-regulation focused on compliance and socially acceptable behaviour. Kopp’s (1982) description of self-regulation included students’ abilities to “comply with a request” (p. 199) and to “generate socially approved behaviour” (p. 199). More recently, it appears researchers are coming to a consensus that self-regulation involves controlling one’s cognitive and emotional processes to achieve a desired goal. McClelland and Cameron (2012) defined self-regulation as the “capability of controlling or directing one’s attention, thoughts, emotions, and actions” (p. 136). Carver and Scheier (2014) described self-regulation as “self-corrective adjustments” (p. 3) that are “needed to stay on track, whatever one’s current purpose is” (p. 3). The notion of achieving a goal implies that planning may be a component of self-
regulation, and so planning appears in some definitions of self-regulation (e.g., Montroy, Bowles, Skibbe, McClelland & Morrison, 2016).

Current concepts of self-regulation are less focused on compliance, but remain multi-dimensional (Duckworth & Kern, 2011). Thus, self-regulation encompasses a range of regulatory behaviours working together to maximize an individual’s potential for achieving a goal. Diamond (2013) described self-regulation as referring to “processes that enable us to maintain optimal levels of emotional, motivational, and cognitive arousal” (p. 152). Boekarts (1997) linked self-regulation in school settings to cognition, motivation, volition, goal setting, effort, and prior knowledge, whereas Moffitt et al. (2011) linked self-regulation with conscientiousness. To further complicate the issue, self-regulation, and how it is manifested, changes depending on the child’s developmental stage. In early childhood, self-regulation is sometimes associated with compliance (e.g., Kochanska, Coy, & Murray, 2001) whereas in adolescence it can be associated with refraining from risky behaviours (Magar, Phillips, & Hosie, 2008). With such a broad range of constructs involved in self-regulation, it is perhaps not surprising that a universally accepted definition eludes us.

The Ontario Ministry of Education (2010) does not define any of the LSWH—including self-regulation. Instead, teachers are given sample behaviours that could be used as indicators for each of the six LSWH. The behaviours listed are:

- sets own individual goals and monitors progress towards achieving them;
- seeks clarification or assistance when needed;
- assesses and reflects critically on own strengths, needs, and interests;
- identifies learning opportunities, choices, and strategies to meet personal needs and achieve goals;
• perseveres and makes an effort when responding to challenges. (p. 11)

These behaviours do not form a definitive list. Teachers are specifically told “the sample behaviours are intended to assist but not restrict teachers” (p. 10, emphasis in the original) in their assessment and evaluation of the LSWH. Thus, Ontario teachers have considerable leeway in how they define and operationalize the six LSWH, and how they distinguish among the different levels of performance.

These sample behaviours align with current, research-based definitions of self-regulation, and could be reasonably perceived as helpful for students to achieve their learning goals. However, some of these behaviours may not be obvious in the classroom. For example, unless a teacher specifically undertakes goal setting activities in class, and monitors those activities, it may be impossible for a teacher to know if students are setting goals, and monitoring progress towards them. While goal setting can be a public, transparent process, it can also be latent. Of these listed behaviours, only “seeks clarification or assistance when needed” is a directly observable behaviour. Even constructs such as “perseverance” or “making an effort” can be difficult to assess (Linn & Miller, 2005).

Since self-regulation is a complex behavioural construct, especially in the context of learning in a classroom, it is difficult to assess (Credé & Kuncel, 2008; Dinsmore, Alexander & Loughlin, 2008; Stroud, 2013). To tackle this complexity, researchers have developed myriad assessments of self-regulation including self-report inventories, observation, structured interviews, think-alouds, error detection tasks, online exercises, and teacher judgments (Galla et al., 2014; Stroud, 2013). Of these methods, self-report measures have enjoyed the greatest popularity (Winne & Perry, 2000). Self-report measures are popular for assessing self-regulation because they can be useful for accessing internal thought processes and because they are
economical and easy to implement (Winne & Perry, 2000). At the classroom level, there is evidence that many Ontario teachers have students complete self-assessments of their LSWH, but those self-assessments are not considered when teachers make LSWH grading decisions (Merchant, 2016).

Unfortunately, self-report tools are known to be less than ideal measures of self-regulation. Respondents, especially children, may not be reliable reporters of their own learning strategies. Lapses in memory, distorted self-image, changing self-perception, and different ways of interpreting inventory items are all factors that can reduce the validity and reliability of self-report measures (Hirschfeld, Thomas & McNatt, 2007). As an example, one of the most popular self-report measures, the Motivated Strategies for Learning Questionnaire (MSLQ) has weak to middling psychometric properties with Cronbach’s alpha ranging from $\alpha=0.52$ to $\alpha=0.80$ for the different subscales (Pintrich, 1991). Further, self-regulation, as measured by the MSLQ has a low correlation ($r = 0.32$) with academic grades (Pintrich, Smith, García & McKeachie, 1993).

Interviews provide another method to assess self-regulation. As with questionnaires, they involve self-report, but interviews allow researchers to explore a student’s thought processes in much greater depth than questionnaire items. The most popular interview protocol appears to be the Self-Regulated Learning Interview Schedule (SRLIS). Zimmerman and Pons (1986) developed the SRLIS as a means of assessing self-regulation in high-school students. The SRLIS is essentially a count of self-regulation strategies and participants are awarded a point for each strategy they reported they used. Initial results showed that 91% of the students in the validation study could be classified correctly as being in a normal or low achieving stream by the SRLIS results. It could be argued these results are not very impressive as there was a substantial difference in achievement between the two groups. In standardized testing, the high achievement
group scored 2.63 standard deviations higher in math achievement, and 1.75 standard deviations higher in English achievement than the lower achievement group. These very large differences in achievement make it unclear if the SRLIS is sensitive to comparisons between students with more similar levels of achievement. In addition, the participants were streamed into higher and lower achieving groups based partially upon their GPA. Given that teachers commonly include factors such as effort and self-regulation into their grading (Nash & McMillan, 2000; Randall & Engelhard, 2010), it seems likely that students in the low achieving group would have likely have been placed there at least partially because of their lower self-regulation skills. Still, interviews are a form of self-regulation assessment that teachers could incorporate into their practice. While I could find no research indicating to what extent these types of interviews take place in classrooms, it seems likely that, at an informal level, some teachers use their conversations with students as assessment of self-regulation.

Another assessment method reported in the literature is teacher or parent report. Given that teachers have extended ongoing contact with students, it may seem they are ideally positioned to assess students’ self-regulation. However, there are concerns about teachers’ abilities to be reliable raters of students’ self-regulation (Winne & Perry, 2000). In particular, there is a question as to how well teachers can separate self-regulation from other constructs such as achievement and motivation. For example, as part of their validation of the SRLIS, Zimmerman and Martinez-Pons (1988) used teacher reports as a comparison tool and found that teachers assessed different facets of self-regulation as a unitary construct. According to the authors, this provides “compelling evidence that teachers view students' self-regulated learning as a single, theoretical entity” (p. 288). While the teachers in this sample did not assess different facets of self-regulation independently, they also did not include irrelevant factors such
as verbal expressiveness in their ratings. Thus, this study provides limited evidence that teachers assess self-regulation in a holistic fashion but are able to separate self-regulation from confounding constructs.

Instead of using self or others to report on students’ self-regulation, it is possible to design specific tasks requiring self-regulation and observe students’ performance on the task. The academic diligence task (ADT) developed by Galla et al. (2014) is an example of this type of measure. The ADT asked participants (high-school seniors) to solve a series of single digit subtraction problems using an online software tool that has built in distractors. The distractors were Tetris and YouTube videos. The findings were that the time participants stayed on task before predicted both senior year GPA ($\beta = .07, p = .02$) and junior year standardized math scores ($\beta = .07, p = .01$). While these regression coefficients are statistically significant, they are low, and lower than other predictors in the model such as gender, race and attitude towards mathematics. Correlations with other measures of self-regulation were low. As an example, the strongest correlation ($r = .15, p < .01$) found in their data set was between time on task and grit (as measured by a self-report inventory). These results make it difficult to ascertain what the ADT is measuring. Further questions arise. For instance, is the self-regulation demonstrated by participants during the ADT typical of their ability to self-regulate? While the researchers attempted to motivate participants by telling them completing the ADT would enhance their intelligence, it is unknown whether this information served to motivate the participants to try their hardest. On a surface level, it is difficult to see why high-school students would be motivated to persist in performing simple subtraction problems. Thus, while the ADT, and other task-based measures of self-regulation allow researchers to observe self-regulation in situ, there remains a problem of generalizability. Not only is it not clear how well decontextualized
measures of self-regulation apply to complex environments such as classrooms, it is also unknown if the performance observed is indicative of that student’s typical performance over time. The applicability of task-based measures to classroom assessment is called into further question by the lack of evidence that teachers use specific tasks or assignments to assess self-regulation. Rather, it appears that secondary teachers assess self-regulation by observing students during normal day-to-day classroom interactions (Merchant, 2016).

Hadwin, Nesbit, Jamieson-Noel, Code, and Winne (2007) used trace data to solve the problem of a decontextualized task, and to provide fine-grained data related to students’ use of self-regulated learning strategies. The authors developed a software tool called gStudy that gave students an online, hyperlinked learning environment which presented information using text, graphics, and video. As students navigated the learning environment, the software collected information (referred to as “trace data”) about the strategies they used to learn the material. This information was stored in a log file and analyzed. Analysis included statistics such as frequency of use of different strategies, diversity of strategies used, and which strategies tended to precede or follow others. The trace data provided rich information about students’ study habits, and because the students had also completed the MSLQ, the authors were also able to compare students’ self-report of their study strategy use to their actual study strategies. The overall finding was that agreement between the two measures (what the authors refer to as “calibration”) was only 27 percent. While trace data can help answer important questions about students’ study strategies, there are some important questions about the utility of such data for teachers. Firstly, teachers are not in position to engage in the types of data analysis outlined by the authors, although it would be possible to have automated software tools analyze the data for teachers and provide summaries and possible interpretations. A more important limitations is that much
learning in a classroom does not take place in an online environment, and how trace data could be collected and analyzed in a face to face learning environment is not obvious, and therefore it is not clear how useful trace data would be in a classroom assessment context.

The measures of self-regulation described above give an indication of the difficulties of assessing self-regulation with acceptable levels of reliability and validity. Self-report inventories are quick, easy to score and tap into internal thought processes that may not be evident when observing behaviours. However, these inventories also have validity and reliability problems due to participants’ changing moods, lack of self-knowledge, and differing interpretations of the items (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Paulhus & Vazire, 2007). Interview protocols allow respondents to expand upon their answers and explore a range of responses beyond what was originally conceived by the researcher, but they are time consuming and still rely upon the respondent’s memory. With their better ability to control the environment, task-based assessments can provide reliable data; however, there is a question as to how well the results generalize, or how meaningful the results are in a school context. Trace data provides detailed, fine-grained data about students’ study habits, but collection and analysis are problematic in most classroom environments.

If researchers who are expert in assessment struggle to assess self-regulation, is it reasonable to expect teachers to create and administer quality assessments of self-regulation? The answer may be yes. While teachers lack psychological and assessment expertise, they have sustained interaction with their students, allowing them to offer a context-rich perspective that is missing from assessments of a single event such as think aloud processes or task-based measures. Thus, it is possible that teachers are in a privileged position to assess self-regulation (Zimmerman & Pons, 1988). In addition to seeing their students in a variety of contexts over
time, teachers are also exposed to a range of self-regulatory capabilities among their students, giving them a sense of how a student’s self-regulation compares with norms of the current classroom and the teacher’s prior classrooms. Further, teachers can assess self-regulation using a variety of tools. Nothing prevents teachers from using questionnaires or interview protocols with their students. This type of data could be supplemented with day-to-day observational data and student self-assessments, such as reflections or journals.

The concerns raised by some researchers (e.g., Hoge, & Butcher, 1984; Winne & Perry, 2000) about teachers’ abilities to assess self-regulation are worthy of consideration but these concerns are counterbalanced by research that indicates that teachers have the potential to be competent assessors of self-regulation (Klug, Bruder, Kelava, Spiel & Schmitz, 2013; Zimmerman & Martinez-Pons, 1988). Given that self-regulation is context dependent, and changing as students develop, it is reasonable to conclude that a single measure of self-regulation will likely be inadequate. Teachers are positioned to give students multiple assessments, of different types, over an extended period of time, potentially allowing them to develop a complete picture of a student’s self-regulation.

Method

The central research question for this study is how Ontario secondary teachers’ conceptions of self-regulation align with definitions of self-regulation found in the literature. This question was addressed by investigating teachers’ definitions of self-regulation, and what student behaviours they considered as evidence of self-regulation. A mixed methods approach was adopted, with three phases of data collection. The first phase used interviews with 26 high-school teachers from three school districts. These teachers had a broad range of teaching and subject area experience. The interview protocol focused on teachers’ assessment practices
surrounding the six learning skills and work habits, including self-regulation. Interview questions probed assessment practices, construct definitions, and grading decisions. Interviews were recorded and transcribed verbatim. These transcriptions were the same ones used for the study reported in Chapter 3, but only data related to how the participant defined self-regulation was included. These data were combined with responses from an online survey (described later in this section) that asked respondents to identify student behaviours they associate with self-regulation.

The second phase of data collection included a large set of report card grades from the school district where most of the interviews took place. This district is different from the two districts that provided the data for the study reported in Chapter 4. Each set of grades included the student’s achievement in a single course. A set of grades included the achievement grade, six LSWH grades, number of absences, and number of lates. Further, there were examination results from a standardized Grade 9 mathematics examination that all students in Ontario must write. The raw data included N = 26,076 sets of grades. Data cleansing reduced the sample to 22,962 sets of grades. Grade sets were eliminated when there was no grade for self-regulation within the set. This tended to occur if either no grade information was entered for the student or the academic grade was present but the LSWH grades were not. The deleted data did not match the retained data in a few important aspects. Firstly, 83.6% of the stricken data was from Grade 12, compared to 51.5% of the data that was retained. Secondly, 64.7% of the stricken data had a subject achievement grade that was either missing or zero. Ignoring the missing or zero achievement grades, the mean grade of the stricken data ($M = 76.1, SD = 15.72$) was significantly lower ($t = -3.32, p = .001$) than the mean grade of the retained data ($M = 77.9, SD = 14.55$). While statistically significant, the absolute difference was small.
The final phase of data collection used a 38-item online survey that examined how teachers valued and operationalized learning skills and work habits. The survey underwent two phases of development. Using data from the interviews, a set of approximately 100 potential survey items was developed. In consultation with a survey expert, items were selected and refined to create an initial draft of the survey. This draft contained 61 items. Twenty teachers trialled this initial draft of the survey. Based upon their feedback, and further consultation with a survey expert, the survey was shortened to 38 items. Further changes included edits to the wording of four items and reordering the items. The resulting final draft of the survey was deployed online (see Appendix B for quantitative items and results not reported in the body of this paper).

The intent of the survey was to investigate five dimensions of teachers’ LSWH assessment practice. These dimensions were valuation, operationalization, conceptualization, teacher practice, and policy. Valuation items measured how much respondents valued the LSWH, while the operationalization items investigated how the respondents assigned the grades. The conceptualization items were intended to measure respondents’ belief the LSWH represented distinct constructs, and the teacher practice items investigated what student behaviours they used to assess the LSWH. Finally, the policy items measured the influence of policy on respondents’ LSWH assessment practice. While the survey was designed to explore teachers’ assessment of all six LSWH, only results relating to the assessment and grading of self-regulation are reported here.

Survey participants were recruited through social media. This method of recruitment allowed for a broad sample in terms of geographical location, but also meant that many surveys were not usable. Of the 300 surveys received, only 108 were usable. A further 112 surveys were
completed by elementary teachers, and the remaining 80 surveys were not complete enough to incorporate into the statistical analyses. Survey results were analyzed using descriptive statistics for the quantitative items. The qualitative items were coded using the same themes that emerged from the interview data.

Results

Results are presented in two sections: teachers’ definitions of self-regulation, and how teachers assess self-regulation. As demonstrated in the reporting of these results, these sections are not independent, as teachers’ definitions very likely impact their methods of assessment.

Definition and Dimensionality of Self-Regulation

Data from all three phases of data collection were used to address the research questions. Two broad themes emerged from the data. First, teachers struggled to articulate a definition of self-regulation, and second, they viewed self-regulation as a construct that overlapped the other five LSWHs they assessed. As one example of this struggle for articulation, a beginning history teacher directly stated, “I don’t understand self-regulation,” and went on to explain that this lack of understanding made it difficult for him to assess it. He further explained this was because when it comes to assessing self-regulation, “I don’t know what I am looking for.” Similarly, a mid-career English teacher also reported not being able to define self-regulation, but this was not an issue for him as he placed very little value in these assessments.

The second broad theme encompassed the belief amongst many of the teachers that self-regulation was the foundational construct upon which the other five LSWH rested. As one special-education teacher noted, self-regulation “overlaps with almost all of them [the LSWH] because if a student can’t self-regulate most of the other [LSWH] suffer.” A mathematics teacher who described self-regulation as the “toughest” LSWH to assess characterized self-regulation as
“almost the cumulation of them [the LSWH]. It’s the summative of the other 5 [LSWH].” A geography teacher viewed self-regulation as “compliance-based” and expressed the view that students who scored high on self-regulation were complying with teachers’ wishes and expectations, and therefore likely to score high on all the LSWH.

During the interviews most teachers reported that self-regulation overlapped with the other five LSWH, but the exact nature of that overlap depended on the teacher. Across the 26 participants there was no consensus on where the overlap lay. For example, three teachers saw self-regulation and responsibility as overlapping constructs. An experienced technical education teacher stated that,

Self-regulation definitely ties into responsibility. They work hand in hand. If you are self-regulating, you have, usually, huge responsibility, you know, you’re a person of responsible practices. You’re cleaning up, you’re doing this, you’re doing that, and you’re cleaning your area.

A mid-career Chemistry teacher also saw self-regulation as tied to responsibility. When describing what LSWH late assignments would impact, he responded with, “maybe it is a responsibility thing, maybe it is a self-regulation thing.” An English teacher also tied missing or late work to responsibility and self-regulation. When students did not submit formative work (e.g., rough drafts of essays) she would lower their self-regulation grade, and when they did not submit summative work, she lowered the responsibility grade. This teacher also reported that the boundary between self-regulation and initiative was “blurry”.

Blurred boundaries between self-regulation and the other five LSWH existed for many teachers. A civics teacher saw self-regulation as overlapping with initiative, and a late career
business studies teacher linked self-regulation with independent work. An English teacher who was close to retirement described self-regulation using the same student behaviours he used to describe initiative and organization. Another English teacher reported making no attempt to distinguish among the six LSWH. When completing report cards, he assigned students the same grade across all six LSWH.

Seven of the teachers interviewed used the sample behaviours given by the Ministry of Education to help them define self-regulation. Only two of these teachers reported using the Ministry of Education assessment policy documents to help them define self-regulation, while the remaining five teachers independently linked self-regulation to goal-setting, metacognition, or goal achievement. Two teachers associated seeking help with self-regulation, and another two teachers incorporated reflection into their construct of self-regulation. Thus, while there were instances in which teachers’ concept of self-regulation aligned with controlling cognitive and emotional processes to achieve a desired goal, this was not the dominant view.

There were two teachers who did not provide any definition of self-regulation or give examples of student behaviours that served as indicators of self-regulation. Seventeen teachers in the interview sample described self-regulation in terms of students complying with behavioural and academic expectations set by the teacher. Examples included not distracting others, submitting work in a timely manner, and understanding when you need to take a break. To illustrate, a mid-career humanities teacher said that, “self-regulation is knowing that I really am feeling antsy right now, and maybe I should just ask to use the washroom and say Mr. X can I go for a walk?” A mid-career civics teacher linked self-regulation with coming to class regularly, on time, and ready to learn. He asked himself “Are they [students] regulating their behaviour, or are they coming to class stoned?” An English teacher thought that “part of self-regulation is that they
are respecting the rights of others in the classroom,” and that it needed to be assessed “through their behaviours, and what they say and do.” A different English teacher noted that girls tend to achieve higher self-regulation grades because some girls have a mindset of “I am a follower, I am a pleaser, I need to do everything the way it is expected.” One other teacher in the interview sample mentioned that girls achieve higher self-regulation grades than boys.

The report card data did demonstrate that girls receive higher self-regulation grades than boys. In Grade 9, the mean self-regulation grade for girls \((M = 3.42, SD = 0.84)\) was higher than for boys \((M = 3.10, SD = 0.94)\). This difference was statistically significant \((t(11,130) = -19.08, p < .001)\). The Grade 12 data results were similar with girls’ self-regulation grades \((M = 3.40, SD = 0.86)\) being higher than boys’ \((M = 3.09, SD = 0.98)\). Again, t-tests revealed the difference to be statistically significant \((t(11,816) = -18.62, p < .001)\). The effect sizes for gender differences in Grade 9 \((d = .36)\), and in Grade 12 \((d = .34)\) were both small (Cohen, 1992). The report card data also demonstrated that girls receive higher achievement grades than boys both in Grade 9 \((t(10,980) = -16.4, p < .001)\) and Grade 12 \((t(11,812) = -16.6, p < .001)\). To test whether gender differences in self-regulation grades remained after controlling for academic achievement linear regression was used. Self-regulation was the dependent variable and the achievement grade and gender were used as predictor variables. When the achievement grade was entered into the model first, it was found that gender nearly disappeared as a predictor variable. The standardized coefficient for the achievement grade \((\beta = 0.650, p < .001)\) was much higher than for gender \((\beta = 0.079, p < .001)\), and the inclusion of gender improved the variance accounted for by only 0.6%. Collinearity diagnostics (tolerance = 0.98, VFI = 1.02) were found to be well below problematic levels (Field, 2013).
This view of self-regulation as compliance aligns with findings from the survey data which showed that teachers associate productive classroom behaviours with high self-regulation. One survey item used branching logic within the survey so that when respondents selected self-regulation as the most valuable LSWH, the next question asked them why. Twenty-two of 100 respondents selected self-regulation as the LSWH most important for success in their subject area. Of these respondents, 18 further explained why. The dominant theme from these responses was that self-regulated behaviour allowed students to learn better in the classroom. One teacher wrote that students “cannot address academic deficits until [they] can control behaviour enough to engage in class activities,” and another wrote that self-regulation was central to students’ abilities “to focus and work well with others.” The notion of focus or staying on task was put forward by half the respondents to this item. Only four of the 18 responses related to the sample behaviours of self-regulation put forward by the Ministry of Education. All four of those responses addressed metacognition and reflection.

Some survey respondents commented on the difficulty of defining self-regulation. One teacher wrote, “It [self-regulation] is hard to define and generally overlaps with most other learning skills. Many teachers don’t seem to quite understand what it is to begin with, how can they assess it?” Another commented that “definitions of it vary.” Relevant to these two comments, one survey item asked directly if respondents thought their colleagues shared the same definition of self-regulation. Thirty-nine of 75 agreed that they shared the same definition of self-regulation as their colleagues, and only 19 disagreed (the remaining 14 answered “neither agree nor disagree”). Other items designed to inform how teachers define self-regulation included a more general item that asked if the six LSWH represented distinct constructs. Of the
75 respondents, 34 agreed the six LSWH were distinct from each other, and 40 disagreed (one respondent answered “neither agree nor disagree”).

Report card data cannot directly address the question of how teachers define self-regulation, but exploratory factor analysis was used to examine the extent to which the six LSWH were assessed as separate constructs. Because a Kolmogorov-Smirnov test demonstrated the distribution of LSWH grades was not normal \( D(22,904) = 0.30, p < .001 \) for self-regulation, principal axis factoring was determined to be the best algorithm to extract the factors (Fabrigar & Wegener, 2011). The factor analysis revealed that the six LSWH grades could be represented by a single factor that accounted for 85.25% of the variance. Self-regulation and initiative had the highest factor loadings, with values of 0.92 for both, but the other four LSWH also had high loadings that ranged between 0.87 and 0.91. While exact factor loadings changed slightly, the overall factor structure was invariant across grades and genders. These results suggest that across this district secondary teachers were not assessing self-regulation independently from the other five LSWH. This finding is consistent with the interview data and survey findings that many teachers find self-regulation difficult to define distinctly from the other LSWH.

Assessing Self-Regulation

All the interview participants reported following district and Ministry of Education policies when assessing academic achievement. Every teacher used the results of specific, summative assessments to determine academic achievement grades. These assessments were identified to students on the course outline and marking rubrics for the summative assessments were public. Further, all teachers in the sample recorded grades in a gradebook and used the recorded grades to determine the final grade. This consistent approach to grading, and adherence to policy, was not found in teachers’ assessments of self-regulation. Part of the reason for this
may have been the difficulty in defining the construct. The interview and survey data both suggest Ontario secondary teachers find self-regulation difficult to assess. During her interview, an early career business studies teacher described assessing self-regulation as “really tough” while a beginning history teacher echoed this sentiment saying, “in terms of assessing, I find that one [self-regulation] a bit trickier.” When I asked a mid-career geography teacher how she assessed self-regulation she responded by saying she “would have to read the little bullets underneath” to remind herself of what behaviours should be considered when assessing self-regulation. An English teacher also found self-regulation difficult to assess, and so decided to tie the self-regulation grade to unauthorized use of mobile phones in the classroom. By using this narrow definition of self-regulation, the teacher was able to create a construct that was easy for him, and his students, to understand.

To test whether Ontario secondary teachers found self-regulation more difficult to assess than other LSWH, a survey item asked teachers to rank the LSWH in order of easiest to assess. Self-regulation was rated as more difficult to assess than independent work, \( t(75) = 3.25, p < .01 \), but was not significantly different from the other four LSWH. Respondents who selected self-regulation as the most difficult LSWH to assess were given a follow-up question asking why it was so difficult to assess. All 14 persons who selected self-regulation as most difficult to assess responded to this item. Five responses focused on the difficulty of observing self-regulation as not only is it a latent construct, but it can manifest itself in different ways in different contexts. One teacher noted that “it is really hard to see that it [self-regulation] is actually happening” and another commented that “it is hard to know if a student is lazy or exhausted from working five nights a week.” A further five responses centered on the difficulty of finding time to assess self-regulation. One teacher wrote, “It is difficult to make time to hear from students about their
thinking around their learning, particularly in groups that aren't able to write much/well.” In addition, two teachers responded that they struggled to assess self-regulation because they did not have a good, actionable definition of self-regulation.

To help them collect data on self-regulation some teachers reported using student self-assessment. Half (13 of 26) of the interviewed teachers said they used student self-assessment as part of their LSWH assessment practice. These teachers described using student self-assessment as a way of assessing the student’s metacognitive abilities, and of seeing how well the student’s self-assessment aligned with the teacher’s assessment. Self-assessment served as an opportunity for teachers to engage in dialogue with their students about their LSWH, but it did not enter into the grading decisions. No interview participant reported using self-assessments to impact the LSWH grade. Mismatches between the teacher’s assessment and the student’s self-assessment were seen as evidence of a student’s distorted self-image, and not as a potential misjudgement by the teacher.

Survey respondents also reported using student self-assessment to collect data about students’ self-regulation, but the percentage was much higher. Sixty-two of 76 respondents (82%) indicated they had students self-assess their LSWH, a proportion that was significantly higher than that of interview respondents ($\chi^2(1) = 9.93, p = .002$).

To better understand the behaviours teachers consider when assessing self-regulation, a list of typical classroom behaviours was created. This list was created based upon behaviours mentioned by teachers during the interview phase. Survey respondents were asked to indicate what LSWH they associated with different classroom behaviours and could choose any number of the LSWH to be associated with each behaviour. Thus, a respondent could indicate that a behaviour impacted zero, one, two, or even all six LSWH. Results from this item are shown in
Table 1. As can be seen from the table, the top four behaviours teachers associated with self-regulation were disruptive behaviours, suggesting teachers were using negative behaviours to identify the lack of self-regulation.

Table 5.1. Proportion of respondents who associated each behaviour with the individual LSWH

<table>
<thead>
<tr>
<th>Behaviour</th>
<th>Self-Regulation</th>
<th>Organization</th>
<th>Responsibility</th>
<th>Initiative</th>
<th>Independent Work</th>
<th>Collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texting in class</td>
<td>0.83</td>
<td>0.10</td>
<td>0.49</td>
<td>0.13</td>
<td>0.27</td>
<td>0.14</td>
</tr>
<tr>
<td>Talking out of turn</td>
<td>0.83</td>
<td>0.04</td>
<td>0.23</td>
<td>0.00</td>
<td>0.07</td>
<td>0.39</td>
</tr>
<tr>
<td>Distracting another student</td>
<td>0.77</td>
<td>0.04</td>
<td>0.40</td>
<td>0.07</td>
<td>0.27</td>
<td>0.34</td>
</tr>
<tr>
<td>Arriving late to class</td>
<td>0.56</td>
<td>0.46</td>
<td>0.80</td>
<td>0.19</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>Being polite towards the teacher</td>
<td>0.54</td>
<td>0.01</td>
<td>0.37</td>
<td>0.04</td>
<td>0.04</td>
<td>0.26</td>
</tr>
<tr>
<td>Not completing an in-class assignment</td>
<td>0.47</td>
<td>0.43</td>
<td>0.73</td>
<td>0.31</td>
<td>0.71</td>
<td>0.13</td>
</tr>
<tr>
<td>Complaining about working in a group</td>
<td>0.46</td>
<td>0.04</td>
<td>0.30</td>
<td>0.08</td>
<td>0.08</td>
<td>0.85</td>
</tr>
<tr>
<td>Constantly asking for help</td>
<td>0.41</td>
<td>0.17</td>
<td>0.34</td>
<td>0.43</td>
<td>0.63</td>
<td>0.06</td>
</tr>
<tr>
<td>Submitting Homework on time</td>
<td>0.41</td>
<td>0.61</td>
<td>0.93</td>
<td>0.38</td>
<td>0.44</td>
<td>0.04</td>
</tr>
<tr>
<td>Refusing to work with another student in a pair group</td>
<td>0.37</td>
<td>0.04</td>
<td>0.32</td>
<td>0.11</td>
<td>0.03</td>
<td>0.94</td>
</tr>
<tr>
<td>Seeking help after class</td>
<td>0.31</td>
<td>0.17</td>
<td>0.67</td>
<td>0.92</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>Not bringing required supplies to class</td>
<td>0.28</td>
<td>0.66</td>
<td>0.76</td>
<td>0.18</td>
<td>0.20</td>
<td>0.08</td>
</tr>
<tr>
<td>Asking a relevant question</td>
<td>0.18</td>
<td>0.00</td>
<td>0.32</td>
<td>0.92</td>
<td>0.10</td>
<td>0.24</td>
</tr>
<tr>
<td>Helping a fellow student</td>
<td>0.10</td>
<td>0.03</td>
<td>0.34</td>
<td>0.62</td>
<td>0.00</td>
<td>0.96</td>
</tr>
</tbody>
</table>
It can also be seen that behaviours such as seeking help and asking questions were selected by few teachers, even though these behaviours align with the Ontario Ministry of Education’s suggested behaviours such as seeking assistance or clarification when needed.

The interview participants fell into three groups when describing what behaviours would impact students’ self-regulation grades. Eleven teachers were placed into the first group. These teachers struggled to define what behaviours impacted self-regulation grades. This difficulty stemmed from not being able to clearly define or distinguish self-regulation from the other five LSWH. The second group of nine teachers used policy documents to determine which behaviours they would consider when assessing self-regulation. These teachers used behaviours such as time management, goal-setting, and planning as part of their assessment process. The third group of six teachers linked self-regulation with good behaviour in the classroom. A technical education teacher in this group linked self-regulation negatively with “horseplay” in the shop, and a mathematics teacher saw it as “behaviour management.” Other behaviours teachers in this group considered included not distracting peers, and “respecting the rights of others.”

The report card data had few indicators of student behaviour, but attendance data provided an opportunity to test the association between self-regulation grades, student absences, and tardiness. More than half the survey respondents indicated they associated coming late to class with self-regulation, although few of the interview participants mentioned this behaviour directly. The Pearson correlation coefficient between the self-regulation grade and the number of absences was low but statistically significant ($r = -0.28, p < .001$). This was similar to the value obtained between the self-regulation grade and number of times late to class ($r = -0.23, p < .001$). These results come from an aggregate of the Grade 9 and 12 data.
During interview, nine teachers reported using academic achievement to inform their LSWH grades, and another three stated they directly linked the LSWH grades to academic achievement. On the survey, 66 of the 75 survey respondents (88%) agreed that poor academic performance (below 50% achievement) makes it likely that a student would receive a grade of “needs improvement” for at least one of the LSWH. At the other end of the achievement scale, 28 of 74 teachers (38%) agreed that achievement above 80% made them reluctant to award “needs improvement” or “satisfactory” on any LSWH. These data suggest that many teachers conflate academic achievement and self-regulation in their grading decisions. To further test this hypothesis, the strength of association between self-regulation grades and academic achievement was calculated. The Pearson correlation coefficient was moderately high and statistically significant ($r = 0.65, p < .001$). This coefficient is an aggregate that includes both Grade 9 and Grade 12 students and all subject areas.

Certainly, these correlations could be further evidence of conflation; however, they could also represent an important relationship between positive self-regulation and higher academic achievement. To further test if teachers conflate academic achievement and self-regulation, the associations amongst self-regulation grades, teacher awarded achievement grades, and standardized examination scores were investigated. If the correlation between teacher awarded achievement and self-regulation grades was solely due to better learning, the correlation should be equally strong between self-regulation grades and the standardized examination score. Since the only suitable standardized examination that Ontario secondary students write is in Grade 9 mathematics, the analysis was restricted to this course. The sample for this analysis was $N = 736$. The correlation between the self-regulation grade and teacher awarded achievement in Grade 9 mathematics ($r = 0.69, p < .001$) was similar to the overall result reported above, while the
correlation with the standardized mathematics examination was lower ($r = 0.48$, $p < .001$). The difference between the two coefficients was found to be statistically significant using Fisher’s $r$ to $z$ transformation ($z = 6.11$, $p < .001$). This provides some evidence that teachers may be conflating achievement and self-regulation when awarding grades.

Discussion

The finding that teachers award girls higher self-regulation grades than boys is consistent with other studies (e.g., Chapple, Vaske, & Hope, 2010; Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006), although the effect sizes reported here appear to be lower than what other researchers have found. As an example, Duckworth et al. (2015) reported an effect size ($d = 0.61$) that is double what is reported here. This large difference in effect size may be because teachers in this study assessed self-regulation using their own methods, whereas in Duckworth et al. teachers assessed self-control using a standardized instrument. Further Duckworth et al. conducted their study at a prestigious private school and had a much smaller sample size.

The data here do not explain what the underlying causes of this gender difference may be. There is evidence from neuroscience that during the teen years, girls’ and boys’ brains develop differently, and that marked areas of difference include the prefrontal cortex and inferior frontal gyrus—both areas are involved with executive function and social cognition (Blakemore & Choudhury, 2006; De Bellis et al., 2001), and are therefore likely to influence self-regulation. Another possibility comes from interview data which demonstrated that some teachers allow a student’s academic achievement grades to influence the LSWH grades. In this sample, girls have higher academic achievement grades than boys, and it is possible they are receiving higher self-regulation grades as a result. This interpretation is supported by the linear regression results demonstrating the gender difference in self-regulation grades almost disappears when academic
achievement is controlled. Another possibility is that girls are more compliant than boys. There is some evidence for this (e.g., Carlo & Randall, 2002; De Wied, Branje, & Meeus, 2007; Eisenberg, Hofer, Sulik, & Liew, 2014), and as many teachers appear to be defining self-regulation in terms of compliant behaviour, it follows that girls would receive higher self-regulation grades than boys. A better understanding of the underlying causes in gender differences in self-regulation grades could be useful for better understanding what future actions and interventions may help close the achievement gap between boys and girls in high school.

The three phases of data collection provide a consistent, if somewhat blurry, picture of how Ontario secondary teachers define, assess, and grade self-regulation. Ontario secondary teachers have differing definitions of self-regulation, but the most common conception of a self-regulated student is one who demonstrates productive, compliant classroom behaviours. This conception of self-regulation most closely matches the definition provided by Kopp (1982) and is less consistent with more recent definitions that have a stronger focus on goal setting, and acting to achieve that goal (e.g., Carver & Schreier, 2014; Diamond, 2013). This is despite advice from the Ontario Ministry of Education (2010) assessment guide, and from Ontario based researchers such as Shanker (2010). The source of the disagreement may be that teachers are associating self-regulation with student behaviours that are most helpful to creating a classroom environment that teachers believe is most conducive to teaching and learning. Teachers seem to prefer orderly, disciplined classrooms, in which students submit assignments on time, and the students work productively together during class. While such a classroom is a reasonable goal for teachers to achieve, it is not a goal that originates from a student perspective. By contrast, current conceptions of self-regulation have the individual student’s goals as their foundation. A student may be self-regulated but have goals that are at odds with the teacher’s expectations. The
class clown might be one example, where the student has goals of being liked, and having peers laugh at their antics, and takes appropriate steps to achieve those goals. If these student goals run counter to those of the teacher, it seems likely the teacher would view the student’s behaviour as evidence of poor self-regulation.

The small number of teachers using the sample behaviours provided by the Ontario Ministry of Education (2010) to anchor their definition of self-regulation is of importance for two reasons. The first is that the sample behaviours provide a framework for teachers to use in assessing self-regulation. Consistent use of this framework could potentially lead to more consistent definitions of the construct across teachers and better clarity for students, parents, and other stakeholders on how to interpret the self-regulation grades. The second reason is that it indicates teachers are either unfamiliar with LSWH assessment policies, or do not follow these policies closely. This is very different from how interview participants described their assessment of academic achievement. Teachers described their assessment of achievement as being strongly aligned with district and provincial policy. All interview participants reported adhering to assessment policies such as explicitly stating which assignments were summative assessments at the beginning of the year, not deducting marks for late work, and giving students rubrics with the assessments.

The current data cannot be used to determine why teachers follow assessment policies for academic achievement more closely than for LSWH. It is reasonable to hypothesize that one possible cause is the lower emphasis placed on LSWH assessment by school and district administrators. Only one interviewed teacher recalled being questioned about her LSWH grades by an administrator (the teacher happened to teach the child of that administrator, and the questioning took place during a parent-teacher interview). Further, only two teachers reported
being given formal training on assessing LSWH, although a further five had (with administrator support) started their own professional learning community surrounding the assessment of LSWH.

Another possible reason teachers are not following assessment policies related to LSWH is these grades are not used for high stakes decisions. This was mentioned by 17 of the 26 interview participants. As an example, many Ontario universities consider skills like self-regulation when making admissions decisions, but they do not use the LSWH grades on report cards as a source of information. Instead, they ask teachers to assess and report upon these skills through reference letters. In short, the LSWH grades do not appear to be used for any purpose. This practice, coupled with little to no accountability for this portion of their assessment practice, may account for teachers’ willingness to depart from policy guidelines. Further research would be needed to ascertain whether this was true.

Assessment experts have repeatedly called for classroom assessment policies that allow grades to be “pure” measures of achievement (e.g. Brookhart, 1994; Guskey, 2006; Marzano & Heflebower, 2011A). This call has spurred the separate reporting of non-achievement factors, such as student self-regulation, from academic achievement. While the rationale behind separating achievement and non-achievement factors is clear, teachers have thus far been given little direction or advice on how to effectively assess and grade non-achievement constructs such as self-regulation. The results presented here indicate the existence of a few pages in a large assessment policy document is not sufficient to counteract the lack of training teachers have received on how to assess complex constructs such as self-regulation. This implies that improving classroom assessments of self-regulation will require teachers to be given support and training that goes beyond a single policy document. This support and training should focus on
defining self-regulation, and what classroom behaviours will be used to assess it. Further, there should be a consensus on what separates the different levels of achievement. Given what we know about teachers’ professional learning, a single day workshop will not be sufficient to improve teachers’ practice, although it may initiate some important conversations (Garet, Porter, Desimone, Birman, & Yoon, 2001). Improving teachers’ classroom assessment of self-regulation will require not only input and training from experts, but also leadership from school administrators. Fortunately, there is evidence that training can result in teachers improving their assessments of learning skills (Klug et al., 2013), but how do we implement such training? One suggestion is to have school administrators identify teachers who are strong at assessing self-regulation and use them as leaders within the school, as this has been shown to be effective in improving teacher practice (Guskey, 2003). Unfortunately, this suggestion is predicated on the assumption that school administrators are able to identify effective self-regulation assessment practices, and there is no evidence this is the case.

Summary

Self-regulation is a construct which has many definitions within the research literature. Just as researchers in psychology and education have struggled to arrive at a shared definition of self-regulation, so too have teachers. While some teachers are direct in stating they do not understand what self-regulation is, it appears from interview and survey data that Ontario secondary teachers generally associate self-regulation with compliant, productive behaviours in the classroom. Teachers use a variety of classroom behaviours to assess self-regulation, but these behaviours are not consistent across teachers. Interview and survey data also revealed that many teachers incorporate student self-assessment into their self-regulation assessment practice. Self-report is a popular form of assessing self-regulation among researchers, as it potentially yields
information on latent thought processes (Winne & Perry, 2000), but the results from interviews indicate that teachers do not incorporate self-assessment information into their grading decisions.

Given the different definitions of self-regulation, the high correlation of self-regulation grades with other LSWH grades, and the variety of behaviours and assessment practices used to inform self-regulation grades, it becomes difficult to ascertain what these grades represent. This makes using self-regulation grades problematic for formative purposes, such as helping to improve students’ self-regulation, or summative purposes, such as informing admission into special programs or post-secondary institutions.

Improving teachers’ assessment practice will require strong direction both from administration at all levels (school, board, and ministry), and from educational researchers. Current research into classroom assessment and teacher grading focuses on subject area achievement and not on other constructs that teachers may be expected to assess such as effort, participation, or collaboration. Research into how teachers can better operationalize these constructs is necessary for developing policies and practices that will allow classroom assessment and reporting practices to provide meaningful, actionable information.
CHAPTER 6: GENERAL DISCUSSION

This chapter serves to provide a broader summary and analysis of the research questions first presented in Chapter 1 and examined in Chapters 3, 4, and 5 supported by the literature reviewed in Chapter 2. I will briefly summarize the findings from the three studies presented earlier and examine how the three studies relate to the four research questions introduced in Chapter 1. Broad themes that span the studies will be noted, along with implications of the work and potentially important areas of further study.

Classroom assessment and teacher grading practices are well studied areas of educational research (Brookhart et al., 2016). However, prior research in these areas has focused almost exclusively on assessment and grading of subject area achievement. This is perhaps not surprising given the important role academic grades play in students’ program decisions, future opportunities, and self-concept (e.g., Arens et. al., 2017; Newton, 2007). However, Learning Skills and Work Habits (LSWH; or similar skills) have been assessed and reported upon in Canadian classrooms for decades, yet I could not find research that directly addressed how they are assessed. This is an unfortunate and important gap in the research, especially given the current interest in these skills as highly valued traits. Not only have business groups such as the Conference Board of Canada (2015) and education advocates (e.g., People for Education, 2016) called on schools to make developing these skills a priority, but there is extensive evidence from economics and educational psychology that skills like the LSWH are positively associated with myriad benefits including better learning, increased happiness, and better employment prospects (Almlund, Duckworth, Heckman, & Kautz, 2011; Duckworth & Seligman, 2005; Galla & Duckworth, 2015; Heckman & Kautz, 2012).
My research begins to address this gap in the classroom assessment literature. Ontario is an ideal place to undertake research on how teachers assess non-achievement constructs. The six specific LSWH that Ontario K-12 teachers must assess have been in place since 2010, and a similar set of skills was present before then. Thus Ontario educators have had time to develop and refine assessment practices related to these constructs, and school systems within the province have had ample opportunity to develop policies, training, and supports to help teachers complete these assessments. As first presented in Chapter 1, my research focused on four overarching research questions:

1. How are Ontario secondary teachers defining each of the six LSWH? What student behaviours do they use to assess them?
2. To what extent are the LSWH assessed independently of each other?
3. How strongly are the LSWH associated with academic achievement?
4. How do teachers’ conceptions of the LSWH align with established definitions of self-regulation?

Three separate studies were conducted to address these research questions. These three studies are reported as distinct manuscripts in Chapters 3, 4, and 5 and they begin to address the lack of research on the measurement of other skills beyond academic achievement in secondary schooling. Each of these studies makes a unique contribution to the research literature. The first study provided qualitative baseline data to inform the research design and analyses for the other two studies. The 26 interviews completed for this initial study yielded a rich dataset that revealed a breadth of approaches, beliefs, and practices among the participants. The second study used report card data from two school districts to examine the dimensionality of the six LSWH grades, along with their relationship to achievement. The third study focused on a single LSWH, self-
regulation. Self-regulation was chosen because it is the LSWH with the strongest research base and theoretical foundation. The data for this manuscript included interview data from the first study, report card data from a third school district, and an online survey designed to investigate Ontario teachers’ LSWH assessment practices. Combined, the findings from these three studies provide important insights into the larger research questions of the study. Each of these questions will be first addressed individually and then broader themes and implications will be discussed.

RQ1 – Teachers’ definitions of LSWH

The results presented in Chapters 3 and 5 demonstrate that Ontario secondary teachers are not defining the LSWH consistently, nor are they consistent in selecting which student behaviours they associate with each LSWH. As examples, both the interview data and the survey data demonstrate that teachers consider specific behaviours such as texting in class, or not completing an in-class assignment, as indicators of more than one LSWH. The lack of consistent definitions or the operationalization of each LSWH may have roots in teachers not having solid, internalized conceptions of the LSWH. As mentioned in Chapter 3, fewer than half the interview participants could name all six LSWH. Without a concrete understanding of the components of each LSWH or of the LSWH themselves, it seems the teachers are relying on general impressions of the student. Supporting this conclusion is the finding that independent of how much assessment information was collected, all the interviewed teachers reported using informal observation as their primary LSWH assessment tool.

Interview participants provided few examples of specific activities to assess LSWH, and none of the reported activities were designed as an assessment of LSWH. Instead, they were designed as assessments of curriculum expectations, and teachers reported they also found them useful for assessing one or more LSWH. There are different possible explanations for this
finding. Teachers may not have a strong conceptual understanding of the LSWH, and this lack of understanding of the underlying constructs likely hinders their efforts to assess these skills. This may explain why teachers reported using activities they designed as assessments of subject area achievement as LSWH assessments. Alternatively, teachers may not believe the LSWH can be separated from the curricular expectations, and so use assessments of the curriculum also as opportunities to assess LSWH. Lastly, teachers may consider the LSWH as components of a more global construct in which the individual components are not easily distinguished but the overall construct itself can be assessed holistically through continued observation. Support for each of these explanations was found in the data. Of importance, the inconsistent manner in which Ontario secondary teachers operationalize the LSWH highlights a critical finding for my research. This finding also largely parallels what other grading researchers have found with respect to teachers’ grading of academic achievement. Brookhart (1991) used the term “hodgepodge” (p. 36) to describe teachers’ grading practices, while Nash and McMillan (2000) used the term “flexible” (p. 9). While there has been much focus on what components of learning and achievement should be included in grades since that time, my findings demonstrate little may have changed—teachers’ grading practices are not consistent.

While inconsistencies were found in terms of both methods of assessment and conceptions of LSWH, the most commonly reported method of observation to assess LSWH may have merit. Despite the ad hoc nature of incidental observation, there is some support for it as an assessment method. Gardner (1992) wrote that “assessment ought to become part of the natural learning environment” (p. 90), and Merrell (2001) argued that naturalistic behavioural observation should be the primary strategy for assessing social skills in children and adolescents. Further support for incidental observation could come from future performance ratings that
secondary students may face in the future. As an example, prior research has shown that raters of workplace-based assessments demonstrate similar diversity of philosophy and practice as reported by the teachers in my first study (Govaerts, Van de Wiel, Schuwirth, Van der Vleuten, & Muijtjens, 2013). If skills such as the LSWH are assessed holistically in employment contexts, this may provide a justification for high school teachers to also assess them in this way. In addition, there is a reasonable argument to be made that incidental observations made over time give teachers a stronger sense of students’ typical performance on the LSWH, including indicators of the range of student performance on these skills. Specific, identified LSWH assessment activities may result in assessing maximal performance, or minimal performance, or any level in between, giving a limited, and potentially incorrect, measurement of a student’s ability to collaborate, self-regulate or be organized. Assessments of typical performance over time likely best represent the “habit” portion of LSWH and may provide more important information than assessments of maximal performance. Galla and Duckworth (2015) argued that because habits do not require effortful control, they are easier to maintain. This is an important point because when a student behaviour is habitual, it is not only likely this behaviour will continue in the future, but also likely the behaviour will be present even when the student’s emotional state and motivation are changing. Thus incidental observations over time may have the potential to be better predictors of future LSWH performance than single measures such as a questionnaire or task-based measure. Ongoing observations over time also allow for assessment of LSWH in both high-stakes and low-stakes situations, and for teachers to report what is a one-off or unusual behaviour for a student, and what is normal.

Counter to arguments in support of incidental observation as an assessment tool are calls for classroom assessment tasks to be clear in purpose and varied in design (e.g., Klinger,
McDivitt, Howard, Munoz, Rogers, & Wylie, 2015). In addition, in cases of conflict (e.g., with parents), ratings based upon incidental observation would be difficult to justify. This leads to the question of whether Ontario secondary teachers should be creating specific, varied assessment tasks for the LSWH, or if it is reasonable for teachers to rely so heavily on incidental observations. As an example, three of the teachers interviewed (see Chapter 3) noted that negative behaviours or incidents tend to stand out, and therefore exert undue influence on LSWH grades. Unless observations are structured, it becomes very difficult for teachers to know if they are only recording behaviours that are atypical, and therefore notable, or are recording a representative sample of students’ behaviours. This threat to the validity of the interpretations of the LSWH assessments made through incidental observations is something that must be addressed if these assessments are to be useful.

While informal observation may have some advantages as an assessment tool, it should be supplemented with additional assessment methods. From the self-regulation literature, we see different assessment methods applied to self-regulation including structured observation, interviews, and self-report measures. Those trying to measure collaboration also use self-report and structured observations, and often include peer-report measures too (Greenwald & Zukoski, 2018). These different assessment methods could all be used by teachers to give a more complete picture of students’ LSWH. Such a strategy is aligned with the Ontario Ministry of Education’s (2010) requirement that assessments should be “varied in nature” (p. 6). The high prevalence of self-assessment reported by teachers in Study 1 and Study 3 (Chapters 3 and 5) suggests teachers are willing to include diverse methods of LSWH assessment into their practice, especially when they see educational value in those assessments.
RQ2 – The Independence of the LSWH Constructs

The second research question focused on the extent to which the LSWH are assessed as independent constructs. All three studies provided findings relevant to this question, although Study 2 most directly addressed this question. The interview participants reported they struggled to distinguish among the six LSWH, and this finding was consistent with the survey results, although a significant minority of the survey respondents believed the six LSWH were distinct constructs. As a specific example, one of the survey questions read “the six different learning skills and work habits are distinct from each other. i.e. there is little overlap.” Of the 72 respondents who expressed an opinion, 40 disagreed to some degree and 32 agreed. Survey respondents were also very confident they could explain the difference between organization and responsibility to students, with a mean score of 6.13 (out of 7) on that item. In contrast, 12 of the 26 interview participants initially reported being able to distinguish between responsibility and organization, but when asked to make those distinctions, only 6 could. Four participants recanted their initial answer, and two gave further responses that did not clarify the differences between the two constructs. It seems likely the differences in responses between interview participants and survey respondents was due to the interviews allowing deeper exploration of teachers’ conceptions of the LSWH, highlighting challenges not easily identified through the survey.

Teachers’ difficulties in assessing the six LSWH as independent constructs was best exemplified by the report card data presented in Study 2. Perhaps the most striking finding from the report card data is that approximately half of the students received the same grade across all six LSWH in both districts. Exploratory Factor Analysis of LSWH grades from one district identified that those grades were best represented as a unidimensional construct, and this was confirmed for a second district using CFA. Further, this factor structure was consistent across
gender, grade, and subject area. LSWH grades from a third district (reported in Study 3) also had a unidimensional factor structure. This single factor accounted for over 80% of the variance in LSWH grades for both districts in Study 2, and over 85% of the variance for the district in Study 3.

These results are consistent with Ferrito’s (2015) finding that a unidimensional factor structure best modeled Grade 4 teachers’ ratings of seven different aspects of social-emotional learning. Much earlier research by Grolnick and Ryan (1989) asked Grade 3 and 6 teachers to rate students’ “academic competence.” Academic competence was operationalized as a mixture of achievement, motivation, and the ability to work independently. Teachers’ ratings on the scale were found to be unidimensional. More recently, Matthews, Ponitz, and Morrison (2009) used the Childhood Behaviour Rating Scale as a teacher-report instrument in assessing the self-regulation skills of Kindergarten children. The authors of this scale claimed a two-dimensional factor structure (Bronson, Goodson, Layzer, & Love, 1990), but when used by the teachers in the Matthews, Ponitz, and Morrison study, the ratings were unidimensional. Similarly, high school teacher ratings of four dimensions of collaboration were found to be unidimensional (Wang, MacCann, Zhuang, Liu, & Roberts, 2009). These prior studies, together with my findings, lead to the question of how well teachers assess different skills as separate constructs. Zimmerman and Martinez-Pons (1988) suggest that teachers’ familiarity with students gives them the potential to be good assessors of their students’ self-regulation, but it is possible that familiarity leads to strongly formed holistic impressions of the student, hindering teachers from assessing different constructs independently.
RQ3 – LSWH and Academic Achievement

Studies 1 and 3 reveal that teachers believe LSWH grades and achievement grades should be correlated, and the report card data reported in Study 2 further illustrates the extent to which this is realised. General correlations (including all subjects and grades) between LSWH grades and achievement grades ranged from $r = 0.72$ to $r = 0.81$ depending on the district. Of the subjects that were examined individually, Grade 9 Academic English had the highest correlation between LSWH grades and achievement grades with $r = 0.84$. The bivariate correlation coefficients between LSWH grades and teacher awarded grades ranged from $r = 0.77$ to $r = 0.80$ for Grade 9 mathematics. These values are consistent with prior research (e.g. Duckworth, Shulman, Mastronarde, Patrick, Zhang, & Druckman, 2015; Nota, Soresi, & Zimmerman, 2004; Zimmerman & Kitsantas, 2014). My research builds upon previous research by including standardized test scores in addition to LSWH grades and teacher awarded grades.

The finding that LSWH grades are more strongly correlated to teacher awarded grades than to standardized test scores is perhaps not surprising; however, it suggests that policies designed to ensure that academic achievement grades and LSWH grades are independent may not be working as intended. Certainly, other potential causes for the higher correlation between LSWH and teacher awarded academic grades exist and further study is needed to more deeply explore this finding. For example, the higher correlation between LSWH and teacher awarded grades may result from teachers using a broad variety of assessment methods (e.g., inquiry projects) in which success requires strong LSWH. It is also possible that the Grade 9 standardized mathematics test examines a narrower portion of the curriculum than teacher assessments. While these explanations are possible, the most likely explanation is that Ontario secondary teachers are conflating achievement and non-achievement factors. Not only is this
explanation consistent with other grading research (e.g., Duncan & Noonan, 2007), but there is supporting evidence from my research. During interviews, only two teachers directly reported that LSWH grades were not correlated to achievement, and two reported directly linking the LSWH grades to achievement. The remainder reported that achievement and LSWH grades should be correlated to some extent. In addition, the survey finding reported in Study 3 that 88% of respondents believed low achievement would lead to low LSWH ratings provides further evidence that the conflation of academic achievement and LSWH is at least partially responsible for this phenomenon.

The finding that some teachers use achievement to inform their grading of LSWH is an important addition to the literature on teachers’ grading practices. It has been well established that teachers allow non-achievement factors to colour their judgment of academic achievement (e.g., Bowers, 2011; Brookhart, 1991; McMillan, 2001) but it seems the reverse is also true. If Ontario secondary teachers are using their holistic impressions of students to inform LSWH grading decisions, it appears that achievement is one factor influencing those impressions. Thus there is good reason to believe that asking Ontario teachers to separately assess achievement and non-achievement factors is not achieving the intended goal of removing construct irrelevant variance from both sets of grades.

RQ4 – Teachers’ Conceptions of Self-regulation

The fourth research question focused on how teachers define self-regulation. Self-regulation was chosen because it the LSWH with the greatest research base from which to draw, and because interview data indicated that some teachers considered self-regulation as fundamental to the other five LSWH. Findings from the three different methods were generally consistent, and the interview, report card, and survey data all indicated that Ontario secondary
teachers struggle to distinguish self-regulation from the other five LSWH. Interview and survey participants stated this directly, and the unidimensionality of the LSWH grades found in the report card data provided further supporting evidence.

Two of the interview participants stated that self-regulation is a foundational construct underpinning the other LSWH, and this idea was supported by the survey findings. Self-regulation and responsibility were the two LSWH associated with the widest range of behaviours by teachers on the survey. This finding suggests that teachers believe many different observable behaviours can serve as indicators of self-regulation (and responsibility). Both the interview and survey data also demonstrated that teachers tend to associate self-regulation with classroom behaviours related to compliance or disruption. The interview data presented in Study 1 indicate that responsibility is most commonly described as meeting obligations such as deadlines. Among the 25 qualitative survey responses defining responsibility, the most common themes were attendance and submitting required work. Taken together the interview and survey data suggest that most Ontario secondary teachers define self-regulation in relation to how each student’s behaviour impacts the smooth functioning of their classroom, and not to how individual students are trying to set and achieve long-term goals. These teachers used cooperative, compliant behaviour as an operational definition of self-regulation. Students meeting deadlines, completing homework, and coming to class on time could all be reasonably construed as compliant behaviours. It is possible that compliance is the foundational construct underlying teachers’ assessments of LSWH. I recognize the term compliance has negative connotations, but I do not mean compliance as subservience, I mean compliance as meeting a teacher’s explicit and latent expectations. The exact nature of these expectations is going to be unique to every teacher, but
common expectations likely include obeying classroom rules, respecting deadlines, asking relevant questions, interacting positively with others, and performing well academically.

Admittedly, these findings were not universal. Eight of the 26 interview participants mentioned goal setting as being related to LSWH assessment and two of the survey respondents mentioned goal setting or goal achievement as being related to self-regulation. This indicates that there are Ontario secondary teachers who link self-regulation to planning and organisation to support learning, and take deliberate, concrete actions to assess self-regulation using this definition of the construct. According to the interview data, this subsample of teachers engaged students in goal setting exercises, reflection, and self-assessment.

The finding that most teachers associate self-regulation with compliant, productive behaviour has implications for interpreting teacher reports of self-regulation. To what extent can these ratings be considered ratings of self-regulation? The published psychometric properties of teacher-report scales such as the Cognitive Behavior Rating Scale (CBRS) demonstrate good reliability coefficients (e.g., Bronson, Goodson, Layzer, & Love, 1990) but this is not evidence that teachers are responding to the items as researchers intended. There is evidence that teacher-report measures of self-regulation do not correlate well with other measurement methods. For example, Duckworth and Kern (2011) found teacher reports of self-control correlated weakly with a task-based measure of the construct, with Pearson correlation coefficients ranging from $r = 0.14$ to $r = 0.21$. Similar numbers have been reported in other studies (e.g., Gioia, Isquith, Guy, & Kenworthy, 2000; Toplak, West, & Stanovich, 2013). There are several potential causes for these low correlations. The task-based instruments may ask students to complete tasks that are different from what they normally do in school, leading to different levels of self-regulation. Another possibility is that if students are aware their self-regulation is being assessed on a single
task, they self-regulate at higher levels, and can do so because their self-regulation does not need to be sustained over a long period. Teacher-report measures, on the other hand, likely measure typical performance over time, or, if teachers best remember negative behaviours and incidents, worse than typical performance. Further, performance-based measures “bypass the whole issue of rational goal pursuit” (Toplak, West, & Stanovich, 2013, p. 137-8). The goals of the task are unlikely to be set by the student and may not be relevant to other learning in the course. This would likely impact how the student performs on the task. A goal-oriented, self-regulated student who sees no value in the task may exert minimal effort in completing it. A student who is normally disengaged in the course, may find the task engaging because it is different from the types of activities normally provided by the teacher.

My findings suggest that construct-irrelevant variance in teacher reports is a potential reason teacher-reports do not correlate well with other measures of self-regulation. It seems that teachers’ holistic impressions of students are influencing their ratings of specific skills. When a student is capable, or polite, or keeps their notebook tidy, a halo effect exists, and that student is rated highly across a broad range of constructs. Further, the studies presented here demonstrate that, in the case of self-regulation, teachers and researchers defined the construct differently. Thus, it is possible that teacher-report measures of self-regulation are not measures of self-regulation, but of a different construct (e.g., compliance). Further work would need to be done investigating how teachers complete observer-report scales such as the CBRS to determine if this is true.

The Gender Divide

Gender comparisons in LSWH performance were not part of the original purpose of my research, but the report card data allowed testing for gender differences in LSWH and
achievement grades. Consistent with previous research (Duckworth et al., 2015), the present research found that girls receive higher LSWH grades than boys. As noted in Chapter 4, there are several potential reasons girls may be awarded higher LSWH grades than boys. In a Swedish study, Lekholm and Cliffordson (2009) found that girls received higher grades than boys because they had better LSWH (Lekholm and Cliffordson use the term “common grade dimension”), and girls’ better LSWH were a result of their higher motivation to learn. This may be what is occurring in Ontario secondary schools. It is also possible that teachers perceive girls to be more compliant than boys, leading teachers to form more positive perceptions of girls, which are then reflected as higher LSWH grades.

While the report card data do not provide insight into the underlying causes for gender differences in LSWH grades, the inclusion of a single-gender course provided a context in which to examine gender differences in LSWH performance. Smaller gender differences in LSWH grades in Grade 9 Physical Education (a single-gender course) than in other Grade 9 courses is an interesting finding with several possible explanations. One possible reason is that Ontario secondary teachers use a combination of norm- and criterion-referencing in making grading decisions. If boys have lower LSWH than girls, the norms of Grade 9 boys’ Physical Education classrooms would be lower, and this would raise boys’ mean LSWH grades. Alternatively, boys may have a higher interest in Physical Education (or girls have a lower interest) than in other Grade 9 courses, leading to changes in LSWH performance. Certainly, both alternatives are possible and merit further attention and exploration. Nevertheless, such findings highlight the lack of guidance Ontario teachers receive on what separates different levels of performance on the LSWH. If teachers are criterion-referencing their grades, those criteria are set by the teacher,
and it seems likely those criteria would at least be partially based upon norms observed by the teacher during their career.

The finding in Study 3 that gender differences in LSWH grades nearly disappear when academic achievement is controlled lend further complexity to the phenomenon of gender-based differences in LSWH grades. On the one hand, given that LSWH grades and achievement grades are so strongly correlated, it is not surprising that when academic achievement is controlled, gender differences nearly disappear, as much of the variance in those grades has been removed. However, this finding also highlights the importance of understanding the causation underlying the strong association between LSWH and achievement grades. If, as many researchers have found, high LSWH lead to higher achievement grades because teachers include LSWH in their achievement grading decisions, it suggests that narrowing the achievement gap between boys and girls would be best addressed by trying to improve boys LSWH. However, if the causal direction is in the opposite direction, where girls’ higher academic achievement leads to them being awarded higher LSWH grades, it suggests that other strategies would be needed to close the gender achievement gap. Further research into the underlying causes of gender differences in LSWH grades may be helpful in better understanding how to minimize gender achievement gaps.

Research into gender differences in cognitive development has not found consistent differences in overall intelligence between boys and girls (e.g., Ardila, A., Rosselli, Matute, & Inozemtseva, 2011; Miller & Halpern, 2014), although there are consistent findings that girls enjoy an advantage when it comes to reading and writing (e.g., Hyde, & Linn, 1988; Logan & Johnston, 2009; Solheim & Lundetraa, 2018). There is evidence from university students that girls are more motivated to learn, and adopt better study strategies (Marrs & Sigler, 2012).
Evidence of gender differences in study habits at high school demonstrates that differences exist, but how they relate to differences in academic performance is not clear (Rosander & Bäckström, 2012). In total, while we know that gender differences in academic achievement are common, and exist across nations, we do not have consistent evidence as to why this is the case.

Distinguishing among levels of performance

One of the findings that emerged from this research was that not only did teachers find it difficult to create clear, distinct definitions of the six LSWH, they also found it difficult to provide clear distinctions of performance within each LSWH. The interview data demonstrate this directly, and it is possible to argue the report card data support this conclusion as well. At the low end of the rating scale there was consensus among interview participants that ‘Needs improvement’ represented a serious concern regarding a student’s LSWH. ‘Needs improvement’ reflected LSWH performance so poor that student learning, and potentially the learning of others, was negatively impacted. Survey data yielded a similar finding. In contrast, there was much ambiguity about what level of LSWH performance was best represented by the terms ‘Good’ and ‘Satisfactory.’ The teachers I interviewed universally agreed that ‘Satisfactory’ meant below expectations, but how far below was not consistent. One teacher described this grade as “dire” whereas most teachers seemed to feel that satisfactory performance did not drastically impede the student’s learning. Good was typically seen as meeting expectations but no more. Expectations were always set by the teacher, as no teacher reported being given standards by which to grade students’ LSWH. The fact that teachers are not provided with standards by which they can judge levels of LSWH performance makes it impossible to use the grade to infer how the student is performing on the LSWH. Every teacher will have their own,
often latent, standards by which the students are judged. Further, it is possible that teachers’
standards will vary with the norms of the classroom.

The lack of published standards may explain why the most commonly reported method of
distinguishing amongst different LSWH levels was frequency of behaviour. A grade of
‘Excellent’ was typically used for students who always completed their work on time and were
never tardy. A grade of ‘Good’ was used for students who usually completed their work on time,
and rarely came late. ‘Excellent’ was also described by about half the teachers as exceeding their
expectations, but there were two teachers who admitted that any student who is doing well will
get LSWH grades of ‘Excellent’ across the board. Another teacher awarded ‘Excellent’ if they
did not notice the student, as it must mean the student is getting on with their work without
problems.

The Ontario Ministry of Education (2010), does not guide teachers on how to
differentiate performance levels across the LSWH, but frequency of observed behaviour has
been previously used by researchers (e.g., Hadwin, Nesbit, Jamieson-Noel, Code, & Winne,
2007; Job, Walton, Bernecker, & Dweck, 2015; Zimmerman & Martinez-Pons, 1988). Further,
other provinces (e.g., Manitoba, Saskatchewan, New Brunswick, and Newfoundland) use
frequency as a method of reporting performance of similar skills on their report cards (Merchant,
Klinger, & Love, 2018). The use of frequency of demonstration as a means of grading LSWH
performance has some value since it is easy to implement, but it fails to account for variation in
quality. Consider as an example, two students, one who contributed many ideas during group
work although none were implemented, and a second who made a sole contribution that was
adopted and found to be effective. Would frequency of contribution be an appropriate measure of
collaboration in this case? It could be argued the student who made the contribution that
enhanced the work of the group is the best collaborator, and that this student’s silence during other portions of the conversation was also effective collaboration because it allowed other student’s contributions to be heard. Another perspective would be that the contributions made by the frequent contributor could also have been effective, had they been implemented, or that his suggestions, while not adopted, spurred other ideas, enhancing the performance of the group. Frequency of an observed behaviour may be easily observed and recorded (although how well a teacher can do this accurately and reliably for all students in the class is open to question), but it may not sufficiently capture the complexity of these skills. This method of grading ignores the “skill” component of LSWH and privileges the “habit” component.

Contradictions within the data

While the findings from the three studies were generally consistent, some important contradictions were present. For instance, interview respondents were unanimous in reporting they struggled with distinguishing among the LSWH performance levels, and the report card data seems to support this finding. However, survey data contradicted this finding, with 91.8% of respondents strongly indicating they could articulate the difference between ‘Excellent’ and ‘Good’ to students. While this is a very high percentage, it should be examined critically. The survey data were self-report, and there were instances from the survey data in which respondents’ perceptions of their practices were not supported by large scale report card data. For example, only 32.0% of respondents agreed that ‘Excellent’ was the LSWH grade they most commonly awarded, and yet, the actual percentages from the report card data were much higher. Report card data from all three districts demonstrate that depending on the LSWH, grade, and district, 70 to 90 percent of all LSWH grades awarded were either ‘Excellent’ or ‘Good’. Further, for every LSWH, in every district, at every grade level, ‘Excellent’ was the most frequently
awarded LSWH. The strong negative skew indicates the LSWH grades have weak ability to discriminate among levels of performance, especially for students with strong LSWH.

Implications for practice and future research

If the LSWH grades are not useful in their current form, there are several possible approaches that could be taken to improve their utility. The first would be to report upon a set of narrow, easily observed and quantified constructs. This is already done for absenteeism and tardiness but could be expanded to include other easily observable events such as late or unsubmitted assignments, disciplinary actions, or assistance to other students. Such a system would create clearly defined measures, although an obvious concern would be the narrowing of attributes deemed to be important for inclusion. Do these easily observed behaviours serve as good indicators of broader underlying constructs such as self-regulation? They certainly serve as indicators of compliance, but if we wish to adequately measure more general constructs such as collaboration or self-regulation, it is likely more sophisticated measures will be needed. This leads to the question as to what we can reasonably expect teachers to assess well in classrooms.

Five of the interview respondents mentioned they felt that six LSWH was too many for them to assess. The findings presented here offer some support for that feeling. It is possible that reducing the number of LSWH that teachers must assess would allow them to more clearly define each LSWH, and what separates the levels of achievement. This would potentially give greater clarity to students, parents, and other stakeholders about the meaning of the LSWH grades. Note that the specific LSWH would not need to be consistent across subject areas or teachers, although some level of coordination would be desirable to ensure that students are assessed on all the LSWH across their teachers. Teachers could choose two LSWH they felt were most critical to success in their subject area and assess only those LSWH. The success of such a
change would likely be dependent on efforts to support changes in teachers’ practices. If teachers continue to use holistic impressions to assess the LSWH, no improvement would result.

Attention to these skills might also be enhanced if teachers were obliged to write two or three sentences for each LSWH, describing how it was assessed. This description would be common to all students in the class, but it would give parents some context in which to place the LSWH grade. For example, a teacher might write,

Initiative was assessed through a student’s willingness to ask questions, work on assignments without prompting, and help clean up after laboratory work. The chemical elements research project also figured into these assessments, as high initiative grades were awarded to students who independently found a variety of sources to support their research and found novel ways of presenting their findings.

A comment such as this could assist parents in interpreting what an Initiative grade represents. Further, it is possible the act of writing these sentences would prompt teachers to think more carefully about their LSWH assessment practice. At the very least, the transparency would provide some level of accountability that could spur teachers to ensure their LSWH grading decisions are defensible and made in a reasoned, systematic manner.

My findings suggest that teachers tend to award relatively high levels of attainment for LSWH. Given that many LSWH grading decisions are made in a manner that is difficult to defend, this could indicate that teachers err on the side of high grades in cases of uncertainty. High LSWH grades may help avoid conflict with students or parents, which may be especially important if there is no documentation to support the grading decision. There is evidence from prior research that in cases of uncertainty or doubt, teachers tend to award the higher grade
(Randall & Engelhard, 2010). If high, uniform LSWH grades are the observable effect of teachers’ varied, uncertain definitions of the constructs and unclear distinctions between LSWH performance levels, what are the root causes of this uncertainty? One potential reason is the lack of concrete policy guiding LSWH assessment and grading, but other possibilities include districts not offering effective training on how to assess LSWH, Ontario Bachelor of Education programs not including LSWH assessments in their assessment courses, and even the fact that high school teachers are subject area specialists, whose initial undergraduate education (e.g., English, mathematics, history) would not have prepared them to define and assess constructs such as the LSWH. Further, because there do not appear to be any consequences or stakes attached to the LSWH grades, there is little accountability for the LSWH grades, possibly reducing teachers’ motivation to improve this portion of their practice. These reasons are not mutually exclusive, they could all be contributing factors.

Improvement to practice is important because Ontario secondary teachers’ assessment and grading of the LSWH is so varied and idiosyncratic. Such variations make it challenging to use these grades for improvement, program selection, or admissions decisions. Of importance, these assessments, and resulting grades, do not meet required standards of validity to be useful for high stakes decisions. Perhaps more troubling is that it is difficult to see how students or parents could use these grades for improvement purposes. From the interview data it emerged that teachers’ formative use of LSWH assessment happens immediately, and in class. Teachers do not wait until report card time to address problems. Instead, the problematic behaviour or incident (e.g., distracting another student or submitting an assignment late) is addressed directly with the student in the moment. While this formative use of LSWH assessment data is potentially helpful to the student, it seems unlikely these types of incidents are regularly reported to parents.
Unless the teacher has contacted a parent prior to report cards to discuss a concern, parents must wait until the report card to know how well their child is performing on the LSWH; however, receiving a LSWH grade at report card time is unlikely to be useful to parents seeking to support the development of LSWH in their child. The parent needs to know how the teacher is defining, assessing and distinguishing each LSWH in order to further support their child’s development of these skills. Overall, it seems unlikely that a single grade could be a useful reporting tool for the LSWH. The skills themselves are complex, and student performance on them can change depending on context. This may explain why during interviews, teachers reported that students and parents value the report card comments more than the LSWH grades.

If the LSWH themselves are complex, accurate reporting of these skills likely requires a method that allows for communicating this complexity. Currently, report card comments and parent-teacher interviews are two such methods that are employed, although not necessarily on a mandated or consistent basis. Alternate methods of reporting performance on the LSWH is an area that would be worthy of future study. For example, what types of comments do teachers make on report cards related to LSWH? Are those comments general in nature, or do they focus on specific LSWH? What about in parent-teacher conferences? Does the discussion include LSWH? If so, what is the focus? Is it improvement? Is it relating specific events that occurred within the classroom? A study of alternate reporting methods would help broaden our understanding of how teachers assess and report upon skills such as the LSWH.

Many teachers struggle to define the LSWH as constructs. This was demonstrated through the interview, report card, and survey data. As a result, a single event or student behaviour can impact multiple LSWH grades. Improving this situation will require teachers to receive effective and focussed training and professional learning. The current guidance given to
teachers on how to assess and grade the LSWH is low. Consequently, teachers must decide for themselves what separates the different performance levels (e.g., How is ‘Good’ initiative different from ‘Satisfactory’ initiative?), and how to weight different behaviours and events that influence the LSWH grade. If “demonstrates curiosity and interest in learning” (Ontario Ministry of Education, p. 11) and “recognizes and advocates appropriately for the rights of self and others” (p. 11) are both behaviours that teachers should consider when assessing initiative, how is a teacher supposed to grade a student who demonstrates great curiosity in their learning, but consistently fails to recognize the rights of his peers?

The issues identified here highlight the need to provide teachers with specific, concrete guidance on how to assess the LSWH, and how to distinguish among levels of achievement. This guidance could be provided at a provincial, district, or school level, but without it, it is difficult to see how teachers’ LSWH assessment and grading practices could improve. As school systems increasingly incorporate constructs such as LSWH, 21st Century skills, socio-emotional learning, and character into their assessment and reporting, policy and research needs to support their ability to do so effectively. While all teacher education programs in Ontario are required to include training in assessment and evaluation, an online search of course descriptions within Ontario faculties of education revealed a single elective course at one university that mentioned assessment and evaluation of the LSWH. A more formal study conducted by Poth (2013) found that none of the 14 teacher education programs in Western Canada included assessment of non-achievement constructs in their assessment courses, even though teachers in Western Canada are required to assess and report upon such constructs. Even the most recently trained teachers are unlikely to have received formal instruction in how to assess LSWH or related skills. After teacher candidates graduate and move into practice, there is anecdotal evidence that some
districts include assessing LSWH as part of their induction into the profession. For example, the Limestone School District Board in Eastern Ontario includes one to two hours of instruction on how to assess LSWH as part of its two-day induction program.

On a broader scale, resources exist to help teachers learn how to assess skills like the LSWH (e.g., Marzano & Heflebower, 2011B; Stanley, 2014), but they are far less common than resources on how to assess subject area achievement. There have been calls for policies and resources to support teachers in helping them better assess skills such as the LSWH (e.g., National Research Council, 2013; Stecher & Hamilton, 2014), and it appears these calls are being heeded. Several states including California, Massachusetts, and Illinois have introduced statewide standards for social-emotional learning, and the Partnership for 21st Century Learning (2007) has published an assessment guide for teachers on assessing 21st Century Skills. Still, more work is needed. The Joint Committee for Standards in Educational Evaluation most recent edition of the classroom assessment standards do not directly address classroom assessment of constructs like the LSWH (Klinger et. al., 2015), and only two Canadian provinces have policy documents giving teachers direction on how to assess and report upon non-achievement constructs (Merchant, Klinger & Love, 2018). To make substantial improvements in teachers’ LSWH assessment practice, good policy, resources such as books and assessment standards, and workshops will likely need to be supplemented with advice and guidance on the job. Currently, many Ontario school districts have specialists in literacy, mathematics, or other subject areas who can travel around the district and support teachers in developing their practice. A similar scheme may be beneficial for LSWH assessment.

Perhaps the most profound implication for practice arising from my research may be that proper assessment of the LSWH requires that students have opportunities to demonstrate them.
This is not the case in some classrooms. One French teacher who was interviewed reported that initiative was impossible for her to assess because students in her class have minimal opportunity, or desire, to demonstrate initiative. Instead, students are expected to follow directions. To what extent is this true in other classrooms? Survey data revealed that initiative is most often assessed as asking questions or seeking help. While these may be good indicators of initiative for some students, it may inadvertently punish students who are very capable in the subject. A similar problem occurs with self-regulation. If goal-setting and maintaining focus on that goal are critical aspects of self-regulation, then teachers need to give students opportunities to set their own goals, and opportunities to demonstrate how they are continually taking actions to achieve that goal. To what extent that occurs in classrooms is unknown, but the interview and survey data presented here suggest it is the exception and not the norm.

It may be that the best opportunities to assess the LSWH are not inside the classroom but outside. Clubs, student council, fund-raisers, school teams, and other extra-curricular activities may serve as the best contexts to assess the LSWH. This is because these activities can be where students are given more autonomy and independence than they have in the classroom. An interesting extension to my research would be to compare students’ LSWH in different contexts. These contexts could include different subject areas, extra-curricular activities, and employment situations. These different contexts may also be useful for further studying gender differences in LSWH.

Limitations

The multiple methods used in this dissertation generally produced consistent findings, increasing confidence in the veracity and reliability of the findings. Nonetheless, limitations exist within the studies that may challenge the generalisation of the findings. For the studies presented
in Chapters 3 and 5, both the interviews and survey sampling methods likely yielded samples that were not representative of the Ontario secondary teaching population. As noted in Chapter 3, participation in the interviews was voluntary, and it seems probable that teachers with an interest in assessing LSWH would be more likely to volunteer. Consequently, it is possible that teachers in the interview sample had thought more deeply about, and were more conscientious about, their LSWH assessment practices than most Ontario secondary teachers. There is some evidence for this in the fact that five of the teachers came from a collaborative inquiry group focused on the assessment of LSWH. Nonetheless, because some of the teachers were snowball sampled from a teachers’ hockey league, it seems probable the sample of 26 teachers included a broad range of LSWH assessment philosophies and practices that reflected the diversity of LSWH assessment and grading practices within Ontario secondary teachers.

As with the interviews, the survey respondents were also volunteers. The recruitment strategy of using social media met the desired objective of having respondents from geographically diverse areas, but also likely led to the high number of unusable surveys. Of 300 survey responses, only 108 were usable, leading to a smaller than desired sample. This is an important limitation for two reasons. One is the small sample size precluded more advanced statistical analyses from being conducted. For example, exploratory factor analyses designed to test the dimensionality of the survey did not yield interpretable results. Further, the small sample precluded breaking down the respondents into smaller groups such as by subject area or years of teaching experience. However, the large number of elementary teachers who responded to the survey provides a data set that allows investigation of the differences between elementary and secondary teachers’ approaches to assessing LSWH.
Finally, both the interviews and survey responses likely suffered from common problems with self-report data. Responses to interview questions were all retrospective, and while it would have been preferable to use think aloud protocols as teachers completed their report cards, ethical considerations made this approach problematic. Another potential problem with the interview and survey data is social desirability in survey response. There is no direct evidence social desirability influenced the survey data, but some mismatches between the report card data and the survey and interview data suggest it is possible. As an example, only 50.7% of survey respondents agreed that students tend to receive the similar grades across all six LSWH, but report card data showed that, depending on the district, between 62.7% and 67.0% of students received not just similar, but identical grades across all six LSWH. In addition, report card data demonstrated that ‘Excellent’ was the most commonly awarded grade for every LSWH in every district, but only 32.0% of the survey respondents reported awarding ‘Excellent’ more commonly than the other ratings. Beyond social desirability in survey response, other possible explanations for these mismatches include inaccurate memories of grading practices, and a survey sample that was not representative of the general population of Ontario secondary teachers.

The report card dataset was very large, giving confidence the samples represented the populations of each district, but it was also limited in several important respects. Due to confidentiality reasons, it was not possible to obtain data with school or teacher identifiers. This precluded more complex analyses such as hierarchical linear modeling which would have allowed the relative importance of district, school, teacher, and student factors to be examined. Also, the lack of standardized examinations in Ontario meant that comparisons between teachers awarded grades, LSWH grades, and examination scores was limited to Grade 9 mathematics. It would have been illuminating to have been able to perform similar analyses for other subjects.
Summary and Conclusions

Vague, varied, or inconsistent definitions of a construct lead to poor validity (Bass, 2005), and it appears the LSWH are especially problematic in this regard. While the Ontario Ministry of Education (2010) has given teachers example behaviours that could be used to assess each LSWH, my research indicates that most Ontario secondary teachers neither explicitly nor implicitly use these behaviours, even if they are aware of them. Instead, it appears that teachers are creating their own latent definitions and operationalizations of the constructs. These latent definitions appear to be one-dimensional and related to compliance with the teacher’s expectations for student behaviour in the course. Consequently, it is difficult to imagine how a parent, or other stakeholder, could extract information from the LSWH grades beyond the teacher’s general impression of the student.

Skills such as the LSWH have been of interest to researchers and educators for some time. From the earliest days of intelligence and standardized testing, it has been recognized that constructs such as IQ, or subject area achievement represent important, but limited, information about a student (Binet & Simon, 1916), and the calls for schools to develop and assess skills such as the LSWH are becoming stronger (Kautz, Heckman, Diris, Ter Weel, & Borghans, 2014; Levin, 2012). Further, such calls are being made not just in North America, but globally (Ananiadou & Claro, 2009). Assessment of these skills is a recognized challenge. The difficulty in defining self-regulation is well documented (Boekaerts & Corno, 2005), and this difficulty applies to all the LSWH. Further, different assessment methods all have their weaknesses (Stroud, 2013), leading to the conclusion that devising valid assessments of the LSWH, or related skills, is likely going to require varied assessment methods delivered over time. Assessing LSWH in this way can help avoid important assessment issues such as maximum vs.
typical performance, and method-based variance. Zimmerman and Martinez (1988) argued that teachers are well positioned to provide high quality assessments of student self-regulation, and this argument could be extended to all the LSWH. The sustained interaction that teachers have with their students over time gives them many opportunities for implementing different types of assessments, whereas other authors (e.g., Hoge, & Butcher, 1984; Winne & Perry, 2000) have expressed concerns about teachers’ abilities to assess constructs such as the LSWH. Recent evidence from Duckworth et al. (2015) and Ferrito (2015) suggests these concerns are well founded, and the studies presented here add to that evidence. The studies presented here build upon prior work in several ways. First is the examination of what specific student behaviours are incorporated into these assessments. Secondly, the inclusion of a single gender course, along with mixed gender courses, allows for additional insights into both teacher assessment and gender differences in LSWH. Finally, the large sample size of the report card data, including standardized tests, gives robust evidence for associations between LSWH grades, achievement grades and standardized test results.

Another important contribution to the literature is the finding that some teachers use academic grades to inform their LSWH grading decisions. Calls from assessment and measurement experts (e.g., Brookhart, 2004; Guskey, 2006) to separate achievement and non-achievement factors are well intentioned but this research shows that some of the implicit assumptions in this recommendation may not hold true. As an example, policies requiring teachers to report upon achievement and non-achievement factors are founded upon the assumption that by giving teachers the opportunity to separate achievement and non-achievement factors they will do so. The interview data demonstrate that some teachers not only do not separate achievement and LSWH—they equate them. Whether by survey or by interview, many
teachers reported feeling uncomfortable giving high LSWH grades to students with low achievement grades and vice-versa.

These findings call into question the assumption that separately reporting on non-achievement factors will give a more complete picture of student performance. Instead, the studies presented here demonstrate that interpreting LSWH grades, and extracting actionable information from them is very difficult, if not impossible. Rather than providing information about student performance on specific skills, LSWH grades most likely reveal a teacher’s general impression of the student’s ability to meet the expectations of the classroom. This is not dissimilar to what other researchers have found is communicated by the achievement grade (e.g., Bowers, 2011; Lekholm & Cliffordson, 2008). This is not to say that separating achievement and non-achievement factors is a bad idea. If we are going to ask teachers to assess skills such as the LSWH, we need to support them in this endeavour and help them move beyond using subject area achievement, or general impressions of the student, to award LSWH grades.
REFERENCES


secondary teachers of academic subjects: Implications for instruction in

mediate the relationship between self-control and positive life outcomes. *Journal of
personality and social psychology, 109*, 508-525.

Galla, B. M., Plummer, B. D., White, R. E., Meketon, D., D'Mello, S. K., & Duckworth, A. L.
(2014). The Academic Diligence Task (ADT): Assessing individual differences in effort
on tedious but important schoolwork. *Contemporary Educational Psychology, 39*, 314-
325.

professional development effective? Results from a national sample of

B.R., & O’Connor, M.C. (Eds.), *Changing assessments: Alternative views of aptitude,
achievement and instruction* (pp. 78-119), London: Kluwer.

Gestsdottir, S., von Suchodoletz, A., Wanless, S. B., Hubert, B., Guimard, P., Birgisdottir, F.,


Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and


APPENDIX A: INTERVIEW PROTOCOL FOR CHAPTER 3

- Initial small talk about something to break the ice (e.g. weather, traffic, summer is coming soon, the way the classroom is decorated, a funny teaching story etc.).

- Introduce the study, the LOI and the consent form.
  
  Thank you so much for participating in my study. Your willingness to contribute to the study is enormously helpful for me, but I also hope that you will find today’s session useful. You should have already read the Letter of Information and signed the consent form. If not, here they are and take your time to read through them. The main points I would like to stress are that this session will be audio recorded and transcribed, but all data will be confidential. Once the transcripts are created, names are removed, so the only identifying information retained will be your subject area. E.G. you might be known in the study as “Science teacher #2”. You should also know that you can withdraw from the study at any time (up until results are submitted for publication), and that you are under no obligation to answer every question. If a question makes you uncomfortable or you do not want to answer it for any reason, just say so and we will move on.

- Ensure the teacher has their gradebook and is comfortable. Give them the $20 gift card. Ask for permission to start recording and once permission is granted, begin recording.

- Begin the think aloud / interview session.

Think aloud questions and prompts:

1. Select student #8 from your gradebook and describe the process you went through to assign the LSWH grades to that student.
   - What evidence did you consider?
   - What do you record in the gradebook to help you make the decisions?
   - What memories or events informed this decision?
   - Do you consider things that happen outside the classroom in your LSWH grades? E.G. a student initiates and organizes a fund raising drive for Syrian refugees?
   - Are there specific activities you do in class that you think are good opportunities to assess the LSWH?

2. Compare/contrast the process of deciding upon an academic grade vs. a LSWH grade.

3. Thinking about this same student, when you awarded this grade what were your thoughts about how the grade would be received and/or interpreted by the student.
   - Did you imagine it as a wakeup call?
   - As encouragement?
• As a pure reflection of their learning skills and work habits achievement?
• Was this communication primarily to the student, to parents or both?

4. Please select your top student in terms of the learning skills and work habits and describe what characteristics make them the top student.

5. Please select another very good student in terms of the learning skills and work habits and describe what characteristics make them very good, but not the top student.
   • What distinguishes “good” from “excellent” for the LSWH?
   • Which levels are the most difficult to distinguish between? (e.g. good vs. excellent or good vs. satisfactory) Why?

6. Tell me about a student who achieved very highly in their academic work but demonstrates low LSWH. What are they like? What allows them to be successful academically? Which LSWH are they weakest in? Strongest?
   • Tell me about a student who has strong LSWH, but is weak academically. Are they strong in all LSWH? What prevents them from achieving higher?

7. I noticed student #8 got a “XXXX” in independent work. How is independent work different from initiative?
   • What are the different things you look for when assessing independent work and initiative?
   • How is “organization” different from “responsibility”?
   • Can you have a student who is organized but not responsible? What would this look like?

8. In your opinion what should this portion of the report card look like?
   • What should be assessed?
   • What kind of scale would you use?
   • Should it be comments only? Grades only? A combination?

9. Do you have any additional comments you would like to add?
APPENDIX B: QUANTITATIVE SURVEY ITEMS AND RESULTS

Please rank the learning skills in order of importance for student success in the class you indicated in the first question. Rankings should be from most important (1) to least important (6).

<table>
<thead>
<tr>
<th>LSWH</th>
<th>Mean rank (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration</td>
<td>3.94 (0.17)</td>
</tr>
<tr>
<td>Independent Work</td>
<td>3.63 (0.16)</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.31 (0.16)</td>
</tr>
<tr>
<td>Organization</td>
<td>3.89 (0.15)</td>
</tr>
<tr>
<td>Responsibility</td>
<td>2.84 (0.14)</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>3.39 (0.18)</td>
</tr>
</tbody>
</table>

Thinking of the same class, please rank the learning skills according to how difficult they are to assess. The easiest learning skill to assess should have a rank of 1, and the hardest should have a rank of 6.

<table>
<thead>
<tr>
<th>LSWH</th>
<th>Mean rank (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collaboration</td>
<td>3.78 (0.22)</td>
</tr>
<tr>
<td>Independent Work</td>
<td>2.76 (0.18)</td>
</tr>
<tr>
<td>Initiative</td>
<td>3.63 (0.18)</td>
</tr>
<tr>
<td>Organization</td>
<td>3.40 (0.20)</td>
</tr>
<tr>
<td>Responsibility</td>
<td>3.72 (0.18)</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>3.72 (0.19)</td>
</tr>
</tbody>
</table>

Rate your level of agreement with each statement. (Strongly disagree = 1, Strongly agree = 7)

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students ask questions or share concerns about their learning skills and work habits grades</td>
<td>2.88 (1.77)</td>
</tr>
<tr>
<td>Parents want to discuss the learning skills and work habits grades with me</td>
<td>3.82 (2.01)</td>
</tr>
<tr>
<td>Reporting learning skills and work habits is as important as reporting achievement</td>
<td>5.46 (1.87)</td>
</tr>
<tr>
<td>Strong learning skills and work habits are necessary for success in post-secondary education</td>
<td>6.70 (0.59)</td>
</tr>
<tr>
<td>I have materials on display in my classroom (e.g. posters) about the learning skills and work habits</td>
<td>4.66 (2.19)</td>
</tr>
<tr>
<td>Learning skills and work habits are a good predictor of future success</td>
<td>6.00 (1.33)</td>
</tr>
<tr>
<td>Universities should consider the learning skills and work habits grades in their admissions decisions</td>
<td>5.13 (2.03)</td>
</tr>
<tr>
<td>Students will ask questions or share concerns about their academic grades</td>
<td>6.41 (0.82)</td>
</tr>
<tr>
<td>Parents want to discuss the academic (percentage) grades with me</td>
<td>6.43 (1.04)</td>
</tr>
<tr>
<td>Academic achievement is a good predictor of future success</td>
<td>4.25 (1.79)</td>
</tr>
</tbody>
</table>
Rate your level of agreement with each statement. (Strongly disagree = 1, Strongly agree = 7)

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have specific activities I use to assess specific learning skills and work habits</td>
<td>5.12 (1.81)</td>
</tr>
<tr>
<td>I have students self-assess their learning skills and work habits</td>
<td>5.66 (1.62)</td>
</tr>
<tr>
<td>Learning skills and work habits should be incorporated into the achievement grade, and not reported separately</td>
<td>4.41 (2.11)</td>
</tr>
<tr>
<td>Students typically receive a similar grade across all six learning skills and work habits</td>
<td>4.11 (1.74)</td>
</tr>
<tr>
<td>I only give a grade of ‘N’ if I have a serious concern</td>
<td>4.78 (1.82)</td>
</tr>
<tr>
<td>I use school, district, or ministry policy documents to guide my assessment of the learning skills and work habits</td>
<td>5.13 (1.82)</td>
</tr>
<tr>
<td>Learning skills and work habits grades are best assessed holistically</td>
<td>5.13 (1.32)</td>
</tr>
<tr>
<td>I consider events or behaviours outside the classroom when I assign the learning skills grades (e.g. participation in sports or performing arts)</td>
<td>2.79 (2.06)</td>
</tr>
<tr>
<td>I keep written records to help inform my learning skills and work habits grades</td>
<td>4.96 (1.86)</td>
</tr>
<tr>
<td>The most common learning skills and work habits grade I give is “Excellent”</td>
<td>3.49 (1.65)</td>
</tr>
</tbody>
</table>

Rate your level of agreement with each statement. (Strongly disagree = 1, Strongly agree = 7)

<table>
<thead>
<tr>
<th>Item</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can explain to students what differentiates “Excellent” from “Good” for each learning skill and work habit</td>
<td>5.93 (1.26)</td>
</tr>
<tr>
<td>I can explain to students how “Organization” is different from “Responsibility”</td>
<td>6.13 (1.17)</td>
</tr>
<tr>
<td>The six different learning skills and work habits are distinct from each other. i.e. there is little overlap</td>
<td>3.80 (1.83)</td>
</tr>
<tr>
<td>My colleagues and I define “Self-regulation” in a similar way</td>
<td>4.49 (1.63)</td>
</tr>
<tr>
<td>Achievement below 50% makes it likely a student will receive &quot;Needs Improvement&quot; for at least one of the learning skills</td>
<td>5.89 (1.39)</td>
</tr>
<tr>
<td>I am reluctant to give S or N for the learning skills and work habits if the achievement grade is over 80%</td>
<td>3.66 (1.95)</td>
</tr>
<tr>
<td>The 6 learning skills and work habits are also job skills</td>
<td>6.14 (1.03)</td>
</tr>
<tr>
<td>I award a &quot;Needs Improvement&quot; only when a student rarely displays a learning skill</td>
<td>4.34 (1.79)</td>
</tr>
<tr>
<td>I use rubrics to assess the learning skills and work habits</td>
<td>4.30 (2.13)</td>
</tr>
<tr>
<td>I have a section in my course outline that explains to students how I assess the learning skills and work habits</td>
<td>4.83 (2.23)</td>
</tr>
<tr>
<td>What distinguishes different levels of achievement in the learning skills is the consistency with which the student demonstrates the learning skill.</td>
<td>5.49 (1.21)</td>
</tr>
</tbody>
</table>
February 09, 2017

Mr. Stefan Merchant
Ph.D. Candidate
Cultural Studies Program
Queen's University
Kingston, ON, K7L 3N6

GREB Ref #: GEDUC-843-17; TRAQ # 6020088
Title: "GEDUC-843-17 Learning Skills and Work Habits: Patterns of assessment in Ontario Secondary Schools"

Dear Mr. Merchant:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GEDUC-843-17 Learning Skills and Work Habits: Patterns of assessment in Ontario Secondary Schools" for ethical compliance with the Tri-Council Guidelines (TCPS 2 (2014)) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (405.001), your project has been cleared for one year. You are reminded of your obligation to submit an annual renewal form prior to the annual renewal due date (access this form at http://www.queensu.ca/traq/signon.html, click on "Events", under "Create New Event" click on "General Research Ethics Board Annual Renewal/Closure Form for Cleared Studies"). Please note that when your research project is completed, you need to submit an Annual Renewal/Closure Form in Romeo@traq indicating that the project is 'completed' so that the file can be closed. This should be submitted at the time of completion; there is no need to wait until the annual renewal due date.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period (access this form at http://www.queensu.ca/traq/signon.html, click on "Events", under "Create New Event" click on "General Research Ethics Board Adverse Event Form"). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example, you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To submit an amendment form, access the application by at http://www.queensu.ca/traq/signon.html, click on "Events", under "Create New Event" click on "General Research Ethics Board Request for the Amendment of Approved Studies". Once submitted, these changes will automatically be sent to the Ethics Coordinator, Ms. Gail Irving, at the Office of Research Services for further review and clearance by the GREB or GREB Chair.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Sincerely,

[Signature]

John Freeman, Ph.D.
Chair
General Research Ethics Board

c: Dr. Don Klinger, Supervisor
Dr. Richard Reeve, Chair, Unit REB
Ms. Erin Rennie, Dept. Admin.
May 11, 2017

Mr. Stefan Merchant
Ph.D. Candidate
Faculty of Education
Queen’s University
Duncan McArthur Hall
511 Union Street West
Kingston, ON, K7M 5R7

GREB Ref #: GEDUC-850-17; TRAQ # 6820764
Title: "GEDUC-850-17 Assessing Learning Skills and Work Habits in Ontario Secondary Schools - A survey of teacher attitudes and practices"

Dear Mr. Merchant:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GEDUC-850-17 Assessing Learning Skills and Work Habits in Ontario Secondary Schools - A survey of teacher attitudes and practices" for ethical compliance with the Tri-Council Guidelines (TCPS2 (2014)) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (405.001), your project has been cleared for one year. You are reminded of your obligation to submit an annual renewal form prior to the annual renewal due date (access this form at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Annual Renewal/Closure Form for Cleared Studies"). Please note that when your research project is completed, you need to submit an Annual Renewal/Closure Form in Romeo/traq indicating that the project is 'completed' so that the file can be closed. This should be submitted at the time of completion; there is no need to wait until the annual renewal due date.

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period (access this form at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Adverse Event Form"). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example, you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To submit an amendment form, access the application by at http://www.queensu.ca/traq/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Request for the Amendment of Approved Studies". Once submitted, these changes will automatically be sent to the Ethics Coordinator, Ms. Gail Irving, at the Office of Research Services for further review and clearance by the GREB or GREB Chair.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Sincerely,

John D. Freeman, Ph.D.
Chair
General Research Ethics Board

c: Dr. Don Klinger, Supervisor
   Dr. Richard Reeve, Chair, Unit REB
   Mrs. Eris Rennie, Dept. Admin.
April 05, 2016

Mr. Stefan Merchant
Ph.D. Candidate
Faculty of Education
Queen's University
Duncan McArthur Hall
511 Union Street West
Kingston, ON, K7M 5R7

GREB Ref #: GEDUC-804-16: Romeo # 6018165
Title: "GEDUC-804-16 How do Teachers Assess Learning Skills and Work Habits?"

Dear Mr. Merchant:

The General Research Ethics Board (GREB), by means of a delegated board review, has cleared your proposal entitled "GEDUC-804-16 How do Teachers Assess Learning Skills and Work Habits?" for ethical compliance with the Tri-Council Guidelines (TCPS 2 (2014)) and Queen's ethics policies. In accordance with the Tri-Council Guidelines (Article 6.14) and Standard Operating Procedures (401.001), your project has been cleared for one year.

You are reminded of your obligation to submit an annual renewal form prior to the annual renewal due date (access this form at http://www.queensu.ca/trag/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Annual Renewal Form forApproved Studies").

You are reminded of your obligation to advise the GREB of any adverse event(s) that occur during this one year period (access this form at http://www.queensu.ca/trag/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Adverse Event Form"). An adverse event includes, but is not limited to, a complaint, a change or unexpected event that alters the level of risk for the researcher or participants or situation that requires a substantial change in approach to a participant(s). You are also advised that all adverse events must be reported to the GREB within 48 hours.

You are also reminded that all changes that might affect human participants must be cleared by the GREB. For example, you must report changes to the level of risk, applicant characteristics, and implementation of new procedures. To submit an amendment form, access the application by at http://www.queensu.ca/trag/signon.html; click on "Events"; under "Create New Event" click on "General Research Ethics Board Request for Amendment of Approved Studies". Once submitted, these changes will automatically be sent to the Ethics Coordinator, Ms. Gail Irving, at the Office of Research Services for further review and clearance by the GREB or GREB Chair.

On behalf of the General Research Ethics Board, I wish you continued success in your research.

Sincerely,

John Freeman, Ph.D.
Chair
General Research Ethics Board

c: Dr. Don Klinger, Faculty Supervisor
   Dr. Liying Cheng, Chair, Unit REB
   Ms. Erin Wicklan, Dept. Admin.