

A MULTI-METHOD STUDY OF YOUTH DEVELOPMENT AND WELLBEING:  
A META-NARRATIVE ANALYSIS AND PROGRAM EVALUATION STUDY

by

Sofia A. Puente-Duran

Master of Education, Ontario Institute for Studies in Education, University of Toronto (2013)

Bachelor of Arts (Honours), Ryerson University (2011)

A dissertation

presented to Ryerson University

in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

in the program of

Psychology

Toronto, Ontario, Canada, 2018  
© Sofia A. Puente-Duran, 2018

## AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A DISSERTATION

I hereby declare that I am the sole author of this dissertation. This is a true copy of the dissertation, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this dissertation to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this dissertation by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my dissertation may be made electronically available to the public.

A Multi-Method Study of Youth Development and Wellbeing:  
A Meta-Narrative Analysis and Program Evaluation Study

Doctor of Philosophy, 2018

Sofia A. Puente-Duran

Psychology

Ryerson University

**Abstract**

Youth are embedded within multiple environmental systems, developing within families, neighbourhoods, and multicultural communities. Such systems influence the formation of identity and wellbeing. It is important to monitor the wellbeing of youth across their environments, given the complexities of diverse youth experiences. Accordingly, the present dissertation comprises two studies to address the topic of youth development and wellbeing using multiple data collection techniques. In Study 1, a meta-narrative analysis was undertaken examining concepts of youth wellbeing across multidimensional indices. A search was performed across the grey literature base and seven indices fit the search criteria. Data were extracted using a codebook to guide a thematic analysis and critical appraisal to compare, contrast, and critique indices. Results showed three key findings. (1) Indices had some overlap to conceptualize wellbeing, using an average of six dimensions. (2) Data collection used similar levels of population-level statistics and self-reported data. (3) A large proportion of measures focused on youth deficits, with less focus placed on positive attributes. In Study 2, an evaluation was conducted assessing the impact of a school-based art program on the socio-emotional wellbeing of adolescents from three grade 8 classrooms, within one inner-city, multicultural neighbourhood. A mixed-method, multi-informant evaluation design was employed, and implementation processes of the program were

assessed. Survey data and open-ended responses were collected from 74 students at three time-points (pre, post, follow-up), using multilevel modeling to examine time-points nested within students. Responses were also collected post-program from six artist facilitators and three teachers. Program implementation results showed high levels of fidelity, and high quality ratings. Results from multilevel models showed significant variation at the between-student level. Across students, significant improvements were found over time for art skill, self-expression, and confidence presenting. Qualitative data revealed themes across informants regarding the positive impact of the program on student growth. Findings also indicated the importance of a safe space for adolescents to learn about themselves, and be vulnerable. These two studies shed light on the multiple ways in which youth development and wellbeing are assessed, and the ways in which a local-level community program can support their wellness.

## Acknowledgments

I would first like to acknowledge my sincere appreciation to my doctoral supervisor, Dr. Kelly McShane, who made this endeavour possible. Thank you for your ongoing support and encouragement. I admire your passion to incorporate community-based partnerships into research, and strive to highlight the community voice, particularly in working with vulnerable groups. Thank you for being so accepting and flexible in supporting my academic and life paths. I would also like to thank my second reader, Dr. David Day, who spent much time guiding me through my dissertation with intellectual discourse and editing help over the past two years.

Thank you to Dr. Morton Beiser, for whom I hold great admiration and who has been an important mentor throughout my academic endeavours. Thank you for recruiting me as your research assistant first almost ten years ago (!) and again five years ago. I value our shared passion in helping to shed light on the ever-pressing needs and supports for marginalized communities, including immigrant and refugee populations.

I would like to thank the staff and my peers in the Department of Psychology, and more specifically my colleagues in our lab – we were only a trio, and I so very much enjoyed your company! I also appreciated learning from and collaborating with my lab-mates at OISE – such a valuable time for my academic and personal growth.

Thank you to my family, local and abroad, who provided me with ongoing love and support. I could not have done this without your encouragement. First, to my parents: listening to your music, poetry, and words of wisdom provided me with so many important and reflective moments

of peace and relaxation. To my brother and sister: I admire you always! To my dearest nephew and niece: you brighten my days and add valuable insight into the work that I do, by showing me the important variability that can exist in the paths of child growth. Also, thanks to my family and friends in Chile who were a constant shoulder to lean on in general, and during my practicum placement in Santiago.

Lastly, I would like to thank my important and favourite humans: my best friends and significant other for their ongoing encouragement and always reminding me that life is about a balance. The cumulative support has been invaluable.

## Table of Contents

Declaration.....	ii
Abstract.....	iii
List of Tables.....	vii
List of Figures.....	viii
List of Appendices.....	ix
Chapter 1: General Introduction.....	1
Approaches to Monitoring Youth Wellbeing.....	4
Summary and Rationale for Current Studies.....	14
Research Questions.....	16
Chapter 2: Study 1. Meta-Narrative Analysis.....	17
Introduction.....	18
Methods.....	34
Results.....	45
Discussion.....	66
Chapter 3: Study 2. Program Evaluation Study.....	74
Introduction.....	75
Methods.....	96
Results.....	123
Discussion.....	146
Chapter 4: General Conclusion and Implications.....	156
Appendices.....	160
References.....	175

## **List of Tables**

### **Study 1**

Table 1. Selected List of Indices, Index Number, Name, Year, Age Ranges, and Number of Methodological Papers per Index.....	48
Table 2. Definitions of Population Samples Across Indices.....	50
Table 3. Definitions of ‘Wellbeing,’ and the Theoretical Framework to Conceptualize Wellbeing Across Indices.....	51
Table 4. Domain Clusters and Labels Across Indices.....	55
Table 5. Number of Domains, Indicators, and Measures Per Index and the Mean Across Indices.....	56
Table 6. Overlap of Indicators Within Domains and Overall Overlap Across Indicators, Measures, and Sources of Data .....	58
Table 7. Overall Rate of Measurement Across Age Ranges Per Domain.....	60
Table 8. Rates of Respondent Category (Unit of Analysis) and Positive, Negative, and Neutral Measures Across Domains.....	61

### **Study 2**

Table 9. Areas of Interest for Adolescents Quantitative Measure Items, Sub-Domains, and Internal Consistency in the Present Study.....	112
Table 10. Descriptive Statistics of Participants Overall and Between Schools.....	125
Table 11. Implementation Outputs and Rates Across the Activity Checklist and Observed Student Activity Completion.....	127
Table 12. Implementation Outputs Across Self-Reported Variables (Means and Standard Deviations).....	128
Table 13. Implementation Outputs of Program Enjoyment and Attendance Rate.....	131
Table 14. Means and Standard Deviations of Outcome Variables at Different Time-Points...	139
Table 15. Final Model (Model 4) of the Multilevel Models for each Study Variable Across Time-Points.....	140

## List of Figures

### Study 1

Figure 1. Conceptual framework and phases in the meta-narrative analysis.....	36
Figure 2. Flow diagram of search strategy.....	47

### Study 2

Figure 3. Conceptual framework of the evaluation guiding the study.....	98
Figure 4. Equation of final model of the multilevel models for each outcome.....	121
Figure 5. Time-point analysis for art skill across adolescents in the art program.....	135
Figure 6. Time-point analysis for self-expression, confidence presenting, and self-esteem across adolescents in the art program.....	137

## **List of Appendices**

Appendix A. Search Strategy for Indices, Inclusion and Exclusion Criteria.....	160
Appendix B. Details of the Codebook, Areas of Inquiry, and Thematic Comparison Questions to Assess Indices in the Meta-Narrative Analysis.....	161
Appendix C. List of Eleven Methodological Papers in the Analysis Process.....	162
Appendix D. Ryerson Research Ethics Board (REB) Email Correspondence.....	163
Appendix E. List of Key Informant Interview Questions.....	164
Appendix F. Program Logic Model.....	165
Appendix G. Table of Data Sources, Measures, Question Items, and Psychometric Properties for the Corresponding Areas of Interest.....	166
Appendix H. Student Consent Form.....	169
Appendix I. Student Surveys.....	171

## **Chapter 1: General Introduction**

Canadian youth develop within multiple levels of environmental influence, operating within families, neighbourhoods, and multicultural communities that have an impact on development and wellbeing (Belsky, 1984; Benson, 1997; Bronfenbrenner, 1979; Lerner, 2006; Lerner, Fisher, & Weinberg, 2000). Recent data in Canada indicated that youth between the ages of 15 to 34 years old were becoming increasingly diverse when compared with prior generations (Statistics Canada, 2018). For example, Canadian youth today are more multicultural, educated, and more frequent users of the Internet than prior years. However, these youth are also more likely to experience new challenges when compared with other generations, such as being more likely to report discrimination, having fewer full-time employment opportunities, and subjected to Internet bullying (Statistics Canada, 2018). Consequently, such aforementioned challenges increase the risk for youth to develop mental health problems, sometimes compromising their meaningful engagement in society (Evans, 2006; Lerner, Lerner, Almerigi et al., 2005; McCay, Quesnel, Langley et al., 2011; Merikangas, He, Burstein et al., 2010; Phelps, Balsano, Fay et al., 2007; Statistics Canada, 2018; Wilkinson & Pickett, 2011). In fact, suicide acted as the second leading cause of death for youth in Canada ages 15 to 24 years old (Statistics Canada, 2014).

Despite the value of research on youth development and wellbeing at a national level, measuring diverse youth experiences in Canada has become increasingly multifaceted and complex. As a result, data on youth wellbeing outcomes often show equivocal findings. This variability can sometimes occur because of the challenges in assessing how measurements related to the state of youth at national levels accurately reflect youth within diverse local community contexts (Bradshaw, Noble, Bloor et al., 2009; Eccles & Gootman, 2002; Lerner et al., 2000; 2005; McGregor, Coulthard & Camfield, 2016). It is also the case that development and wellbeing are typically conceptualized across many dimensions, such as socioeconomic status,

education, civic and community participation, relationships, and physical or psychological health (OECD, 2008). Assessing development and wellbeing across these multiple dimensions can add complexity in producing consistent findings of outcomes for young people in a given region (Brown & Corbett, 2003; Moore, Theokas, Lippman et al., 2008; Pollard & Lee, 2003). It is important to assess the state of youth at national levels, in addition to appraising how to best measure youth wellbeing and their growth within and across neighbourhoods (Bean & Forneris, 2006; Benson, 1997; Eccles & Gootman, 2002; Lerner et al., 2000; 2005; Tiffany et al., 2012; Wright et al., 2006). It is critical then, to monitor the development and wellbeing of these diverse youth across dimensions of their environmental systems, considering data collection techniques, and multi-method perspectives to assess a more comprehensive understanding of youth outcomes.

The present dissertation comprised two studies that sought to capture the multiple ways in which youth wellness is understood. Accordingly, this dissertation emerged out of a partnership with two community organizations focused on supporting youth wellbeing: Laidlaw Foundation (in study 1) and Lakeshore Arts (in study 2), situated in the urban centres of Toronto, Canada. Study 1 assessed concepts and measures related to youth wellbeing by systematically analyzing, comparing, and contrasting wellbeing indices using a meta-narrative analysis approach. Study 2 included youth within a local community context and evaluated the impact of a school-based arts program on the socio-emotional wellbeing of these adolescents using a multi-method, multi-informant evaluation design. The present studies were positioned to leverage an understanding of the ways in which youth development and wellbeing have been conceptualized and measured, as well as identifying areas where variability or consistency has been found. Identifying important facets of youth development and wellbeing can add insight into important

factors that enable youth to thrive within their environments, and ways in which to capitalize on their strengths and assets in order to better understand youth who come from various socioeconomic, demographic, and cultural backgrounds, reflecting Canada's diversity.

### **Approaches to Monitoring Youth Wellbeing**

Data collection to measure and monitor youth development and wellbeing at national levels is often based on a system of empirical inquiry that stems from quantitative methodology and large-scale databases (e.g., national achievement scores; medical records (Ben-Arieh, 2000; Lerner, 2006). Indeed, it is well documented in the literature that such methodologies are informative and efficient at providing a snapshot of how youth are doing at a given moment within regions and across nations (Booyesen, 2002). At the regional level, data are collected and used to assess wellbeing across multiple dimensions, such as employment, education, or health (Moore et al., 2008; O'Hare, 2012; OECD, 2008). Subsequently, data across dimensions are aggregated, and can be used to rank youth from one nation relative to international standards (CYP, 2013; OECD, 2015; UNICEF, 2013). For example, one international report ranked Canadian youth between the ages of 15 to 29 as second among 54 countries in having the highest overall *youth development* (Youth Development Index, 2013). Another report ranked young people in Canada between the ages of 0 to 17 as 17<sup>th</sup> among 29 countries in having moderate *overall wellbeing* (UNICEF, 2013). It appears that Canadian youth can be positioned into differing ranks of such reports, adding challenges to understand the state of youth development and wellbeing in Canada.

Despite efforts to monitor youth wellbeing and their growth at national and international levels, reports on the state of youth development and wellbeing often show equivocal findings (Brown & Corbett, 2003; Hur & Testermann, 2012; O'Hare, 2012). A notable challenge is that

young Canadians are monitored across criteria using widespread, socially-constructed definitions of youth, and are often assessed using data across multiple dimensions within their environments (Benson, 1997; OECD, 2008). Indeed, evidence suggests that one standardized conceptualization of the dimensions related to development and wellbeing for young people across nations does not exist (Ben-Arieh & George, 2001; Booyesen, 2002; Brown, 2008; Hauser, Brown, & Prosser, 1997; McAuley & Rose, 2010; O'Hare & Gutierrez, 2012; OECD, 2008). It is important to acknowledge the diversity to incorporate and define dimensions of development and wellbeing, as it is often in response to a particular theoretical framework, the values and norms placed upon wellbeing within a given region, community or organization, or the availability and access to data (McGregor et al., 2016). It is also critical to disentangle the variability in concepts and measures related to assessing youth because using different measures can contribute to variability in outcomes and can implicate differing societal consequences in the availability of policies, services, and program to support youth needs (Moore et al., 2004).

Data also derive from varying units of analysis, including macro-level data (e.g., national-level statistics) and micro-level data (e.g., self-reported responses). Although valuable, macro-level, large-scale data derive from an objective perspective (Ben-Arieh, 2005; Brown, 2008; Land, 2012). Data that derives from the micro-level incorporates the subjective perspective, which is self-reported from the young samples themselves (Bastos, Fernandes, & Passos, 2004; Ben-Arieh, 2005; Fernandes, Mendes, & Teixiera, 2012; Moore et al., 2008), and in this case, data that are youth-reported. Lastly, data on development and wellbeing are often based on identifying indicators of risk, prevention, and maladaptive attributes among youth, such as data related to economic deprivation, drug and alcohol consumption, or rates of violence, suicide, or death (Lerner, 2004; Moore et al., 2008; Phelps, Balsano, Fay et al., 2007). However,

it is also important to conceptualize youth wellbeing by focusing on the positive characteristics and assets that youth bring to a society (Eccles, Barber, Stone, & Hunt, 2003; Damon, 2004; Larson, 2000; Lerner, 2004). This approach uses indicators that are strengths-based, such as data related to post-secondary completion rates, economic prosperity, engagement in civic or community activities, or having access to good quality services and programs (Lerner, 2004; Moore et al., 2004; 2008). A long-standing focus on negative facets has often necessitated that resource allocation, policymaking, and governmental expenditure be ascribed to youth programming and services that ameliorate or prevent problem behaviours (deficit-reduction), rather than providing resources that help youth navigate through everyday challenges, capitalize on desirable behaviours and abilities, and actualize their potentials (Lerner, 2004; Lerner et al., 2000). Thus, it is critical to monitor youth using positive indicators of development, rather than focusing solely on indicators that monitor youth risks or the absence of negative attributes (Moore et al., 2004).

The aforementioned findings are further hindered by the fact that Canada is regarded as a highly culturally and demographically diverse country (Statistics Canada, 2018). That is, there is significant heterogeneity among youth across ethnicities and religions, in addition to demographics and socioeconomic status (SES). It is possible that large-scale data may not be adequately reflective of diverse youth experiences in Canada. This is because such data emerge from a particular set of assumptions – often using quantitative and objective metrics – which indicate possible complexities to fully capture relevant or meaningful perspectives from youth within their local community contexts (Billari, 2004; Hur & Testerman, 2012; McGregor et al., 2016). Other factors affecting a comprehensive understanding of youth within communities include the importance placed on societal norms and values, institutional factors, or cultural

expectations within and between nations (Billari, 2004). Importantly, large-scale databases and national statistics across nations and within Canada have been shown to under- (or mis-) represent particular cohorts of individuals who are often brushed to the margins of society, such as those in low-income, among Indigenous communities, newcomer immigrants and refugee populations, youth in homeless conditions, or youth in detention centres (Bradshaw, Hoelscher, & Richardson, 2006; McCreary, 2015; Stepping Up, 2016). Such missing perspectives elucidate the challenges with data collection across sectors of society, and the increased need to gather data from these marginalized communities. Studies suggest that gathering data, such as self-reported surveys, from these populations is often limited by factors, such as language barriers, temporary housing conditions, or legal restrictions (Bradshaw, Hoelscher, & Richardson, 2006; McGregor et al., 2016).

### **Theoretical Framework to Conceptualize Youth Wellbeing**

Given the multiple complexities that surround a youth's environment, the overarching theoretical framework that is used in the present dissertation to conceptualize development and wellbeing stems from a socio-ecological systems perspective (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 1998). This framework characterizes individuals who are embedded within multiple layers of environmental systems, and who develop across proximal and distal layers of their surroundings. In the context of youth wellness and their growth, data reflect the systems that make up these environments. The most proximal environments (i.e., the microsystem) are the most direct and influential systems, closely connected to the individual youth, including family, peers, households, and schools. Next, the mesosystem describes the interactions and connections that take place within the youth's influences in the microsystem, such as interactions between parents and schools. Next, the exosystem incorporates influences

that indirectly affect the youth, such as neighbourhood quality, parental workplace conditions, or access to services and programs. The most distal environment (i.e., the macrosystem) encompasses layers of the environment that are indirectly related to the individual, including governmental bodies, societal norms and values, policies, and overall economic conditions (Bronfenbrenner, 1979). Such a framework is important in gathering and assessing data that stem from these multiple systems and across dimensions. In the present dissertation, it was also important to consider Bronfenbrenner's socio-ecological systems framework, as it would provide a foundation by which to elucidate the comprehensive and multidimensional ways in which youth development and wellbeing is conceptualized and measured.

### **Concepts and Measures Related to Youth Development and Wellbeing**

Evidence suggests that some measures related to a youth's environment within a nation show good validity and reliability, such as a measure of SES (Boyce, Torsheim, Currie, & Zambon, 2006). For example, Boyce and colleagues (2006) found that the *Family Affluence Scale* (FAS), a four-item measure of SES, was a good representation of economic wealth among youth ages 11, 13, and 15 years old across 35 countries. Although there was substantial variability in FAS scores across countries, the authors found that the FAS had good internal consistency, indicating that the scores for the four items in the FAS highly correlated with one another. In addition, the FAS showed good criterion validity, indicating that it was an accurate and relevant measure to assess family wealth and income for these youth. This is because the FAS strongly correlated with other traditional national assessments of economic wellbeing and wealth of a country, including the Gross Domestic Product (GDP; "the total value of products/services per person exchanged globally" p. 477). Additionally, this measure has been shown to positively correlate with other measures of wellbeing, such as with health, indicating

that the lower the FAS score, the lower the health score within a country (Boyce et al., 2006; Torsheim, Currie, & Boyce et al., 2004).

Conversely, other studies suggest that particular concepts or measures can represent vastly different ideas across contexts, acknowledging the variability in contextual and demographic understandings among youth (Batista-Foguet et al., 2004; Schwartz & Melech, 2000). For example, such differences may be a result of culture and ethnicity, being situated within an urban or rural setting, or the importance placed upon obtaining economic wealth within a society (Batista-Foguet et al., 2004; Schwartz & Melech, 2000). In one study, Batista-Foguet and colleagues (2004) found weak associations when comparing items on the FAS across countries, and suggested that FAS items are not equally relevant in each country and, as a consequence, do not always contribute equally to international comparisons. Such findings indicate that some measures may be more or less reflective of youth within nations. It is also possible that many measures could be revised to reflect contexts relevant to the regional level.

Similarly, it has been shown that the ways in which concepts related to youth wellbeing are measured and understood can vary greatly within particular societies or cultures (Schwartz & Melech, 2000). For example, Schwartz and Melech (2000) examined the psychometric properties of a measure of *Worry* within and across 14 countries, in response to a 33-item self-report survey among university students and teachers. Question items were related to objective (macro) levels and subjective (micro) levels of worry. Macro items included factors, such as “worsening of destruction of the environment” or “damage to nature (forests, animals, etc.),” and micro items included factors, such as “my own death” or “my parents dying” (Boehnke et al., 1998; Schwartz & Melech, 2000). First, the authors supported findings from a previous study (Boehnke et al., 1998), and found that subjective-level worries were more positively associated with mental

health dysfunction (negative affect and anxiety) across all cultures. In contrast, objective-level worries showed no associations with poor mental health and weak associations with positive wellbeing (positive affect and satisfaction with life; Boehnke et al., 1998; Schwartz & Melech, 2000). These findings suggest the possibility that micro worries worsen one's capacity to cope with perceived highly threatening situations and thus reduce the importance to attend to macro worries. Alternatively, macro worries, consisting of concern for others or the environment, seem to remove concern for the self, and indicate greater importance for larger socio-cultural issues (Boehnke et al., 1998). Schwartz and Melech (2000) also found that, while all participants reported some form of worry, there were types of worries that were more or less relevant to particular nations over others (Schwartz & Melech, 2000). For example, it was shown that the macro-level worry of destruction to the environment in the environment domain was most pronounced among those in densely populated and highly polluted societies. While, the micro-level worry of fear of not having enough money to live on in the economic domain was found particularly among those in countries with greater income inequality (Schwartz & Melech, 2000). Additionally, in some cultures that place importance on the individual self rather than the collective society as a whole, micro level worries were shown to be more prevalent than macro level worries (Schwartz & Melech, 2000). These findings also suggest that individuals may show differences in attributing valid meaning to latent (unobserved) concepts related to wellness (e.g., "worry," "happiness," or "quality of life") as a function of age, social, economic, and cultural factors (Cummins, 2000; Schwartz & Melech, 2000).

### **The Need to Capture Youth-Driven Data**

Data in research on development and wellbeing are gathered from multiple units of analysis, including objective and subjective levels of data (Booysen, 2002). Indeed, although

objective and subjective measures can independently be useful and reliable in assessing wellbeing, or quality-of-life (QOL), it has been shown that these response types at times can show weak associations (Cummins, 2000). Cummins (2000) articulated that correlations between objective and subjective measures are often a function of the extent to which national-level standards align with personal-level perspectives. Thus, data derived from an objective, macro-level perspective cannot always fully capture perspectives at the micro, subjective-level that are relevant to youth.

A growing body of literature indicates the importance of assessing data derived from the subjective-level perspective (i.e., self-reported; Bastos et al., 2004; Ben-Arieh, 2005; Fernandes, Mendes, & Teixiera, 2012; Moore et al., 2008; 2014). Scholars suggest that the inclusion of self-reported data in research provides a form of assessment where respondents are treated as the subjects, rather than objects in the understanding of their own experiences (Ben-Arieh, 2005; 2007; Hagerty & Land, 2007; Lamb & Land, 2013). It is important to identify measures of wellbeing that consider perspectives from youth in order to offer data that provide a more comprehensive picture of youth wellness and the factors that affect their lives. In line with a theory of human agency (Bandura, 1982; Martin, Sugarman, & Thompson, 2003), studies indicate that incorporating subjective-level data can elucidate youth perspective and participation in the analysis of their own development, and can highlight facets of self-identity and personal agency (Bandura, 1982; Lerner, 2004). That is, using self-reported data focuses on empowering the youth perspective, and is closely linked to a rights-based approach (Ben-Arieh, 2005; 2007; OECD, 2008). In addition, Ben Arieh (2005) suggested that the adolescent subjective perspective does not always reflect what parents might have to say about them (i.e., parental respondents). Thus, it is important to incorporate data from youth perspectives in the understanding of their

development and wellbeing. Using such data makes it possible to incorporate youth-level influence over future decisions across social, policy, and governmental reforms that have an impact on their growth.

Scholars have also suggested that some dimensions are not fully captured in measurement from a youth's perspective, or constructs that lack valid and reliable measures, such as measures related to youth-focused mental health challenges and supports, social service availability, as well as community program provisions, accessibility, and participation (Hur & Testerman, 2012; Moore et al., 2004). Particularly lacking have been measures that are youth-centred and that provide a positive outlook on these individuals nested within their communities (Eccles & Gootman, 2002; Lamb & Land, 2013; Moore et al., 2004). These findings add complexity to fully comprehend the strengths, challenges, and needs of young people.

### **The Need to Capture Positive Youth Development**

Another important effort of research on development and wellbeing has been to identify relevant measures that capture not only negative youth outcomes, but also focus on positive youth development (PYD; Catalano et al., 2002; Damon, 2004; Larson, 2000; Lerner, 2004; Moore, Lippman, & Brown, 2004). PYD indicates taking a strengths-based approach in understanding the assets of youth and focuses on their interests, unique attributes, and potentials (Catalano et al., 2002). In addition, this approach seeks to highlight the interactions between youth and their communities (Damon, 2004; Eccles & Gootman, 2002; Lerner, 2004). Indeed, studies suggest that data collection in development and wellbeing research has placed more focus on the deficits and negative attributes of young people (Bastos & Machado, 2009; Fernandes, Mendes, & Teixiera, 2012; Larson, 2000; Moore et al., 2004; 2012; 2014). Such data focus on the maladaptive behaviours, adversities, and the increased risk for youth to capitulate to struggles

and who show adverse outcomes later in life. Accordingly, government resources, policies and programs are primarily allocated to services that focus on reducing or preventing problem behaviours (Lerner, 2004). Although many youth face challenges in their everyday lives, it is not simply the deficits that define them. That is, youth wellness and growth encapsulate environments that are fluid and malleable, rather than only on attributes that are static and hereditary (Lerner et al., 2005; Moore et al., 2012). Thus, youth positions in society, their challenges, and their interactions with the greater community may be used as a way to build on strengths, help develop their identities, and actualize their potentials (Eccles & Gootman, 2002; Larson, 2000; Lerner et al., 2005).

Proponents of PYD articulate that this approach comprises of five overarching categories. *Competence* refers to the skill development in social and cognitive tasks. *Confidence* includes positive self-worth, self-esteem and efficacy. *Connections* indicates positive bonds with external environments including to people, schools, and communities. *Character* refers to the extent to which an individual shows integrity, morality, and respecting values and norm within a society. Lastly, *care and compassion* indicates showing respect, empathy, and acceptance of diverse others (Lerner et al., 2005; Eccles & Gootman, 2002; Erdem et al., 2016; Roth & Brooks-Gunn, 2003).

It has been shown that capturing PYD with relevant measurement is related to personal agency, strengths, and achievements, as well as positive bonds and attachment with important others, community participation and engagement (e.g., in programs, hubs), and acknowledges efforts for communities to promote safe and supportive environments (Bean & Forneris, 2016; Catalano et al., 2002; Tiffany et al., 2012). It is particularly important to assess the impact of community services and program provision on positive youth outcomes within at-risk, inner-city

neighbourhoods. In this way, studies can incorporate relevant measures that more fully capture youth wellbeing within their communities. These assessments can also indicate possible routes for improvement and where more supportive and structural resources (both emotional and physical supports) are needed, such as community programs or hubs, support networks (from peers or adults), and educational resources (Holloway & LeCompte, 2001; Wright et al., 2006).

### **Summary and Rationale for the Studies**

In light of the above, the present doctoral dissertation sought to address gaps in the literature and explore areas of influence relevant to youth wellbeing from two environmental levels: A macro-level perspective and a micro-level perspective. Additionally, the present studies emerged out of partnerships with two community organizations, Laidlaw Foundation (2015) in study 1 and Lakeshore Arts (2016) in study 2. Both organizations work with youth in the urban centres of the Greater Toronto Area (GTA). It was important to establish community partnerships with organizations that are youth-centred in order to engage with stakeholders who could provide additional and unique insight into youth awareness from multiple perspectives. It is well documented in the literature that building and fostering trust across academic-community partnerships can advance research and evaluation for given topics of interest (Masuda et al., 2011; Minkler, & Wallerstein, 2011), as well as for marginalized, or vulnerable populations (Christopher, Watts, McCormick, & Young, 2008; Chung, Jones, Jones et al., 2009). In the case of the present study, it was invaluable to incorporate such partnerships in the advancement and understanding of research focused on youth development and wellbeing.

Specifically, the present dissertation sought to:

- (1) Examine youth measurement from a macro-level perspective. The goal of study 1 was to undertake a meta-narrative analysis with a critical appraisal of national-level reports

(using indices) to examine data collection that focus on assessing the state of youth wellbeing. A comprehensive analysis of each report would be used to identify variability across indices in the conceptualization and measurement of wellbeing. Lastly, the goal was to provide a synthesis of findings, critique component of indices and possible recommendations.

- (2) Examine measurement from a micro-level perspective. The goal of study 2 was to conduct an evaluation of a school-based community art program to address the concepts of youth development and wellbeing among diverse, multicultural youth situated within inner-city areas. The focus of study 2 was on the implementation process of the program, and the ways in which a program could provide beneficial outcomes related to socio-emotional growth, school experience, and identify the supportive networks in their environments.

It was important to incorporate a multi-method design in the present dissertation by using both quantitative and qualitative data techniques. Using multiple methods can illustrate a more comprehensive understanding of the study questions and findings related to youth development and wellbeing (Caracelli & Greene, 1993; Cook & Campbell, 1979; Creswell & Clark, 2007). In the case of the present dissertation, both studies incorporated quantitative metrics and numerical statistics, such as from systematic analyses, self-reported survey data, and observational measures. Both studies also incorporated qualitative data, such as from stakeholder interviews and open-ended responses from participants under study. A multi-method design has been shown to strengthen study questions using advanced statistical methods, numerical outcomes, as well as incorporating discourse with stakeholders, community partners, and under-served populations (multicultural youth situated in areas of high need). A comprehensive assessment adds

understanding to important community issues, as well as enhances knowledge of best practices for study development, and the benefits of high quality community programs (Grimshaw et al., 2000; Wright et al., 2006).

### **Research Questions**

The present dissertation addresses two broad research questions:

- (1) *How is youth wellbeing conceptualized and measured at the macro-level, and across dimensions in multidimensional indices (within reports)?*

Using a synthesis of findings from reports (using indices), it is hypothesized that understanding youth wellbeing within macro-level measurement can provide insight into existing gaps and variability in concepts and measures across wellbeing dimensions.

- (2) *How is youth wellbeing supported at the micro-level, local community context?*

It is expected that understanding youth contexts within a local community program can identify themes and environments that contribute to socio-emotional wellbeing and growth – constructs that can be elucidated by addressing the perspectives of diverse, multicultural youth within their local schools and communities.

## **Chapter 2: Study 1**

**A critical appraisal of youth wellbeing and development across indices**

**informed by a meta-narrative analysis**

## **Introduction**

An emerging effort for developmental researchers and youth advocates has been to fill gaps in the collection and availability of data related to youth wellbeing (Lerner et al., 2000). Defining youth development and wellbeing is multifaceted and complex, given that youth are becoming increasingly demographically diverse and multicultural (Statistics Canada, 2018). For example, national-level data indicated that Canadian youth between the ages of 15 to 34 years old accounted for twenty-five percent of the country's population (Statistics Canada, 2018). In addition, the number of youth between the ages of 25 to 34 years old showed significant increases compared with previous years, while the number of youth aged 15 to 19 years old decreased (Statistics Canada, 2018). Over one quarter of Canadian youth also indicate being a member of an ethnic minority group, or indicate having been born outside of Canada (Statistics Canada, 2018). Thus, in order to generate credible understandings of development and wellbeing, it is important to more fully understand the concepts and measures related to monitoring these diverse youth.

With that said, one way to aggregate and convey data on youth wellbeing is through national-level reports, or in the form of an index (Land, Lamb, Meadows, & Taylor, 2007; Land, Lamb, & Mustillo, 2001). A wellbeing index provides evaluations of the state of optimal growth and development for individuals living in given nations (Ben-Arieh, 2000; Moore, Theokas, Lippman et al., 2008; O'Hare, 2012). Such an index also typically incorporates a multidimensional combination of features that define the wellbeing of a particular population sample, such as incorporating data related to the individual's personal-, social-, or community-level environments (Ben-Arieh, 2007; Land, Lamb, & Zheng, 2011; Lee, 2014; O'Hare & Gutierrez, 2012; Pittman & Irby, 1997). For example, dimensions of wellbeing relate to how a

young person is doing in the areas of health, education, or economic wellbeing (Ben-Arieh, 2000; 2006; Land, Lamb, & Zheng, 2011; Moore et al., 2008). Such data across dimensions are subsequently aggregated to formulate the index, and provide big picture data that are readily available, offering a snapshot of outcomes and a depiction of societal-level trends (Ben-Arieh, 2007; Bradshaw et al., 2006; Lerner, 2000; 2006). This information becomes relevant for youth advocates, policy-, and decision-makers to monitor trends and outcomes of young people and to promote discussions on the investment and importance that a society places on supporting youth needs (McGregor, Coulthard, & Camfield, 2016; Moore et al., 2003).

Traditionally, however, there has been variability in the ways in which different indices conceptualize and measure wellbeing (Bradshaw et al., 2006; Fernandes et al., 2012; McGregor et al., 2016; O'Hare, 2012; O'Hare & Lee, 2007). That is, indices differ in the ways in which wellbeing is defined, and in the selection of index features related to wellbeing, including the domains, indicators, and measures that make up a multidimensional index. Consequently, indices may offer different conclusions on wellbeing outcomes contingent upon the index components that are selected by an organization that creates the index. Indeed, disparities across conceptualizations of wellbeing can offer important and unique contributions to how young people are doing within and across nations (Moore et al., 2003; McGregor et al., 2016). However, such discrepancies also contribute to increased variability in the understanding of youth. Here, some indices may report that youth in a region are fairing well, while other indices may indicate that youth in that same region are falling behind, relative to international standards. Lastly, to date, research in wellbeing indices has primarily focused on data related to younger children sometimes grouped with youth (e.g., birth to 18 years old; UNICEF Canada, 2016), or youth grouped with older adults (e.g., ages 18+; Canadian Index of Wellbeing, 2016). Less empirical

attention has been placed on the extent to which data on wellbeing in large-scale indices are relevant and specific to youth. It is critical then, to address the gaps in data collection that seeks to monitor the wellbeing of these diverse youth. Accordingly, the present study examined concepts and measures related to dimensions of youth wellbeing across multidimensional indices.

## **Background**

Studies have shown the common and ongoing use of wellbeing indices across nations (Ben-Arieh, 2000; Ben-Arieh & George, 2001; Land, Lamb, Meadows, & Taylor, 2007; Land, Lamb, & Mustillo, 2001; Moore, Theokas, Lippman et al., 2008; O'Hare, 2012). First, indices provide data on the state of wellbeing at a given moment, and use measurements that are readily available to a given society as unitary quantitative snapshots of data on wellbeing (Bradshaw et al., 2006; Land et al., 2007; 2011). Second, wellbeing in indices is conceptualized across multiple dimensions according to particular disciplines or organizations that create the index and incorporate available measures that assess such dimensions (O'Hare & Gutierrez, 2012). Indices also provide information to monitor societal-level trends at one point in time, and can discern similarities and differences between nations across dimensions of wellbeing for these young people (Ben-Arieh, 2007; Land et al., 2001; 2007; 2011). Importantly, such data can highlight outcomes on youth, as well as promote discourse among policymakers and youth advocates on the strengths and gaps across societies in supporting the needs of youth. Such discussions become relevant to elucidate areas where the provision of policies, supports, and services are most available or particularly lacking to enable these individuals to thrive (Ben-Arieh, 2000; Pittman & Irby, 1997). While valuable, indices have been criticized for their conceptual and methodological limitations (Ben-Arieh & George, 2001; Bradshaw et al., 2006; Fernandes et al.,

2012; O'Hare, 2012; O'Hare & Lee, 2007). First, the concept of wellbeing across indices is not standardized (Ben-Arieh, 2006; O'Hare & Gutierrez, 2012; Moore et al., 2008). This indicates that conceptualizing development and wellbeing within an index is contingent upon the values, ideologies, and importance given within a particular organization that creates the index in incorporating some dimensions and measures over others (Ben-Arieh, 2000; Moore et al., 2008; O'Hare, 2012; OECD, 2008). Second, discrepancies in conceptualizations of wellbeing and domain selections are also limited by the availability of data (O'Hare, 2012). Third, indices use a combination of measurement units that place more focus on objective-levels of data, including societal (e.g., national statistics) or household (e.g., parental/caregiver respondents) units of analysis, rather than subjective-levels of data that are self-reported from young people (Bastos, Fernandes, & Passos, 2004; Ben-Arieh & George, 2001; Fernandes et al., 2012; Goldin, Patel, & Perry, 2014; O'Hare & Gutierrez, 2012). Consequently, many indices do not always provide data on the subjective perspective from youth, and thus lack the ability to deliver a comprehensive, multi-informant assessment from multiple perspectives of wellbeing outcomes across different units of measurement (Bastos & Machado, 2009). Lastly, indices often focus on defining young people through a deficits-based framework, which emphasizes weaknesses and challenges, rather than incorporating measures that focus on positive development and strengths (Bradshaw et al., 2006; Damon, 2004; Larson, 2000; Masten et al., 2005; Moore et al., 2007; O'Hare, 2012; O'Hare & Lee, 2007). Given such limitations, it is critical to disentangle the variability in conceptual and methodological understandings of youth wellbeing data across indices.

Although previous research has examined index variability, less empirical attention has examined the extent to which data in wellbeing indices focus on youth (Land et al., 2007; 2012). No research to date has examined how wellbeing is conceptualized for youth by comparing and

contrasting themes across indices, and the variability that exists in the selection of index features, including domains (e.g., economic wellbeing, education), indicators (e.g., youth (un)employment, educational achievement), and measures (e.g., rate of youth suicide, average test scores in reading or math). The objective of the present study was to examine the ways in which the authors of these indices approach the topic of youth development and wellbeing from different perspectives across domains. In order to achieve this objective, the study used a meta-narrative method (Greenhalgh & Wong, 2013). A meta-narrative approach synthesizes findings across different perspectives and uses a thematic codebook of questions related to the topic at hand as it derives from varying disciplines or bodies of literature (Greenhalgh & Wong, 2013). In the case of the present study, the meta-narrative method was used to examine consistencies (and lack thereof) in conceptual and methodological approaches to the study of wellbeing. This analysis was important as it permitted the assessment of variability across conceptual and theoretical frameworks, processes in the collection of data, and measurement strategies. Lastly, a critical appraisal of indices identified gaps, and areas of dimensions that are given more or less attention, as well as recognizing priorities in the assessment strategies, with a particular focus placed on a youth-centred framework.

### **Conceptualizations of Youth Wellbeing Across Indices**

Empirical evidence suggests that conceptualizations of wellbeing across indices can be multidimensional or unidimensional (Booyesen, 2002; Decancq & Lugo, 2013; O'Hare & Gutierrez, 2012; OECD, 2008). For example, a multidimensional index of wellbeing incorporates and aggregates domains of health, education, and economic wellbeing. Alternatively, a unidimensional index focuses solely on one aspect of wellbeing, using an epidemiological, educational, or economic perspective. In the present study, it was of interest to

include indices that were multidimensional in order to compare the various ways in which dimensions of wellbeing are selected and aggregated by the organization. Considering wellbeing across multiple dimensions can also necessitate the inclusion of data from multiple perspectives and areas of inquiry (Bradshaw et al., 2006; Cummins, 2000; OECD, 2008). That is, researchers from different disciplines come together, and gather data in their areas of expertise across wellbeing dimensions in order to write reports related to the topic at hand. Such data on youth wellbeing include indicators, which are selected and weighted to fit into and define respective overarching domains, such as educational achievement. Measures are also included as the precise assessment tool for a given indicator, such as average test scores in mathematics (Bradshaw et al., 2006; Land et al., 2007; Moore et al., 2007; O'Hare, 2012). Lastly, the data source is from where the measure derived, such as the OECD Programme for International Student Assessment (PISA) educational testing results. Thus, a structure is created that incorporates domains, indicators, measures, and data sources, which make up the features that encompass a wellbeing index (O'Hare, 2012).

Many scholars have articulated the importance of considering approaches to conceptualize wellbeing that combine multiple dimensions related to a youth's environmental influences in the construction of an index (Bradshaw, 2006; Fernandes, Mendes, & Teixeira, 2012; Hagerty & Land 2007; Land et al., 2007; 2012; Marmot & Wilkinson, 1999; McGregor et al., 2016; Moore et al., 2014). Studies reviewing multidimensional indices have shown that wellbeing is conceptualized across a variety of developmental perspectives that inform the empirical approach or theoretical lens upon which wellbeing is based (three approaches are briefly described below; Bronfenbrenner, 1979; Cummins, 1996; Marmot & Wilkinson, 1999; Moore et al., 2014). First, a socio-ecological perspective identifies youth who develop within

nested systems, including among family, peers, and communities, and thus informs an approach to the study of wellbeing and data collection within an index that encompasses the distal and proximal dimensions that have an impact on the individual (Bronfenbrenner, 1979). Additional perspectives include the social determinants of health or quality-of-life (QOL) approaches and inform the gathering of data related to the determinants that can have an impact on personal-level health or quality of life, respectively (Cummins, 1996; Marmot & Wilkinson, 1999; Wallander & Koot, 2016).

Evidence has also highlighted some consensus and some challenges associated with the selection and number of domains that make up wellbeing (O'Hare & Gutierrez, 2012; OECD, 2008). For example, in a recent comparative analysis paper of nineteen multidimensional child wellbeing indices from international, national, and regional levels, O'Hare and Gutierrez (2012) found some overlap in the number of domains used to represent wellbeing. The study found that most indices were constructed using six or seven domains. Yet, some indices included as low as two, and as high as eight domains. Such findings indicate that while some agreement exists in the inclusion of six to seven domains to conceptualize wellbeing, there is still substantial variability. Moreover, studies have found common themes emerge for domain inclusion, with most consensus related to the broad dimensions of "Economic Wellbeing," "Education," and "Health" (Fernandes, Mendes, & Teixeira, 2012; Lamb & Land, 2013; O'Hare & Gutierrez, 2012). However, studies have also shown that the labeling of domain names can largely vary (Brown, 2008; Hauser et al., 1997; O'Hare & Gutierrez, 2012). For example, O'Hare and Gutierrez (2012) found that the domain labels used to examine the overarching dimension of "Education" across indices included labels that ranged from "Education," "Learning," to "Education Attainment." Thus, such findings indicate the variations of labeling overarching clusters of

dimensions, even when assessing similar items.

Other studies suggest that conceptualizing wellbeing across domains is influenced and restricted by the availability of data (Ben-Arieh, 2000; Bradshaw et al., 2006; 2009; Lippman et al. 2011; Moore et al. 2008). That is, additional consistencies and discrepancies in domain selection exist due to the strategies used in the access to, and collection of, data. Data collection to assess the wellbeing of young people are often driven by population-level data (national statistics), or household-level data (parental/caregiver respondents), with less focus placed on data derived from the sample under study (i.e., subjective-level, self-reported data; Ben-Arieh, 2000; Fernandes et al., 2012). Indeed, studies highlight that incorporating self-reported data provides a youth perspective and adds a holistic approach in the understanding of what the individual reports about their own environments (Ben-Arieh, 2005; Lamb & Land, 2013).

Although recent efforts have been made to incorporate such data into each domain across wellbeing indices, one critical review paper found that data on this unit of analysis are still scarce (Fernandes, Mendes, & Teixeira, 2012). Fernandes and colleagues (2012) identified the extent to which indices included self-report measures and showed that, although some indices included this unit of data, analyses were largely based on an aggregation of measures across multiple units of analysis, such as combining population-level statistics and subjective-level self-reports, into one outcome. These findings suggest a current limitation in the creation of wellbeing indices, whereby a youth-reported perspective is hindered due to the aggregation of varying units of analysis for one outcome. In addition, studies have also shown that self-reported data is often focused on younger children, or combines data across broad age ranges (Fernandes, Mendes, & Teixeira, 2012; Moore et al., 2012; 2014; OECD, 2008). No studies to date have focused on the extent to which self-reported measures are specific or relevant to youth. Indeed, given the

scarcity of wellbeing indices or reports focused solely on youth, much of the above work (e.g., Ben-Arieh) refer to their research population as ‘children,’ in accordance with the definition of a child (between the ages of 0 to 18 years old) from the Convention on the Rights of the Child (CRC; United Nations, 1989). Given the inclusion of broad age ranges in indices, and of measures using the term ‘youth,’ it was decided to incorporate such previous research in the advancement of the topic in the present study.

### **A Strengths- vs. Deficits-Based Approach**

Indices have traditionally incorporated indicators of wellbeing that focus on the deficits of young people (Moore, Lippman, & Brown, 2004). Indeed, researchers have made efforts to conceptualize the topic of development and wellbeing through a strengths-based approach, indicating the importance of incorporating positive measures in order to capitalize on youth assets, interests, unique attributes, and potentials (Damon, 2004; Larson, 2000; Lippman et al., 2011). In line with models of positive youth development, studies suggest that the indicator movement has been focused on using measurement that assesses the interactions between environmental, systemic, and personal experiences among youth (Eccles & Gootman, 2003; Larson, 2000; Lerner et al., 2005).

Scholars also articulate that a push toward incorporating positive indicators is at the heart of a rights-based approach to the study of wellbeing, and can also validate youth by providing them with agency over their lives (Larson, 2000). However, scholars suggest that positive measures can sometimes be harder to describe and measure than negative facets, and thus necessitates further conceptual clarity and psychometric rigour (Ben-Arieh & George, 2001; Moore, Lippman, & Brown, 2004). Indeed, recent efforts have been made to define and validate positive indicators, such as by engaging stakeholders, gathering in conferences, promoting

discussions among researchers and policymakers, and outlined in a compendium of youth development (cited in Moore et al., 2004). In addition, Ben-Arieh and George (2001) articulated that definitions can be deemed as follows, “negative indicators are indicators that are measuring the existence of harmful aspects in children’s lives or their absence (i.e. child abuse or injuries). Positive indicators are indicators that measure the existence of desired and positive aspects of children lives (i.e. success in school, supporting family, and leading a healthy life style.” (p. 611). Similarly, studies focused on positive youth development list examples of measures that take a strengths-based approach, and include items such as post-secondary completion, entrepreneurial endeavours, positive interpersonal connections, and perceptions of respect and positive youth support, such as from the government or community, within a given society (Lerner 2004; Lerner et al., 2005; Moore et al., 2004; 2012; 2014; Pittman & Irby, 1997).

For example, Moore and colleagues (2012) identified the importance of incorporating positive indicators into an index based on the wellbeing of children (ages 6-11) and adolescents (ages 12-17). The authors found significant variability in the rates at which these young individuals are affected by both positive and negative factors in their surrounding. Findings indicate the importance of addressing both forms of data in the understanding of youth health and progress. Other studies have found that, although indicators attempt to conceptualize wellbeing through a positive development framework, it is often limited in the availability of data for such indicators, and thus cannot always fully capture or measure positive outcomes (Moore et al., 2004; 2014). However, no studies to date have examined the rate at which measures include positive indicators across indices, particularly as assessed for the socially-malleable term of youth cohorts.

### **Discrepancies in Definitions of Youth**

Studies have shown that difficulties exist to define the wellbeing of young people, which is often a function of conceptual definitions and available measures according to particular disciplines (Bradshaw et al., 2006; 2009; Cummins, 2000; Eccles et al., 1993; 2003; Erickson, 1968; Land et al., 2007; 2011; Zarrett & Eccles, 2006). Yet, a common, widespread term is used for youth. Studies suggest that the term ‘youth’ can be conceptualized via an interaction between one’s chronological age and the biological, cognitive, social, and cultural processes that can vary over time and across environments (Aapola, 2002; Clark-Kazak, 2008; Erickson, 1968; Laslett, 1989). Here, youth may be categorized according to a prescribed age range using numerical designations (time). Youth definitions may also be influenced by governing or institutional systems and corresponding to regional legislation, such as within education, health care, or legal and judicial systems (environment; Aapola, 2002; Clark-Kazak, 2008). For example, young individuals in the education system are expected to complete particular grades in line with their chronological ages. Accordingly, measurement is used to assess youth wellbeing in the education system for prescribed age groups, such as attendance or completion rates of high school or post-secondary education. However, age inclusion across sectors of governmental bodies and legislations can largely vary, such as for voting participation, military service, driving, getting arrested, or engaging in paid employment (Cohen, 1980). Additionally, the term youth is also based on the social, cultural, and contextual norms of a given society, indicating possible variations in definitions across cultures. In some cultures, youth-hood may start at puberty or marriage; in other cultures, youth ages may not have any specific criteria (Clark-Kazak, 2008).

Age criteria are also varied across institutional systems and at times are contradictory (Aapola, 2002; Cohen, 2002). For example, young individuals may receive distinct health care services after the age of 12, or judicial systems will accuse youth under the age of 18. Youth are

also defined as a function of developmental and cognitive processes that often work in tandem with chronological age. This form of development signifies a certain intellectual or behavioural maturation that may indicate shifts in expectations of roles and responsibilities, and denotes an arbitrary period between childhood and adulthood (Aapola, 2002; Cohen, 2002).

The existing variability in age conceptualizations of youth extends to the Canadian context. Such varied age ranges across measurement perpetuate equivocal and unclear understandings of youth and can depend on the context, or from where measurement is derived (sources of data). For example, national-level survey data from Canada define youth between the ages of 16 to 28 years old (Statistics Canada, 2011; 2013). Additionally, another government body from the Human Resources and Skills Development of Canada includes youth between the ages of 15-24 years old (Statistics Canada, 2011). Meanwhile, the Ministry of Children, Community, and Social Services indicates ‘no clear definition’ for young individuals obtaining these services (Statistics Canada, 2011). Moreover, different age divisions exist for children, youth, and adults across provincial levels (Statistics Canada, 2011; 2013). Here, the division between childhood and adulthood may begin after 18 or 19 years old depending on the legislations of each province (CIC, 2016). Accordingly, there is a period between childhood and adulthood that denotes the term ‘youth,’ but the ages that are placed upon “children and youth,” “youth and adults,” or “young adults” often overlap (Statistics Canada, 2013).

In spite of such complexities, the lack of a clear, universal definition of ‘youth’ adds variability to the study of youth wellbeing. Indeed, having the term youth grouped into other socially prescribed age categories would naturally indicate that a youth encompasses multiple categories, with multiple conceptual understandings. For example, wellbeing measures may group broadly defined age ranges, and thus provide outcome data on broad groups of young

individuals, such as those from birth to age 18, or individuals ages 15-49. Here, a 15-year-old, for example, might be placed into two largely varying developmental age ranges that can each result in distinct outcomes and implications, such as the importance placed upon policymaking in some areas over others, or public- and private-sector resource allocations to some services over others. Such examples beg the questions of ‘What is a youth?’ and ‘How and in what ways are youth conceptualized and measured across a nation?’

### **The Present Study**

The present study included indices from reports on ‘youth’ and ‘young people’ insofar as they were labelled as such in a given index or measure, in order to identify similarities and variability across indices, terminology, and measurement. Given the gaps and existing variability in conceptualizations related to youth wellbeing and their age ranges, as well as across dimensions of wellbeing, it was critical then, to appraise how different bodies of literature address the topic of interest. Accordingly, the present study used a meta-narrative approach to examine the topic at hand across indices in order to systematically map out, interpret and synthesize findings across index contexts (Greenhalgh et al., 2009; Wong et al., 2013). Here, indices were used as the documents of interest, as they would provide diverse sources of information to appraise how wellbeing has been conceptualized across multiple disciplines and dimensions for youth.

It was important to undertake a meta-narrative analysis in the present study because such an approach teases out “the meaning and significance of the literature rather than producing an encyclopaedic inventory of every published paper on the topic” (Greenhalgh et al., 2009, p. 731). In the present study, a meta-narrative method was particularly suitable for three reasons. First, this approach encourages collaboration with stakeholders and community partners in formulating

the research questions and contributing to discourse throughout each phase in the analysis. Second, it was possible to use a systematic way to search the literature for documents across *multidisciplinary* standpoints, but that are not necessarily found in peer-review journals, rather in the grey literature base (Greenhalgh & Wong, 2013). Lastly, in using a meta-narrative approach it was possible to provide a way to make sense of the existing literature via a critical appraisal component to equate findings. A critical appraisal component could enable a form of dialogue that systematically organizes limitations and inconsistencies across indices, and introduces a way of assessing the given topic using a narrative method (Greenhalgh et al., 2009; Prilleltensky, 1997). Thus, the analysis in the present study sought to thematically explore current trends and variability across wellbeing indices (Collins & Hayes, 2010; Greenhalgh et al., 2009; Wong et al., 2013). This form of comprehensive review to examine variability across existing youth wellbeing indices has not previously been conducted.

More specifically, the present study sought to address gaps in the variability of conceptualizations of youth wellbeing across multiple standpoints within multidimensional wellbeing indices. In addition, the present study focused on extracting conceptual or theoretical approaches used to define wellbeing, and the consideration of incorporating a strengths-based approach to reflect positive indicators of youth development (Eccles et al., 2003; Larson, 2000; Lerner et al., 2005). It was also important to extract the similarities and differences in the selection and inclusion of domains, indicators, and measures used to make up an index. Indices would be compared and contrasted, with a focus placed on: (1) age range inclusion and definition of ages, where available, (2) index composition, including domain and indicator selection (i.e., the dimensions and definitions of wellbeing), (3) the validation process to selection index features (4) theoretical frameworks used to conceptualize wellbeing, (5) precise measurement

used (i.e., the assessment strategy, youth-focused data, and sources of data from where measures were derived), and (6) whether indicators focus on a deficits- vs. strengths-based approach to development. The present study also used a synthesis of data to critique the strengths and limitations related to age contexts, domain, indicator, and measurement selections, where data are misaligned, and where possible improvement efforts could be made.

## **Study Objectives**

In light of the above, the focus of the present study employed two overarching objectives. First, particular attention was paid to assessing the variability across indices that contained measures that use the term “youth” and second, presenting a critical appraisal of the findings in order to equate outcomes based on strengths and limitations of indices. Specifically, the structure of the present study highlights the following research questions to assess key elements across indices:

- (1) What is the context of the index (international, national, provincial)?
- (2) What is the age range inclusion for the sample populations in each index?
- (3) What is the rationale to define “youth” or “young people” across indices?
- (4) What is the theoretical lens and definition of “youth wellbeing” across indices?
- (5) How have different approaches across indices shown to create an index?
  - a. How were the index features selected or validated?
  - b. What are the domain selections and labels that are used to make up youth wellbeing?
  - c. What is the rate of overlap across domain and indicator selection and labeling?
  - d. What is the rate of overlap across measures and sources of data?

- (6) What is the rate of measurement that includes youth-specific data?
- (7) What is the unit of analysis within and across measures?
  - a. Where do the data come from, and what is the collection technique (i.e., national-level statistics, family/parental data, self-reported by youth)?
- (8) What is the rate across measures that take a strengths-based approach versus a deficits focus?
- (9) Using a synthesis of findings that reflect upon such variability, a critique component assesses strengths and limitations of indices related to:
  - a. Ages, domains, indicators, and measurement inclusions.
  - b. Where is most focus placed across indicators and domains?
  - c. What are some areas that show to be underdeveloped or could be improved?

## **Methods**

### **Community Partnership**

The present study emerged out of a partnership with Ryerson University and Laidlaw Foundation (LF), beginning in 2015. LF is a charitable community organization providing funding to youth who propose projects focused on youth development, wellbeing, and community engagement in the urban centers of Toronto, Canada. The goal of the partnership was to help LF strengthen the ability to document social impact on youth development and wellbeing. In addition, findings from the present study could better position LF to identify areas where they are concentrating funding resources that might be most fitting with addressing youth needs. Over time, this project could help LF identify grantees in showing their impact in critical domains, where previous data have been limited, or absent. Thus, over multiple stakeholder meetings with the Executive Director, staff, and with the Board of Directors, it was possible to help LF improve internal processes used to identify progress and impact of grantee projects related to youth development and wellbeing within Toronto communities. Importantly, Laidlaw was incorporated into each phase of the analysis and in formulating the research questions. This collaboration would permit findings in the present study to be particularly suitable and pragmatic to the goals of LF and in strengthening their philanthropic processes.

### **Study Design**

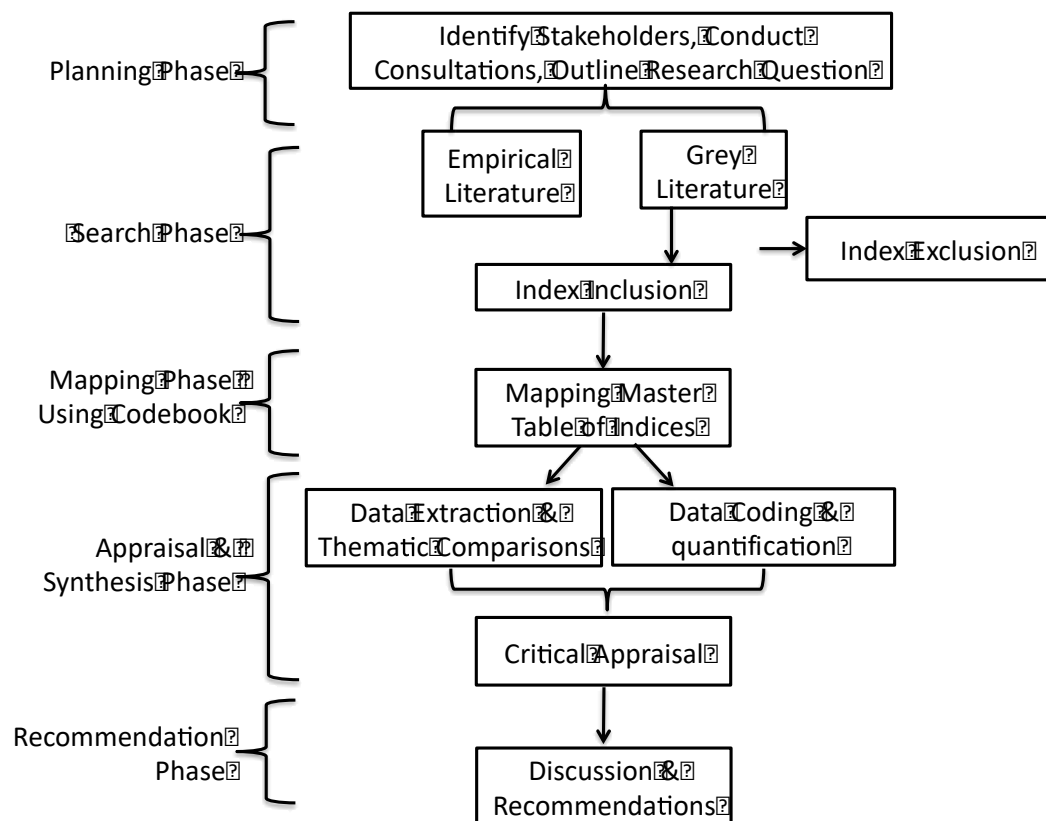
The present study was based on a meta-narrative analysis of youth wellbeing indices. As previously noted, this method was considered most suitable as a way to systematically analyze relationships (similarities and differences) across research question and areas of interest. In addition, using a meta-narrative approach was best suited in the present study as it provided a mixed-method assessment of themes and variables by using qualitative and quantitative data for

a topic that has not been synthesized and tested in this way (Collins & Hayes, 2013; Greenhalgh et al., 2005; 2009; Greenhalgh & Wong, 2013). The present study adhered to the guidelines in RAMESES standard protocol for a meta-narrative approach (Realist And MEta-narrative Evidence Syntheses: Evolving Standards; Wong et al., 2013).

Importantly, a meta-narrative analysis was used in the present study, as opposed to other traditional Cochrane-like approaches, such as systematic reviews or meta-analyses, in that the present study did not seek “to address the questions ‘what works?’ and ‘what is the effect size?’ but to illuminate and clarify a complex topic area and highlight the strengths and limitations of the different research approaches to that topic” (Wong et al., 2013, p. 21). That is, the present study did not look to exhaustively summarize data from one discipline (based on search engine optimization and a targeted-based approach), nor solely focus on an overarching integration of patterns or effect sizes, such as from substantive variables or numerical outcomes of a particular controlled intervention or trial (i.e., psychometrics, clinical trials, and effect sizes) to determine what is most effective. In addition, Otte-Trojel & Wong (2016) indicated that “relying solely on evidence generated from systematic Cochrane-like reviews that expressly filter out contextual influence and human factors may give decision- and policy makers only partial, or even misleading, information on which to base decisions” (p. 276). Thus, the present analysis was interested in comparing broad themes in literature bases regarding the topic of interest in assessing the ways in which commonly used multidimensional reports (using indices) similarly or differently approached (conceptualized and measured) the topic of youth development and wellbeing (Greenhaugh & Wong, 2013; Wong et al., 2013).

Thus, the present study used a systematic search of the grey literature to map out and interpret findings across indices related to youth wellbeing in order to identify the body of

literature, differences, similarities, strengths, and limitations within and across indices (for a conceptual framework guiding the study see Figure 1 below; Greenhalgh et al., 2009). The meta-narrative process incorporated discussion and collaboration with stakeholders and the community partner to formulate study questions, which also included ongoing iterative revisions. This analysis also allowed for the assessment of quantitative data related to methodologies across indices, and also supported the ability to tease apart the content and concepts of each separate index in a qualitative, thematic manner (Wong et al., 2013). Data were analyzed across themes in a codebook to address the study questions and assess detailed information in order to more fully examine variability across indices. Details of the six phases are further described below.



*Figure 1.* Conceptual framework and phases in the meta-narrative analysis (Figure modified from Greenhalgh et al., 2009).

## **Planning Phase**

In the planning phase, key stakeholders were included in order to incorporate a combination of organizations from diverse disciplines in the addition of perspectives for the present topic of interest. It has been suggested that having a team of organizations with diverse backgrounds can contribute to formulating the key research questions, as well as provide a more comprehensive analysis and appraisal of findings (Greenhaugh et al., 2005).

First, as previously noted, a partnership with Laidlaw Foundation was imperative in the exploration of indices on youth wellbeing, given that the focus of the organization was centred on assessing key indicators of youth wellness and where to allot grant funding for youth projects in the community. In line with community-based participatory research (CBPR; e.g., Christopher et al., 2008), it was imperative to discuss and integrate ongoing suggestions and feedback from Laidlaw at each phase of the study. In addition, Laidlaw provided the funding for the project to take place. Accordingly, all decisions and iterations of the present study, including the research questions, index selection, methodology, methods, findings and deliverables were developed and discussed alongside the organization, the author of the present study, and the author's supervisor. Additional ongoing support and involvement in iterations was provided by one member of the Faculty of Arts at Ryerson University, who has done extensive work cultivating academic-community partnerships in the Greater Toronto Area. Thus, index selection was based on a set of inclusion/exclusion criteria (see Appendix A) that was established with Laidlaw, in conducting the systematic search of the grey literature. Subsequent discussion and agreement was obtained for the final list of indices to be used.

Second, a complimentary step in the planning phase was to conduct key stakeholder consultations with two additional organizations. Consultations were conducted with UNICEF

Canada and the Students Commission of Canada who shared similar goals and were interested in collaborating on discourse to discuss approaches to the study of variability across wellbeing indices particularly for youth. Ongoing consultation between the Fall and Summer of 2015-2016 led to the identification of one document that listed previous indices and reports that have been used in the topic at hand. The document was shared and consisted of seventeen listed references.

### **Search Phase**

A search was performed across multiple databases for both academic and grey literatures to collect and analyze documents related to youth wellbeing and conceptual foundations. This phase of the meta-narrative analysis was also iterative, and thus the course of the search for documents in this study was three-fold and described below.

**Search for empirical articles.** An exploratory search for empirical articles was performed in the topic area (i.e., articles reviewing reports or indices on the topic of development and wellbeing among young people) as a way to identify seminal papers, provide context, and gauge background information (i.e., historical trends and particular methodological approaches) in the topic. The search for empirical articles was conducted across two electronic databases and included Google Scholar, and the OECD iLibrary. Such databases were selected as they did not focused on any one particular discipline and could provide a broad overview of the literature from various perspectives of researchers and experts in the topic of interest. Noteworthy in the search is that it was not an exhaustive list of all seminal papers on the topic of youth wellbeing and of previous reviews of indices. However, it was a relatively comprehensive inclusion of commonly cited articles (Greenhalgh et al., 2009). The search thus identified leading articles that focus on the review of reports and indices of wellbeing, as well as providing information on historical trends and preferred methods in the creation of such indices (e.g., Ben-Arieh; O'Hare;

Moore).

**Search for grey literature.** After gauging background information, a search for documents in the grey literature base was an imperative step in the collection of indices for the meta-narrative analysis. It has been argued that, “Grey literature is literature produced by any organisation whose central purpose is not publishing. These organisations include governments, businesses, not-for-profits, health organisations and associations, and more” (Retrieved from: <http://dal.ca.libguides.com/systematicreviews/greylit>; Hartling et al., 2017). Thus, in the present analysis, it was most suitable to search for documents in the grey literature base given that such reports (and indices) are most often found within a given webpage of the organization that formulated the index, and often intended for policy-makers and government levels, in addition to interested individuals or youth advocates. In addition, the grey literature could support the focus of the multidisciplinary nature of indices across organizations, and varied conceptualizations of wellbeing across indices in the meta-narrative review process. Thus, a systematic search for reports or indices was performed using the advanced settings option in Google.

In addition to consultation with stakeholders, the list of keywords and inclusion/exclusion criteria were selected based on discussion and agreement with the community partner. Please refer to Appendix A for a breakdown of the search criteria used to identify reports and indices (i.e., inclusion criteria, search terms, date range, dates of search, exclusion criteria), and further details are described below. A search was performed between 2015 and Spring 2016. In the advanced setting option in Google, documents were limited to English, and were selected based on a data range from the past ten years (i.e., 2006 to Spring 2016). All keywords were inputted using ‘AND’ between each word, and could be found ‘anywhere in the text.’ In addition, a search function was used to exclude ‘all explicit information.’

**Index Inclusion Criteria.** *Age.* As previously noted, given that the word ‘youth’ is a socially-malleable term with no clear definition, it was important to conceptually deem youth in line with previous findings, which at times overlap with ages of childhood and adulthood. Thus, the present analysis sought to incorporate the inclusion of indices insofar as they mentioned the terms “youth” within the index. The inclusion of ages across measures had to include outcomes on young people over 12 years old (or as young as grade 7; so as not to include indices on only those typically regarded as “young children”), as well as those under 30 (so as not to focus on individuals typically defined as being in the stage of “middle adulthood;” Whitbourne et al., 2014).

*Population sample.* Population selection was based on the inclusion of reports that incorporated data on North American youth from Canada and the United States (USA). The focus was placed on identifying indices that included data from these countries for three reasons. First, the decision to include indices with data on youth in Canada and the USA was discussed and agreed upon in consultation with the community partner. Second, assessing the variability in monitoring youth wellbeing across these North American countries was considered the most relevant to the Canadian context, and for the goals of Laidlaw. Lastly, indices that incorporated data on the Canadian and the USA contexts helped manage the scope of data collection and findings. This decision was important for LF in considering service provision for local youth with possible needs that have been either met or overlooked. It was possible to consider that these indices could be used as relevant examples in the understanding of youth wellbeing in Canada. In addition, reports were removed if they only focused on one sub-population of youth, such as youth in juvenile detention centres, youth in foster care, or immigrant/refugee youth.

*Date range.* Inclusion of indices was limited to only those that were published in the past

ten years. It was important to limit the date range of index inclusion given the importance to compare more recent, and commonly used indices, in addition to limiting the scope of findings from all reports published since the inception of indices. Importantly, recently used indices could also elucidate ever-evolving trends and changes that have been described in seminal papers that allude to historical contexts. Thus, the identified indices could add critical information to the progress in constructs, terms, and data collection as a function of advances in scientific exploration and methodology (Clark-Kazak, 2008; Land, Lamb, & Zheng, 2011).

*Organization/stakeholder.* After the search for indices, a list of the organizations or stakeholders who work together to create and disseminate the report or index was identified. Indices were included so as to only select one report per organization. Thus, if two reports were found from a given organization, the most recent report was included, and prior reports were removed.

In sum, the selection was based on multidimensional, international or regional reports that included data in Canada and/or the USA, and that included the terms ‘youth’ and ‘wellbeing.’

**Additional methods papers.** The search subsequently led to an additional literature search of corresponding empirical papers related to the methodological and statistical designs of each index. Additional relevant methodological papers were used to provide a focused assessment of existing youth wellbeing and development metrics in line with each index. Such studies were retrieved using a snowball search (Greenhalgh & Wong, 2013) from the reference sections or footnotes of each index, when available. This inclusion would help strengthen the assessment and empirical inquiry of the conceptual approach, methodology, and statistical analyses found among indices, as well as add to the comprehensive analysis for each index (Wong et al., 2013).

## Mapping Phase

The mapping phase in the present study included teasing apart each index and extracting data using a codebook with areas of interest and thematic questions. Proponents of the meta-narrative approach indicate that questions of interest can be selected and adapted based on discussion with experts in the field and advocates in the topic of interest (Greenhalgh et al., 2005; 2009). Accordingly, thematic comparison questions were identified and agreed upon with the community partner. In addition, in line with standards of a meta-narrative review (Greenhalgh et al., 2005), it was important to incorporate four particular dimensions across areas of inquiry or “lenses” in the codebook. Accordingly, the codebook was used to appraise and synthesize findings based on these four dimensions, and included: conceptual, theoretical, methodological, and instrumental dimensions of each index, further described below.

First, the *conceptual lens* identified the facets that are important to the topic at hand and included, the population inclusion (age ranges) and their definitions where available, as a rationale for including specific age ranges of the population under study. In addition, it was important to identify the selection of index components (i.e., selection of domains, indicators, measures, and source of data) as a way to gauge how each index conceptualized development and/or wellbeing. Next, the *theoretical lens* of each index identified the framework used to conceptualize wellbeing, as well as the definition of ‘wellbeing’ where available in each report or within a corresponding methodological paper. Third, the *methodological lens* focused on addressing the rationale in the selection of index features (and by whom they were selected or validated), and the data collection strategies (data availability and unit of analysis). Lastly, the *instrumental lens* included identifying the total list of measures used across indices, which were compared in various ways (described below in the ‘Appraisal Phase’ section).

## **Appraisal Phase**

An appraisal of each index corresponded to manually extracting information related to each question of interest in the codebook (For details of the codebook see Appendix B; Collins & Hayes, 2010; Greenhalgh et al., 2005; 2009). After reports were pulled, data were organized and synthesized into an Excel spreadsheet based on: The context (international, national, provincial), the organizational/stakeholder affiliations, and the year of data collection. Next, data were extracted in order to tease apart the main research questions. First, the population sample was appraised by searching for the age range inclusion and definitions of the sample, where available. Second, the theoretical lens or approach used to encapsulate wellbeing was extracted by searching for the words ‘theoretical’ ‘lens’ or ‘approach.’ Third, it was important to appraise information regarding the process of selecting the index features, or by whom they were selected, and the approach in selecting and validating such features. Fourth, domains were appraised by assessing the domain selection, labels, and number of domains per index. Next, the rates of overlap of index features were assessed in various ways. First, the rate of overlap of domains across indices was assessed by grouping domains with sometimes differing labels into an overarching category, such as placing domains of “material wellbeing” and “family economic wellbeing” into the overarching category of “Economic Wellbeing.” Second, the indicator overlap was assessed by identifying indicators that addressed the same topic, such as indices that used indicators related to “healthy eating” within the health domain. Third, the rate of overlap was assessed across the total list of measures in three ways. First, measures that focused on youth-specific data were assessed. This analysis was performed by categorizing ages and comparing each age range group across all measures. Measures across age ranges were subdivided as follows. (1) Measures specific to youth (youth-specific), indicated measures that

included data only on youth over the age of 12 (or grade 7 and above) and under the age of 30.

(2) Measures not specific to youth (non-youth specific) was defined as a measure that did not include data specifically focused on youth (I.e., under age 12 or early childhood; or measures that did not disaggregate data by age, e.g. combined outcome of ages 10-19, or Grades 6-10; or used parental, family/household-level data with dependents 0-18).

(3) Measures that provided data on both youth and non-youth outcomes and were separated by ages, insofar as data on ages in the index were disaggregated. For example, measures that included separate outcomes for those 11, 13, and 15 years old within a measure. Here, an 11-year-old could be placed in “non-youth specific” and a 13- and 15-year-old could be placed in “youth-specific.”

(4) Measures that did not indicate the age specification (age not specified), indicated using measures either at the population- or government-level or self-reported surveys where age was not indicated in the index. Thus, data among measures would only be considered youth-specific when age inclusion reflected samples between the ages of 12 and 30.

Second, the data collection technique and source of data were assessed. Here, measures were subdivided by units of analysis and respondent levels, including national/governmental levels, household/parental levels, or self-reported data.

Third, the rate at which measurement looked at deficits or positive development was assessed. Here, measures were categorized by taking a strengths-based (positive) or deficits-based (negative) approach. In line with previous empirical investigations (e.g., Ben-Arieh & George, 2001; Moore et al., 2004; Pittman & Irby, 1997), strengths-based measures included items such as graduation from post-secondary education or early stage entrepreneurial activity. Deficits-based measures included items such as rates of youth suicide or violent crime. Some measures were also assigned a neutral label, such as youth in the labour force with estimates of employment and unemployment. Importantly, following the principles of the meta-narrative

approach, coding and reliability in the data extraction process were conducted through an iterative process using both pragmatic and reflexive approaches over multiple meetings and dialogue with all parties involved in informing the analysis phase. In addition, given that the present documents (indices) include an inherent structure, and they are typically pre-organized into sections or categories, the data extraction process involved little room for subjective interpretation. With that said, extraction of data and findings in the narrative process were discussed through ongoing iterations in the appraisal and synthesis processes.

### **Synthesis Phase and Analysis Plan**

Information was extracted from each index, and was synthesized and partitioned into tables to provide a narrative (thematic) component corresponding to the questions of interest in the codebook. Each question item and associated variables were extracted and analyzed qualitatively in an Excel spreadsheet. Data were subsequently assigned codes using numerical allocations (e.g., *international* = 1, *national* = 2, *provincial* = 3, where required) and quantified using SPSS (IBM, 2017). Thus, data were compared across each area of interest and descriptive statistics and frequencies were assessed to examine the rate of measures across questions of interest. Data were then synthesized in order to formulate a critical appraisal component assessing the overall strengths and limitations of indices, areas where most focus is placed, and areas for improvement.

Recommendations. Recommendations for best practices were presented to the community partner for internal use in a “Final Report” document. Thus, a separate report containing detailed information on recommendations has been supplied to the organization and presented to the Board of Directors.

## **Results**

## **Index Inclusion**

The breakdown of sources that contributed to the final meta-narrative analysis is shown in Figure 2 below. An initial empirical literature search phase resulted in articles that were included as background and contextual information in the topic of conceptualizing and measuring wellbeing for young people in indices. Additionally, stakeholder consultations (multiple meetings and emails between the Fall 2015 and Summer 2016) were conducted with UNICEF Canada and The Students Commission of Canada in 2015. This discussion resulted in the addition of one non-empirical document, which highlighted a non-comprehensive list of seventeen key population-level youth surveys and indices. From these searches, a total of four indices were found and appraised that matched the search criteria (indices 3, 4, 5, and 6; see Table 1). An additional grey literature search resulted in 99 documents, whereby the titles and webpages were reviewed. Six documents fit the inclusion criteria; however, three indices were removed as they were duplicates from the stakeholder document (indices 3, 4, and 5). Thus, the grey literature search resulted in the addition of three indices appraised (index 1, 2, and 7). Lastly, from the reference sections or footnotes of each index, eleven empirical papers were included for further detail on the conceptualizations and methodologies of each index. For a list of corresponding methodological papers see Appendix C. Thus, a total of 18 sources were used for the meta-narrative analysis.

All indices included at least some data on youth over the age of 12 and under age 30. Indices were found at the international ( $n = 4$ ), national ( $n = 1$ ), and provincial ( $n = 2$ ) levels, and all included data on youth in Canada and/or the USA contexts. Data on Canadian youth were included in five indices (with the exception of indices 2 and 5), and data on youth in the United States were included in five indices (with the exception of indices 6 and 7). Among indices, four

were international comparative indices, one was a national index (United States), and two indices were provincial within Canada (British Columbia and Ontario). Indices were included as they combined multiple domains (equal to, or more than four) to conceptualize wellbeing, and used more than one indicator to construct a given domain.

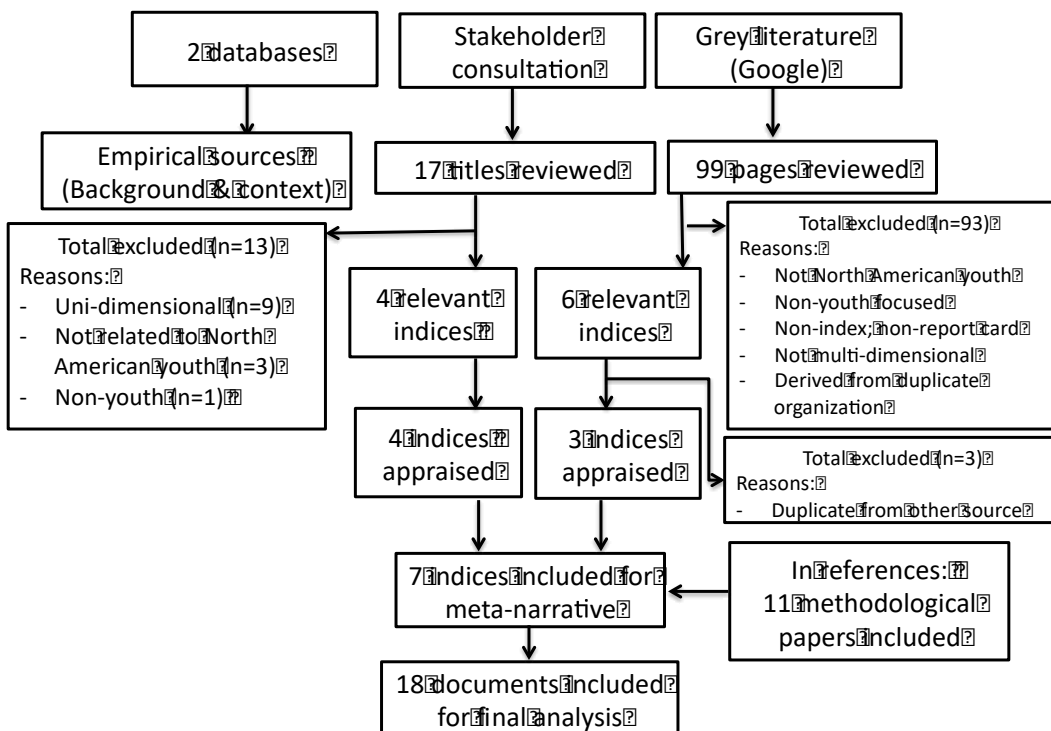


Figure 2. Flow diagram of search strategy.

Table 1

*Selected List of Indices, Index Number, Name, Year, Age Ranges, and Amount of Methodological Papers per Index*

	Index (# and Name)	Abbrev	Year	Age Range	Method
International	1) Youth Development Index	YDI	2016	15-29	1
	2) Global Youth Wellbeing Index	GYWI	2014	10-24	1
	3) UNICEF Innocenti Report Card 11: Child Wellbeing in Rich Countries	UN11	2013	0-17	4
	4) World Health Organization: Social Determinants of Health and Well- Being Among Young People	WHO	2009-10	11, 13, & 15	1
National	5) National Child and Youth Well-being Index (United States)	CWI	2013	Birth to 17	2
Provincial	6) McCreary Report: Growing Up in British Columbia (BC)	Mc BC	2015	12-22	1
	7) Stepping Up: Ontario (ON)	SU ON	2016	12-25	1

*Note.* Abbrev = Abbreviations; Method = Additional methodological papers.

### **Age Range Inclusions and Definitions of the Population Samples**

**Age ranges.** The maximum age for data outcomes was age 29 (index 1), age 25 (index 7), age 24 (index 2), followed by age 22 (index 6). Other data on outcomes reached a maximum age of 17 (index 3 and 5) and 15 (index 4).

**Sample definitions.** Results show that all indices used the term ‘youth’ at least once across measures. Although age range categories were incorporated, it was shown that four indices provided a clear description to define their population sample, deriving from indices 1, 2,

4 and 7 (YDI, GYWI, WHO, and SU ON, respectively). Refer to Table 2 below for definitions of youth. All other indices did not provide clear definitions for their population samples or rationale for age range inclusion criteria.

Results across definitions showed that the four indices denoted ‘youth-hood’ (or adolescence) as a period of transition that fell between childhood and adulthood. Additionally, it was found that the term ‘youth’ was described as a social construct and thus largely varied according to societal disciplines, including political, social, and cultural contexts. The term ‘youth’ was also marked as a period of maturation, semi-dependency, and increasing autonomy, which were used as proxies in identifying personal-level wellbeing.

Table 2

*Definitions of Population Samples*

Index	Sample Definitions
YDI	“Definitions of youth are more contextual, dependent as much on formal nomenclatures as on informal factors such as culture, tradition and socio-economic conditions in a country or community. In simple terms, youth is a period of transition during which children and adolescents gradually come to be recognized as adults. Falling between childhood and adulthood, it is a period of semi-dependency during which young people try to achieve personal autonomy while still remaining dependent on their parents or the state. The length of periods of dependency varies hugely across socio-economic and political contexts.” (p. 6-7)
GYWI	“Youth is a stage of life: As much a social construct, as much a term of science to discuss, evaluate, and assess a heterogeneous population group that shares characteristics of transition (Often a synonym with young people).” (p. 6)
WHO	“Young people aged between 11 and 15 years face many pressures and challenges, including growing academic expectations, changing social relationships with family and peers and the physical and emotional changes associated with maturation. These years mark a period of increased autonomy in which independent decision-making that may influence their health and health related behaviour develops. Behaviours established during this transition period can continue into adulthood, affecting issues such as mental health, the development of health complaints, tobacco use, diet, physical activity level and alcohol use. HBSC’s findings show how young people’s health changes as they move from childhood through adolescence and into adulthood. They can be used to monitor young people’s health and determine effective health improvement interventions.” (p. 2)
SU ON	"Adolescence is recognized as a distinct developmental stage that occurs between childhood and early adulthood. There are approximately 2.47 million young people between 12 and 25 years living in Ontario. This represents 18.3 per cent of our province's overall population." (webpage)

*Note.* SU ON definition found in Stepping Up webpage:

<http://www.children.gov.on.ca/htdocs/English/professionals/steppingup/youngpeople.aspx>)

### **Definitions and Theoretical Approach to Conceptualize Wellbeing in Each Index**

For a description of theoretical lenses and definitions of ‘wellbeing’ refer to Table 3 below. Across theoretical lenses to conceptualize wellbeing, two indices approached the topic of youth wellbeing from a socio-ecological perspective (Indices 6 and 7). One index described the

framework broadly as a multidimensional approach to wellbeing, but elucidated that the theoretical framework was based on previous empirical work related to fundamentals of an ecological perspective (Index 3). Two indices approached the topic of wellbeing from a capabilities and asset-based approach and a positive youth development framework (Indices 1 and 2). Another index approached the topic of wellbeing from a social determinants of health perspective (Index 4). Lastly, one index derived from a framework that used a quality-of-life approach (Index 5).

### Table 3

*Definitions of 'Wellbeing,' and the Theoretical Framework to Conceptualize Wellbeing Across Indices*

Index	Definitions of Wellbeing and/or Development	Theoretical Framework
YDI	Youth development can be understood in a variety of ways. It was initially perceived and understood from within the psychological perspective, which tends to explore the ways young people grow emotionally and construct identities. A broader perspective on development, often linked to sociological insights, focuses on young people's experiences and the ways in which these experiences provide opportunities for life management and distinct, culturally-shaped, perspectives on the world and their place within it. Crucially, this 'asset-based approach' to youth development recognises the ability and agency of young people themselves to influence development outcomes. Youth development can also be impacted by the transmission of inequalities between generations, as some find pathways blocked because of the socio-economic status of their families, while inherited privilege opens the door to opportunities for others.	Capabilities Approach/ assets-based
GYWI	This framework for youth wellbeing is rooted in theories of change and human development for both individuals and countries. It draws upon the body of work measuring quality of life, economic and social progress, as well as the discipline of youth development. In non-technical terms, wellbeing has been defined as "the state of being happy, healthy, or prosperous"... [In more technical terms] development literature define wellbeing associating it with welfare, utility functions, and multidimensional measures of societal growth and progress... [This index] approach to wellbeing has been informed by measures of poverty and development that have moved from an interpretation based on income or economic growth to one that is multidimensional and includes objective as well as behavioural and subjective elements.	Draws on the principles and fundamentals of positive youth development (assets-based framework)
UN11	All dimensions focus mainly on children's microsystem, i.e., on the children themselves and the different subsystems that directly impact on their life. Their objective is to represent the conditions children find for their development and participation in society and child outcomes. Belonging to the same system the dimensions are interdependent and interrelated.	Multidimensional approach to wellbeing (draws on socio-ecological perspectives)
WHO	HBSC focuses on understanding young people's health in their social context – at home, at school, with family and friends. The study of social determinants looks at factors outside what could traditionally be defined as "health" areas but which	Social determinants of health perspective:

	nevertheless have an enormous impact on health and well-being. Researchers in the HBSC network are interested in understanding how these factors, individually and together, influence young people's health as they move into young adulthood.	
CWI	Children and youth live unique lives; each experiences a range of social conditions at different points. The Index comprises Key Indicators associated with different stages of the first two decades of life. Different Indicators capture children and youth at different stages. Objective was to measure the circumstances of children's lives in a way that reflects their well-being – to assess their quality of life – and to track changes in well-being and over time.	Quality-of-life approach
Mc BC	Well-being is shaped by relationships and the wider environment.	Socio-ecological understanding of development
SU ON	These themes are based on an ecological model of development and a person-centred approach that is also consistent with Aboriginal ways of knowing. Important early interactions for youth include the smaller circles (their personal health and development, their family and friends). As they age, they grow to become members of the broader community (through education and employment, engagement and participation).	Socio-ecological, positive-assets based

## Index Composition Across Indices

**Index selection and methodology.** Some overlap existed across indices in the methodological approach to formulate the index and to select domains, indicators, and measures. All indices indicated using some combination of multiple stakeholders (e.g., government bodies, policymakers) and experts (e.g., leading researchers, academic or community experts, experts across sectors) to decide upon index composition. It was also indicated that index validation was combined with a search of empirical literature in three indices (indices 2, 3, 5). Three indices indicated incorporating youth voices in the construction of the index and in the decision to select domains that were identified as important to youth (indices 2, 6, and 7). In all indices, there was mention of the limitations in data collection as a function of the availability of data sources.

**Domain selection and labels.** For domain clusters and labels used in each index see Table 4. Each index was made up of a unique number and combination of domains (range = 4 – 7). The mean number of domains across seven indices was 5.7. No two indices combined identical domains or labels to conceptualize wellbeing. Across seven indices examined, there was no clear consensus in the labeling of domains. Domains that were placed into overarching “clusters” were based on similarity of indicators used to make up the domain, even though they did not use the exact same label. Among exact use of labels, results show most agreement for the label of “Education” (43%), followed by 29% for each of “Health,” “Family Economic Wellbeing,” and “Family and Peer Connections.” All other domain labels within clusters differed in terms of the specific names used to capture a similar dimension.

Results showed overall consensus across seven indices in assessing domains related to Economic Wellbeing, Education, and Health. Five of seven indices (71%) included domains related to Community/Civic Engagement. Four of seven indices (57%) assessed Risk Behaviours, and four of seven indices (57%) measured Social Relationships/Connections. Three of seven indices (43%) assessed Child Safety, and four of seven indices (57%) used a domain category that did not overlap with any other index, such as “Information and Communication Technology” or “Emotional and Spiritual Wellbeing”. For the amount of domains, indicators, and measures per index see Table 5.

Table 4

*Domain Clusters and Labels Across Indices*

Domains Clusters and Labels							
	YDI	GYWI	UN 11	WHO	CWI (US)	Mc (BC) <sup>1</sup>	SU (ON)
Eco	Employment & Opportunity	Economic Opportunity	Material Wellbeing	Fam. Affluence (SES) <sup>2</sup>	Fam. Economic Wellbeing	Fam. Economic Well-Being	Employment & Entrepreneurship
Edu	Education	Education	Education	School	Education Attainment	Child Learning	Education, Training & Apprenticeship
Hea	Health & Wellbeing	Health	Health & Safety	Behaviors & Outcomes <sup>3</sup>	Health	Child Physical & Mental Health	Health & Wellness
Com Eng	Civic Participation Political Participation	Citizen Participation	--	--	Community Eng	Community Connections	Civic Engagement & Leadership
Risk	--	--	Behav. & Risks	Risk Behav.	Safe/ Risky behav.	Child Behavior	--
Soc Rel	--	--	--	Family & Peers <sup>4</sup>	Social Relat.	Family, Peers Connections	Supportive Friends & Family
Saf	--	Safety & Security	--	--	--	Child Safety	Diversity, Social Inclusion & Safety
Oth	--	Information & Comm. Technology	Housing & Environ.	--	Emotional/ Spiritual Wellbeing	--	Coordinated & Youth-Friendly Communities

*Note.* Abbreviations: Eco = Economic; Edu = Education; Hea = Health; Com Eng = Community Engagement; Risk = Risky Behaviours; Soc Rel = Social Relationships & Connections; Saf = Safety Behaviours; Oth = Other domains with no overlap; Behav. = Behaviours; Environ. = Environment; Comm = Communication.

<sup>1</sup>This index only included 6 domains, which combined “Family, Peers, and Community Connections” – However, for the purposes of fluidity, and overlap of indicators with other domains, the domain was subdivided into “Community” (Community Engagement Domain), and “Family and Peers” (Social Relationships Domain).

<sup>2</sup> Family Affluence was not incorporated and labeled as one of the four key domains, but was tested and measured in the index, and was thus used as an indicator, which added one measure (FAS) in the present study.

<sup>3</sup> This index included two separate domains of “Health Outcomes” and “Health Behaviours;” however, because many of the indicators overlapped with other indicators under the same domain, these two domains were integrated within the “Health Domain”.

<sup>4</sup> This index included a domain of “Social Context” which consisted of Family, Peers, and School Contexts. Thus, for the purposes of fluidity and integration into overarching categories, this domain was subdivided into “Family and Peers” (Social Relationships Domain) and “School” (Education Domain).

Table 5

	YDI	GYWI	UN11	WHO	CWI	Mc	SU	Total	<i>M</i>
# of Dom.	5	6	5	4	7	6	7	40	5.7
# of Indic.	18	40	26	30	28	30	57	229	32.7
# of Meas.	18	40	30	35	28	65	57	273	39

*Number of Domains, Indicators, and Measures Per Index and the Mean Across Indices*

*Note.* Abbreviations: Dom. = Domains; Indic. = Indicators; Meas. = Measures; *M* = Mean.

**Indicators.** The number of indicators across all domains ranged from eighteen to fifty-seven ( $M = 33$ , total = 229). Two indices showed agreement on the precise number of indicators to use ( $N = 30$ ). However, no consensus was found for the number of indicators to use within a given domain. The range in the number of indicators used to make up a domain across all indices was between two (i.e., found in Civic Participation in index 2, Social Relationships in index 5, and Family Economic Wellbeing in index 6), and eleven (i.e., found in Health and Wellness in index 7). Results also showed little consistency in indicator labels, despite measuring similar items. It was found that some indicators included multiple similar items and thus were further amalgamated into one overarching indicator cluster. For example, the WHO-HBSC index included multiple indicators (items), including “Eating breakfast daily,” “Eating fruits/veggies daily,” and “Consuming soda or sweets”, and were thus amalgamated into an overarching indicator of “Eating behaviour.” Another indicator cluster of “Families in low income” was a combination of “Child poverty”, “Household income”, and “Family affluence.” Thus, after further amalgamation, overall indicator count was 129. For an overall breakdown of percentages and number of indicator overlap within a domain and between at least two indices see Table 6.

Results for overall indicator overlap across domains showed a 33% agreement rate. Results also showed that there was a moderate level of overlap across all domains, ranging from 19% to 67% overlap rate. The highest rate of indicator overlap was found in the domain of ‘Risk’ (67%,  $N = 8$  indicators) followed by ‘Health’ (48%,  $N = 13$  indicators). The lowest rates of overlap were found in ‘Community/Civic Engagement’ ( $N = 3$  indicators) and ‘Safety’ ( $N = 3$  indicators), each with 19% indicator agreement rate.

**Measures.** There was a total of 273 measures across all indices. Results show that there was a 40% overlap rate among measures that assessed similar indicators (items; See Table 6). It was shown that measures that assessed similar items were sometimes found within different domains, and thus a breakdown of measurement similarity per domain was not shown. For example, five measures were found for the indicator of Exercise/Sedentary Behaviour, whereby one was found in the domain of Risk Behaviour, and four were found in the domain of Health. Additionally, three measures were found for the indicator of Food Security, whereby two were found in the Economic Domain, and one was found in the Social Relationships Domain. Measures that assessed similar items most often derived from different sources of data. Results show that there was a 7% agreement rate for data sources used to measure similar items. For example, the measure of ‘Youth Not in Education, Employment, or Training (NEET) Rate’ was assessed in five indices, and among those, three sources derived from OECD data. Additionally, twelve measures assessed the indicator of ‘Academic Achievement’ for various grade levels, where only two measures showed overlap for mathematics derived from the data source of PISA testing scores.

Table 6

*Overlap of Indicators Within Domains and Overall Overlap Across Indicators, Measures, and Sources of Data*

	Total	Total	Indicator Overlap
	Indicators	Measures	Rate
Domain Clusters	<i>N</i>	<i>N</i>	<i>N</i> (%)
Economic	15	31	4 (26.7)
Education	18	42	7 (38.9)
Health	27	63	13 (48.1)
Civic/Community	16	33	3 (18.8)
Risky	12	46	8 (66.7)
Social Relationships	9	18	3 (33.3)
Safety	16	24	3 (18.8)
Other Domains	16	16	--
Total	129	273	--
Total Indicator Overlap	--	--	42 (32.6)
Total Measure Overlap	--	--	108 (39.6)
Total Data Source Overlap			18 (6.6%)

**Youth-specific age range inclusion among measures.** For a breakdown and analysis of age range inclusion among measures across domains see Table 7. Results showed that, overall, the largest proportion of age range inclusion for measures across domains was found in the youth-specific category, with the exception of measures in the ‘Social Relationships’ and ‘Other’

domains. The highest rate of youth-specific measures was found in the ‘Risky Behaviours’ domain (96%), while the highest rate of non-youth specific measures was found in the ‘Social Relationships’ domain (72%). Overall, the highest rate of measures where ages were not specified was found in the ‘Other’ domain category, as well as ‘Community/Civic Engagement’ with measures largely related to societal (national statistics), or policy-related data reflecting no particular age grouping (e.g., measures of *% population with access to clean water, or level of air pollution in society*).

Table 7

*Overall Rate of Measurement Across Age Ranges Per Domain*

Domain Clusters	% Youth-Specific (N)	% Non-Youth Specific (N)	% Age not specified (N)
Economic	58.1 (18)	38.8 (12)	9.7 (3)
Education	80.9 (34)	30.9 (13)	4.8 (2)
Health	68.2 (43)	49.2 (31)	3.2 (2)
Community	63.6 (21)	12.1 (4)	24.2 (8)
Risk Behaviours	95.7 (44)	32.6 (15)	--
Social Relationships	55.6 (10)	72.2 (13)	--
Safety	54.2 (13)	25.0 (6)	20.8 (5)
Other Domains	18.8 (3)	43.8 (7)	37.5 (6)
% Measures (N)	68.1 (186)	36.9 (101)	9.5 (26)
(Of Total = 273)			

*Note.* Measures that were disaggregated by age groups were combined into their respective category (e.g., age 11 placed into non-youth specific, age 15 placed into youth specific group), and thus the count can exceed 100%; Age not specified (due to societal/policies, and self-reported) measures were combined.

## **Data Collection Strategy**

The effort of each index to collect data was based on a combination and utilization of population-level statistics, household/parental data, and self-reported responses. Most indices included a combination of the aforementioned data sources from these three units of analysis. Index 4 was only based on self-reported survey data.

**Respondent unit of analysis.** Among measures, 49% ( $N = 134$ ) of responses were self-reported by the population under study, 9% ( $N = 25$ ) responses were at the household (parental) level, and 42% ( $N = 114$ ) were at the societal (population) level data (See table 8 below). Results show that the domain that included the largest rate of self-reported data among measures was Risky Behaviours (85%,  $N = 39$  measures). The domain with the most parental/household data was Economic Wellbeing (39%,  $N = 12$  measures), and the domain with the largest amount of societal (population-level) data was Educational Wellbeing (86%,  $N = 36$  measures). Among youth-specific measure ( $N = 186$ ) across indices, 65% were self-reported from youth between the ages of 12 and 30 years old.

Table 8

*Rates of Respondent Category (Unit of Analysis) and Positive, Negative, and Neutral Measures Across Domains*

Domains	% Self- Report (N)	% Parent/ House (N)	% Pop./ Societal (N)	% Positive (N)	% Negative (N)	% Neutral (N)
Economic	9.7 (3)	32.3 (10)	58.1 (18)	32.3 (10)	61.3 (19)	6.5 (2)
Education	14.3 (6)	--	85.7 (36)	66.7 (28)	4.8 (2)	28.6 (12)
Health	54.0 (34)	4.8 (3)	41.3 (26)	20.6 (13)	66.7 (42)	12.7 (8)
Community	57.6 (19)	--	42.4 (14)	57.6 (19)	21.2 (7)	21.2 (7)
Risky	84.8 (39)	--	15.2 (7)	15.2 (7)	78.3 (36)	6.5 (3)
Social	72.2 (13)	27.8 (5)	--	50.0 (9)	33.3 (6)	16.7 (3)
Safety	54.2 (13)	--	45.8 (11)	29.2 (7)	50.0 (12)	20.8 (5)
Other	56.3 (9)	31.3 (5)	12.5 (2)	62.5 (10)	31.3 (5)	6.3 (1)
% Total (N)	49.8 (136)	8.4 (23)	41.8 (114)	37.7 (103)	47.3 (129)	15.0 (41)
Total Youth- Specific (N = 186)	64.5% (120)	--	--	35.4% (66)	48.4% (90)	15% (28)

*Note.* Abbreviations: House = Household; Pop. = Population.

### **Strengths vs. Deficits Measures**

Data extraction results showed that overall all indices indicated the importance of incorporating positive indicators. Across the 273 measures, 47% took a deficits approach to measurement (e.g., youth unemployment, mortality rate, binge drinking, drug use, poverty), and 38% focused on positive outcomes and strengths (e.g., taking part in volunteer activities, self-

rated emotional wellbeing, or high school completion). An additional 15% focused on outcomes that were neither positive nor negative and were identified as neutral (e.g., residential mobility, minimum age for voting participation, self-rated health). For the percentage and number of positive, negative, and neutral measures across domains see Table 8. Among youth-specific measures ( $N = 186$ ), 48% were negative indicators, 35% were positive.

## **Critical Appraisal**

### **Strengths and Limitations**

**Age Critique.** Findings showed that some measures provided data that separated ages into respective cohorts of younger and older individuals under study. However, most of these measures derived from one report (WHO-HBSC, 2010). A large proportion of measures grouped largely disparate age categories into one outcome, such as grouping data on an 18 year old alongside younger children as young as birth, or with early childhood (e.g., under age 12), or alongside those in middle-adulthood (ages 15-49). Such amalgamations of largely different ages make it difficult to disentangle age-outcome relationships. In addition to age disaggregation, only one index considered disaggregating differences between genders (i.e., male and female) and family affluence (i.e., socioeconomic status; WHO-HBSC, 2010).

It was also found that most data of measures included multiple age groups to assess indicators in each domain. Nevertheless, it was shown that the largest proportion of measures was in the youth-specific category. Interestingly, it appears that there is an effort across indices to provide data on youth over age 12 and under age 30. Moreover, although some survey data incorporated young individuals with multiple ages, some measures appeared only to use a subset of age groups to include in a given outcome. For example, some outcomes in the McCreary BC

Report obtained data from the ‘BC School Satisfaction Survey,’ which was an annual survey of students in Grades 3-4, 7, 10 and 12. However, further analysis showed that some measures and outcomes only included data on children in grades 3-4.

Although indices indicated using samples within particular age ranges, it was found that some measures were in fact outside of the denoted age range inclusion. For example, although the UN 11 index indicated using a sample of ages between birth to 17 years old, this index in fact included outcomes on youth up to 19 (e.g., for the Educational Participation measure). Additionally, the GYWI indicated using samples between the ages of 10-24 years old. However, it was shown that one measure for HIV rate included outcomes on a sample of those up to the age of 49 years old.

**Domain Critique.** There was important overlap of domains where all indices measured (1) Economic Wellbeing, (2) Health and (3) Education. However, results showed that no two indices used the same combination of domains to make up an index. Additionally, in some instances it was shown that different domain categories across indices used a similar indicator (item) to represent two different domains. For example, the indicator of ‘Not in Education, Employment, or Training,’ was incorporated in two different domains, including Economic Wellbeing and Education, which depended on the index. There were also discrepancies in labeling of domains. For example, the domain cluster of Economic Wellbeing included labels, such as Material Wellbeing and Low Income. Also, the domain cluster of Community Engagement included labels, such as Civic/Political Participation and Community Connections.

**Indicator Critique.** Indicator selection within a given domain showed variability across different indices. Similarly, the total number of indicators per index also showed to largely vary, ranging from eighteen to fifty-seven. In addition, an indicator used within one domain in one

index, was sometimes found in another domain of another index. For example, this overlap was found across the “Health” and “Risk” Domains. Here, substantial overlap of indicator selection indicated that a total of seven indicators were found in both “Health” and “Risk,” such as the indicators of Teenage Pregnancy, Tobacco Use, Alcohol Use, and Illicit Drug Consumption.

Additionally, across indices, all argued for the importance of taking a strengths-based approach and the need to include positive indicators. Indeed, at least one indicator across indices used a positive measure. Similarly, although the largest proportion of indicators were measured as a function of their deficits, it was found that there was also a moderate level of positive measures across indicators. The highest number of positive indicators was found in the Education Domain, while the highest number of negative indicators or deficits was found in the Risky Behaviour domain.

**Measurement Critique.** Measures were important assessment tools for a given indicator. There was some agreement on measures across indices. However, there was little consensus on data sources, and data collection methods, ranging from large-scale population-level statistics, to parental (household) surveys data, or self-reported perception data. All indices argued for the importance of a child-centred approach and all indices included at least some measures that were self-reported by the sample. In fact, it was found that almost 50% of measures used “self-reported” data. However, it was also found that most outcomes across domains were a combination of data from varied sources. This means that, for any given domain, there may have been a combination of any three aforementioned units of analysis, providing both objective and subjective levels of data. Only one index provided data solely focused on self-reports. Such data is important for providing information based on how young people feel about their own wellbeing. Additionally, this data can be used to inform comparisons in outcomes across units of

analysis in order to show how youth see themselves when compared with how they are portrayed objectively

## **Discussion**

This study examined the variability across multidimensional indices in the conceptual and methodological approaches to assess the topic of youth wellbeing. Findings from the study revealed that across indices, some similarities were found, but considerable differences were also shown. That is, it was found that indices varied in the amount and selection of domains, indicators, and measurement across indices. Here, no two indices combined the same sets of domains. Additionally, there was little consistency across indices in indicator selection for one domain cluster. It was also found that measures varied depending on the data collection technique and availability of data sources within a given domain. As well, variability was found in the use of positive and negative indicators. Thus, consistent with previous findings, it appears that there is variability in the construction of wellbeing indices, providing unique conceptualizations, and showing varying combinations of domains, indicators, and measures as strategies for assessment (Booyesen, 2002; O'Hare, 2012). The following sections highlight a range of findings and points of discussion that reflect upon such variability.

### **Selection of Index Components**

Findings from the present study are in line with previous research showing consistency across all indices in the inclusion of domains related to material/economic wellbeing, education, and health (Fernandes et al., 2012; O'Hare & Gutierrez, 2012). Nevertheless, little consensus was found in domain labeling. Additionally, no two indices used the exact same set of domains, or number of domains to create a composite index comprised of multiple domains (Hagerty & Land, 2007; Moore et al., 2012; 2014; O'Hare & Gutierrez, 2012). Moreover, an average of 5.7 domains were represented across seven indices. Indeed, this finding supports a previous study, indicating a possible agreement among scholars in the inclusion of six to seven dimensions

related to wellbeing (O'Hare & Gutierrez, 2012). Findings from the present study thus show a possible growing consensus related to the number of dimensions of wellbeing, but also highlight the challenges associated with conceptualizing wellbeing across domains. It is critical then, to identify consensus across labeling in order to better conceptualize the understanding and inclusion of dimensions related to youth wellbeing and development (O'Hare & Gutierrez, 2012).

These findings may be explained in two ways. First, taking a bottom up approach to the topic of wellbeing, it appears that indices are created as a function of the precise data sources used and also the availability to access these data (Booyesen, 2002; McGregor, Coulthard & Camfield, 2016; Moore et al., 2012; 2014). That is, data at each level of analysis may be more or less accessible for some domains or measures over others. Thus, in line with previous findings, it appears that data are readily available on measures related to the domains of economic wellbeing, health, and education. Second, taking a top down approach, it is possible that some domains may be easier to conceptualize than others, such as by using a precise theoretical lens or approach to conceptualize the domain (Booyesen, 2002; Hagerty & Land, 2007; Lamb & Land, 2013). Such data would make a domain easier to construct and define, and could thus show greater agreement across indices of indicator selection within a domain. In the present study, it appeared that there was most agreement in the conceptualization of the domains of Risk, Health, and Education. Conversely, it appeared that other domains were vague and showed little agreement, where less indicator overlap was found in the domains of Community Engagement, Safety, and Social Relationships.

Such aforementioned findings on indicator overlap suggest that it could be easier to conceptualize factors related to the risks that surround an individual youth, identify rates of

health and illness, or assess how well youth are doing in school. In addition, it is possible that agreement exist across societies and international organizations in contributing to defining dimensions and accessing standardized data for some indicators over others. In this case, it is possible that some factors in domains are more universally agreed upon, such as laws and regulations that govern criminal acts of illicit drug use or violent crime (as risky behaviours), assessing eating behaviours, weight, or accessing medical records of life expectancy or death rates (as health), or delivering annual standardized education testing, such as in math, science, and reading (education). In fact, a global call for education testing was informed by the OECD (PISA testing system) in assessing math, science, and reading for various school grades. These tests thus provide nations with national-level data that are readily available, and thus can be used to inform standard measures across indices for the Education Domain. In addition, such data can allow nations to assess how their students are ranking compared with international standards (Taut & Palacios, 2016). Conversely, the domains of community engagement, safety, and social relationships may be more ambiguous depending on the context of a given nation or region. That is, such domains may be assessed as a function of the possible differences in conceptualizations depending on the organization who creates the index, the context of the index (e.g., international vs. provincial) or data source availability, adding challenges to discern clear understandings of these domains.

Indeed, although index creation is highly dependent on data availability (Bastos et al., 2004; Ben-Arieh, 2000; Booysen, 2002; Decancq & Lugo, 2013; Fernandes et al., 2012; O'Hare, 2012), indices would benefit from emphasizing the need for a standard multidimensional construct that is consistent across domain and indicator use and labels. For example, studies highlighting important facets in conceptualizing wellbeing indeed show agreement that a socio-

ecological perspective provides an accurate reflection of the range of possible environmental influences that surround the lives of young people (Ben-Arieh, 2007). Nevertheless, such a perspective does not always capture perspectives from the individual, or subject (Ben-Arieh, 2007). That is, indices often fail to capture self-reported data from the young population sample. Moreover, in the present study there were general discrepancies found in the number and selection of indicators that make up a domain. Thus, indicators could also benefit from a consistent model (use and amount), particularly for domains where consensus has been lacking across indices. For example, research related to the Canadian Index of Wellbeing (CIW) for adults uses eight domains and eight indicators within each domain (CIW, 2016). The CIW was developed using a strategic framework of best practices and involved consultation and discussion groups with participants on their perceptions on the important dimensions of standards of living and quality of life (CIW, 2016). Thus, such a model could inform the precise standards across indices for domain and indicator inclusion. In the present study, only three indices incorporated a youth voice through focus groups in order to identify and select domains that were important to them.

Additionally, efforts could be made to more clearly articulate indicator agreement within a given domain. For example, findings from the present study identified that indicators for community engagement showed most overlap for ‘volunteering’ and ‘political voting participation.’ However, other indicators did not show any overlap, but still provide important information for youth within communities, such as ‘sense of belonging to community’ and ‘rates of donating.’ In line with studies on community engagement for youth, it is possible that indices would benefit from gathering data at the local level as it reflects availability and engagement of community programs or services focused on youth (Larson & Angus, 2011). That is, perhaps an

index could elucidate rates at which community programs are provided and in what areas of services to meet the needs of youth. Indeed, data that is collected at the community-level can identify youth perspectives on the programs that have most impact on their wellbeing, their participation rates, and self-reported program effectiveness overall.

### **Methodology and Data Collection Strategies to Create an Index**

It was found that methodology in index creation and selection of domains and indicators were most often formulated by the organization, whereby groups of people and committees come together to decide upon the selection of index features, including government bodies, leading researchers, or academic and community experts (Fernandes, Mendes, & Teixeira, 2012; Goldin, Patel, & Perry, 2014; Lamb & Land, 2013). That is, consensus was found across sectors to highlight the importance of collaborative agreement in index formulation for decisions to be made. Indeed, Decancq and Lugo (2013) suggest that a good selection and representation of indicators in multidimensional indices is incorporating a combination of available data, consultation from experts in the field, in addition to the subjective individual opinion within a given region. However, it is also noteworthy to point out that what many indices are missing in the index creation process is the involvement of a youth voice and participation in the understanding of their own wellbeing and development (CIW, 2016; CYP, 2016). Indeed, findings from the present study showed that only three indices incorporated youth input in domain selections and creation of indices that assess wellbeing (i.e., GYWI, 2013; Mc BC, 2015; SU ON, 2016). That is, as it stands, it appears that youth are not embedded in most stakeholder committees across indices as it relates to the processes involved in formulating an index. Thus, it appears that less emphasis is placed on capturing dimensions of youth wellbeing from a youth-integrated perspective. Indeed, it is possible that such findings can be attributed to the limited

time or possibility to access youth samples. However, three indices showed that through focus groups and surveys it is possible to incorporate a youth voice in the deconstruction of their own wellbeing. Such findings suggest some acknowledgement related to recognizing the importance of human agency and rights among youth.

### **Measurement Approaches to Conceptualize Youth Wellbeing**

Interestingly, although it was expected that most data would be focused on population- or societal-level statistics, findings from the present study showed that in fact there was a large proportion of self-reported data across indices. Here, a similar range of measures used data from both self-report and population levels, and less data were gathered at the household or parental-levels. It is possible that such findings can be attributed to the increasing awareness across organizations to include self-reported measures from the sample under study in order to capture individual-level perspectives (Ben-Arieh, 2005; Fernandes et al., 2012; Hagerty & Land, 2007; Moore et al., 2012; 2014). This finding was shown in the present study whereby indices indicated the importance of the use of self-reported data, and with the most self-reports found in Risk Behaviours. Importantly, data from this unit of analysis can be used as a measurement strategy to align or contradict the level of wellbeing observed in the environment through objective measures across youth outcomes (Moore et al., 2007). Additionally, it was found that one index solely focused on self-reports from the population at hand (WHO, 2010). Indeed, although findings showed that most indices placed importance on the inclusion of a large amount of self-reported data across all domains, one index made efforts to incorporate the exclusive use of self-reported data.

Another finding showed that a significant proportion of indicators defined youth wellbeing through a deficits framework. That is, in line with previous findings, it appears that

youth wellbeing indicators are most often measured and driven by data related to youth weaknesses and vulnerabilities (Moore et al., 2012; 2014). Fewer indicators addressed wellbeing using a strengths-based approach and examined positive indicators as a way to assess wellbeing and development. In the present study, all indices indicated some importance to incorporate positive developmental indicators. Thus, in line with previous findings, it indeed shows that there is an increasing awareness and need to focus on strengths-based, positive youth outcomes in the creation of multidimensional composite indices.

### **Variability in Youth Age Ranges**

Lastly, it was found that age discrepancies existed across indices in the inclusion criteria and cutoffs for youngest and oldest samples. Findings indicate that difficulty remains to clearly identify age range inclusion and capture “*Who are ‘youth’?*” Indeed, although outcomes were often a combination of multiple age groupings, it was found that a large proportion of measures were catered specifically to youth. This finding was surprising, and highlights a possible emerging trend to capture data specifically catering to older cohorts of youth (Land, Lamb, & Mustillo, 2001). Nevertheless, many measures combined broad age ranges for one measure, adding challenges to disentangle age-specific outcomes.

In the present study, only one index separated all outcomes by age and gender, thus providing distinct outcomes for each category (WHO, 2010). Indeed, it is beneficial to pay more attention to measures that separate ages by specific cohorts across wellbeing categories. A focus on age disaggregation is important in yielding critical information in the assessment of environmental impacts that affect youth for different age groups. Such findings could also help elucidate the contexts in which some youth flourish and environments in which others fail (Lerner, 2006; Lerner et al., 2005). Lastly, it is possible that data could be strengthened by

incorporating a longitudinal approach in the data collection strategy. This approach is important as a way to identify ongoing changes in optimal growth and wellbeing, in response to community or societal trends that have an impact on the wellbeing of youth. In light of the above, finding consistencies across indices in data collection methods and conceptualizations of youth wellbeing can contribute to a more credible and reliable comparison of youth within and across nations (Ben-Arieh, 2005; Moore et al., 2007; 2014).

### **Chapter 3: Study 2**

**Evaluating the impact of a school-based arts program on adolescent socio-emotional wellbeing, development, and school experience among inner-city, multicultural communities**

## Introduction

Adolescence is a sensitive developmental period encompassing complex and transitioning environments (Eccles & Gootman, 2002; Merikangas, He, Burstein et al., 2010). During this period of development, adolescents are expected to form their identities (Erikson, 1968; Holloway & Lecompte, 2001), while navigating challenges across multiple layers of environmental influence, such as within the family (Belsky, 1984), at school and with peers (Kochel, Ladd, & Rudolph, 2012; Masten, Roisman, Long et al., 2005), and among broader neighbourhood contexts (Bronfenbrenner, 1979; Lerner et al., 2005). Adolescents are also embedded within important transitions and maturation, which can add stress, such as changing from middle school to high school (Ferguson, & Snipes, 1997; Holloway & Lecompte, 2001). Consequently, if not adequately supported, such difficulties can accumulate for these adolescents, who can show increased susceptibilities to develop emotional and behavioural problems (Phelps, Balsano, Fay et al., 2007), experience academic challenges (Masten et al., 2005), and can suffer long-term consequences, such as identity confusion (Erikson, 1968; Merikangas, He, Burstein et al., 2010). In spite of these challenges, there is a need to better understand the supportive structures that protect these youth from experiencing adversity and isolation. Studies have shown that arts-based programs can provide a way for adolescents to thrive within their schools, social circles, and neighbourhoods (Coholic, Eys, & Lougheed, 2012; Hampshire & Matthijsse, 2010; Remer, 1996; Upitis, 2011; Upitis & Smithrim, 2003; Wright et al., 2006). Such programs have been shown to benefit adolescents by providing a supportive and inclusive infrastructure consisting of developmentally appropriate tasks (Hampshire & Matthijsse, 2010; Wright et al., 2006); opportunities to experience positive bonds with facilitators (Holloway & LeCompte, 2001); improving access to the program through integration

into existing classrooms and meaningful engagement with peers (Baker & Harvey, 2014; Eccles & Gootman, 2002; Holloway & LeCompte, 2001). Through these experiences and opportunities, such programs support adolescents to develop a sense of mastery and personal efficacy over their own experiences (Wallace-DiGarbo & Hill, 2006). Art programs can also help adolescents improve socio-emotional wellbeing and reduce maladaptive behaviours (Wright, John et al., 2006; 2006; Douglass, 2011; Reynolds, Nabors, & Quinlan, 2000), allow young people to gain skills in various art techniques (Wright, John et al., 2006), and develop confidence, self-expression (Holloway & LeCompte, 2001), and a connection with school (Lima, 1995; Mynaříková, 2012).

However, although studies have examined the relationship between arts-based programs and adolescent outcomes, less attention has been paid to the processes through which these programs have an impact on their participants. In other words, there has been a dearth of research evaluating the processes that make a program effective. Such an evaluation builds on a structure of program implementation ideals, including adherence to activity implementation, adequate delivery of components from sensitive facilitators in the program, and having an organized infrastructure (Carroll, Patterson, Wood et al., 2007; Patton, 1994; 1997). Another factor that requires further attention is the ability for art programs to be accessible for its participants among inner-city, underserved communities. In the present study, it was important to provide a program that reduced typical barriers of program attendance, such as concerns with accessibility, cost, and safety (Hampshire & Matthijsse, 2010). Thus, the community art program took place inside the classroom. The present study used a mixed-method evaluation design (i.e., integrating quantitative and qualitative data; Caracelli & Greene, 1993) to examine the impact of a school-based art program on the socio-emotional wellbeing, positive development, and school

experiences of adolescents in their grade 8 classrooms. Additionally, focal concerns included better understanding the effectiveness of the art program and the ways in which it provides adolescents with a positive infrastructure among an inner-city, multicultural community.

### **Arts-Based Programs as Supports for Adolescents**

Arts-based programs have been shown to be important in supporting wellbeing and development for youth (Holloway & LeCompte, 2001; Mantie, 2008; Mynaříková, 2012; Wright et al., 2006; Vygotsky, 1971). Studies suggest that this form of art activity provides adolescents with innovative and creative outlets through which to express themselves (Lima, 1995; Mynaříková, 2012) and allows young people to organize abstract ideas or symbols into meaningful experiences (Holloway & LeCompte, 2001; Matthews, 1989; Vygotsky, 1971). Additionally, art programs create a space for adolescents to learn about themselves within a supportive environment (Baker & Harvey, 2014; Holloway & LeCompte, 2001; Wright, John, Alaggia, & Sheel, 2006). Given the value of programs that foster art-based activities, there is a need to better understand some underlying processes through which such programs may serve to be effective for adolescents and have a positive impact on their socio-emotional outcomes.

### **An Ecological Systems Theoretical Framework**

Taking an ecological systems perspective to adolescent development (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 1998), it was important to articulate the ways in which adolescents are embedded within the multiple systems in their environment, such as within families, peers, schools, and communities. In the case of the present study, it was important to assess the role of the community and school systems, as well as connections with influential others, which interacts with and sets the stage for the adolescent youth to thrive and mature in

their environments. Additionally, the beneficial role of these systems is to provide a supportive infrastructure particularly for adolescents who are making meaning of their external environments, as well as of their personal experiences and identity (Eccles et al., 2002; Erikson, 1968). Accordingly, the present study focuses on the interactions between the adolescent and the important role of an arts program that goes inside the classroom. The program also incorporates influential others in order to address how the program has an impact on the socio-emotional wellbeing of these young adolescents. Similarly, tied to this perspective are scholars who propose that the effectiveness of arts-based programs for adolescents is largely based on: (1) providing adolescents with a supportive infrastructure that is relevant, inclusive, and developmentally appropriate (Durlak et al., 2010; Eccles & Gootman, 2002; Mynaříková, 2012; Wright et al., 2006), (2) establishing a positive rapport with artist facilitators (Holloway & LeCompte, 2001; Wallace-DiGarbo & Hill, 2006; Waller, 2006), and (3) reducing barriers to access the program inside the classroom and creating a meaningful experience with peers (Baker & Harvey, 2014). Such environments nurture a space where adolescents can strengthen self-efficacy, which consists of creating a sense of mastery and agency over their art creations, as well as making sense of their identities and personal experiences through art (Bandura, 2000; Wallace-DiGarbo & Hill, 2006).

### **The Benefits of a Supportive Environment**

Studies suggest that adolescents can benefit from taking part in arts programs that provide a supportive infrastructure across neighbourhoods, as well as within neighbourhoods that are demographically diverse (McCay et al., 2011; Owen et al., 2008; Wright et al., 2006). Benefits are related to the context in which art-making takes place, including an environment that is inclusive, youth-friendly, and takes a non-traditional approach in order to attract a broad

range of participants (Eccles et al., 2002; Hampshire & Matthijsse, 2010; Mantie, 2008). For example, art programs have been shown to benefit young participants by taking an approach that does not conform to one standard model, and does not require participants to have a priori skillsets where failure may be possible (e.g., advanced singing skills; Hampshire & Matthijsse, 2010; Mantie, 2008). That is, a program can be particularly inclusive and supportive if art learning is non-judgmental, non-discriminatory, and where participants are given the tools to succeed in achieving the task at hand (Hampshire & Matthijsse, 2010; Mantie, 2008). In other words, having an art program that is flexible can reduce or remove structured norms in art-making, thus further creating an environment that is inclusive to participants. Such programs create significant potential for personal growth and participation amongst demographically diverse populations (Mantie, 2008). Conversely, studies have shown that some participants who feel they lack particular dispositions or skillsets to thrive in an art program show program disengagement (Hampshire & Matthijsse, 2010; Mantie, 2008).

### **The Important Role of the Artist Facilitator**

Studies have shown that artist facilitators in art programs can lead to positive youth outcomes (Forman et al., 2009; Holloway & LeCompte, 2001; Wallace-DiGarbo & Hall, 2006; Waller, 2006). Indeed, it has been shown that the artist facilitator plays a vital role in contributing to adolescent wellbeing (Holloway & LeCompte, 2001). That is, a sensitive facilitator enables a form of instruction that acts as a guide, yet also supports the opportunity for participants to think independently and critically and gives participants the space to build skills and personal mastery (Holloway & LeCompte, 2001; Vygotsky, 1971). For example, a facilitator may provide participants with guidance and tools to complete particular tasks, with the ultimate goal of enabling these participants to increase their skills in independent decision-making and

critical thinking (Holloway & LeCompte, 2001; Luke et al., 2007). Thus, the facilitator poses as a model or mediator that supports participant growth, and not only a supplier of knowledge.

Particularly noteworthy is that facilitators have been shown to provide a supportive space that makes the program relevant for adolescents (Eccles et al., 2002). For example, characteristics of the facilitator include supporting participants by taking the time to work with them one-on-one while teaching them new tools and perspectives in various forms (Eccles et al., 2002; Mynaříková, 2012; Waller, 2006). Additionally, facilitators can provide a space that breaks down the power dynamic between the adult facilitator and student relationships (Forman et al., 2009; Wallace-DiGarbo & Hall, 2006). That is, facilitators who lead an activity by first expressing themselves and sharing their personal stories and experiences can instil a positive rapport and show young participants the powerful expression of being vulnerable. In this way, facilitators act as role models and provide young participants with new tools to express themselves. As well, facilitators can enable young individuals to feel better equipped to engage in activities that require reflection and expression of personal experiences. Facilitators may also be selected based on particular indicators that help adolescents feel more comfortable to learn from them (Forman et al., 2009; Van Lith, Fenner, & Schofield, 2009). For example, having facilitators that show similar socio-cultural demographics in which these adolescents are embedded can be an essential component to foster participant engagement and expression in various tasks at hand (Coholic et al., 2012). In other words, the diversity of student demographics in a given program is reflected in the facilitators who are also diverse in their socio-demographic backgrounds. An additional benefit to fostering diversity among facilitators is found within schools, given that the school context is a particularly safe space that builds on the adolescent's systems of the neighbourhood-level diversity (Van Lith et al., 2009).

Similarly, artist facilitators in art programs have been shown to give adolescents a space to develop a sense of mastery and personal agency over their own experiences (Freire, 1996; Holloway & LeCompte, 2001; Vygotsky, 1971). Here, it has been shown that sensitive facilitators can foster a space that enable adolescents to make sense of personal experiences, build skills in confidence and expression, and gives adolescents a voice and independence to make decisions (Forman et al., 2009; Wallace-DiGarbo & Hill, 2006; Wright et al., 2006). Additionally, learning various art techniques can be powerful because they provide participants with new tools to share experiences and express themselves (Holloway & LeCompte, 2001). Importantly, studies suggest that such skill development can be particularly valuable for adolescents who are from diverse socio-demographic contexts, and high-needs areas who may otherwise have a marginalized voice (Freire, 1996). In this way, art programs that are offered to adolescents in these areas can enhance program participation and provide these adolescents with resources and tools, that may otherwise be limited or absent in their environments (Mantie, et al., 2008; Wright et al., 2006).

Lastly, studies have shown the importance of developing such skills for adolescents during a particularly important transition stage in their lives, such as forming their identities during the move from middle school to high school (i.e., grade 8 to grade 9; Holloway & LeCompte, 2001). For example, Holloway and LeCompte (2001) showed that facilitators in art programs provide this age cohort with new mechanisms, such as using art techniques for self-expression, to use as practical tools moving forward into high school. These tools support adolescents with steps toward maturation, such as feelings of agency, empowerment, and a sense of accomplishment, as well as feeling more confident, self-aware, and having enhanced images of themselves (Holloway & LeCompte, 2001).

## **Reducing Barriers and Creating Meaning by Bringing a Program Inside the Classroom**

Indeed, a noteworthy factor that has been shown to create a meaningful program experience has been the opportunity of bringing a program inside the classroom (Eccles et al., 2002; Forman et al., 2009; Holloway & LeCompte, 2001). Eccles and colleagues (2002) suggest that an in-school program becomes an extension of the classroom, and fosters an inherent supportive infrastructure and safe space for young individuals who are engaging in new forms of learning. It has been shown that in-school programs are important for two reasons. First, these programs can increase participation and inclusiveness by reducing typically documented barriers to access the program, such as the burdens of travel, cost, or safety concerns (Eccles et al., 2002; Wright et al., 2006). Second, the classroom setting becomes a foundation upon which adolescents can learn with familiar others in order to support a collective identity amongst peers and teachers (Holloway & LeCompte, 2001; Vygotsky, 1971).

First, given the inherent challenges associated with program accessibility and participation for young individuals, studies suggest that finding ways to reduce deterrents to engage in a program are critical (Hamphsire & Matthijsse, 2010; Mantie, 2008; Wright et al., 2006). For example, Hamphsire and Matthijsse (2010) examined the effects of a music program for participants ages 9-11 years old across three socio-demographically diverse neighbourhoods. The authors found that, while many participants expressed interest in the program, participants within lower socioeconomic conditions showed lower levels of involvement in the program (Hamphsire & Matthijsse, 2010). Importantly, for these participants, there were greater barriers to access the program, including fear of neighbourhood safety and parental concerns to travel to the program, which outweighed the opportunity to engage in this extra-curricular opportunity (Hamphsire & Matthijsse, 2010). Thus, it is critical that a program provides an accessible and

relevant context for its participants in order to enhance beneficial program effects and participation (Hamphsire & Matthijsse, 2010; Mantie, 2008; Wright et al., 2006). Indeed, studies have shown that bringing a program inside the classroom is an important way to increase program accessibility and participation (Durlak & Wells, 1998; Eccles & Gootman, 2002). That is, programs can be particularly successful inside the classroom, given that participants do not have to rely on transportation or financial costs, particularly when communing across communities where transportation may be inaccessible or inefficient. Such conditions have also shown to reduce attrition rates among program participants (Durlak & Wells, 1998).

Second, studies have shown that engaging in arts programs in the classroom enhances communication amongst peers and teachers either verbally or expressively through art (Baker & Harvey, 2014; Holloway & LeCompte, 2001). That is, learning amongst familiar others can promote the ability to develop connections and trust with supportive others, and enables the ability to share difficult feelings or personal experiences through art expression. This environment can provide adolescents with a sense of relief of expressing themselves and feeling empowered within a group dynamic (Baker & Harvey, 2014). For example, Baker and Harvey (2014) examined the effects of a music program offered to twelve students in a grade 3/4 classroom. Quantitative and qualitative data revealed that the classroom environment was a meaningful space to support collective learning and increase dialogue among peers. Additionally, the program fostered skills in decision-making, shared behaviours, and reduced student insecurities to create art on their own. This study suggests that the collective space of the classroom creates a form of safety, as well as social skills training, thereby strengthening focus and attention to the task at hand with familiar others. However, other studies suggest that if peer supports are not in place, or if the safe space is impaired, it is possible that goals of the program

can be hindered for young participants engaging in art programs (Hamphsire & Matthijsse, 2010).

### **The Effect of Art Programs on Adolescent Development and Wellbeing**

It is well documented in the literature that arts-based programs can support the development and wellbeing of adolescents across a range of domains and across contexts (Chilton, 2013; Fraser & Keating, 2014; Hamphsire & Matthijsse, 2010; Mantie, 2008; Matarasso, 1996; Mynaříková, 2012; Wright et al., 2006; 2006). For example, various art techniques, such as music, painting, and dance, have been shown to provide an effective therapy to reduce maladaptive behaviours and improve psychosocial health for marginalized populations of adolescents at risk, including for refugee youth who suffer from trauma (Rowe, Watson-Ormond, English et al., 2017), drawing and collage exercises with mindfulness methods for individuals in child protection services (Coholic et al., 2012), for young survivors of abuse (Douglass, 2011; Reynolds et al., 2000), and among those with special needs, such as autism (Epp, 2008). Arts-based learning has also been described as having a positive impact on neurological functioning as a function of reducing anxiety and stress reactions (Chilton, 2013).

Although art programs have been shown to support development and wellbeing across a range of areas, the present study focused on art-based programs that foster positive development among adolescents from diverse socio-demographic communities (Hampshire & Matthijsse, 2010, Mantie, 2008; Wright et al., 2006; 2006), as well as within their classrooms (Baker & Harvey, 2014; Holloway & LeCompte, 2001; Kisida, Bowen, & Greene, 2016; Lampert, 2013; Matarasso, 1996; Mynaříková, 2012). Art programs can also take a strengths-based approach to wellbeing in order to foster positive outcomes among youth (Holloway & LeCompte, 2001). In line with a positive youth development framework, scholars articulate that programs that focus

on youth strengths and assets can be particularly beneficial in fostering socio-emotional development and wellbeing over time (Larson, 2000; Lerner et al., 2000; 2005). Thus, in the present study, it was important to define wellbeing in the context of an art program that focuses on positive psychosocial facets of adolescent development related to psychological and social outcomes, respectively (Erdem et al., 2016). That is, identifying psychosocial outcomes for adolescents could highlight the value of a program that fosters their positive growth. Particular attention was paid to exploring self-efficacy in an academic context, as well as the effect of a program in providing a supportive infrastructure, and the social relationships that are important during this stage of growth. As previously noted, this framework of wellbeing is grounded in a structure of five important Cs of development, including competence, confidence, connections, character, and care and compassion (For further details of the 5 Cs refer to p. 9 or p. 93; Erdem et al., 2016; Lerner et al., 2005), and could thus bring awareness to the positive impact of an art program on adolescents situated in diverse, multicultural neighbourhoods.

For example, Holloway and LeCompte (2001) captured the effects of a theatre arts program, taking place in a middle school for adolescent girls. Using a qualitative, case study design, results showed that the theater program contributed to increases in self-identity and self-expression among these adolescents. It was found that, through components of the program, including the use of theatre techniques, a focus on performance strategies, and becoming another character, participants were better able to gain insight and skills into expressing themselves and creating new images of themselves and their future (Holloway & LeCompte, 2001). It appeared that other activities including critiquing each other's work and collaborative decision-making also contributed to their positive growth. While valuable, data from this study came from a small sample and single informants of five girls using qualitative interview data.

Similarly, studies have also shown that art programs benefit young individuals in their classrooms and help facilitate the development of social skills and collaborative learning amongst peers (Baker & Harvey, 2014; Durlak et al., 2010), enhance skills in critical thinking and inquiry (Kisida, Bowen, & Greene, 2016; Lampert, 2013; Luke et al., 2007), as well as provide adolescents with an opportunity through which to increase interest in and motivation for learning in schools (Iwai, 2002; Luftig, 1994; Mynaříková, 2012; Upitis, 2011).

It has also been shown that such skill development can also extend to learning and academic success in some other subject classes (Holloway & LeCompte, 2001; Mynaříková, 2012). For example, Mynaříková, (2012) showed the value of an art program in a classroom using drawing, music, and drama. Through pre- and post-evaluation surveys completed by participants in grade 5, the author was able to identify measurable changes in these young individuals who showed increases in communication skills, self-confidence in expressing themselves, and more accepting of peers within their school, as well as increased positive attitudes toward various school subjects. The above study also suggested ways in which the art program supported these children, including through enhanced communication with and support from their teachers, and fostering a positive, collaborative, and safe class climate (Mynaříková, 2012). However, while valuable, there are limitations to the generalizability of this study, given that data were of younger participants ages 11-12, and situated in one city in the Czech Republic.

### **Remaining Challenges for Arts-Based Programs**

Despite identifying the value in arts-based programs, few studies have identified the evaluation processes of the program that make it effective for its participants (Newman, Curtis, & Stephens, 2003). In other words, less attention has been paid to the corresponding evaluation processes of art programs that provide effective assessments to identify program success, such as

through the use of implementation fidelity testing (Carroll et al., 2007; Durlak & DuPre, 2008; Howell & Yemane, 2006; Patton, 1997; Quinby, Fagan, Hanson et al., 2008; Sidani, 2015).

Accordingly, a central component to identify the quality of the program in the present study was the evaluation of program implementation processes.

Previous studies have shown the importance of integrating empirically-based and rigorous approaches to program evaluation in order to assess the implementation process and quality, contributing to program success (Durlak & DuPre, 2008; Quinby, Fagan, Hanson et al., 2008). That is, an analysis of implementation fidelity identifies the extent to which a program is delivered as intended and can effectively provide its services to have an impact on its population base. Studies have shown that key aspects to assess a comprehensive understanding of program implementation include, the dosage of and adherence to activities and the rate of activity completion (e.g., activity completion checklists), characteristics of program staff and facilitators (e.g., preparedness, organized, clearly explain instructions), and program organization levels (e.g., organized infrastructure; Carroll et al., 2007; Durlak & DuPre, 2008; Teague, Bond, & Drake, 1998). Additional ways to assess implementation include measures at the participant-level, such as observing the extent to which participants completed the program activities as intended, as well as assessing attendance rates and participant satisfaction (Carroll et al., 2007; Teague et al., 1998). These variables include such measures as the number of sessions attended (participant dosage), and the degree to which participants were satisfied with the program (Carroll et al., 2007; Durlak & DuPre, 2008; Teague et al., 1998). Studies have shown that both self-reports and observational measures can be used to assess different aspects of the implementation process (Durlak & DuPre, 2008). Importantly, Carroll and colleagues (2007) identified the importance of assessing the quality of facilitators in delivering the program and the

program uniqueness (differentiation). These components can be identified as the sensitivity used by facilitators to deliver the program, as well as the important aspects that make the program unique, respectively. Thus, sensitive delivery and program uniqueness can further highlight the ability to accomplish positive program effects (Carroll et al., 2007).

In a review of over 500 studies, Durlak and DuPre (2008) showed the importance of capturing implementation fidelity on program outcomes and success. Results showed that, through fidelity testing, including assessing the dosage of adherence to activities, the review of studies was able to confirm the importance of implementation testing to show that a program could be particularly effective for its participants when it had been implemented as it was intended (Durlak & DuPre, 2008). In fact, programs that show high levels of implementation fidelity have been shown to have significantly greater positive effects, often measured by the effect size, for participant outcomes than programs that show low implementation fidelity or that do not monitor such levels (Durlak & DuPre, 2008; Taut & Aiken, 2011). Conversely, if a program is effective but has not shown implementation fidelity, it is difficult to conclude why it is effective, and cannot fully provide a valid assessment of the success of the program (Carroll et al., 2007).

Studies suggest the importance of evaluating art-based programs (Newman et al., 2003). However, evaluations addressing implementation fidelity and program quality have been largely lacking. Indeed, such evaluations would benefit art programs by providing an important assessment of adherence to the program components and showing whether the program was delivered in the manner in which it was intended. Additionally, evaluations of arts programs can be important in assessing facilitator sensitivity in adapting a program to the specific context, such as delivering the instructions in a kind, caring, and compassionate way. In other words, it is

important that components are not just delivered, but tailored in the way they are delivered to a specific context, being sensitive to demographically diverse populations (Harn et al., 2013). Indeed, scholars suggest that components of a given program may not need to be delivered exactly the same way every time (Carroll et al., 2007; Durlak & DuPre, 2008; Harn et al., 2013). For example, Carroll and colleagues (2007) articulated that, “adherence may not require every single component of an intervention to be implemented. An intervention may also be implemented successfully, and meaningfully, if only the "essential" components of the model are implemented” (p. 44). Indeed, it has been shown that an acceptable level of implementation fidelity for programs to be effective can range from 60% to 80%, with some adaptability of the delivery of the program to fit specific contexts and participant needs (Durlak & DuPre, 2008; Harn et al., 2013; Sidani, 2015). In addition, Harn and colleagues (2013) showed that over 88% of programs that assessed implementation fidelity indicated the use of some modification in program delivery. These findings suggest that a good facilitator takes the time to show care and concern for the participants, and adapts the program components to the given context. Implementation fidelity variables have also been shown to interact and are identified as moderators between program objectives, activities, and outcomes (Carroll et al., 2007). Thus, implementation fidelity in the present study was critical as a precursor variable to examine the program impact on adolescent outcomes and to identify overall program success. That is, a fidelity assessment was considered key to explain specific mechanisms through which the program achieved – or failed to achieve – beneficial outcomes and program success (Carroll et al., 2007; Dreeszen, 2003; Durlak & DuPre, 2008; Patton, 1997; Taut & Aiken, 2011).

### **The Interplay Among Neighbourhoods, Art Programs, and Adolescent Outcomes**

Capturing adolescent outcomes within a local neighbourhood context can be an important

way through which to identify particular strengths and limitations of a given community. Additionally, it is important to identify the benefits of programs nested within schools in the community that may foster adolescent development. There is substantial evidence linking neighbourhood characteristics with adolescent socio-emotional wellbeing. Studies have shown negative associations between communities in lower socioeconomic conditions and a greater risk for young individuals to develop behavioural and emotional problems among ages 12-17 years old (Aneshensel & Sucoff, 1996), among children and youth (Bradley & Corwyn, 2002; Evans, 2006; Leventhal & Brooks-Gunn, 2000), and are more likely to show longitudinal challenges with difficulties in schools (Masten et al., 2005) than adolescents from higher income areas. For example, neighbourhood contexts that have been associated with greater vulnerabilities for adolescents can often be attributed to reduced neighbourhood safety, such as high crime and violence rates, as shown through self-perceptions and observational data, as well as limited resources found in the community, such as limited access to local community hubs, safe public spaces, and lower sense of community cohesion (Leventhal & Brooks-Gunn, 2000; Sampson, Raudenbush, & Earls, 1997).

Indeed, if a program is accessible, and achieves its goal to be implemented as intended, studies suggest that such program outcomes can produce measurable community-wide effects, such as improving engagement in the community and positive bonds with community members (Catalano et al., 2002; Chung, Jones, Jones et al., 2008; Hawkins, Catalano, & Miller, 1992; Wright et al., 2006). In this case, art programs that are relevant and accessible to youth within a community can be effective in strengthening adolescent community engagement and help reduce maladaptive behaviours, such as violence and crime (Wright et al., 2006; 2006). However, less focus has been placed on identifying the relationship between neighbourhood characteristics,

program implementation, and positive adolescent outcomes. Thus, the present study focused on exploring a neighbourhood that can be characterized as inner-city and diverse in both multiculturalism, as well as in socioeconomic conditions, using corresponding census data (Aneshensel & Sucoff, 1996; Statistics Canada, 2011). Having a better understanding of processes involved in programs dedicated to arts-based learning can be an important step toward identifying the value of arts for adolescents (Newman, Curtis, & Stephens, 2003), the impact and characteristics of program effectiveness on participant outcomes (Catalano et al., 2002), as well as the extent to which the program fosters positive youth development (Catalano et al., 2002; Damon, 2004; Larson, 2000).

### **The Present Study**

The present study focused on the effect of a school-based arts program<sup>5</sup> on adolescent outcomes. The program provides students with an opportunity to take part and learn about two important creative art techniques, including photography (digital media) and poetry (spoken word). Participants were students in grade 8 from three classrooms situated in a multicultural and diverse neighbourhood in Toronto, Canada. In addition to an assessment of adolescent outcomes, including socio-emotional wellbeing and school experiences, it was also important to identify program implementation processes that support adolescent wellbeing.

The present study identified program fidelity by assessing adherence to the processes of the program and whether activities were being implemented as intended, as indicated by facilitators of the program. Additional data were used from the artist facilitators and classroom teachers to capture program organization and infrastructure. It was important to incorporate this fidelity assessment in the present study for four reasons: first, implementation fidelity has been

---

<sup>5</sup> Details of the program, "*Shazaam: In Focus*" can be found in the 'Methods' section.

largely lacking in previous studies for art programs. Second, in line with a systems perspective (Bronfenbrenner, 1979), multiple environments contribute to the formation of this program, thus adding complexities in the layers of influential systems that contribute to the program model. In this way, the community organization forms a partnership with school boards and goes inside the classroom to work with teachers and students who form the program environment. Third, artist facilitators are incorporated and selected based on characteristics of those who represent the neighbourhoods among which the program takes place, adding another influential layer to student outcomes (Remer, 1996). Thus, the community organization is involved in the initial organization and infrastructure of the program, and sets the stage for artist facilitators and classroom teachers to have the supportive and organized structure for students and for the program to occur. Lastly, at the student level, these participants take part in the program and its activities. Thus, it is critical to assess whether students are able to complete the tasks at hand, as well as identify their overall program engagement and satisfaction, using observational and self-reported measures. Using multiple forms of evaluation measures and indicators can enable a way to identify the impact of the program on student outcomes (Durlak & DuPre, 2008). It was also possible to assess relationships between implementation fidelity variables and participant engagement, showing whether outcomes were related to intervention effectiveness (Abbott et al., 1998).

## **Study Objectives**

The present study used a short-term longitudinal (over three time-points), mixed-method design, with quantitative (surveys, observational) and qualitative (interviews, open-ended responses) data. Data were gathered from artist facilitators, teachers, and students. This study sought to assess the relationship between program processes and outcomes of an arts-based

program, as well as how it can contribute to positive development among adolescents in their grade 8 classroom.

The present study had four broad objectives:

- (1) To establish an evaluation approach that identifies implementation processes using six criteria. The implementation goals were to assess: the dosage and adherence of delivering the program activities as indicated in the program manual; the extent to which students completed the activities, as instructed; whether the program provided a supportive infrastructure and was organized; the extent to which students received clear instructions; student attendance rates; participant satisfaction with the program. First, it was hypothesized that the program components and activities would be delivered as expected and would show high levels of consistency of activity adherence across schools. It was also believed that a certain degree of modifications of activities would take place given facilitator-level sensitivity and adaptability in delivering the program components as they fit to the specific classroom context. Second, if the program was delivered as indicated, it was expected that students would show high levels of activity completion, indicating that they could adequately complete their instructed tasks. Third, it could be expected that the program would also show high ratings from facilitators and teachers of organization levels and infrastructure, as well as high student-rated instruction clarity. Lastly, given the classroom context, it was expected that students would show high rates of attendance and participants would indicate satisfaction with the program.
- (2) To evaluate the impact of the program on student outcomes, as identified by facilitators, teachers, and students, using a model of 5 Cs in positive youth development (Larson, 2000; Lerner et al., 2000; 2005). It was also important to assess student demographics,

taking into account diversity of gender and ethnicities. Using multilevel modeling to examine time-points nested within students, it was expected that adolescent self-reports of their socio-emotional outcomes and school experiences would show improvements overtime. It was also expected that teachers and artist facilitators would report on the positive impact of the program for these adolescents, observing student transformations in confidence-building and self-expression. Lastly, using a regression analysis, it was expected that a significant factor contributing to adolescent enjoyment of the program would be related to the support and safe space provided by program facilitators.

- (3) To collect feedback from key users of the program in order to identify the extent to which the program was beneficial and relevant in general, and to identify areas for improvement. Using mixed-methods from multiple informants, participants would elucidate perceived challenges and successes of the program, and suggested improvements. It was also expected that teachers and artists would support the structure and components of the program.
- (4) To explore neighbourhood profiles using Census data. These data would enable the identification of neighbourhood characteristics as it compares with student demographics. Additional relationships will be explored between community-level socio-economic status (SES) and ethnicities as compared with participant-level demographics. Here the goal would be to provide comparisons of objective neighbourhood data in which these students are embedded, in order to reduce the collection of subjective-level data that may otherwise be invasive or unknown to adolescents (e.g., household income). It was expected that the neighbourhood profile would be closely linked with student demographics, and show high levels of multiculturalism and socio-demographic

variability within the neighbourhood.

## Methods

### Evaluation Approach

A partnership was established between Ryerson University and Lakeshore Arts to evaluate the program, *Shazaam! In Focus* with respect to program implementation, delivery, processes, and outcomes. In order to maintain respect of the partnership, an Evaluation Agreement document was created, consistent with best practices for a partnership between academic and community partners for an evaluation study (Patton, 1997), following principles adhered to by both parties including, to maintain mutual respect and accountability between the parties, to recognize the expertise and responsibilities of each party, to respect the individual privacy rights of the community partner's program participants, to recognize the value of capacity building for project members, and to support the processes involved for each party to complete program requirements, and meet partnership tasks. The study procedure and abstract were submitted to Ryerson University's Research Ethics Board (REB) in the Spring 2017 and were deemed as quality improvement work and exempt from REB approval. The ethical guidelines put forth by the Canadian Evaluation Society (CES) were adhered to in this project. The REB email correspondence to undertake this evaluation can be found in Appendix D.

This evaluation study used a utilization-focused evaluation (UFE) approach (Patton, 1997; 2008) with a developmental evaluation component (Patton, 1994; 2010) to examine the effects of the art program. Specifically, this study incorporated a mixed-method evaluation and time series design to estimate the effect of time on outcomes among participants in the program (Cook & Campbell, 1979; Grimshaw et al., 2000). The UFE and developmental evaluation designs were used in the present evaluation process for two reasons. First, the program was implemented over two phases and thus could be evaluated at both time-points allowing a

developmental process to occur incorporating ongoing feedback and consultation with key stakeholders (Patton, 1994). Second, the program could become highly useful and relevant to its participants (users), thus becoming better able to serve the schools within communities (uses), as a practical and relevant program (Patton, 1997; 2008; 2011).

### **Utilization-Focused Evaluation (UFE) Design**

This UFE was designed to address specific key concerns related to program *effectiveness* and *areas of improvement*, using quantitative and qualitative data from key users (Patton, 1997). Focal items would address the following questions: (1) whether program management were highly involved in the program and effectively facilitated the preparation and dissemination of the program (implementation processes), (2) whether particular program components had a significant effect on student participants (value), (3) whether the program had a positive impact on the students in various outcome areas of interest (impact), and (4) address feedback from stakeholders, students, teachers, and artist facilitators to inform improvement efforts of the program. The present UFE was designed and conducted taking steps of a UFE in line with best practices (Patton, 1997; 2008; 2013). For a conceptual framework of the evaluation steps guiding the study see Figure 3 below. After establishing a community partnership, the next step was to identify program impetus and objectives (via key informant interviews, logic model). The next steps were to identify best practices in tools for delivery, and develop measures informed by a mixed-method design. After tool creation, it was possible to distribute and collect data, and assess program processes (implementation fidelity). Next, data were analyzed to identify outcomes of key users. Lastly, a developmental evaluation component was used to identify any changes and improvements in program delivery over the course of two implementation phases, including the Spring (Phase 1) and Fall (Phase 2) sessions (Patton, 1994). In the present

evaluation design, it was important to ensure ongoing feedback and consultation from program administrators (Patton, 1997).

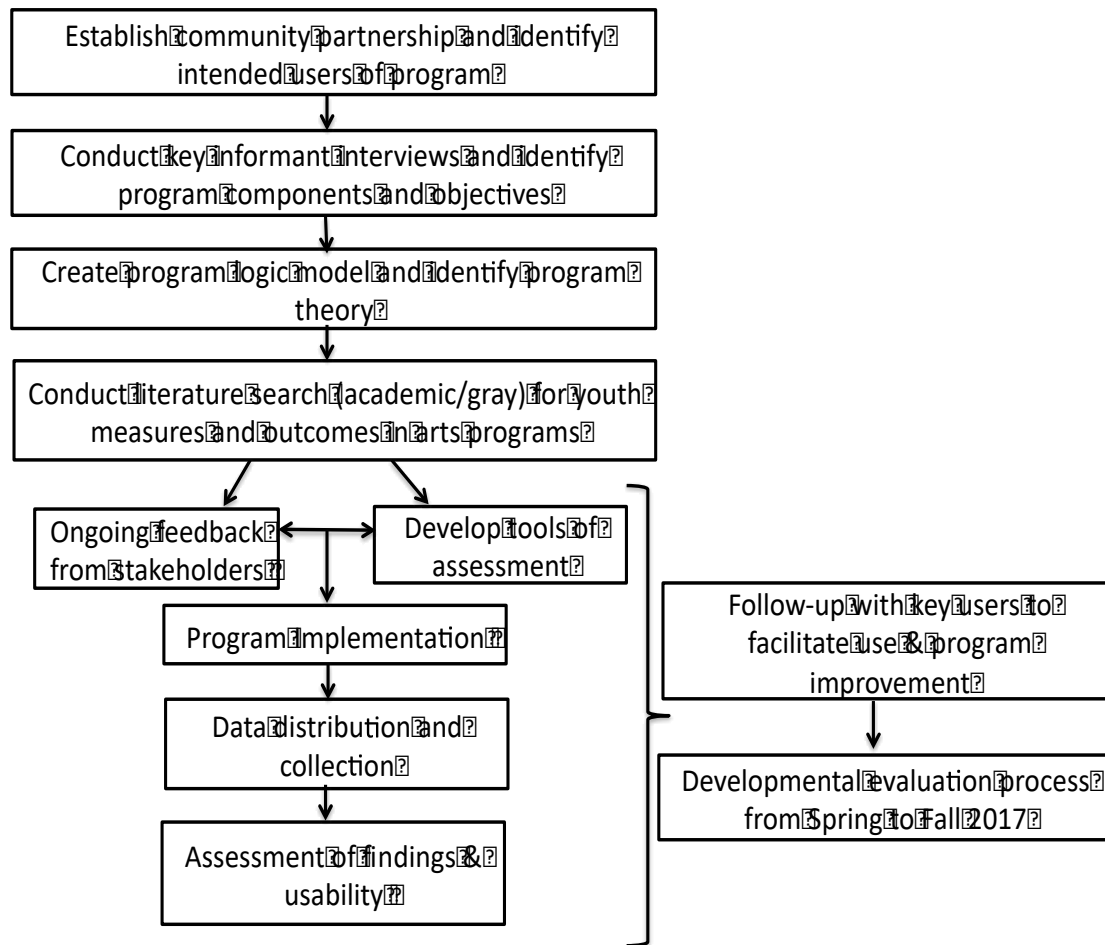


Figure 3. Conceptual framework of the evaluation guiding the study (Adapted from Patton, 2013).

**Mixed-methods design.** In the present study, quantitative and qualitative data would be used in various ways (e.g., Caracelli & Greene, 1993; Creswell & Clark, 2007). First, quantitative data were employed using surveys administered over three time-points, and included at baseline on Day 1 of the program (Time 1), post-program on Day 6 of the program (Time 2), and follow-up, four weeks post-program (Time 3), consisting of Likert-scale response categories (e.g., 1 = *completely disagree* to 7 = *completely agree*), as well as observational measures from

program facilitators and teachers. Second, qualitative data were employed using open-ended responses and allowing participants of the program to incorporate written responses of personal thoughts and feelings related to program components, as well as overall experiences and valuable lessons learned within the program. Interviews with key informants were also used to enable the provision of a comprehensive understanding of the program impetus and objectives. It was important to employ both forms of data collection in the present study in order to elucidate valuable assessments using numerical outcomes as well as verbal or written responses, adding insight to an objective perspective of program components and processes, as well as perceived successes and challenges of the program (Creswell & Clark, 2007).

### ***The Shazaam! In Focus Program***

The program was created by Lakeshore Arts as an in-school program for students in their first period grade 8 classroom. The program has been implemented for over ten years, and has established partnerships with, and support from, local school boards (i.e., the Toronto District School Board (TDSB), and Catholic District School Board, TCDSB) in the inner-city, high-needs areas of Toronto. The program was implemented as a way to provide an opportunity to schools where such art activities and resources have been lacking, or removed from the school boards. The program took place during the Spring and Fall semesters of the school year, over the course of six days in two weeks (two hours per day, and three days per week), and occurred in three different classrooms per semester. The program is always facilitated by one professional photographer and one spoken word artist per classroom, and incorporates an artist assistant to assist with organization of tasks. Importantly, the program incorporates the classroom teacher as an additional support to the students. It is expected that the teacher is highly present and engaged across the six days of the program.

The program emphasizes the importance of self-reflection, expression, and confidence, and provides students with a meaningful opportunity to examine their own life experiences and personal identities through art. Over the course of six days, the students engage in various art tasks. These activities include daily creative writing exercises, a collage creation, a combination of individual and group work, as well as the two primary activities addressed in the present study, whereby each student creates (1) a photograph and (2) poem. First, the metaphorical self-portrait (the photograph) was taken, which was a digital photograph related to an image that is a metaphorical representation of the student (e.g., a tree, the lake, an object in the classroom), but students were not allowed to take a photo of themselves. Each student is also given extensive instruction in camera etiquette, and given the opportunity to edit their photos using Photoshop alongside a photography artist. The metaphorical photograph was accompanied by a written reflection explaining the metaphor, in an activity titled, 'Metaphorical Image Reflection.'

Second, students also completed several creative writing exercises that led into one final Spoken Word (poetry) piece. This activity was related to a self-reflection and titled, 'Reflections of Me,' which they performed on the last day in the form of a Poetry Slam. Students also learn tips and tricks related to giving strong presentations and photography techniques. The program was supported by a 3-year 'Growth Grant' from the Ontario Trillium Foundation (OTF) to undertake this evaluation. OTF is a government-level granting agency that supports organizations in nonprofit sectors (OTF, 2018). The grant was an important addition to the program in order to strengthen *Shazaam: In Focus* in various ways. First, the grant fostered the implementation of an evaluation design in order to enable the development and continued implementation of the program, adding merit to the expansion and improvement of the program. Second, the grant enabled an evaluation, which could track service provision, outputs, and outcomes. The goal of

the grant was to evaluate the program and to be able to offer it to more schools in areas of high-need.

### **Key Informant Interviews**

Consultation interviews with key informants were conducted ( $N = 4$  interviews) to create a program logic model (LM), and to inform an evaluation approach to capture relevant data and help assess outcomes of the program. Specifically, the interviews with key informants were conducted with program administrators ( $N = 2$ ) and artist facilitators ( $N = 2$ ). Each interview lasted 30-60 minutes. These interviews focused on the program's Theory of Change, which was based upon a framework to align the stakeholder objectives with program outcomes (Patton, 1997; 2008). In particular, question items were related to program goals, program impetus, identifiable outcomes, and goals for student participant growth. For details on question items for key informant interviews see Appendix E. These items were in line with the standards of an interview protocol put forth by the Canadian Evaluation Society (Evaluation Canada, 2017; Yarbrough et al., 2011).

### **Program Logic Model and Theory of Change**

The program logic model was used to clearly articulate components and outcomes of the program, including program resources/input, activities, outputs, and outcomes (short, intermediate, long-term/impact). For a detailed overview of the program logic model in this evaluation see Appendix F. The purpose of the LM is to categorize each constituent of the program, and illustrate how each section aligns with others, thus providing a visual roadmap of the program (Hulett, 1997; Schmitz & Parsons, 1999; W. K. Kellogg Foundation, 2004). The LM has been defined as, "a systemic and visual way to present and share relationships among the resources an organization has to operate programs, the activities an organization plans, and the

changes or results an organization hopes to achieve” (WKKF, p. 1). Thus, the LM in the present evaluation was a beneficial way to provide a visual representation of the art program for both stakeholders and evaluators, and a useful illustration to clearly articulate program components and objectives (WKKF, 2004).

After identifying constituents of the program, and creating the logic model, it was possible to identify a theoretical framework upon which the program transpired. As previously noted, the art program used modalities of art that enabled the adolescent participants to engage in art-making that primarily focused on self-reflections (i.e., using photography, collages, and creative writing), to work alongside peers, as well as to take part in multiple presentations of their work. Thus, the impetus of the program was focused on how adolescents in grade 8 can learn to have a voice, articulate their thoughts and feelings, as well as express themselves in a meaningful way. It was also found that key underlying processes for the program’s effectiveness included: (1) providing adolescents with a safe space to engage in art creations and express themselves within their classroom, (2) incorporating sensitive artist facilitators who reflect the socio-demographic characteristics and reside within these neighbourhoods in which these adolescents are embedded, and (3) empowering these adolescents to use vulnerability as a powerful tool to reflect on themselves and make meaning of their personal stories and experiences, particularly as they soon move forward into high school (grade 9). It was expected that through these activities, the program objectives encouraged skill building in developing confidence, self-expression, autonomy, and peer respect. Additionally, the program fostered skill development in using verbal and body language as powerful modalities to express themselves, and to better equip students in decision-making, sharing experiences, being heard, and helping form their identities.

## **Develop Tools for Assessment**

In line with guidelines of a UFE (Patton, 1997), tool creation in the present evaluation would best occur through ongoing stakeholder consultation and feedback. Tool development and question item selection were informed by two processes. First, an assessment of consultation interviews identified key program objectives. Second, an extensive search of the literature within academic and grey documents identified evidenced-based measures previously used to evaluate educational and community programs. The latter methodological component involved consulting a literature base related to existing tools, and an assessment of such previously validated measures. Previous measures were assessed based on psychometric properties (reliability and validity), where available. First, in the present study it was important to use a reliability test based on internal consistency, which indicated that items within a scale were sufficiently correlated with one another in measuring a construct (Streiner, 2003). Internal consistency for scales using Chronbach's alpha ( $\alpha$ ) is considered acceptable when values are above .70. A value above .84 is considered high, and below .70 is considered moderate-low. Second, face validity was used in the present study to identify measures that were relevant to and reflective of each program goal (Holden, 2010; Nevo, 1985). Using these criteria, it was possible to incorporate relevant measurement items that best aligned with program goals.

In response to key stakeholder feedback and concerns, tool development was in line with the priority objectives set forth by the program, and items were selected taking into account (1) items that were relevant and sufficiently captured the different areas of interest in the measures (i.e., using face validity and internal consistency), (2) a condensed length and scope of measures for program users (i.e., less than twenty items selected overall, and 2-3 items per sub-area of interest), and (3) a limited time allotted during the program to complete the self-reported

measures (i.e., 5-10 minutes). Therefore, tool creation in the present study was an adapted construction of previous items, and incorporated a limited selection of items from multiple measures, insofar as it provided enough detail to capture different program objectives. For further details on the measures used in the present study refer to the “Measures” section below.

### **Developmental Evaluation Design**

The developmental evaluation (DE) component included ongoing communication with the community partner, and continuous program and tool improvement after the initial implementation phase and preliminary results from the Spring session. Patton (2010) argues that a sufficient amount of stakeholder engagement and feedback is required for a systematic and sensitive developmental evaluation to take place, and for useful changes to occur. In other words, an effective relationship with ongoing feedback with stakeholders can meet a threshold that enables the enhancement and program improvement that can take place over multiple implementation phases of the program (Patton, 2010). Thus, in the present evaluation study, it was important to integrate feedback from program users, in addition to ongoing stakeholder discussion, in order to enhance tool design, program components, and materials of the program.

The DE component was primarily incorporated as a way to test feasibility from the initial phase (Phase 1), and to enable the testing of logistics in the implementation of new approaches of tool use and methodologies, which could adapt and change for the second phase (Phase 2) in response to data-driven findings (Patton, 1994; 2010). It has been articulated that a DE enables program goals that are “emergent and changing rather than predetermined and fixed, time periods are fluid and forward-looking rather than artificially imposed by external deadlines, and purpose is learning, innovation, and change” (Patton, 1994, p. 318). Thus, the DE design in the present study was used as a way to test feasibility and for the program to become more useful for

users through the ongoing testing and improvement of subsequent phases. Please note that Phase 1 pilot findings of student outcomes were not included in the present analysis, but rather were only used for internal purposes for the community partner. Thus, in the below section of ‘Phase 1 Key Findings’ findings relate to the ways in which the program developed and improved over time, becoming more useful to stakeholders, and program participants.

### **Data Collection Methods**

**Recruitment of schools and artist facilitators.** Recruitment was conducted by the community organization, which has previously engaged a diverse group of schools and artist facilitators across lower-income, high-needs areas in Toronto. In the present evaluation, it was important to build upon the existing collaborations between the community organization, the school boards, as well as with the artist facilitators, in order to reinforce these long-standing partnerships. This form of recruitment stems from an ecological systems approach (Bronfenbrenner, 1979), related to incorporating the overlapping socio-ecological systems that surround the youth’s environment (Epstein, 2001). Such partnerships have also shown to strengthen community-building and foster collaborative environments (Epstein, 2018; Remer, 1996).

Partnerships between the community and school, as well as with the artist facilitators, have shown to be particularly beneficial in such lower income neighbourhoods in order to provide novel programs and resources that support youth development (Remer, 1996). It was also important to select facilitators based on their diverse representations of ethnic diversity, genders, and socio-demographically, and who represent leaders in their fields of photography and spoken word poetry, residing within the neighbourhoods in which these students are situated. The artists’ strong influence in such neighbourhoods largely contributes to the program’s goals

to have a positive impact on students and can positively influence artist-student dynamics (Remer, 1996). In addition, the facilitators received extensive training over multiple training meetings prior to each program session. Thus, students in selected classrooms were given the opportunity to take part in the 6-day program as it corresponded with an agreement from the schools and teachers.

**Consent procedure.** Student consent forms were distributed by Lakeshore Arts on the first day of the program (session 1). The consent forms provided students with information regarding the community and academic partnership, the purpose of the evaluation, and details of the survey measures that would be distributed during the program. The consent document can be found in Appendix H.

### **Phase 1 Implementation**

During the initial evaluation phase (Phase 1), and after tool design, an initial test of the survey tools were implemented in the Spring 2017 across three classrooms in two schools. Study participants included students ( $N = 80$ ), facilitators ( $N = 6$ ), and teachers ( $N = 3$ ). Each informant responded to Likert scale question items and provided open-ended responses related to priority outcomes, as well as perceptions of program effectiveness, strengths, challenges, and suggestions for improvement. Such data were used to inform program stakeholders on whether initial program goals aligned with participant outcomes, and whether the program achieved its intended purpose. The data were distributed and collected by the community program, and analyzed by the evaluation team.

After data collection and analysis, as well as feedback from facilitators, teachers, and adolescents, data were subsequently shared with stakeholders in a collaborative manner (Patton, 1994). Consultation meetings and ongoing feedback with the community partner occurred across

multiple meetings between the Spring and Summer 2017 (meeting were either in-person or by telephone) in order to engage and reflect upon Phase 1 study findings. Questions were addressed reflecting input on tool comfort, relevancy, and areas for program improvement. Thus, data could inform whether the program had an impact on its participants (effectiveness), as well as ongoing reflection of how to improve sessions for future programming (improvement; Patton, 1997; 2010). Results from Phase 1 would set the stage for the second implementation phase and the present study (Phase 2).

### **Phase 1 Key Findings**

Key findings for improvement efforts included, 1. To include artist facilitators with diverse socio-demographic and ethnic backgrounds, and diverse genders. 2. To better explain activity instructions to students. 3. To reduce the amount of “filler” activities, and thus allow students to spend more time on the major program components. 4. To adapt online surveys for artist facilitators and teachers, and make in-person or verbally-recorded responses available. 5. To make class sizes each semester more equal and not too large (e.g., classes with < 30 students). 6. To better systematize student IDs and align them with activity workbooks as a way to maintain program organization. 7. Emphasize the importance of instruction clarity at the annual facilitator training session. Thus, in consultation, revisions, and updates were made over the Summer 2017 for the Fall 2017 sessions.

### **Phase 2 Implementation**

Analyses of the present evaluation study were based on Fall 2017 data. The present study included 74 adolescent students in grade 8 during the Fall 2017 academic semester from three different classrooms in two schools and in one neighbourhood (Statistics Canada, 2017). Study participants also included artist facilitators ( $N = 6$ ), and classroom teachers ( $N = 3$ ).

## **Quantitative Implementation Indicators**

Measures for the implementation assessment were based on core indicators of program implementation, and included the following assessments: (1) fidelity of program delivery, (2) observer rated student activity completion, (3) program infrastructure – organization and preparedness, (4) student-rated clarity of rules and expectation, (5) participation rate of students, and (6) participant satisfaction.

**Fidelity of program delivery.** An activity checklist was used to measure adherence to the program manual and assess the extent to which activities were implemented and completed as intended. The checklist was completed by program facilitators at the end of each day of the program, and included a table of daily activities as indicated in the manual ( $N = 58$  activities across 6 days), an indication of activity completion (*Yes/No*), by whom an activity was completed, and a description of any reasons for modifications or change. Facilitators also completed one open-ended item, indicating one notable modification used in the delivery of the program in order to assess possible facilitator sensitivity to adjust as needed (Harn et al., 2013).

**Independent-observer rated of student activities.** A measure of student activity completion assessed the extent to which students were able to complete activities as instructed. The activity checklist was rated by an independent observer (a research assistant) using a codebook to assess the completion of two major component of the program: (1) metaphorical imagery and (2) self-reflections. First, the creative writing related to metaphorical imagery of a photograph indicated whether students would provide a description of their photograph using metaphorical language and relate it to themselves (i.e., use descriptive words that are representations or symbols of personal reflections). In line with the facilitator instructions manual and objectives for student activity completion, response categories were rated as 1 =

*Metaphorical language and related to the self*, 2 = *Metaphorical language but no relation to the self*, and 3 = *No metaphorical language used*. Second, the self-reflection spoken word assessed the extent to which students related their pieces to themselves (i.e., use the word “I”), and whether they gave examples of personal, meaningful experiences with a description (i.e., use words such as “family,” “friends,” “goals,” “interests”), and expressed language of states of emotion (“sad,” “happy,” “angry,” “anxious,” etc.). In line with the instruction manual and objectives of the task, response categories were rated as, 1 = *Reflective and related to the self*, 2 = *Reflective but no indication of the self*, and 3 = *Not reflective and not of the self*. A research assistant rated student activities independently. Additionally, 20% of student activities were rated together with the author of the present study until complete rating agreement was met. An additional random selection of 20% of activities was discussed with the supervisor of the author of the present study. Any remaining discrepancy in ratings were discussed and re-assessed among the three researchers until ratings were reviewed and achieved agreement.

**Organizational/ infrastructure fidelity variables.** Facilitators and teachers were asked Likert-scale questions related to program and classroom organization and infrastructure. Facilitator questions included, ‘How prepared was the Lakeshore Arts staff for this program?’ ‘How clear was the Lakeshore Arts staff in communicating the objectives of this program?’ ‘and ‘How prepared were the teacher(s) and students in this classroom?’ For teachers, questions included, ‘How organized was the Lakeshore Arts staff for this program?’ ‘How organized were the *Shazaam! In Focus* Artists?’ and ‘How clear were the artists in presenting the instructions over the 6-day *Shazaam! In Focus* program?’. Questions were based on five-point response categories and adapted for each item, ranging from ‘*not very*,’ ‘*neither/nor*’ to ‘*very*’.

**Self-reported clarity of delivery.** Student participants were asked a Likert scale question related to clarity of program rules, and included ‘Program rules and expectations were clear’ which consisted of a seven-point response category from 1 = *Not true at all* to 7 = *very true*.

**Participant attendance rates.** The number of sessions attended by students was assessed using an attendance list completed by the teacher on a daily basis.

**Participant satisfaction.** Participants were asked whether they were satisfied with the program. Student participants were asked how much they enjoyed participating in *Shazaam*, on a scale of 1 to 10. Artist facilitators were asked about their overall satisfaction working with the program, and responded on a 5-point Likert scale ranging from very dissatisfied to very satisfied. Teachers were asked about their overall experience taking part in the program, with 5-point Likert responses ranging from extremely negative to extremely positive.

### **Quantitative Survey Measures**

Items were developed based on two processes: First, stakeholder interviews were used to identify program objectives, components, and the goals for the program and for student outcomes. Second, a model of positive youth development was used, encompassing a strengths-based approach to the selection of items (Larson, 2000; Lerner et al., 2000; 2005). Measure selection of items from the original sources were adapted and modified to be relevant to the art program context. Quantitative items were in the form of Likert-scales, and were completed by students, artist facilitators, and teachers. Surveys included question items related to program-specific processes and adolescent socio-emotional outcomes.

**Student surveys.** The baseline (pre-program) survey at Time 1 included four demographic questions and thirteen Likert-scale items, taking up to 5-10 minutes, depending on the variability of individual-level response rates. This survey was completed at the start of the

first session of the program (session 1). A post-program survey included 23 Likert scale items, and was completed on the final day of the program (session 6). The follow-up survey consisted of 14 Likert scale items, and was completed four to six weeks post-program delivery, as an adapted version of the pre and post-program surveys. Student surveys can be found in Appendix I.

Questions were related to demographics, program delivery, and socio-emotional outcomes. Socio-emotional outcomes were categorized as close as possible to the construct definitions of the five Cs of positive youth development (Erdem et al., 2016) including, competence (art skill development, critical thinking), confidence (self-esteem, confidence in presenting, self-expression, and having a voice/autonomy), connections (connectedness with school, teachers, and peers, and support from artist facilitators), care and compassion (respecting peers), and character (engagement or enthusiasm in program). See Table 9 for the list of areas of interest, sub-areas, and internal consistency for the present study, where indicated (Cronbach's alpha). For a complete list of the data sources, areas of interest, question items, and psychometric properties please refer to Appendix G.

Table 9

*Areas of Interest for Adolescents Quantitative Measure Items, Sub-Domains, and Internal Consistency in the Present Study*

Areas of Interest	Sub-Domains	Chr. alpha
Student Demographics	Age (1 item)	--
	Gender (1 item)	
	Ethnic background (1 item)	
	Language(s) spoken at home (1 item)	
Program Delivery	Clarity of rules/instruction (1 item)	--
	Program participation (enjoyment, 1 item)	--
Competence	Art skill development (1 item)	--
	Critical thinking (1 item)	--
Confidence	Self-esteem (3 items)	0.89
	Self-expression (1 item)	--
	Confidence with presenting (1 item)	--
	Decision-making/autonomy (1 item each)	--
Connections	School Connectedness (7 items)	0.72
	School attachments	
	Teacher connections	
	Peer connections	
	Facilitator support (2 items)	0.71
Care and Compassion	Peer Respect and Inclusion	
	Positive items (2 items)	0.73
	Negative items (2 items)	0.78
Character	Program engagement (7 items)	0.75

*Note.* Chr = Chronbach's alpha. Alphas were generated for the sample in the present study.

*Demographic items.* Self-reported demographic items was derived from question items from Statistics Canada Census data and the National Longitudinal Survey for Children and

Youth (NLSCY, 2009; Statistics Canada, 2011) and included four items related to age, gender (*Male, Female, or Prefer not to say*), self-identified ethnicity/cultural background (*What is your ethnic background?*), and language(s) spoken at home (*List all*).

*Competence.* Competence consisted of two sub-domains: (1) art skill development, consisted of one question item (modified from Wright et al., 2006) of self-reported art ability (e.g., painting/drawing, music, writing) rated on a scale, ranging from 1 = *Very poor* to 10 = *Excellent*. (2) Critical thinking, consisted of one item (derived from Luke et al., 2007) including “In the program, I was able to make clear connections to personal experiences.” Response categories ranged from 1 = *Not at all true for you* to 7 = *Completely true for you*.

*Confidence.* Confidence consisted of four sub-domains. First, self-esteem consisted of three items derived from the Rosenberg Self-Esteem Scale (Rosenberg, 1985), and included “On the whole, I am satisfied with myself,” “I feel that I have a number of good qualities” and “I take a positive attitude toward myself.” Response categories were adapted from Rosenberg’s scale in line with other measures and ranged from 1 = *Not at all true for you* to 7 = *Completely true for you*. These three items were selected in the present study based on their focus on positive attributes, as well as showing good face validity. The three items were aggregated at baseline and post-program and showed high internal consistency in the present study ( $\alpha = .89$ ). The items also showed high correlations, and thus due to space purposes only the “positive attitudes” item was used as the follow-up time-point variable. Second, two separate items were related to voice/power in decision-making and autonomy (Bean & Forneris, 2016; Tiffany et al., 2012), and included “I feel a lot voice/power to make my own decisions” and “I was encouraged to take responsibility and make my own decisions about my work.” Response categories ranged from 1 = *Not at all true for you* to 7 = *Completely true for you*. Lastly, two items were created for the

present study in line with the program's objectives: (1) confidence presenting and (2) self-expression. Question items included self-rated "confidence to give presentations to your class" and "Comfort in expressing yourself," respectively. Response categories ranged from 1 = *Poor* to 5 = *Excellent*. These question items were asked at all three time-points, and were standardized across time-points to a 7-point scale, ranging from 1 = *Not at all true for you* to 7 = *Completely true for you*.

*Connections.* Connections consisted of two sub-domains and included positive connections within the school climate (positive attitudes toward school, bonds with teachers and peers), and connections with artist facilitators. First, school connectedness was comprised of seven items derived from the NLSCY (NLSCY, 2009). Items were chosen based on selected relevancy and face validity. Items were related to positive attitudes toward school, perceptions of doing well in school, and positive bonds with teachers and peers. First, question items about school, included "In general, how do you feel about school?" with response categories ranging from 1 = *I hate school* to 5 = *I like it a lot*, and "How well do you think you are doing in your school work?" with response categories ranging from 1 = *Very poorly* to 5 = *Very well*. Second, two questions related to relationships with teachers, included "In general, teachers treat me fairly," with response categories ranging from 1 = *Never* to 5 = *All of the time*, and "I feel like there are teachers that I can talk to about myself, or my problems," with response categories ranging from 1 = *Not at all true for you* to 7 = *Completely true for you*. Three question items related to peer connections, and included "In general, I get along easily with others my age," "I have many friends in this class," and "I feel that there are friends I can talk to when I have a problem." Response categories ranged from 1 = *Not at all true for you* to 7 = *Completely true for you*. All school connection items were standardized on a 7-point scale across time-points, and

showed acceptable internal consistency of  $\alpha = .72$ . Lastly, two question items were related to connections with facilitators in the program and included, “Adults in the program respected me” and “Adults in the program paid attention to what's going on in my life.” Response categories ranged from 1 = *Not at all true for you* to 7 = *Completely true for you*. These two items showed acceptable internal consistency ( $\alpha = .71$ ).

*Care and compassion.* Question items for care and compassion were related to the area of interest of accepting and respecting peers. Question items consisted of positive and negative experiences, with response categories ranged from 1 = *Not at all true for you* to 7 = *Completely true for you*. Positive peer respect included “I learned to accept differences in others,” “I felt supported by my peers.” Internal consistency of the two positive items showed  $\alpha = .73$ . Negative items included, “I did not feel supported by my peers” and “I felt like I didn’t belong.” Internal consistency of the two negative items was  $\alpha = .78$ .

*Character.* The sub-domain for character included program engagement. This area of interest consisted of seven question items related to overall enthusiasm to engage in the program and its activities, such as “I was very involved in program activities,” “I felt safe taking part in the program and its activities,” “I was comfortable with the digital media/photography activity,” and “I was comfortable writing poetry.” The seven items included response categories, ranging from 1 = *Not at all true for you* to 7 = *Completely true for you*, and showed high internal consistency of  $\alpha = .75$ .

**Artist facilitator and teacher surveys.** Survey questionnaires for artist facilitators and teachers were completed post-program with 12 Likert-scale items each, and one item related to the likelihood of recommending the program on a scale of 0 to 10. Each survey took approximately 5-10 minutes to complete. Items on the teacher and artist surveys focused on:

program processes (I.e., clarity of instructions, length of program, overall experience with program), perceived impact of program on the adolescents, and satisfaction of working with the community organization.

### **Qualitative Measures**

In addition to the key informant interviews (previously mentioned), qualitative measures consisted of open-ended responses from students, artist facilitators, and teachers.

**Student open-ended responses.** Five open-ended items were included on the post-program survey (Day 6), related to (1) the most and (2) least favourite parts of the program, (3) one valuable lesson learned from the program, (4) ways to improve the program for future programming, and (5) overall experience with the program. One follow-up open-ended response included the most valuable lesson learned from the program.

**Artist facilitator and teacher open-ended responses.** Artist facilitators and teachers responded to open-ended items at the post-program time-point (session 6), and were related to (1) an observed highlight or story of a student's experience, and (2) ways in which the program could improve. An additional two open-ended responses were completed by artist facilitators and included (3) a particular challenge found in the program session, (4) a notable modification that occurred in the delivery of program (as part of implementation fidelity).

### **Neighbourhood Census Data**

A description of each neighbourhood in which the schools were situated were identified. In the context of the education system, it is most often the case that the sector, or Ward, of the school also indicates the neighbourhood in which the adolescent student resides. This is because, in addition to municipal school board regulations (City of Toronto, 2017), families select a school based on proximity and accessibility, which often indicates the same area as the

household. Two neighbourhoods in which the schools are situated include: (1) Neighbourhood 18 (N18) in Ward 6 ( $N = 3$  classrooms in 2 schools). Data came from the 2011 Census Tract for the City of Toronto, in accordance with neighbourhood profile listings (City of Toronto, 2017; Statistics Canada, 2017) and includes descriptions for the two schools situated in neighbourhood N18. Data were derived from the 2011 Census because the subsequent 2016 Census had not been fully reported at the time of the present analysis (Statistics Canada, 2011; 2017). The variables included neighbourhood demographics related to (1) *average household income* (after tax), calculated as the sum of all income from members in the household<sup>6</sup>, (2) *self-identified ethnicity* is an indication of the top five ethnicities, showing possible varied ethnicities of the neighbourhood, and (3) *languages spoken at home*, included percentage of households that indicated having a language learned in childhood and still understood (mother tongue) that is neither English nor French, or speaking predominantly another language in the household.

### **Data Analysis Strategy**

All data were analyzed using SPSS version 25.0 (IBM, 2017). First, descriptive statistics (means, standard deviations, and frequencies) were explored for demographic characteristics of participants. Demographic data were also compared across the three schools using independent samples ANOVA, and chi-square tests, where indicated. It was also important to look at data on outcomes for both Time 2 (post-program) and Time 3 (follow-up). Each time-point provided valuable information on student self-reports, as well as to test program effects on student outcomes, and thus both time-points were included as dependent variables, when available.

---

<sup>6</sup> Low-income measure after tax (LIM-AT) is a fixed percentage (50%) of median adjusted after-tax income of households observed at the person level, where 'adjusted' indicates that a household's needs are taken into account. According to the LIM-AT, a low-income calculation, these rates are: N18 = 24% and CTR = 19%. The

Second, implementation fidelity variables were examined in four ways across the six indicators of fidelity. First, program delivery using the fidelity checklist and student activity completion using an independent-observer rater were analyzed by identifying the percentage (rate) of activities delivered or completed as intended, respectively. This analysis was important to yield the rate (adherence and dose) of each implementation indicator (Carroll et al., 2007). Second, the program fidelity checklist was also assessed using a thematic analysis to identify a list of themes related to reasons for change or modifications of activities. A thematic analysis would allow open-ended qualitative responses of raw data to be synthesized and categorized into given themes that emerge, and subsequently giving themes a numerical allocation and categorical quantification of items (Creswell & Clark, 2007). Third, self-reported Likert scale variables related to program infrastructure, student-rated clarity of rules, participant satisfaction with the program, and number of sessions attended were analyzed using the means (SDs) of each variable, and compared using a correlation matrix to examine the extent to which these variables related with one another (Abbott et al., 1998). Lastly, a regression analysis would show which program-specific component was the most significant predictor of student program enjoyment at Time 2 and at Time 3.

Third, a 2-level linear multilevel model analysis was used to examine how socio-emotional outcomes changed over time across students, using time-points as the level 1 variable and students as the level 2 variable (Heck, Thomas, & Tabata, 2013; Leckie et al., 2014). In the present study, it was important to consider the structure of the multilevel model, and the hesitance of using the classroom or school as levels in the model. In the present study, adding classrooms or schools as levels in the model was not warranted to answer the questions of interest for two reasons. First, the three classrooms were nested within two schools and within

one neighbourhood. In this case, there was not enough clustering of groups at the classroom or school levels to accurately estimate the variation that would be normally distributed around the mean (Leckie et al., 2014). Second, given that the program was run by external facilitators, it was possible to expect that most of the variability in outcomes would be accounted for at the student level (between students) rather than at a teacher- or classroom-level. That is, although the program took place within classrooms, it was not the teachers at the classroom-level who provided the program, and thus the classroom effects would likely have less of an impact on accounting for the variance in outcomes across students (McCormick et al., 2015). Thus, our questions of interest included exploring variability of outcomes across three time-points and between students, as opposed to variability between classrooms. Nevertheless, classroom was entered into the models as a covariate (further details below).

First, the null model (Model 1) would identify the intra-class correlation (ICC) and indicate how much of the variability for each outcome was accounted for at level 2 (i.e., whether there is significant variability between students). This model would also show the grand mean (intercept) of each outcome. Next, to assess whether outcome variables change over time, Model 2 treated time-point as a fixed effect (level-1 predictor) where the intercept of time-point was allowed to vary across students. The time-point analysis was coded using a time contrast model, in line with dummy codes (Heck et al., 2013). First, time 1 was used as the reference category and contrasted with times 2 and 3, which would show whether time 2 was significantly different from time 1, and whether time 3 was significantly different from time 1 (i.e., T1-T2 contrast, and T1-T3 contrast, respectively). Second, all models were rerun using time 2 as the reference category in order to compare whether time 3 contrasted with time 2 (T2-T3 contrast). It is important to note that changing the reference category would not indicate changes to the values

of fixed estimates or residuals in the models, nor the value of the Log Likelihood/BIC criteria, with the exception of the value of the fixed intercept. This estimate was expected to change given the difference to the grand mean intercept as a function of time-levels (Heck et al., 2013). The change in estimate was not expected to be significantly different from one another, and thus the fixed intercept value was used from the time 1 contrast model (Heck et al., 2013).

Next, to assess whether contextual factors accounted for variation in outcomes between students, Models 3 and 4 included covariates at level 2 to control for the effect of covariates on each outcome. First, gender was entered into the model as a fixed effect covariate, with female as 1 and male as 0 (Model 3). Given that all students were born in the same year, it was not expected that age would have a significant effect on the outcome, and thus was not included as a covariate. Next, the classroom variable was entered in the model as a fixed effect using dummy variables for classrooms 1 and 2 contrasted with classroom 3 (Model 4; Heck et al., 2013). Classroom 3 was used as a reference group (1), given that it was the only class in another school, while classrooms 1 and 2 were situated in the same school. Thus, it could be expected that if differences in student outcomes were to occur, it would be more notable across the two schools. All models were tested for diagnostics and demonstrated that model residuals met the normality of residuals assumption, indicating that using a linear mixed model was an appropriate fit. The log likelihood and BIC values were an indication of model criteria, and tested whether the final model was a better fit than the null model (Heck et al., 2013). The BIC was also shown to provide model criteria that were corrected for biases. See Figure 4 below for the model equation of the final model for each outcome: Model 4 (the final model for each outcome is similarly defined).

$$\begin{aligned}
OUTCOME_{ij} &= \beta_0 + \beta_1 TIMECONTR_{ij} + \beta_2 GIRL_{ij} + \beta_3 CLASSDUMMY1_{ij} \\
&+ \beta_4 CLASSDUMMY2_{ij} + u_j + e_{ij}, \\
u_j &\sim N(0, \sigma_u^2), \\
e_{ij} &\sim N(0, \sigma_{e_j}^2)
\end{aligned}$$

*Figure 4.* The written equation of the model indicates that  $OUTCOME_{ij}$  is the average outcome score for Time Contrast  $i$  ( $i = 1,2,3$ ) in student  $j$  ( $j = 1, \dots, 74$ ). The remaining covariates,  $GIRL_{ij}$ ,  $CLASSDUMMY1_{ij}$ ,  $CLASSDUMMY2_{ij}$ , are entered to adjust for the effects of Time on each outcome. The  $u_j$  and  $e_{ij}$  are the student and time specific random effects, respectively. The  $u_j$  are assumed normally distributed with zero mean and constant level-2 between-student variance  $\sigma_u^2$ . The  $e_{ij}$  are assumed normally distributed with zero mean and heteroskedastic level-1 within-student variance  $\sigma_{e_j}^2$ .

Fourth, teacher and artist surveys were examined by identifying means (and SDs) for Likert scale variables and compared across the three classrooms. Fourth, open-ended responses from students, artist facilitators, and teachers related to key outcomes of interest and program processes were analyzed using a thematic coding analysis (as indicated above; Braun & Clarke, 2006; Caracelli & Greene, 1993; Creswell & Clark, 2007). Responses were given themes and a corresponding numerical allocation, thus responses could be quantitatively analyzed using SPSS and compared with the rate of themes across perceptions of outcomes.

Lastly, neighbourhood profiles would be analyzed using an exploratory approach in order to identify relationships between community-level Census data, including SES, ethnicities, and languages, in relation to participant demographics and outcomes using a correlation table. It was important to explore objective-level data on the neighbourhoods in which these students are embedded to incorporate any additional demographic and contextual information.

## **Missing Data**

Participants with missing data for one or more time-points were compared with participants who had data for all three time-points on demographic variables and program attendance rates. Data were compared in two ways: (1) using chi-square analyses for differences across gender and ethnicity, and (2) using independent samples t-tests for differences across age and number of sessions attended.

Data showed that 26% of students ( $N = 19$ ) had missing data for at least one time-point. However, all students attended at least three of six sessions ( $M = 5.7$  sessions attended). Analyses showed that thirteen students had missing data for *only* one of three time-points, and six students had a combination of two time-points missing. Two students had missing data for *only* baseline, four students had missing data for *only* post-program, and seven students had missing data for *only* the follow-up time-point. Please note that no formal documentation took place to record the reason why students were absent. However, through informal discussion with the community partner, it was shared that some absentees of students were due to important attendance at a sporting tournament and students with sickness.

Across variables for adolescents with missing data versus non-missing data, results showed that no significant differences were found for all variables, with the exception of number of sessions attended. First, students with at least one time-point missing showed a similar proportion of males (50%,  $N = 8$ ;  $N = 3$  missing) compared with males with three time-points (56%,  $N = 31$ ),  $X^2(1) = .20$ ,  $p = .65$ . Those with missing data had a mean age of 12.8 years old ( $SD = .39$ ) versus 12.9 years old ( $SD = .34$ ) for those without missing data ( $p > .05$ ). Additionally, proportions of ethnicities for students with missing time-points did not significantly differ from proportions of those with all three time-points ( $X^2(5) = 1.71$ ,  $p = .89$ ). Ethnicities of students with missing data consisted of 44% mixed ethnicity ( $N = 4$ ), 22%

European (non-English,  $N = 2$ ), 22% Asian ( $N = 2$ ), and 11% African ( $N = 1$ ). Lastly, students with missing data attended significantly fewer sessions than those without missing data (5.2 vs. 5.8, respectively,  $p = .001$ ).

Given the comparable demographics across variables of students with and without complete data, it was shown that both groups of students were similar, with the exception of those with missing time-points who missed significantly more sessions. Accordingly, it was expected that outcomes would not differ, given that all students did not miss more than three sessions. Thus missing data were handled using a conservative approach for all three time-points, retaining data of participants, where possible. In terms of multilevel model analyses, it was possible to include participants even with missing time-points in the longitudinal analyses of the data, given the handling of missing data in multilevel modeling.

## **Results**

### **Participant Demographics**

A total of 74 students across three different classrooms in two schools situated within one neighbourhood participated in the Fall session of the program ( $N = 74$  with data for all three time-points). In addition, all artist facilitators ( $N = 6$ ) completed Likert scale questions, five of which completed open-ended responses. All teachers ( $N = 3$ ) completed open-ended responses, two of which completed Likert scale survey data.

In terms of student demographics, the average age of students was 12.9 years ( $SD = .35$ ). A total of 53% ( $N = 39$ ) were male students, and 43% ( $N = 32$ ) were female students ( $N = 3$  missing). Overall, the self-identified ethnicities of students included 31% mixed ethnicity (e.g., Portuguese-Italian, Cuban-Greek), 24% indicated European (non-English, e.g., Polish,

Ukrainian, Italian, Portuguese), 14% Asian (e.g., Tibetan, Filipino), 7% indicated an African country (e.g., Somalia, Ethiopia), 7% indicated Canadian, and 1% came from a Middle Eastern country. Overall languages spoken at home included 51% who spoke only English in their homes, 26% spoke English plus one other language, 1% spoke English plus two other languages, and 12% indicated speaking only another language in the home. Please refer to Table 10 for a breakdown of demographic variables.

In terms of census data, results of neighbourhood characteristics in N18 showed three important features. First, *income levels* showed that the average household income (after tax) was \$52,663 CAD, having a lower average compared with the CTR ( $M = \$70,945$  CAD). The largest proportion of after-tax household income fell in the range of \$20,000-\$49,999, comprising 37% of households, indicating a moderate level of neighbourhood wealth. The percentage of households situated in low-income (under \$20,000) was 24%, indicating particularly larger proportions of low-income households in N18, as compared with the City level ( $M = 16\%$ ). N18 also showed that the smallest proportion of households comprised of income over \$125,000 CAD (7% vs. CTR = 12%). Second, the variable of *self-identified ethnicity* showed 40% of the population in N18 were immigrants. The top 5 birth countries of all immigrants in N18 were Poland, Philippines, other places in Europe, the Americas, and United Kingdom. Recent immigrants (arriving between 2006-2011) primarily came from places in Africa, Philippines, Ukraine, places in Europe, and Saudi Arabia. Third, *language(s) spoken at home* showed that 32% of households indicated having a mother tongue that is neither English nor French. Additionally, the percentage of households that most often or regularly spoke another language at home other than English or French was 19%. The top five non-traditional Canadian languages in N18 were Polish, Filipino, Spanish, Ukrainian, and Russian.

Table 10

*Descriptive Statistics of Participants Overall and Between Schools*

<b>Demographic Variables</b>	Overall	School 1	School 2	School 3	<i>p</i>
Age ( <i>n</i> respond.)	( <i>N</i> = 66)	( <i>n</i> = 19)	( <i>n</i> = 23)	( <i>n</i> = 24)	
Mean Yrs (SD)	12.9 (.35)	12.8 (.38)	12.9 (.29)	12.8 (.38)	.70
Gender ( <i>n</i> respondents)	( <i>N</i> = 71)	( <i>n</i> = 22)	( <i>n</i> = 24)	( <i>n</i> = 25)	
% Male (SD)	52.7 (.50)	65.2 (.48)	62.5 (.49)	33.3 (.49)	.06
% Female (SD)	43.2 (.50)	30.4 (.48)	37.5 (.49)	59.3 (.49)	
% Ethnicity ( <i>n</i> respond.)	( <i>N</i> = 62)	( <i>n</i> = 19)	( <i>n</i> = 22)	( <i>n</i> = 21)	.008**
Mixed	37.1	17.4	45.8	29.6	
European (Non-Eng.)	29.0	34.8	33.3	7.4	
Asian	16.1	26.1	8.3	7.4	
Canadian	8.1	4.3	4.2	11.1	
African	8.1	--	--	18.5	
Middle Eastern	1.6	--	--	3.7	
% Language(s) ( <i>n</i> respond.)	( <i>N</i> = 67)	( <i>n</i> = 19)	( <i>n</i> = 22)	( <i>n</i> = 21)	.59
Only Eng.	56.7	39.1	54.2	59.3	
Eng. + 1 Other Lang.	28.4	30.4	29.2	18.5	
Eng. + 2 Other Langs.	1.5	--	4.2	--	
Only Other Lang.	13.4	17.4	8.3	11.1	

\*Note. Abbreviations: Respond. = respondents; Yrs = Years; Eng = English; Lang. = Language; *p* = significance value. \*sig at the .05 level. \*\*sig at the .01 level. Between schools, ANOVA test was used for age. T-Test was used for gender. Chi-square tests were used for ethnicity and languages.

## Program Implementation Results

**Fidelity checklist of activities.** Results from the fidelity checklist showed that, across schools 79.3% of activities were implemented as planned from the program manual. See Table 11 below for details. Between schools, the rate of activity completion showed 93% for school 1, 69% for school 2, and 76% for school 3. Among activities that were completed when planned, facilitators noted that some alterations occurred (11% for school 1, 18% for school 2, 18% for school 3) and included, starting late/running over time ( $n = 12$  activities), feeling rushed ( $n = 4$  activities), alter activity slightly ( $n = 2$  activities), some students left early/did not complete ( $n = 1$  activity), taking time to calm class/rowdy ( $n = 1$  activity), giving time for a short break ( $n = 1$  activity). Across schools, among activities that were not completed when planned, facilitators noted changing the activity to another day ( $n = 5$  activities), running out of time ( $n = 4$  activities), removing one activity given the class context, and filling time with another activity ( $n = 1$  activity).

Open-ended responses at the post-program time-point of modifications used by facilitators reflected some findings indicated in the fidelity checklist. Facilitators noted, modifying an activity due to student confusions, and adjusting time spent with some students who need more guidance than others. One spoken word artist indicated having to lead an activity usually led by the photography artist. Another facilitator indicated having to instruct an activity on a different day, due to time constraints on the scheduled day. Lastly, one artist indicated that giving instructions was too lengthy and thus had to be shortened due to student disengagement. This artist articulated the importance of finding ways to engage students during instruction,

I realize when you ask them... to come up and participate it kind of creates a different dynamic in the class and the teaching setting. It gives them more power as well, where it's like now they have to use their own skills to figure things out (Facilitator 2).

Table 11

*Implementation Outputs and Rates Across the Activity Checklist and Observed Student Activity**Completion*

Activity Implementation Variables	Overall	School 1	School 2	School 3
Activity Checklist				
Activity completion (%)	79%	93%	69%	76%
Completed with modifications (%)	16%	11%	18%	18%
Not completed as planned (%)	21%	7%	31%	24%
Student Activity Completion	( <i>N</i> = 74)	( <i>n</i> = 23)	( <i>n</i> = 24)	( <i>n</i> = 27)
(1) Metaphorical Image Reflection				
Completion rate (% students)	95%	96%	96%	93%
Metaphorical + self	62.2% (46)	61% (14)	75% (18)	52% (14)
Metaphorical + not self	8.1% (6)	13.0% (3)	4.0% (1)	7.4% (2)
Not Metaphorical	24.3% (18)	22% (5)	16.7% (4)	33.3% (9)
(2) Reflections of Me				
Completion rate, %	89.2%	74%	96%	96%
Reflective + dir. self	45.9% (34)	30.4% (7)	50% (12)	55.6% (15)
Reflective + indir. self	12.2% (9)	8.7% (2)	4.2% (1)	22.2% (6)
Non-reflective	31.1% (23)	34.8% (8)	41.7% (10)	18.5% (5)

*Note.* Abbreviations: Dir. = directly; indir. = indirectly.

Table 12

*Implementation Outputs Across Self-Reported Variables (Means and Standard Deviations)*

Self-Reported Implementation Variables	Overall M (SD)	School 1 M (SD)	School 2 M (SD)	School 3 M (SD)
Facilitator Response ( <i>N</i> = 6)				
Program staff - organized	5.0 (.00)	5.0 (.00)	5.0 (.00)	5.0 (.00)
Program staff - clear objectives	5.0 (.00)	5.0 (.00)	5.0 (.00)	5.0 (.00)
Program staff - supportive	5.0 (.00)	5.0 (.00)	5.0 (.00)	5.0 (.00)
Classroom - prepared	5.0 (.00)	5.0 (.00)	5.0 (.00)	5.0 (.00)
Teacher Response ( <i>N</i> = 2)				
Program staff - organized	4.5 (.50)	--	5.0 (.00)	4.0 (.00)
Artist facilitators - organized	4.5 (.50)	--	5.0 (.00)	4.0 (.00)
Artist facilitators – clear instructions	5.0 (.00)	--	5.0 (.00)	5.0 (.00)
Student Response ( <i>N</i> = 67)				
Clear rules/expectations	6.5 (.86)	6.2 (1.15)	6.7 (.57)	6.5 (.77)

*Note.* Sig. = significance value. \*sig. at the alpha .05 level. \*\*sig. at the alpha .01 level.

Table 13

*Implementation Outputs of Program Enjoyment and Attendance Rate*

Program Participation Implementation Variables ( <i>N</i> students)	Overall M (SD)	School 1 M (SD)	School 2 M (SD)	School 3 M (SD)	Sig.
Program enjoyment T2 ( <i>N</i> = 52)	8.5 (1.47)	7.7 (1.74)	8.8 (1.29)	9.0 (1.11)	.03*
Program enjoyment T3 ( <i>N</i> = 57)	8.3 (1.63)	8.1 (1.99)	8.4 (1.15)	8.4 (1.71)	.74
# of sessions attended ( <i>N</i> = 69)	5.7 (.63)	5.7 (.61)	5.6 (.78)	5.8 (.41)	.72

*Note.* Sig. = significance value. \*sig. at the alpha .05 level. \*\*sig. at the alpha .01 level.

**Independent-observer rated fidelity check.** The fidelity check of the creative writing exercises (*N* = 74 students) showed that an average of 92% of students completed the activities. Additionally, results showed that the largest proportion of students completed the metaphorical imagery and self-reflections activities as instructed. Specifically, results showed that an average of 70% of students (*N* = 52 students) engaged in metaphorical imagery, with 62% of those metaphor completers being related to the self, and 8% not related to the self. Results also showed that 58% of students (*N* = 43 students) engaged in reflective poetry, with 46% of poems directly related to the adolescent self, and 12% indirectly related to the adolescent. For details on results of student activity completion rates see Table 11 above. It was shown that most self-reflections were related to how student backgrounds and adverse life events shaped parts of their identity, as well as disclosing personal information related to family, humanity, interests, relationships, and aspirations.

**Fidelity check of self-reported items.** For fidelity self-report means (SDs) see Table 12

above.

*Facilitators.* Results of Likert scales showed that all facilitators across schools ( $N = 6$ ) felt that the community program staff were ‘extremely prepared’ to put on the program ( $M = 5.0$ ,  $SD = .00$ ), that they were ‘very clear’ in communicating program objectives ( $M = 5.0$ ,  $SD = .00$ ), and that the facilitators felt ‘extremely supported’ by the staff at the community organization over the 6 days ( $M = 5.0$ ,  $SD = .00$ ). Facilitators also indicated that the teachers and classrooms were ‘extremely prepared’ to receive the program ( $M = 5.0$ ,  $SD = .00$ ). One open-ended response from a facilitator indicated the importance of teacher preparedness,

Two students were a bit behind because they missed a day or two. But they were on top of things. I think their classroom teacher also has a lot to do with that. She kind of sets the environment and the tone so then when you come in [to start the program day], they’re already very respectful and they listen to what is asked of them and they deliver (Facilitator 2).

*Teachers.* Results showed that, overall teachers across schools ( $N = 2$ ) felt that the program staff was ‘somewhat’ to ‘very organized’ ( $M = 4.5$ ,  $SD = .50$ ), and that artist facilitators were ‘somewhat’ to ‘very organized’ ( $M = 4.5$ ,  $SD = .50$ ). Additionally all teachers felt that the facilitators were ‘very clear’ in presenting instructions to the students across the program activities ( $M = 5.0$ ,  $SD = .00$ ).

*Students.* Student participants showed to strongly agree that the facilitators communicated clear rules and expectations for the program ( $M = 6.5$ ,  $SD = .86$ ).

**Program Enjoyment.** Results showed that overall on a scale of 0 to 10, student ratings of program enjoyment post-program ( $N = 52$ ) showed that they highly enjoyed the program ( $M = 8.5$ ,  $SD = 1.46$ , range = 3-10). See Table 13 above. A significant difference of program enjoyment was found between schools post-program, and showed that schools 1 and 3 significantly differed ( $F(2, 51) = 4.03$ ,  $p = .02$ ). A similar level of program enjoyment was found

across students ( $N = 57$ ) at the follow-up time-point ( $M = 8.3$ ,  $SD = 1.63$ , range = 3-10), with no significant differences found between schools ( $F(2, 56) = .30$ ,  $p > .05$ ). Additionally, no significant differences were found in average enjoyment scores between times 2 and 3 ( $t(41) = .06$ ,  $p > .05$ ), nor between males and females post-program ( $F(1, 40) = .04$ ,  $p > .05$ ), or at follow-up ( $F(1, 40) = .43$ ,  $p > .05$ ).

**Predictors of Program Enjoyment.** Across program-specific components, including facilitator connection, clear rules and expectations, involved in program activities, and feeling safe in program, results for the regression analysis showed a significant model fit ( $F(5, 44) = 8.52$ ,  $p = .000$ ), with an  $R^2$  of .49, and that the most significant predictor of student enjoyment of the program at Time 2 was feeling safe in the program ( $p = .02$ ). Additionally, regression analyses for program enjoyment at Time 3 showed a significant model fit ( $F(5, 44) = 8.05$ ,  $p = .000$ ), with an  $R^2$  of .48, and that the most significant predictor of continuing to feel program enjoyment over time was feeling respect from facilitators ( $p = .02$ ). However, other variables related to peer support in the program (peer acceptance, felt support from peers, did not feel peer support) did not show a significant model fit to predict post-program enjoyment ( $F(3, 47) = 1.76$ ,  $p > .05$ ), but showed a significant model fit for the follow-up time-point ( $F(3, 47) = 4.02$ ,  $p = .01$ ), with an  $R^2$  of .20, where only peer acceptance was a marginally significant predictor of program enjoyment over time ( $p = .07$ ).

**Attendance rates in the program.** Results showed that average attendance rate for the six sessions of the program was 95% ( $M = 5.7$ ,  $SD = .63$ , range = 3-6), indicating a high level of attendance (See Table 13 above). Across schools ( $N = 69$  students), 70% of students attended 6 days, 18% of students attended 5 days, 4.1% of students attended 4 days, and 1.4% of students attended 3 days. There were no significant differences in attendance rates between schools.

**Correlation matrix.** Due to the limited variability in overall consistent ratings of facilitator and teacher program implementation variables (i.e., maximum response with 5 points), undertaking correlation tests would not provide any meaningful or useable findings, and thus were excluded as possible analyses. That is, the overall high ratings in responses indicate a highly restricted range, where correlations would be extremely high (approx.  $r = +1$ ) as a result.

### **Student Outcomes using Multilevel Modeling**

Descriptive statistics with means, standard deviations, and correlations for study variables across time-points can be found in Table 14. The final model (Model 4) of the multilevel models with estimate and standard errors for each student outcome can be found in Table 15.

**Art skill development.** Using multilevel modeling, first the null model (Model 1) showed that the proportion of variation in art ability between students was approximately 52% (ICC:  $2.09/1.93+2.09 = .52$ ), indicating significant variability accounted for at the student level (level 2). Grand mean of art ability (estimate of the intercept) across all students was 6.48 ( $SE = .20, p = .000$ ). Next, model 2 showed that time was a positive and significant predictor of art ability. Specifically, contrasting each time effect, there was a significant change in art ability from time 1 to time 2 ( $B = .68, SE = .24, p = .007$ ) and time 1 to time 3 ( $B = .90, SE = .25, p = .000$ ), but not from time 2 to time 3 ( $B = .22, SE = .25, p > .05$ ). After entering the level 1 predictor of time, it was shown that 55% of the variance in art ability was accounted for between students (ICC:  $2.18/1.72+2.18 = .55$ ). After controlling for covariates in Models 3 (gender) and 4 (classroom), Time remained significant. Results of the final model showed that only gender had a marginally significant effect on art ability, indicating that females had a slightly higher average rating of art ability than males ( $B = .85, SE = .44, p = .06$ ). In a separate analysis, results of an interaction term for gender\*time indicated that females had a significantly higher change in art

ability when compared to males from Time 1 to Time 3 ( $B = 1.07, SE = .49, p = .03$ ), but not for Time 1 to Time 2 ( $B = .25, SE = .51, p > .05$ ) or Time 2 to Time 3 ( $B = .82, SE = .53, p > .05$ ).

**Self-expression.** Multilevel model results of the null model showed that the grand mean (intercept) of self-expression across students was 5.02 ( $SE = .13, p = .000$ ). The proportion of variation in self-expression between students was approximately 13% ( $ICC: .94/2.23 + .34 = .13$ ), indicating moderate variability accounted for at the student level. Next, when Time was entered into the Model, Model 2 showed that there was significant variation at level 2 ( $B = .49, SE = .21, p = .02$ ). Results showed a significant increase in self-expression scores for T2 contrasted with T1 ( $B = 1.24, SE = .23, p = .000$ ), and T3 contrasted with T1 ( $B = 1.05, SE = .24, p = .000$ ), but not for T2 and T3 ( $B = -.19, SE = .24, p = .43$ ). Models 3 and 4 showed that, after controlling for gender and classroom, there was still a significant effect of Time, but no significant effect of covariates on self-expression. The final model showed an increase in ICC and approximately 22% of the variability was accounted for at the student level ( $ICC: .52/1.81 + .52 = .22$ ).

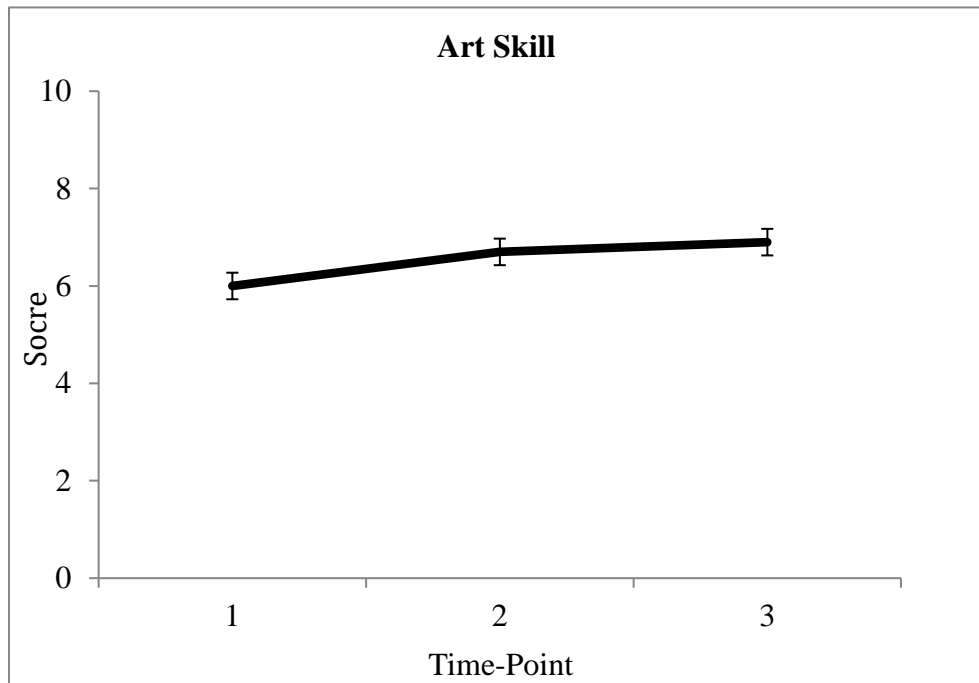
**Confidence presenting.** Multilevel model results of the null model showed the grand mean (intercept) of confidence presenting across students was 4.57 ( $SE = .15, p = .000$ ). The proportion of variation in confidence presenting between students was approximately 42% ( $ICC: 1.13/1.57 + 1.13 = .42$ ), indicating significant variability accounted for at the student level ( $B = 1.16, SE = .30, p = .000$ ). Next, when Time was entered into the Model, Model 2 showed a significant increase in confidence presenting from T1 and T2 ( $B = .58, SE = .21, p = .008$ ), but not from T1 and T3 ( $B = .17, SE = .22, p = .45$ ), or T2 and T3 ( $B = -.41, SE = .22, p = .06$ ). Models 3 and 4 showed that, after controlling for gender and classroom, the effect of time remained significant. There was also a significant effect of gender ( $B = -.68, SE = .32, p = .04$ ), but there was no significant effect of classroom on confidence presenting. In a separate analysis,

there was no significant time\*gender interaction. The final model showed that the ICC remained at approximately 42% of the variability accounted for at the student level (ICC:  $1.09/1.49+1.09 = .42$ ).

**Voice/power to make own decisions.** Multilevel model results in the null model showed the grand mean (intercept) of decision-making across students was 5.58 ( $SE = .11, p = .000$ ). The proportion of variation in decision-making between students was approximately 19% (ICC:  $.36/1.57+.36 = .19$ ), indicating moderate variability accounted for at the student level. Next, Model 2 showed that the entry of Time into the model did not have a significant effect between time-points on the outcome at level 2. When covariates were entered into the Model, Models 3 and 4 showed that, after controlling for gender and classroom there was still a significant effect of Time, but no significant effect of covariates on the outcome. The ICC in the final model showed that approximately 20% of the variability was accounted for at the student level (ICC:  $.39/1.54+.39 = .20$ ).

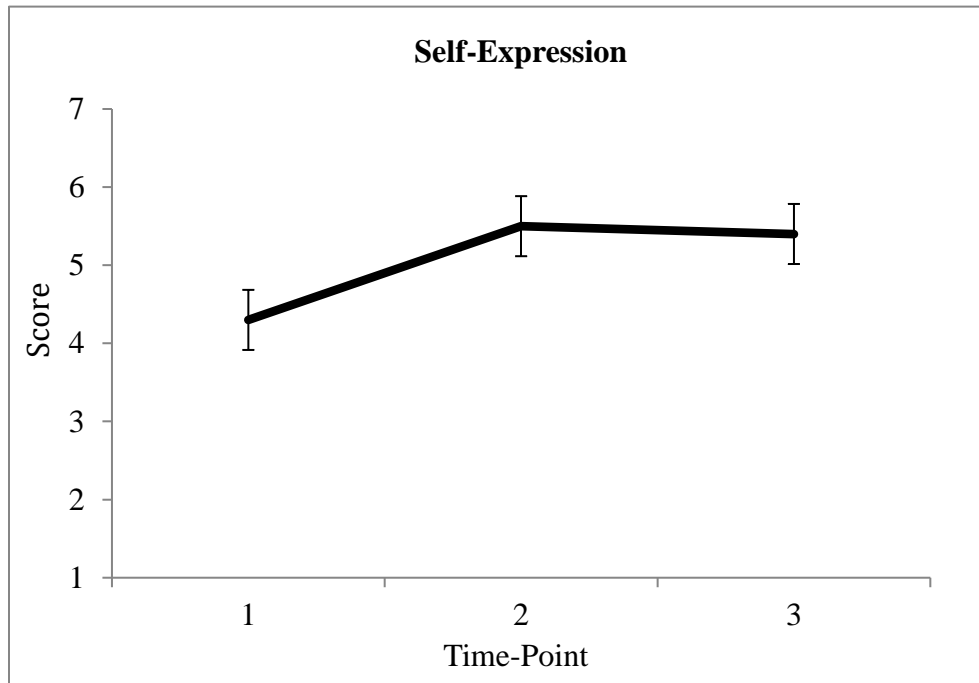
**Self-esteem.** Next, Model 1 of the multilevel model showed the grand mean (intercept) of self-esteem across students was 5.35 ( $SE = .13, p = .000$ ). The proportion of variation in self-esteem between students was approximately 63% (ICC:  $.94/.56+.94 = .63$ ), indicating significant variability accounted for at the student level. Next, model 2 showed that time had a significant effect on self-esteem at Time 2 contrasted with Time 1, showing a significant decrease ( $B = -.33, SE = .13, p = .01$ ). No significant effect was found for Time 3 contrasted with Time 1 ( $B = -.13, SE = .13, p > .05$ ) or contrasted with Time 2 ( $B = .19, SE = .13, p > .05$ ). Model 2 showed that 64% of the variance in self-esteem was accounted for between students (ICC:  $.95/.54+.95 = .64$ ). After controlling for gender and classroom, Models 3 and 4 showed that only classroom had a significant effect on self-esteem. Results showed that classroom 1 had significantly higher rating

of SE than classroom 3 ( $B = .77$ ,  $SE = .32$ ,  $p = .02$ ). In a separate analysis, a significant interaction indicated that, compared with classroom 3, a higher SE score was found for students in classroom 1 only from Time 2 to Time 3 ( $B = .73$ ,  $SE = .28$ ,  $p = .01$ ).

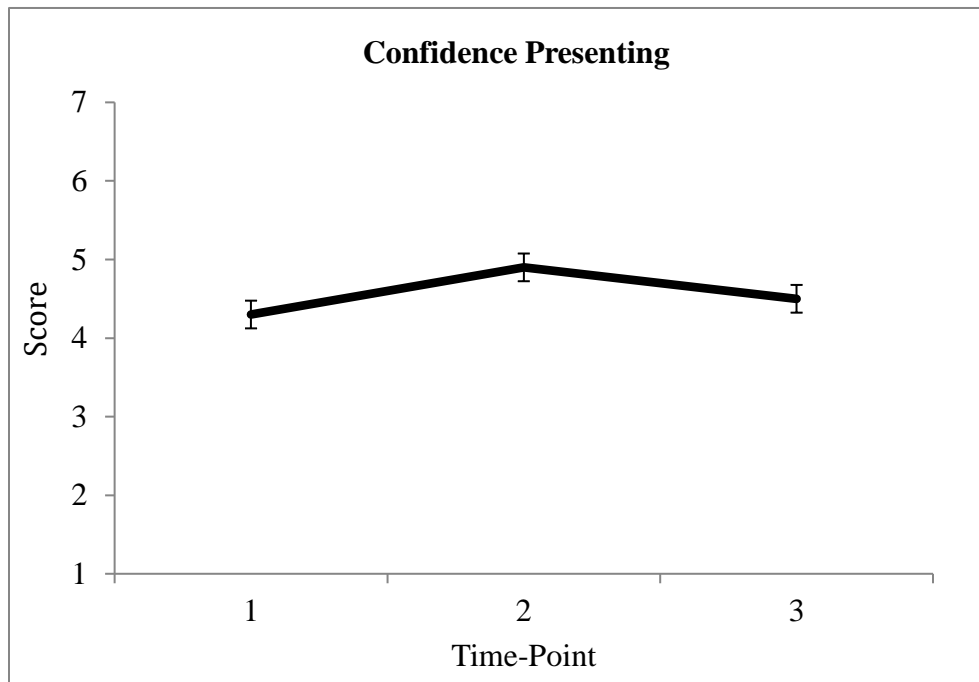


*Figure 5.* Time-point analysis for art skill across adolescents in the art program.

(a) Self-Expression



(b) Confidence Presenting



(c) Self-Esteem

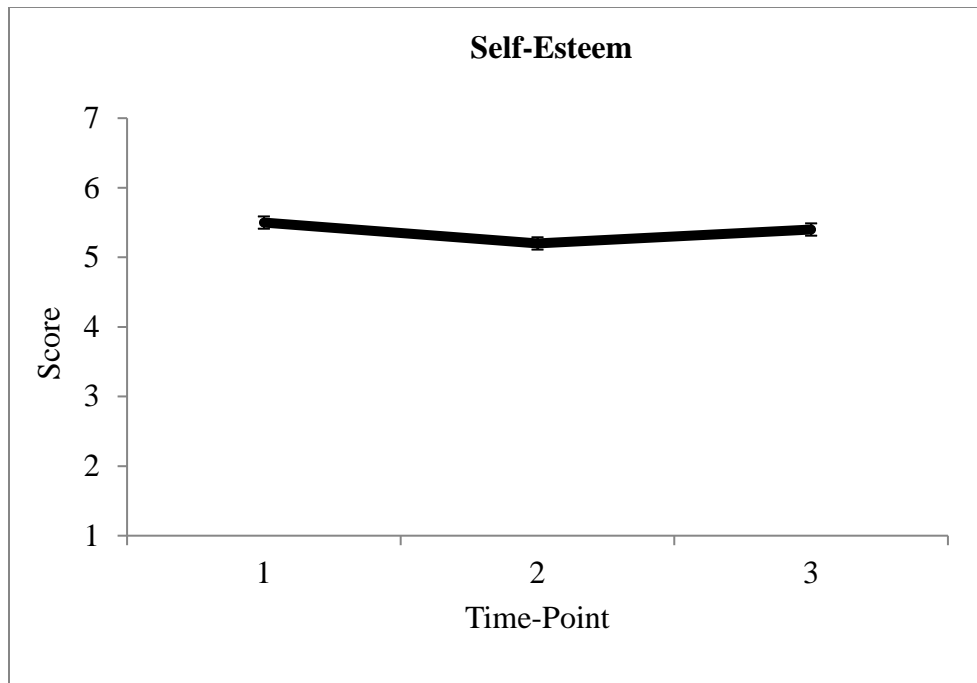


Figure 6. Time-point analysis for self-expression (a), confidence presenting (b), and self-esteem (c) across adolescents in the art program.

### School Experience and Connectedness

**Positive perceptions of school.** Multilevel model results showed that Model 1 showed the grand mean (intercept) of feeling positive about school across students was 4.80 ( $SE = .16$ ,  $p = .000$ ). The proportion of variation in decision-making between students was approximately 67% ( $ICC: 1.53/.75 + 1.53 = .67$ ), indicating significant variability accounted for at the student level. Next, Model 2 showed no significant effect of time-points on the outcome. Models 3 and 4 showed that, after entering covariates into the models there was a significant effect of gender ( $B = .83$ ,  $SE = .33$ ,  $p = .01$ ), indicating that females had a higher average positive feelings about school than males. No effect was found for classroom. In a separate analysis, results showed no significant interaction of time\*gender. The ICC in the final model showed that approximately 65% of the variability was accounted for at the student level ( $ICC: 1.41/.76 + 1.41 = .65$ ).

**Perceptions of schoolwork.** Multilevel model results of the null model showed the grand

mean (intercept) of perceptions of doing well in schoolwork across students was 5.34 ( $SE = .12$ ,  $p = .000$ ). The proportion of variation in doing well in schoolwork between students was approximately 54% ( $ICC: .75/.64 + .75 = .54$ ), indicating significant variability accounted for at the student level. Next, time contrasts in Model 2 showed that time did not have a significant effect on the outcome at level 2. When covariates were entered into the Model, Models 3 and 4 showed that, after controlling for gender and classroom, there was no significant effect of covariates on the outcome. The ICC in the final model showed that approximately 55% of the variability was accounted for at the student level ( $ICC: .79/.64 + .79 = .55$ ).

Table 14

*Means and Standard Deviations of Outcome Variables at Different Time-Points*

Outcome Variables	Time-Points		
	Time 1	Time 2	Time 3
	M (SD)	M (SD)	M (SD)
Art Skill	6.0 (.25)	6.7 (.26)	6.9 (.22)
Self-Expression	4.3 (.20)	5.5 (.15)	5.4 (.17)
Confidence Present.	4.3 (.19)	4.9 (.19)	4.5 (.22)
Decision-Making	5.4 (.17)	5.8 (.15)	5.6 (.19)
Self-Esteem	5.5 (.15)	5.2 (.13)	5.4 (.17)
School Perceptions	4.7 (.18)	4.8 (.19)	4.9 (.19)
Schoolwork	5.3 (.16)	5.4 (.14)	5.4 (.14)

Table 15

*Final Model (Model 4) of the Multilevel Models for each Study Variable Across Time-Points*

Variables	Art Dev.	Self-Exp.	Conf Pres.	Dec-Mak.	Self-Estm.	Sch Perc.	Sch Wrk.
<hr/>							
L2 (Student)							
<i>Fixed Est.</i>	B (SE)	BE (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)
Constant	<b>5.16 (.47)</b>	<b>4.12 (.31)</b>	<b>4.52 (.36)</b>	<b>5.46 (.28)</b>	<b>5.11 (.28)</b>	<b>4.29 (.35)</b>	<b>5.27 (.27)</b>
L1 (Time)							
T1-T2	<b>.75 (.25)</b>	<b>1.22 (.23)</b>	<b>.56 (.21)</b>	.36 (.22)	<b>-.33 (.13)</b>	.02 (.15)	.08 (.14)
T1-T3	<b>.89 (.25)</b>	<b>1.02 (.24)</b>	.17 (.22)	.28 (.22)	-.15 (.13)	.17 (.16)	.20 (.14)
T2-T3	.13 (.26)	-.19 (.24)	-.36 (.23)	-.06 (.23)	.18 (.13)	.15 (.16)	.08 (.15)
Covariates							
Gender	<b>.85 (.44)</b>	-.10 (.27)	<b>-.67 (.32)</b>	.00 (.25)	.09 (.26)	<b>.83 (.33)</b>	.21 (.25)
CL1 Dum	.42 (.54)	.25 (.34)	.09 (.39)	-.29 (.30)	<b>.77 (.32)</b>	.31 (.40)	-.20 (.31)
CL2 Dum	.88 (.51)	.30 (.32)	.22 (.38)	-.03 (.29)	.55 (.31)	-.18 (.38)	-.14 (.30)
<i>Random Est.</i>							
L2	<b>2.05 (.52)</b>	<b>.52 (.22)</b>	<b>1.09 (.29)</b>	<b>.39 (.20)</b>	<b>.88 (.19)</b>	<b>.76 (.10)</b>	<b>.79 (.18)</b>
L1	<b>1.70 (.24)</b>	<b>1.81 (.23)</b>	<b>1.49 (.19)</b>	<b>1.54 (.20)</b>	<b>.55 (.07)</b>	<b>1.40 (.30)</b>	<b>.64 (.08)</b>
-2 Log Lhd.	661.059	706.528	701.518	670.369	555.451	620.890	570.923
BIC	671.259	717.001	711.980	680.831	565.924	631.353	581.395

*Note.* Abbreviations: Art Dev. = Art Skill Development; Self-Exp. = Self-Expression; Conf Pres. = Confidence Presenting; Dec-Mak. = Decision-Making; Self-Estm. = Self-Esteem; Sch Perc. = School Perceptions; Sch Wrk = Schoolwork; L = Level; T = Time; Est. = Estimates; Lhd = Likelihood. Bolded terms are significant at the .05 alpha-level. Fixed intercept value is used from the T1 contrast model. Constant = intercept.

**Relationship with teachers.** Overall, positive correlations were found for student feelings that there were teachers they could talk to at baseline ( $M = 4.6$ ,  $SD = 1.93$ ) and follow-up ( $M = 5.0$ ,  $SD = 1.69$ ,  $r = .68$ ,  $p = .000$ ). A positive correlation was also found for student ratings that teachers treat them fairly ( $M = 4.5$ ,  $SD = .72$ ), with the follow-up time-point ( $r = .38$ ,  $p = .003$ ), but not with the baseline time-point ( $r = .20$ ,  $p = .10$ ).

**Relationships with peers.** Overall, students indicated at the baseline time-point that they

got along easily with their peers ( $M = 5.8$ ,  $SD = 1.08$ ), and that they had many friends in their class ( $M = 6.0$ ,  $SD = 1.14$ ), and these variables were significantly positively correlated ( $r = .49$ ,  $p = .000$ ). These variable were also related to the follow-up time-point for student ratings that they had many friends in class that they could talk to about their problems ( $M = 5.8$ ,  $SD = 1.29$ ), showing positive correlations with getting along easily with peers ( $r = .44$ ,  $p = .001$ ) and many friends in class ( $r = .45$ ,  $p = .000$ ).

**Program engagement.** Overall, students showed significantly greater comfort with photography and editing than poetry and spoken word ( $M = 6.4$  vs.  $5.6$ ,  $p = .000$ ,  $N = 67$ ), and indicated significantly greater interest to take part again in photography over poetry ( $M = 5.7$  vs.  $4.3$ ,  $p = .000$ ,  $N = 62$ ).

### **Qualitative Outcomes**

**Student responses.** *Most and least favourite program component.* Students were asked to respond to open-ended items post-program related to the most and least favourite components of the program. Thematic coding showed substantial overlap across students in their responses, and also reflected some findings from the quantitative results. First, themes across students showed that their most favourable activities included: taking photos/editing (43%), poetry slam/writing (34%), the collage (15%), everything in the program (5%), and watching/listening to peer presentations (2%). Second, themes across students showed that their least favourite components of the program included: writing poetry (24%), presenting in front of the class (18%), the collage (12%), writing about themselves (6%), and taking photos (6%). Additionally, results showed that 23% of students expressed that there was nothing that they did not like (i.e., ‘there is nothing I would change or improve’ or ‘I liked everything’).

*Most valuable lessons learned in the program.* Across students, thematic results showed

that students expressed the most valuable lessons learned from the program included: self-confidence/to be myself (27%), spoken word (14%), self-expression (12%), listening to peers/peer acceptance (12%), taking photos/editing (10%), to try my best (9%).

**Facilitator responses.** Facilitators ( $n = 5$  respondents) were asked to respond to an observed highlight of a student in the program. Thematic results of facilitator responses included: observing student changes in perceptions and attitudes of themselves and of the program over time (transformation and growth;  $n = 4$ ), observing students work hard and engage in activities ( $n = 2$ ), hearing about themes or challenges in student personal lives ( $n = 2$ ), seeing student discovery of new interests ( $n = 1$ ), seeing the students' final performances (poetry slam;  $n = 1$ ), feeling support from the teachers (sets the tone for classroom respect;  $n = 1$ ). One facilitator highlighted a powerful transformation with a student in the program:

Well, one of the girls that wanted nothing at all to do with the entire program, actually ended up saying "Thank You" to the Shazaam crew... She wanted nothing at all to do with it. I had to force her to take a photo. Like, force her. She told me it was boring. She didn't like art. She went on and on and on about how she didn't want to do it. And then, just challenging her, and having her know that, to open herself up to opportunities that come her way, that she might be exposed to something that she wouldn't generally do. Her poem was amazing. Her photo turned out fantastic. And it was from a student that wanted absolutely nothing to do with this.... It's just so cool to see [students] not engage [at the beginning], and then come out with an experience [in the end]. (Facilitator 6)

Another facilitator indicated a highlight of the program and seeing a student excel with some guidance:

The [poetry] slam was really amazing. Also, there was a student who did not want to read, and didn't want to write a poem. And even though they were super nervous, they actually were able to do their poem with a bit of handholding, so that was a huge success. (Facilitator 1)

The above reflections also highlight the ways in which the facilitators provided some form of scaffolding and sensitivity to students in need of assistance.

**Teacher responses.** Teachers ( $n = 3$  respondents) were also asked to respond to an

observed highlight of a student in the program. Thematic results among teacher responses included: using art to express themselves in a new way ( $n = 2$ ), using art to learn in a different way ( $n = 1$ ), to see their students take risks and be creative ( $n = 1$ ), seeing students – who would otherwise typically be defiant – engage in activities ( $n = 1$ ), hearing students talk about the positive impact of the teacher on their lives ( $n = 1$ ), and providing students with opportunities that they do not typically have in a classroom ( $n = 1$ ). One teacher mentioned the importance of the program for this age group, and indicated “*Such a fun experience, but more than that - my students had the chance to learn about themselves. Very important at this age*” (Teacher 3).

Another teacher expressed a highlight of one student transformation in the program:

In general I found that students who aren’t normally engaged in school, the program really helped them to have an opportunity to express themselves in ways that they never usually did. [I had one student] who didn’t do her work all year. Yet when it came to the program, she gave it one hundred and ten percent. We saw a side of her that we never saw before. And [the student] actually came back and performed the poem again... So it gives them [ways] to excel in opportunities that they normally don’t have in the classroom. (Teacher 1)

### **Program Challenges, Impact, and Improvement**

**Artist facilitators.** Likert scale results showed that facilitators felt that the program staff was ‘very friendly’ ( $M = 5.0$ ,  $SD = .00$ ), as well as ‘very friendly’ classroom teacher ( $M = 5.0$ ,  $SD = .00$ ). Facilitators also felt that some teachers were more supportive than others during the 6-day program, ranging from ‘somewhat supportive’ to ‘extremely supportive’ and ‘neither’ ( $M = 4.5$ ). Facilitators also rated the 6-day program length between ‘too short,’ ‘a bit short,’ and ‘just about right’ ( $M = 2.17$ ), and the 2-hour daily length was rated between ‘too short’ and ‘a bit long’ ( $M = 2.5$ ). All facilitators indicated they felt that the program had ‘a great deal of impact on the adolescents’ ( $M = 4.9$ ). On a scale of 0 to 10, the likelihood that facilitators would recommend the program again was 9.8.

**Teachers.** Likert scale results showed that teachers ( $n = 2$ ) indicated that the program staff and facilitators were ‘very friendly’ ( $M = 5.0$ ). Responses for length of program with 6 days ranged from ‘a bit short’ to ‘a bit long’ ( $M = 3.0$ ), and the 2-hour days ranged from ‘a bit short to ‘just about right’ ( $M = 2.5$ ) Results also showed that teachers were ‘extremely likely’ to use the student work as an assessment component for students’ final grades in their report cards ( $M = 5.0$ ,  $SD = .00$ ). All teachers indicated feeling ‘extremely positive’ about the program experience ( $M = 5.0$ ,  $SD = .00$ ). On a scale of 0 to 10, the likelihood that teachers would recommend the program again was 9.7.

**Open-ended responses. Artist facilitators.** Open-ended responses from facilitators supported some findings from the quantitative survey results. First, facilitators indicated some challenges in the delivery of the program, and included feeling limited with time on certain activities, including photography and editing ( $n = 2$ ) and poetry ( $n = 1$ ), feeling rushed in working one-one-one with students and balancing logistics, such as paper work ( $n = 1$ ), and school-related interruptions (e.g., announcements, school assembly). One artist facilitator indicated the challenge with feeling that there is too much structure (limited flexibility), articulating that, “*too much structure can be good to a certain degree, but sometimes limiting*” (Facilitator 5). Second, facilitators also mentioned some improvements that could be made, in line with previous challenges indicated. Improvements included, adding some time to the program ( $n = 3$ ), giving students more choice in activities ( $n = 1$ ), learning how to balance logistics, such as program management/instruction, and activity completion ( $n = 1$ ), adding an initial meeting (e.g., day before program) including all facilitators who will run the particular program session ( $n = 1$ ).

**Teachers.** Teachers ( $n = 3$ ) indicated some challenges and suggestions for improvement

noted within the program. Suggestions included giving students more time for spoken word poetry (as it is a new concept for most;  $n = 1$ ), being mindful of the time taken out of the standard curriculum, as well as falling behind compared to other classrooms ( $n = 1$ ). One teacher indicated that the program was ‘great’ and had no need for improvement.

*Students.* Results showed that 54% of students indicated that there was nothing they would improve (i.e., the program was good as is). Across student responses, results also showed that the top suggestions for program improvement included, for the program to give alternative options for activities (9%), making improvements within themselves (9%), being allotted more time for: photography (6%), art in general (5%), and writing poetry (3%), and lastly, modifying the program length (3%).

## **Discussion**

The present study examined both the delivery of a school-based arts program and the impact of the program on the psychosocial and emotional wellbeing of 74 grade 8 students.

### **Student Socio-Emotional Outcomes**

In terms of adolescent outcomes over the course of the program, multilevel model results showed that there were improvements over time in art skill and some socio-emotional outcomes. However, improvements were not found for other outcomes related to self-esteem and school-related items. First, a significant effect of time was found across students for artistic ability development from baseline to post-program, indicating that students rated themselves as improving in art skill over the course of the six-day program, and this improvement remained at the follow-up time-point. In the case of the present study, the art program provided an opportunity for adolescents to develop important skills related to photography, digital media, and creative writing. Empirical evidence has suggested that programs fostering arts learning shows that every participant can acquire a certain level of mastery over their art (Pitman, 1998). In the present study, it was also possible that learning about art techniques over the course of the program, enabled participants to feel empowered by the creation of their creative art outputs (Holloway & LeCompte, 2001).

In the case of the present study, when gender was added into the model, it was shown that females had a significantly higher average intercept of art skill, indicating that females had an overall higher rating of art skill than males, only for time 1 to time 3. Nevertheless, both genders showed significant improvements. In line with previous studies, the finding that overall all students showed improvements in self-rated art skill from baseline to follow-up indicates the importance of supportive and non-competitive environments across genders to foster

encouragement in arts-building, regardless of baseline art-skills (Upitis, 2011; Upitis & Smithrim, 2003; Wright et al., 2006). Thus, in the present study, both genders improved from baseline to follow-up, although it appears that females show more positive perceptions about their artistic abilities overall.

Second, significant effects of time were shown for some socio-emotional outcomes related to confidence in presenting and self-expression. Results showed significant improvements for both outcomes from baseline to post-program, and this improvement remained for self-expression at the follow-up time-point. The finding that adolescents showed improvements in these outcomes over time is in line with a large literature base of empirical evidence suggesting the importance of fostering such variables within art programs, and the nature of activities that are provided in arts-based learning (Holloway & LeCompte, 2001; Kisida, Bowen, & Greene, 2016; Lampert, 2013; Mynaříková, 2012). In the present study, components of the program were largely based upon self-exploration through daily creative writing as well as ongoing presentation tips to practice sharing their work, which may have added to the improved development of skills specifically related to self-expression and confidence-building. These findings were further validated by open-ended responses indicating the most valuable component of the program for students was learning about self-confidence (to be myself; 27%), and self-expression (12%).

However, significant improvements were not found over time for self-esteem or decision-making. Previous studies have shown that the mechanism through which self-esteem may develop, such as gaining a more positive image of oneself, is more ingrained than are those that relate to other aspects of wellbeing, such as improving skills in confidence to give presentations and to express oneself (Fleming & Offord, 1990). In fact, in the present study there was a

decrease in student-rated self-esteem from Time 1 to Time 2. Nevertheless, these self-ratings showed non-significant increases from Time 2 to Time 3, and thus students showed similar ratings at Time 3 compared with Time 1. It was shown that students in the present study showed overall relatively positive perceptions of themselves, when compared to the grand mean intercept of other outcome variables. It is thus possible that there was a ceiling effect of baseline self-esteem scores, contributing to the lower score in the subsequent surveys. Indeed, it is important to reflect on other factors that may have led to the short-term decrease, such as possible temporal or stressful contexts related to the adolescent environment. For example, it may have been the case that students were experiencing additional stressors during the post-program time-point, such as exam writing, personal relational or familial issues, or challenges with peers. With that said, such alternative possibilities were not documented.

Additionally, it should be noted that the present study did not specifically target understanding those with low self-esteem, or deficits, and thus it was not possible to contrast program success among those with and without certain self-esteem challenges. It is possible that a short-term art program cannot ameliorate self-esteem, nor improve perceptions about school, as no significant changes were found for school-related items of positive perceptions about school and schoolwork. Rather, the benefits of a short-term program can be largely focused on the other, more directly identifiable outcomes.

### **Program Implementation Fidelity**

In terms of the implementation fidelity of the program, findings from the present study showed that overall high rates of fidelity. First, as was demonstrated by the activity fidelity check, there was a high rate of adherence to the program with 79% of activities completed as intended. In line with previous findings, it has been shown that the rate at which good program

fidelity manifests can range between approximately 60% to 80% of activities completed as intended (Durlak & DuPre, 2008; Harn et al., 2013; Sidani, 2015), as well as good quality programs showing sensitive facilitators (Carroll et al., 2007). Similarly, it is also critical to provide a certain level of flexibility and adaptability of program components given the particular context in which participants of the program are embedded (Carroll et al., 2007; Forman et al., 2009; Harn et al., 2013). Additionally, assessing fidelity can contribute to identifying a more direct relationship between participant outcomes and program success (Carroll et al., 2007; Durlak & DuPre, 2008). In this way, previous findings suggest that programs showing high fidelity can more accurately attribute program components and processes with beneficial outcomes among participants, thus accurately identifying that the program was valuable (Dreeszen, 2003; Patton, 1997). Findings from the present study also showed that reasons for change across activities included moving activities to another day and running out of time. In line with previous evaluation studies, it is important to note the reasons for modifications or changes among activities in a program (Durlak & DuPre, 2008; Harn et al., 2013). Such identifications can be important in order to accurately identify what works best and what modifications may be in need of change for the program to be of a good quality delivery and beneficial for participants in a given context.

The present study also showed the importance of assessing student-level activity completion as rated by an independent observer, which showed high levels of completing activities as instructed. Students also indicated receiving clear instructions from facilitators. It has been shown that when activity instructions are adequately provided by facilitators or teachers, who are also sensitive to adjust activities to student-level learning, students can show significantly greater participation in and adherence to completing activities (Abbott et al., 1998).

Additionally important in the present study is the finding that there were highly positive ratings of the program infrastructure and organization, as well as showing high levels of preparedness from staff and from the classrooms. These components provided an essential basis for the program to take place. This finding supports the added importance of structural and staff support, which can contribute to program success (Teague et al., 1998).

Additionally, findings from the present study revealed that the program had a high attendance rate, with an average of 5.7 sessions attended, ranging from 3 to 6 sessions. That is, many students attended all six session, and all students attended at least three sessions. In line with previous findings, studies on youth programming show that there are significant benefits to in-school programs for improving the attendance rate among students (Eccles et al., 2002). This benefit is often shown because the typically documented obstacles in extra-curricular programs contributing to participant dropout or inaccessibility are often removed when programs are placed in a classroom setting (Eccles et al., 2002; Wright et al., 2006). In other words, significant barriers may be diminished or removed for participants to access a program when the program is run in-class time, such as reducing program costs, transportation to/from the program, or challenges associated with recruitment over time. Removing barriers to access a program is also a particularly important component of program success amongst lower-income, high-needs areas (Wright et al., 2006).

Another important implementation factor was shown for program participation and found that students significantly enjoyed the program (8.4 out of 10). This finding indicates that students had overall positive perceptions of the program. Studies suggest that high ratings of program participation and enjoyment can be a significant indicator contributing to high rates of fidelity in the program (Carroll et al., 2007; Dreeszen, 2003). Thus, in the present study it was

possible to consider that high levels implementation fidelity in the program may have been a factor contributing to the high rate of participant enjoyment.

### **Predictors of Program Enjoyment**

In the present study, the program component that was the most significant predictor of program enjoyment for students post-program was feeling safe in the program, as well as feeling respect from facilitators at the follow-up time-point. It has been shown that these two components are important factors that relate with program success (Baker & Harvey, 2014; Holloway & LeCompte, 2001; Wright et al., 2006). First, in line with previous studies, the finding that students felt safe in the program supports the importance of providing a safe and supportive environment in order to largely contribute to participant success and has also been shown to foster socio-emotional health and wellbeing and academic success (Mynaříková, 2012; Remer, 1996; Wright et al., 2006). That is, a supportive environment among art programs in particular is an important way for adolescents to learn. It is possible that in using a novel tool for self-exploration, the presence of supportive others including from facilitators and peers provides adolescents with a form of ease to engage in and share their art creations (Baker & Harvey, 2014; Holloway & LeCompte, 2001). This finding is particularly important for adolescents in high-needs neighbourhoods (Hampshire & Matthijsse, 2010; Wallace-DiGarbo & Hill, 2006; Wright, John, Alaggia, & Sheel, 2006).

Studies have also shown that the role of facilitators is another important component of participant success within programs and is closely linked with the notion of safe and supportive program environments (Forman et al., 2009; Holloway & LeCompte, 2001). Here, the finding of facilitator respect and support shows to act as a positive factor for participant success in the program. This idea is often because facilitators act as guides and mentors that can closely relate

with program participants (Hampshire & Matthijsse, 2010; Holloway & LeCompte, 2001). That is, facilitators may be most effective when they can show sensitivity toward students, balance hands-on support, provide effective instruction, as well as giving participants the space to learn independently and think critically (Holloway & LeCompte, 2001; Pitman, 1998; Vygotsky, 1971).

Similarly, in the present study another important factor was related to the demographics of facilitators who delivered the program. These facilitators were prominent artists in the neighbourhood in which the program took place, and also reflected sensitivity to the diverse demographics of the students. This finding supports previous studies that indicate the added advantage of incorporating facilitators who represent diversity and vulnerability, adding to an environment that is safe and supportive (Coholic et al., 2012; Forman et al., 2009). In the present study, it was possible to consider that having facilitators who show diversity and vulnerability might be one of the ways in which the program benefits the students.

In the present study also found that students showed positive perceptions of their teachers. A previous study has shown the supportive role of the teacher in assisting artist facilitators with an art program in the classroom (Mynaříková, 2012). In this way, the teacher supports students by fostering a positive and collaborative class climate, as well as supporting enhanced communication and respect among teacher-student relationships. In the present study, it was also found through open-ended responses that teacher and artist facilitators felt connected with the students and noticed positive changes within their students over time. Indeed, in the present study, students indicated feeling safe in the program, and this was a key factor for program enjoyment. It is also possible that the factors of feeling safe and feeling adult support in the present study were fundamental to the students in their growth and success across various

outcomes of interest. Here, supporters of positive youth development (e.g., Catalano et al., 2002; Eccles & Gootman, 2002; Erdem et al., 2016; Lerner et al., 2005) believe that such improvements can only be achieved in environments that highlight participant strengths and nurture positive student-adult relationships.

Lastly, another large component of student benefits in the present study was peer presence, support and acceptance of others. Findings from Likert scales showed that students were largely connected with peers in their classroom. Additionally, open-ended responses from students indicated that hearing peer stories and watching peers present their work were largely positive and valuable experiences. In line with previous findings, it has been shown that programs that foster peer respect and encouragement of others can help increase dialogue and communication within the classroom (Baker & Harvey, 2014), in addition to learning more about the stories of one another and appreciation of peer art creation (Wallace-DiGarbo & Hill, 2006). Additionally noteworthy is that in the present study it was encouraged that each student would create the same final products, including presenting a metaphorical self-image and a final poetry piece. In this way, it is possible that the collective nature of activity completion amongst all students contributed to feeling more safe, working toward similar goals, and fostering a mutual respect amongst students within classrooms.

### **Limitations and Future Directions**

A limitation of this study is that the evaluation of the program on adolescent outcomes did not include a control group, which makes it more difficult to attribute child benefits to the program (Patton, 1997). However, it is possible that our data captured such program-outcome relationships through the addition of both quantitative and qualitative methods, as well as observed and self-reported data (Patton, 1997; 2008; Schalock, 1995). Studies suggest that the

inclusion of multiple methods of data collection add a more comprehensive understanding of the program effects on student outcomes (Creswell & Clark, 2007). Indeed, it is also important to consider the sample size and the purposive selection of schools when determining the generalizability of the study. Despite the above, future studies would benefit from the implementation of both intervention and control groups, as well as a larger sample from multiple schools to better understand program impact on diverse student outcomes.

Additionally, it was apparent that students did not attend all sessions. Thus, our time series data were primarily focused on only those students who completed data at all three time-points. However, the advanced statistical methods of multilevel modeling allowed for the appropriate handling of missing data and it was possible to include data from all participants. In addition, the present study had an overall high attendance rate, and participants with missing time-points did not differ significantly on socio-demographic variables. It was possible to thus predict that our findings could be generalized across students overall. However, future studies could benefit from including multiple classrooms per schools and the inclusion of more variability in neighbourhoods to better understand the effect of socio-demographic variability on outcomes.

Lastly, it is noteworthy to highlight that the present study did not use one standardized measure to examine the outcomes of interest. The interest was based on the creation and design of a new tool that could be beneficial for an evaluation of the specific art program, and possible expansion to other similar art programs. In addition, given that the present study relied primarily on adolescent self-reports to understand their socio-emotional wellbeing, it is difficult to fully understand the benefits of the arts program on such variables from multiple perspectives, using subjective and objective data. However, the present study also used an analysis of teacher and

artist facilitator data, as well as activity engagement analysis from the adolescent's activity outputs, giving multiple understandings of the adolescent outcomes. Indeed, future studies could benefit from better understanding the effects of community art programs in general by including an analysis of adolescents with and without behavioural or emotional difficulties at the clinical level, or the inclusion of different art forms thus allowing more options for art creations, such as incorporating music or drawing. It may also be that the program effects may not be seen within the time period of the study; some effects may be delayed and so need a longer follow-up period to fully understand the impact of program effects on students outcomes. Future studies could benefit from increasing the duration and intensity of the art programming to assess whether program length contributes to positive youth outcomes.

## **Chapter 4: General Conclusion and Implications**

*“Children are the living messages we send to the future for a time we will not see.”*

- Neil Postman, 1982

In study 1, identifying the variability that exists in the ways in which indices try to capture youth wellbeing is critical to the study of youth wellbeing and development. It is important for disciplines to use coordinated approaches to identify outcomes on youth across points in time in order to assess consistent progress within and across nations. While indices exist, current conceptual and methodological approaches are often problematic. For example, data can greatly vary within and across indices or, at times, be missing the relevant data collection strategy to fully capture youth needs. Thus, improvement efforts can be made for an integrated and consistent approach to understand wellbeing through increased discourse between community organizations and youth themselves. Indeed, indices can be created that provide relevant and applicable findings to youth across regions.

Moreover, given existing variability in conceptual understandings of youth wellbeing, future direction lies in capturing a comprehensive representation of youth wellbeing. That is, it is important to integrate measures that are youth-centred, in addition to strengths-based. At best, measures at the local level can help assess relevant ways in which to understand the wellbeing of youth in their communities.

The findings provided within this analysis and critique can benefit the goals of service providers and community organizations by identifying core areas that are captured in indices that relate to youth wellbeing and development. This knowledge can be used as a way to develop a framework for assessing the impact of local community or governmental activities, and elucidate the extent to which policies, institutions, and services are meeting the needs of youth today. Thus, the present study can add support to better monitor and understand youth with respect to their

diverse contexts and the gaps that currently exist in the understanding of youth wellbeing. Lastly, this knowledge can help identify societal priorities, as well as where reform, action, or where sometimes limited resources may be allocated in order to more fully support youth and their needs at regional levels.

In study 2, findings can fill the gaps in evaluation studies focused on educational arts programs in three ways. First, it can add to evaluation designs and a model for novel tool delivery, where previous research for arts programs has been lacking. Next, it is possible to capture relevant measures and use available data of program implementation processes, in order to more fully understand program impact on participant outcomes. Lastly, evaluations can more accurately illustrate program success by assessing items of adherence to program delivery, as well as positive adolescent outcomes, and how to create safe and supportive environments for adolescents within their communities.

More specifically, the present study could help enhance awareness of relevant outcomes related to program impact within art programs. In addition, findings could be used to access data and applicable measures that can inform a model for program growth and improvement. Through these findings, implications related to best practices in program delivery can elucidate program effectiveness and success, thus informing social policy reform and government levels to provide monetary and material resources in areas of high-needs. Here, through capturing adolescents and their wellbeing within and across communities, findings indicate the beneficial effects of a supportive infrastructure that can have an impact on positive youth development, which can be strengthened through the power of arts.

Taken together, findings from the present dissertation can inform communities and governments to assess the impact of data collection strategies for youth wellbeing outcomes.

Findings can also support social policy advocacy and capacity building by adding illustration to the complex needs of Canadian youth, and bringing awareness to indicators that reflect the diversity of youth today. It is important to continue investigations related to the psychosocial challenges that encompass youth as their roles shift within and across contexts. The present studies can inform a foundation that focuses on supporting vulnerable populations among youth, contextual influences of wellbeing, and fostering knowledge mobilization across diverse sectors of society.

## Appendix A

### Search Strategy for Indices, Inclusion and Exclusion Criteria

- (1) A search for reports and indices was performed between 2015-2016.
- (2) The search was based on the particular topic of interest on youth wellbeing and included four **overarching inclusion criteria**:
  1. Data inclusion based on young people between the ages of 12-30.
  2. Data inclusion related to the North American context.
  3. Data related to a multi-dimensional conceptualization of wellbeing (i.e., more than one domain tested, and more than one indicator to make-up any given domain).
  4. Reports written in English.
- (3) Using the following **database**: Internet-based grey literature using the Advanced Search setting in Google.
- (4) **Date Range**: Indices were selected based on the most recent index from each organization, and as far back as ten years. Only one index per organization was included among dates from 2006 to 2016.
- (5) And used the following **keywords**:

Note. Keywords were selected so as to encompass enough detail while also limiting the large scope of internet-based entries:

  1. “Youth”
  2. AND “development”
  3. AND “wellbeing” (AND “well-being”)
  4. AND “index”
  5. AND “report”
  6. AND “indicators”
- (6) **Exclusion criteria** includes:
  1. If the report is not in English.
  2. If an index does not indicate age range, nor include data between the ages of 12-30.
  3. If there is no mention of data related to the North American context.
  4. If there is no mention of indicators that make up the domains/index.
  5. If there is no mention of measures that make up indicators.
  6. If there is no mention of how the index was created (methods or data sources).
  7. If there is only one domain used (uni-dimensional).
  8. If there is only one indicator used to construct a given domain.
  9. If duplicate stakeholder/organization (e.g., UNICEF 2013 and 2016).
  10. If index is based on *only* a sub-sample of the population (e.g., disability, First Nations, refugee)

## Appendix B

### Details of the Codebook, Areas of Inquiry, and Thematic Comparison Questions to Assess

#### Indices in the Meta-Narrative Analysis

	Area of Inquiry	Thematic Comparison Questions
<b>Index Context</b>	<b>Context</b>	<i>International? National? Provincial?</i>
	<b>Organization</b>	<i>What is the stakeholder affiliation?</i>
	<b>Year</b>	<i>What the year in which the index was created and disseminated?</i>
	<b>Age Range</b>	<i>What is the age range included in reports?</i>
<b>Themes Across Areas of Inquiry</b>	<b>Index Creation</b>	<i>How were index components selected and validated?</i>
	<b>Theoretical Framework</b>	<i>From what theoretical lens does the index conceptualize wellbeing?</i>
	<b>Population Descriptions</b>	<i>How are youth defined and described?</i>
	<b>Domains</b>	<i>What are the domains selected and how are they labeled? What is the rate of overlap across domains?</i>
	<b>Indicators</b>	<i>What is the variability of indicators within a domain (across indices)? What is the overlap of indicators across domains?</i>
	<b>Measures</b>	<i>What is the overlap or variability in measures used?</i>
	<b>Sources of Data</b>	<i>What is the overlap of data sources used?</i>
	<b>Youth-Focused</b>	<i>What is the rate of measures that are focused on youth-specific data?</i>
	<b>Source of Data &amp; Collection</b>	<i>What is the unit of analysis (respondent)? I.e., National-level statistics, family/parental data, self-reported by youth?</i>
	<b>Positive Measures</b>	<i>What is the rate of measurement that focuses on positive youth development? I.e., A strengths- vs. deficits based approach?</i>
<b>Critique</b>	<b>Strengths &amp; Limitations</b>	<i>Where are domains, indicators, and measures misaligned? Where could more focus be placed for youth-specific contexts? What are some areas for improvement?</i>

## Appendix C

### List of Eleven Methodological Papers in the Analysis Process

Index	Methodological Paper
<b>YDI</b>	OECD. (2008). <i>Handbook on Constructing Composite Indicators</i> . Organisation for Economic Co-operation and Development.
<b>GYWI</b>	Goldin, N., Hyslop, D., Hammond, D., & Patel, P. <i>The Global Youth Wellbeing Index: Methodological Report</i> .
<b>UNICEF</b>	<p>Martorano, B., L., de Neubourg, N. C., &amp; Bradshaw, J. (2013). <i>Child Wellbeing in Advanced Economies in the Late 2000s, Working Paper 2013-01</i>. UNICEF Office of Research, Florence.<a href="http://www.unicef-irc.org/publications/pdf/iwp_2013_1.pdf">http://www.unicef-irc.org/publications/pdf/iwp_2013_1.pdf</a></p> <p>Martorano, B., L., de Neubourg, N. C., &amp; Bradshaw, J. (2013). <i>Child Wellbeing in Economically Rich Countries: Changes in the first decade of the 21st century, Working Paper 2013-02</i>. UNICEF Office of Research, Florence.<a href="http://www.unicef-irc.org/publications/pdf/iwp_2013_2.pdf">http://www.unicef-irc.org/publications/pdf/iwp_2013_2.pdf</a></p> <p>Bradshaw, J., Martorano, B., L., &amp; de Neubourg, N. C. (2013). <i>Children's Subjective Well-being in Rich Countries, Working Paper 2013-03</i>. UNICEF Office of Research, Florence.<a href="http://www.unicef-irc.org/publications/pdf/iwp_2013_3.pdf">http://www.unicef-irc.org/publications/pdf/iwp_2013_3.pdf</a></p> <p>UNICEF (2010). 'The Children Left Behind: A league table of inequality in child well-being in the world's rich countries', Innocenti Report Card 9, UNICEF Office of Research, Florence, p. 3.</p>
<b>WHO</b>	Currie, C. et al. (Eds.) (2011). <i>Health Behaviour in School-aged Children (HBSC) study protocol: background, methodology and mandatory items for the 2009/2010 survey</i> . Edinburgh, Child and Adolescent Health Research Unit.
<b>CWI</b>	<p>Lamb, V. L. &amp; Land, K. C. (2013). Methodologies Used in the Construction of Composite Child Well-Being Indices. In A. Ben-Arieh (Ed.), <i>Handbook of Child Well-Being</i>. New York: Springer.</p> <p>Land, K. C., Lamb, V. L., &amp; Meadows, S. (2012). Conceptual and Methodological Foundations of the Child and Youth Well-being Index. <i>The Well-Being of America's Children: Developing and Improving the Child and Youth Well-Being Index</i> (pp. 13-28). New York: Springer.</p>
<b>MC-BC</b>	Saewyc, E., Stewart, D., & Green, R. (2014). <i>Methodology for the 2013 BC Adolescent Health Survey</i> . [Fact Sheet]. Vancouver, BC: McCreary Centre Society. Available at <a href="http://www.mcs.bc.ca">www.mcs.bc.ca</a> .
<b>SU-ON</b>	Ministry of Children and Youth Services. (2012). <i>Stepping Stones: A Resource on Youth Development</i> . Last modified June 13, 2012. <a href="http://www.ontario.ca/steppingstones">http://www.ontario.ca/steppingstones</a> .

## Appendix D

### Ryerson Research Ethics Board (REB) Email Correspondence

#### Ethics Application - No REB Review Required

rebchair@ryerson.ca

Thu 2017-03-30 4:12 PM

To: Sofia Puente-Duran <spuente@psych.ryerson.ca>; Kelly McShane <Kmcshane@psych.ryerson.ca>; kmcshane@psych.ryerson.ca  
<kmcshane@psych.ryerson.ca>;

Re: REB 2017-111 Evaluating the impact of an educational arts program on adolescent socio-emotional and academic growth among inner-city, high needs schools

Dear Sofia Puente-Duran,

The Research Ethics Board has determined that your protocol does not require its review.

Evaluating the impact of an educational arts program on adolescent socio-emotional and academic growth among inner-city, high needs schools submitted on Mar 30, 2017 10:45 AM

Thank you for your application for ethics review for the above noted study. Based on the information provided the project falls under Article 2.5 of the TCPS2 (the federal guidelines governing research ethics) as the purpose of the data collection is for program evaluation activities and performance reviews which will be used for the purpose of assessment, management and/or improvement of existing educational program activities, and so does not require Research Ethics Board Review. Should your proposed methodology or use of the data change, please reapply to the REB for ethics review and approval. Good luck with the project.

If you have any questions regarding your submission or the review process, please do not hesitate to get in touch with the Research Ethics Board (contact information below).

Record respecting or associated with a research ethics application submitted to Ryerson University.

NOTE: This email account (rebchair@ryerson.ca) is monitored by multiple individuals. If you wish to contact a specific member of the Research Ethics Board, please do so directly.

Yours sincerely,

Dr. Nancy Walton, PhD  
Co-Chair  
Ryerson Research Ethics Board  
416-212-4952  
nwalton@ryerson.ca

AND

Dr. Chris MacDonald, PhD

## Appendix E

### List of Key Informant Interview Questions

---

#### Establishing program theory for the “*Shazaam! In Focus*” program:

##### Interview Questions

The following set of questions will ask about a “program.” Here, we are referring the

“*Shazaam! In Focus*,” run by Lakeshore Arts.

---

##### Question Items

---

1. What is your role in the program?
2. How long have you been working in the program?

##### Program Theory:

3. What was the impetus for the creating the program? What problem is the program addressing?
  4. What are the program’s goals or objectives?
  5. What are the components of the program that are provided? How do they relate to program goals and objectives? (e.g., establish any connections between program activities and goals)
  6. In your own words, describe the program theory of change.
    - a. How does the program work to achieve those outcomes?
    - b. What aspects of the program do you believe will lead to changes and outcomes?
  7. Which program components are most critical to achieving successful outcomes (for students and for the organization)?
-

## Appendix F

### Program Logic Model

			Outcomes & Impact	
Resources/Inputs	Activities	Outputs	Short-Term (< 1 year)	Intermediate (1-3 years)
<p><b><u>Funding</u></b> Ontario Trillium Foundation (3 year grant)</p> <p><b><u>Program Staff</u></b> Executive Director (Susan Nagy) School Program Coordinator (Alessandra Cardarelli) Program Manager (Kim Dayman) Program Officer (Brian Conway)</p> <p><b><u>Program Facilitators</u></b> Artists (x2) Artists Assistants (x2)</p> <p><b><u>Program Support</u></b> Principals/teachers Students Graduate Assistants (x2)</p> <p><b><u>Physical Resources</u></b> • Lakeshore Arts Centre – room • Dedicated classroom • Office Space • Permission Forms • 1x Art Supply Box per school • 1x Camera/student (batteries) • 1x Binder with administrative info per school • 1x Online folder/student • Ipad? • 1x IMac/school • Projector in classroom</p>	<p><b>Dates:</b></p> <ul style="list-style-type: none"> <li>• 6 days over 2 weeks</li> <li>• Spring &amp; Fall sessions.</li> </ul> <p><b>One Week Before:</b></p> <ul style="list-style-type: none"> <li>• Handout PRE OTF survey</li> </ul> <p><b>Day 1:</b></p> <ul style="list-style-type: none"> <li>• Collect PRE OTF survey</li> <li>• <i>Handout/collect Pre-LA program survey</i></li> <li>• Introduction</li> <li>• Familiarize with material</li> </ul> <p><b>Last Day:</b></p> <ul style="list-style-type: none"> <li>• <i>Handout/collect post-program surveys (LA &amp; OTF)</i></li> <li>• Take final photo</li> <li>• Present poetry</li> </ul> <p><b>Teacher &amp; Facilitators</b></p> <ul style="list-style-type: none"> <li>• Complete surveys</li> <li>• Complete Activity Checklist</li> </ul> <p><b>Days Throughout:</b></p> <ul style="list-style-type: none"> <li>• Writing/Presenting exercises</li> <li>• Photos</li> <li>• Poetry/Spoken Word</li> <li>• Self-reflections</li> </ul> <p><b>End of Semester:</b></p> <ul style="list-style-type: none"> <li>• <i>Handout/collect follow-up surveys</i></li> </ul>	<ul style="list-style-type: none"> <li>• # Schools</li> <li>• # Students (overall)</li> <li>• # Permission slips</li> <li>• # Students enrolled/classroom</li> <li>• # Students attended/classroom</li> <li>• # Of students who completed program</li> <li>• # Of students who completed Evaluation Survey (Pre/Post)</li> <li>• # Satisfaction rate by students</li> <li>• # of surveys completed by teachers/facilitators</li> <li>• Areas of improvement identified by students, facilitators and teachers.</li> <li>• # of OTF surveys completed by Family Member.</li> </ul>	<p><b><u>Students</u></b></p> <ul style="list-style-type: none"> <li>• Complete 6-day program</li> </ul> <p><b>Students - Skills:</b></p> <ul style="list-style-type: none"> <li>• Increase Photography/Spoken Word skills</li> <li>• Improve presentation/ communication skills</li> <li>• Given voice to learn how to articulate/express selves</li> <li>• Enhance team building skills</li> <li>• Create supportive environment among peers</li> <li>• Tools to make own decisions</li> </ul> <p><b>Students - Knowledge:</b></p> <ul style="list-style-type: none"> <li>• Build sense of awareness of self-identity within a safe space – At critical age.</li> <li>• Gain confidence – sense of accomplishment</li> <li>• Enhance awareness of language as powerful tool to articulate selves</li> <li>• Build sense of collective identity - peers</li> <li>• Build connection to school/community</li> </ul> <p><b><u>School</u></b></p> <ul style="list-style-type: none"> <li>• Enhanced student engagement in classroom</li> <li>• Create safe space/hubs</li> </ul> <p><b><u>Facilitators</u></b></p> <ul style="list-style-type: none"> <li>• Increased sensitivity/support as mentor</li> <li>• Improved professional skills to facilitate</li> <li>• Mindful to create safe space</li> </ul> <p><b><u>Family/Teachers</u></b></p> <ul style="list-style-type: none"> <li>• Learn about children’s reflections and competencies.</li> </ul>	<p><b><u>Students</u></b></p> <ul style="list-style-type: none"> <li>• Begin high school - Ease transition</li> <li>• Enhanced self-identify/reflection</li> <li>• Ease of navigation, with stronger communication and social skills in school</li> <li>• Engagement in community</li> <li>• Develop a network of friends</li> </ul> <p><b><u>School</u></b></p> <ul style="list-style-type: none"> <li>• Create safe space for students’ to help navigate surroundings</li> <li>• Safe space/hubs help foster student socio-emotional, academic growth</li> </ul> <p><b><u>Facilitators</u></b></p> <ul style="list-style-type: none"> <li>• Increased intellectual and professional property</li> </ul> <p><b><u>Family/Teachers</u></b></p> <ul style="list-style-type: none"> <li>• Maintain improvement in communication and understanding of children.</li> </ul> <p><b><u>Program Competency</u></b></p> <ul style="list-style-type: none"> <li>• Expansion – Increased number of teacher/school take part;</li> <li>• Grow and share program across other communities.</li> <li>• Financial stability for continued offering of program activities.</li> <li>• Offer “Intensive After School” Program.</li> <li>• Improved automation in tracking program outputs.</li> </ul>

## Appendix G

### Table of Data Sources, Measures, Question Items, and Psychometric Properties for the Corresponding Areas of Interest

*\*Note.* Streiner (2003) suggests that internal consistency (Chronbach's alpha) is not a fixed property and thus values can depend on a particular scale used (p. 101). With that said, in social sciences, and as noted in the dissertation (p. 103), internal consistency for scales using Chronbach's alpha ( $\alpha$ ) is considered acceptable when values are above .70. A value above .84 is considered high, and below .70 is considered moderate-low (Streiner, 2003).

Data Source & Measure	Area of Interest, Question Items, Time-Point	Internal Consistency of Original Measure ( $\alpha$ )*
Cycle 8 Survey Instrument - Youth Questionnaire (National Longitudinal Survey of Children & Youth, NLSCY)	<p><b>Student Demographics</b>  <i>What is your age?</i>  <i>What is your gender? (Male, Female, Prefer not to say)</i>  <i>What is your ethnic background?</i>  <i>What language(s) do you speak at home?</i></p> <p><b>2-items from School:</b>  <i>"In general, how do you feel about school?"</i>  <i>"How well do you think you are doing in your schoolwork?"</i>  <i>(Perceived School Achievement, same)</i></p> <p><b>2-items from Teacher Perceptions:</b>  <i>"I feel I can talk to my teachers when I need to." (I feel I can talk to teachers about myself or my problems).</i>  <i>"In general, my teachers treat me fairly."</i>  <i>(Teacher Support, Modified)</i></p> <p><b>3-items from Friends:</b>  <i>"I have many friends."</i>  <i>"I get along easily with others my age."</i>  <i>"Do you have anyone in particular you can talk to about yourself or your problems" ("I have friends I can talk to when I have a problem.")</i>  <i>(Peer Support, Modified)</i></p>	
Adolescent Program Quality	<b>1-item from Appropriate Program</b>	Reliability

<p>Survey (APQS) (Bean &amp; Forneris, 2016)</p> <p>*The APQS is “based off of the National Research Council and Institute of Medicine’s eight program setting features that have been proposed to foster positive youth development.”</p>	<p><b>Structure:</b> “Program rules and expectations were clear” (T2: Program Implementation, same)</p> <p><b>1-item from Empowered Skill-Building:</b> “I was encouraged to take responsibility.” (T2: Decision-Making/Autonomy, modified)</p> <p><b>1-item from Expanding Horizons:</b> “I learned to accept differences in others.” (T2/T3: Peer support, respect and inclusion, modified)</p> <p><b>1-item from Negative Experiences:</b> “I felt like I didn’t belong.” (T2/T3: Peer support, respect and inclusion, same)</p>	<ul style="list-style-type: none"> <li>• Internal consistency for entire measure (<math>\alpha = .86</math>)</li> <li>• Internal consistency for subscales (<math>\alpha = .64-.84</math>)</li> </ul>
<p>Rosenberg Self-Esteem Scale (Rosenberg, 1965)</p>	<p><b>3-items used from T1-T2 (same):</b> “On the whole, I am satisfied with myself.” “I feel that I have a number of good qualities.” “I take a positive attitude toward myself.” (also T3)</p>	<p>Reliability</p> <ul style="list-style-type: none"> <li>• Internal consistency for entire measure (<math>\alpha = .77</math>)</li> </ul>
<p>Tiffany Eckenrode Program Participation Scale (TEPPS) (Tiffany, J.S., Exner-Cortens, D., and Eckenrode, J., 2012)</p>	<p><b>3-items from Personal Development:</b> “Adults at the program respect me.” “Staff at the program pay attention to what’s going on in my life.” (T2: Facilitator Support, modified).</p> <p>“The program’s activities are challenging and interesting.” [Program Participation: Modified “I was comfortable with the digital media/photography activity.” “I was comfortable writing poetry.”]</p> <p><b>2-items from Voice/Influence:</b> “I feel like I have a lot of voice/power to influence decisions about the program” (T2/T3: Decision-Making/Autonomy, modified) “I am very involved in program activities.” (T2, Program Engagement/Enjoyment, same)</p> <p><b>1-item from Safety/Support:</b> “I usually feel safe when I am involved in program activities.” (T2, Program Engagement/Enjoyment, slightly modified)</p>	<p>Reliability</p> <ul style="list-style-type: none"> <li>• Internal consistency for entire measure (<math>\alpha = .87</math>)</li> <li>• Internal consistency for subscales was <math>\alpha = .82</math> (PD), <math>\alpha = .73</math> (SS), <math>\alpha = .66</math> (VI), <math>\alpha = .68</math> (CE).</li> </ul>

<p>Wright et al., 2006</p> <p>Two measures were used from the NLSCY instrument measuring the participants' in-program behavior, including: (1) participation, (2) art skills development. All measures were selected from the National Longitudinal Survey on Children and Youth (NLSCY) – Cycle 1.</p>	<p><b>Art skill Development</b> (T1-T3), modified:  <i>“On a scale of 0 to 10, please CIRCLE how you would rate your artistic ability Such as painting, poetry, acting, playing music, etc.). 0 = No Ability; 10 = Excellent Ability.”</i></p> <p><b>Program Participation</b> (T2-T3, Modified):  <i>“On a scale of 0 to 10, overall, how much did you enjoy participating in Shazaam?” (Circle your answer).</i></p> <p><b>Program Participation, Modified:</b>  <i>“I would participate in Shazaam again.”</i>  <i>“I would be interested in taking another Photography workshop.”</i>  <i>“I would be interested in taking another Spoken Word workshop.”</i></p>	<p>Art skills development: 2-item scale measuring whether adolescents meets goals and shows improvement, <math>\alpha = 0.90</math></p> <p>Participation: 3-item scale measuring enjoyment of and engagement in activities, <math>\alpha = 0.90</math></p>
<p>Critical Thinking (Luke et al., 2007)</p>	<p><i>“Making clear connections to personal experiences.”</i></p>	<p>Reliability Internal consistency for entire measure (<math>\alpha = .70</math>)</p>
<p>4 Likert-Scale questions were created and modified to fit well with the specific objectives of the program.</p>	<p><b>Peer support and respect, modified:</b>  <i>“I supported my peers while they were presenting.”</i>  <i>“I did not feel supported by my peers.”</i></p> <p><b>Confidence Presenting, modified:</b>  <i>“Rate your overall ability to have: Confidence to give presentations to your class.”</i></p> <p><b>Self-Expression modified:</b>  <i>“Rate your overall ability to have: Comfort in expressing yourself.”</i></p>	
<p>Post-Program Open-Ended Responses selected and adapted from CSQ-8, for program-specific questions, in order to assess (1) program participation, (2) program impact, (3) program improvement.</p> <p>(e.g., Client Satisfaction Questionnaire Version 8 (CSQ-8).</p>	<p><b>Program Participation:</b> (Enjoyment/Engagement)  <i>“My most favourite part of Shazaam! Was.”</i>  <i>“My least favourite part of Shazaam! Was.”</i></p> <p><b>Program Impact:</b>  <i>“Please summarize your overall experience”</i>  <i>“One MOST valuable thing I learned from participating in Shazaam was.”(T2, T3)</i></p> <p><b>Program Improvement:</b>  <i>“What is one thing Shazaam could change or improve, and why?”</i></p>	

## Appendix H

### Student Consent Form



#### Consent to Participate in Survey Evaluations about *Shazaam!*

***We appreciate your upcoming participation in Shazaam!. Below is some information about completing the surveys.***

#### What is the **PURPOSE**?

- To find out about the experiences and opinions of students your age who take part in *Shazaam!* The information you provide us with will be used to help improve the future quality of this program so that other individuals, like yourself, can benefit from this experience.

#### Who is **LEADING** the project?

- This evaluation is led by a group at Ryerson University: Kelly McShane, PhD, CPsych, CE and Sofia Puente-Duran (PhD Student). If you have questions, they can be reached at: 416-979-5000, ext 2187.

#### What will you be asked to **DO**?

- You are being asked to complete three surveys (at the start and at the end of the program). The surveys involve rating scales and open-ended questions to understand your opinions of *Shazaam!*. The surveys will take approximately 5 minutes, and will be filled out in your class.

#### What are the potential **BENEFITS** to you?

- **Completing the surveys will not directly benefit you.** However, we hope that these surveys will help us better understand and improve the *Shazaam!* program.

#### What are the potential **RISKS** to you?

- ***You might feel uncomfortable with sharing your identity while providing feedback. However, all surveys are anonymous and your responses will not be tied to your name or any information that identifies you.***
- ***You might have concern about how this might affect your relationship with the program or your school.*** The decision to not answer a question or discontinue participation will in ***no way*** affect you or your relationship with *Shazaam!* As well, ***your participation is voluntary.*** If you feel uncomfortable during any point, you may simply wish not to answer some questions. You are free to stop participating at any time.

#### How will information be **PROTECTED** and **STORED**?

- To further protect your information, all surveys will be stored in protected cabinets/computers, and your individual responses will not be shared with anyone outside our team.

- 
- You understand that you have the right to withdraw from participating at any time and your information is anonymous and protected.
  - Your choice whether or not to participate will ***not*** affect you or your future relations with *Shazaam!* or your school.

Do you *Agree* to participate in the surveys? (You are welcome to change your answer at any time)

*I agree*                      *I do not agree*



**CONFIRMATION OF AGREEMENT:**

Your signature below indicates that you have read the information in this agreement and have had a chance to ask any questions you have about the study. If you would like a copy of this agreement, please let us know.

\_\_\_\_\_  
Name of Participant  
(please print)


\_\_\_\_\_  
Signature of Participant

\_\_\_\_\_  
Date

## Appendix I

### Student Surveys<sup>7</sup>

#### (A) Student Baseline Survey



Form ID: \_\_\_\_\_

### SHAZAAM! In Focus: Baseline Student Survey

Please tell us a little bit about yourself by answering the questions below (Remember, be OPEN and HONEST).

**Brief Questions About Yourself**

- What is your age? \_\_\_\_\_
- What is your gender? (CIRCLE one): ☐ Male ☐ Female ☐ Prefer not to say
- What is your ethnic background? \_\_\_\_\_
- What language(s) do you speak at home? \_\_\_\_\_
- On a scale of 0 to 10, please CIRCLE how you would rate your artistic ability (Such as painting, poetry, acting, playing music, etc.). 0 = No Ability; 10 = Excellent Ability.  
 |-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
 0 1 2 3 4 5 6 7 8 9 10

**For the following statements, please CIRCLE below what best represents your opinion:**

In general, how do you feel about school?	I hate school	I don't like it very much	I like it a bit	I like it quite a bit	I like it a lot
How well do you think you are doing in your schoolwork?	Very Poorly	Poorly	Average	Well	Very Well
In general, my teachers treat me fairly.	Never	Rarely	Some of the time	Most of the time	All of the time

For each statement, please CIRCLE a number that best represents your opinion	Not at all true for you	Somewhat true for you	Completely true for you				
I feel like there are teachers that I can talk to about myself, or any problems.	1	2	3	4	5	6	7
In general, I get along easily with others my age.	1	2	3	4	5	6	7
I have many friends in this class.	1	2	3	4	5	6	7
On the whole, I am satisfied with myself.	1	2	3	4	5	6	7
I feel that I have a number of good qualities.	1	2	3	4	5	6	7
I take a positive attitude toward myself.	1	2	3	4	5	6	7

Please circle your overall ability to:	Poor	Fair	Good	Very Good	Excellent
Confidence to give presentations to your class.	1	2	3	4	5
Comfort in expressing yourself.	1	2	3	4	5
Voice/power to make your own decisions.	1	2	3	4	5

**THANK YOU!**

<sup>7</sup>In the present Appendix section, the surveys have been presented with some formatting alterations and has been placed as an ‘image’ so as to fit the page margins.

## (B) Student Post-Program Survey

ID   /  /  /  /  /  

What is your gender? (CIRCLE one): Male Female Prefer not to say



### SHAZAAM! In Focus: Student Post Survey

We appreciate your talent, time and interest in *Shazaam! In Focus* and we hope that you had fun. Please help us evaluate the program by answering the questions below (Remember, be OPEN and HONEST. All opinions are helpful!).

On a scale of 0 to 10, overall, how much did you enjoy participating in *Shazaam! In Focus* (Please CIRCLE your answer). 0 = *Not at all*, 10 = *Amazing*!

|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
 0        1        2        3        4        5        6        7        8        9        10

Please CIRCLE the number below that best represents how you feel....	Not at all true for you		Somewhat true for you			Completely true for you	
<i>On the whole, I feel satisfied with myself.</i>	1	2	3	4	5	6	7
<i>I feel that I have a number of good qualities.</i>	1	2	3	4	5	6	7
<i>I usually take a positive attitude toward myself.</i>	1	2	3	4	5	6	7
<i>I feel happy when I'm at school.</i>	1	2	3	4	5	6	7
<i>I think I am doing well in my schoolwork.</i>	1	2	3	4	5	6	7
<i>In Shazaam! In Focus, rules and expectations were clear.</i>	1	2	3	4	5	6	7
<i>Adults in Shazaam! In Focus respected me.</i>	1	2	3	4	5	6	7
<i>Adults in Shazaam! In Focus paid attention to what's going on in my life.</i>	1	2	3	4	5	6	7
<i>I was very involved in program activities.</i>	1	2	3	4	5	6	7
<i>I was comfortable presenting in front of the class.</i>	1	2	3	4	5	6	7
<i>I supported my peers while they were presenting.</i>	1	2	3	4	5	6	7
<i>I respected differences in others.</i>	1	2	3	4	5	6	7
<i>I did not feel supported by my peers.</i>	1	2	3	4	5	6	7
<i>I felt like I didn't belong.</i>	1	2	3	4	5	6	7
<i>I was comfortable with the photography/ digital media activity.</i>	1	2	3	4	5	6	7
<i>I was comfortable writing poetry.</i>	1	2	3	4	5	6	7
<i>I felt like I was able to express myself.</i>	1	2	3	4	5	6	7

<i>I felt a lot of voice/power to make my own decisions in the program.</i>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<i>I was encouraged to take responsibility and make my own decisions about my work.</i>	1	2	3	4	5	6	7
<i>In the program, I was able to make clear connections to personal experiences.</i>	1	2	3	4	5	6	7
<i>I felt safe taking part in the program and its activities.</i>	1	2	3	4	5	6	7

**On a scale of 0 to 10, please CIRCLE how would you rate your artistic ability (e.g., Painting, Writing/Poetry, Acting, Playing Music, etc.) 0 = No Ability; 10 = Excellent Ability.**

|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
0      1      2      3      4      5      6      7      8      9      10

**Please share your opinions below (Remember: All opinions are helpful!).**

**1. My most favourite part of Shazaam! In Focus was:**

---



---

**2. My least favourite part of Shazaam! In Focus was:**

---



---

**3. One valuable thing I learned from participating in Shazaam! In Focus was:**

---



---

**4. Please summarize your overall experience participating in Shazaam! In Focus:**

---



---

**5. What is one thing Shazaam! In Focus could change or improve, and why?**

---



---

**THANK YOU!**

## (C) Student Follow-Up Survey

ID                              



What is your gender? (CIRCLE one): Male Female Prefer not to say

### SHAZAAM! In Focus: Student Follow-Up Survey

We hope that you had fun taking part in *Shazaam! In Focus* Please tell us a bit about your experiences below.

On a scale of 0 to 10, overall, how much did you enjoy participating in *Shazaam! In Focus* (Please CIRCLE your answer). 0 = Not at all; 10 = Amazing!

|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
0          1          2          3          4          5          6          7          8          9          10

Please CIRCLE the number below that best represents how you feel....	Not at all true for you		Somewhat true for you			Completely true for you	
<i>I feel happy when I'm at school.</i>	1	2	3	4	5	6	7
<i>I think I am doing well in my schoolwork.</i>	1	2	3	4	5	6	7
<i>I feel like there are teachers I can talk to when I have a problem(s).</i>	1	2	3	4	5	6	7
<i>I have friends I can talk to when I have a problem.</i>	1	2	3	4	5	6	7
Overall, since participating in <i>Shazaam! In Focus</i> ....	Not at all		Somewhat			Completely	
<i>I feel more comfortable presenting in front of a large group.</i>	1	2	3	4	5	6	7
<i>I feel more confident to make my own decisions.</i>	1	2	3	4	5	6	7
<i>I feel that I can take a positive attitude toward myself.</i>	1	2	3	4	5	6	7
<i>I feel that I have gained new skills to better express myself.</i>	1	2	3	4	5	6	7
<i>I learned to respect differences in others.</i>	1	2	3	4	5	6	7
<i>I would participate in <i>Shazaam! In Focus</i> again.</i>	1	2	3	4	5	6	7
<i>I would be interested in taking another Photography workshop.</i>	1	2	3	4	5	6	7
<i>I would be interested in taking another Spoken Word workshop.</i>	1	2	3	4	5	6	7

On a scale of 0 to 10, Please CIRCLE how you would rate your artistic ability today (e.g., Painting, Writing/Poetry, Acting, Playing music, etc.) 0 = No Ability; 10 = Excellent Ability.

|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|  
0          1          2          3          4          5          6          7          8          9          10

One MOST valuable thing I learned from participating in *Shazaam! In Focus* was:

---

---

THANK YOU!

## References

- Aapola, S. (2002). Exploring dimensions of age in young people's lives. *Time & Society*, 11(2-3), 295-314. doi:10.1177/0961463X02011002007
- Abbott, R. D., O'Donnell, J., Hawkins, J. D., Hill, K. G., Kosterman, R., & Catalano, R. F. (1998). Changing teaching practices to promote achievement and bonding to school. *American Journal of Orthopsychiatry*, 68, 542–552.
- Aneshensel, C. S., & Sucoff, C. A. (1996). The neighborhood context of adolescent mental health. *Journal of Health and Social Behavior*, 37(4), 293–310.
- Baker, W. J., & Harvey, G. (2014). The collaborative learning behaviours of middle primary school students in a classroom music creation activity. *Australian Journal of Music Education*, 1, 3-14.
- Bandura, A. (1982). Self-efficacy mechanisms in human agency. *American Psychologist*, 37(2), 122-147.
- Bandura, A. (2000). Self-efficacy. In A. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 17, pp. 212-213). Washington, DC and New York: American Psychological Association and Oxford University Press.
- Bastos, A., Fernandes, G. L., & Passos, J. (2004). Child income poverty and child deprivation: An essay on measurement. *International Journal of Social Economics*, 31(11/12), 1050–1060.
- Bastos, A., & Machado, C. (2009). Child poverty: A multidimensional measurement. *International Journal of Social Economics*, 36(3), 237-251.
- Batista-Foguet, J. M., Fortiana, J., Currie, C., & Villalbi, J. R. (2004). Socioeconomic indexes in surveys for comparisons between countries. An applied comparison using the Family

- Affluence Scale. *Social Indicators Research* 67(3), 315–332.
- Bean, C., & Forneris, T. (2016). Re-examining the youth program quality survey as a tool to assess quality within youth programming. *Cogent Psychology*, 3, 1-14.
- Belsky, J. (1984). The determinants of parenting: A process model. *Child Development*, 55(1), 83-96.
- Ben-Arieh, A. (2000). Beyond welfare: Measuring and monitoring the state of children—new trends and domains. *Social Indicators Research*, 52, 235–257.
- Ben-Arieh, A. (2005). Where are the children? Children’s role in measuring and monitoring their wellbeing. *Social Indicators Research*, 74, 573–596.
- Ben-Arieh, A. (2007). Measuring and monitoring the well-being of young children around the world. *Paper commissioned for the EFA Global Monitoring Report*.
- Ben-Arieh, A., & George, R. (2001). Beyond the numbers: How do we monitor the state of our children? *Children and Youth Services Review*, 23(8), 603–631.
- Benson, P. L. (1997). *All kids are our kids: What communities must do to raise caring and responsible children and adolescents*. San Francisco: Jossey Bass.
- Billari, F. (2004). Becoming an adult in Europe: A macro (/micro)-demographic perspective. *Demographic Research*, 3, 15-44.
- Boehnke, K., Schwartz, S., Stromberg, C., & Sagiv, L. (1998). The structure and dynamics of worry: Theory, measurement, and cross-national replications. *Journal of Personality*, 66(5), 745-782.
- Booyesen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, 59(2), 115–151.
- Boyce, W., Torsheim, T., Currie, C., & Zambon, A. (2006). The Family Affluence Scale as a

- measure of national wealth: Validation of an adolescent self-report measure. *Social Indicators Research*, 78(3), 473-487.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371–399.
- Bradshaw, J., Hoelscher, P., & Richardson, D. (2006). Comparing child well-being in OECD countries: Concepts and methods. *Innocenti Working Paper No. 2006-03*. Florence, UNICEF Innocenti Research Centre.
- Bradshaw, J., Noble, M., Bloor, K., Huby, M., McLennan, D., Rhodes, D., et al. (2009). A child wellbeing index at small area level in England. *Child Indicators Research*, 2(2), 201–219. doi:10.1007/s12187-008-9022-6.
- Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. Available at <http://dx.doi.org/10.1191/1478088706qp063oa>
- Bronfenbrenner, U. (1979). *The ecology of human development. Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U., & Morris, P. (1998). The ecology of developmental process. In W. Damon and R. Lerner (Eds.), *Handbook of child psychology: Theoretical models of human development* (Vol. 5, pp. 993-1028). Hoboken, NJ: John Wiley & Sons Inc.
- Brown, B. V. (Ed.) (2008). *Key indicators of child and youth well-being: Completing the picture*. New York, NY & London, UK: Psychology Press.
- Brown, B. V., & Corbett, T. (2003). Social indicators and public policy in the age of devolution. In R. Weissberg, L. Weiss, O. Reyes, & H. Walberg (Eds.), *Trends in the well-being of children and youth* (pp. ). Washington, DC: Child Welfare League of America Press.

- Canadian Index of Wellbeing. (2016). *How are Canadians Really Doing? The 2016 CIW National Report*. Waterloo, ON: Canadian Index of Wellbeing and University of Waterloo. Retrieved from, [https://uwaterloo.ca/canadian-index-wellbeing/sites/ca.canadian-index-wellbeing/files/uploads/files/c011676-nationalreport-ciw\\_final-s.pdf](https://uwaterloo.ca/canadian-index-wellbeing/sites/ca.canadian-index-wellbeing/files/uploads/files/c011676-nationalreport-ciw_final-s.pdf)
- Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15(2), 195-207.
- Carey, P., & Sutton, S. (2004). Community development through participatory arts: Lessons learned from a community arts and regeneration project in South Liverpool. *Community Development Journal*, 39(2), 123-134.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2002). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *Prevention & Treatment*, 5(15), 1-112.
- Chilton, G. (2013). Art therapy and flow: A review of the literature and applications. *Art Therapy*, 30(2), 64-70.
- Christopher, S., Watts, V., McCormick, A. K. H. G., & Young, S. (2008). Building and maintaining trust in a community-based participatory research partnership. *American Journal of Public Health*, 98(8), 1398-1406.
- Chung, B., Jones, L., Jones, A., Corbett, C. E., Booker, T., Wells, K. B., & Collins, B. (2009).

Using community arts events to enhance collective efficacy and community engagement to address depression in an African American community. *American Journal of Public Health*, 99(2), 237-244.

CIC (2016, January). *Provincial definitions of a minor*. Retrieved from <http://www.cic.gc.ca/english/resources/tools/refugees/canada/processing/minors-prov.asp>

City of Toronto (2017). *Neighbourhood 18*. Retrieved from <https://www.toronto.ca/ext/sdfa/Neighbourhood%20Profiles/pdf/2011/pdf1/cpa18.pdf>

Clark-Kazak, C. R. (2008). Towards a working definition and application of social age in international development studies. *Journal of Development Studies*, 45(8), 1307-1324.

Cohen, C. (2002). United Nations Convention on the Rights of the Child: Developing international norms to create a new world for children. In K. Alaimo, and B. Klug (Eds.), *Children as equals: Exploring the rights of the child* (pp. 49-72). Lanham, Maryland: University Press of America.

Cohen, H. (1980). *Equal rights for children*. Totowa, NJ: Littlefield, Adams and Co.

Coholic, D., Eys, M., & Loughheed, S. (2012). Investigating the effectiveness of an arts-based and mindfulness-based group program for the improvement of resilience in children in need. *Journal of Child and Family Studies*, 21(5), 833-844.

Collins, P. A., & Hayes, M. V. (2010). The role of urban municipal governments in reducing health inequities: A meta-narrative mapping analysis. *International Journal for Equity in Health*, 9(1), 13.

Commonwealth Youth Programme (2013). *Youth Development Index results report*. Retrieved from [http://www.youthdevelopmentindex.org/cms/cmsyouth/\\_images/197918019952385f3219c75.pdf](http://www.youthdevelopmentindex.org/cms/cmsyouth/_images/197918019952385f3219c75.pdf)

Commonwealth Youth Programme (2016). *Youth Development Index results report*. Retrieved from [http://www.youthdevelopmentindex.org/cms/cmsyouth/\\_images/197918019952385f3219c75.pdf](http://www.youthdevelopmentindex.org/cms/cmsyouth/_images/197918019952385f3219c75.pdf)

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Creswell, J. W., & Clark, V. L. P. (2007). *Designing and conducting mixed methods research*. Wiley Online Library.

Cummins, R. A. (1996). The domains of life satisfaction: An attempt to order chaos. *Social Indicators Research* 38, 303–332.

Cummins, R. A. (2000). Objective and subjective quality of life: An interactive model. *Social Indicators Research*, 52(1), 55–72.

Damon, W. (2004). What is positive youth development? *The Annals of the American Academy of Political and Social Science*, 591(1), 13.24.

Decancq, K., & Lugo, M.A. (2013). Weights in multidimensional indices of wellbeing: An overview. *Econometric Reviews*, 32(1), 7-34.

Douglass, D. (2011). *Self-esteem, recovery and the performing arts: A textbook and guide for*

- mental health practitioners, educators and students*. Charles C Thomas Publisher.
- Dreeszen, C. (2003). *Fundamentals of arts management* (4th ed.). Amherst, MA: The Arts Extension Service.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327.
- Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta- analysis of after- school programs that seek to promote personal and social skills in children and youth. *American Journal of Community Psychology*, 45(3-4), 294-309.
- Durlak, J. A., & Wells, A. M. (1998). Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. *American Journal of Community Psychology*. 26(5), 775-802.
- Eccles, J. S., Barber, B. L., Stone, M., & Hunt, J. (2003). Extracurricular activities and adolescent development. *Journal of Social Issues*, 59(4), 865-889.
- Eccles, J. & Gootman, J. A. (Eds.) (2002). *Community programs to promote youth development*. National Research Council and Institute of Medicine, Washington, DC: National Academy Press.
- Eccles, J. S., Midgley, C., Buchanan, C. M., Wigfield, A., Reuman, D., & MacIver, D. (1993). Development during adolescence: The impact of stage/environment fit. *American Psychologist*, 48, 90–101.
- Epp, K. M. (2008). Outcome-based evaluation of a social skills program using art therapy and group therapy for children on the autism spectrum. *Children & Schools*, 30(1), 27-36.
- Epstein, J. L. (2001). Building bridges of home, school, and community: The importance of

- design. *Journal of Education for Students Placed at Risk*, 6(1-2), 161-168.
- Epstein, J. L. (2018). *School, family, and community partnerships: Preparing educators and improving schools*. Routledge.
- Erdem, G., DuBois, D. L., Larose, S. De Wit, D., & Lipman, E. L. (2016). Mentoring relationships, positive development, youth emotional and behavioural problems: Investigation of a meditational model. *Journal of Community Psychology*, 44(4), 464-483.
- Erikson, E. (1968). *Identity: Youth and Crisis*. New York, NY: Norton.
- Evans, G. W. (2006). Child development and the physical environment. *Annual Review of Psychology*, 57(1), 423–451. doi:10.1146/annurev.psych.57.102904.190057
- Ferguson, R., & Snipes, J. (1997). Adapting Erikson to understand stages of engagement and identity development in the YouthBuild program. In R. Ferguson and P. Clay (Eds.), *YouthBuild in Developmental Perspective* (pp. 245-318). Cambridge, MA: MIT.
- Fernandes, L., Mendes, A., & Teixeira, A. A. C. (2012). A review essay on the measurement of child well-being. *Social Indicators Research*. doi:10.1007/s11205-011-9814-9.
- Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2009). Evidence-based interventions in schools: Developers' views of implementation barriers and facilitators. *School Mental Health*, 1(1), 26.
- Fraser, C., & Keating, M. (2014). The effect of a creative art program on self-esteem, hope, perceived social support, and self-efficacy in individuals with multiple sclerosis: A pilot study. *Journal of Neuroscience Nursing*, 46(6), 330-336.

- Freire, P. (1996). *Pedagogy of the Oppressed (revised)*. New York: Continuum.
- Goldin, N., Patel, P., & Perry, K. (2014). *The Global Youth Wellbeing Index*. Rowman & Littlefield. Landham: MD.
- Greenhalgh, T., Potts, H. W., Wong, G., Bark, P., & Swinglehurst, D. (2009). Tensions and paradoxes in electronic patient record research: A systematic literature review using the meta-narrative method. *Milbank Quarterly*, 87(4), 729-788.
- Greenhalgh, T., & Wong, G. (2013). Training materials for meta-narrative reviews. *Version, 1*, 406.
- Grimshaw, J., Campbell, M., Eccles, M., & Steen, N. (2000). Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family Practice*, 17(suppl. 1), S11-S16.
- Growing up in B.C.* (2015). Office of the Provincial Health Officer. Retrieved from <http://www2.gov.bc.ca/assets/gov/health/about-bc-s-health-care-system/office-of-the-provincial-health-officer/reports-publications/special-reports/guibc-2015.pdf>
- Hagerty, M. R., & Land, K. C. (2007). Constructing summary indices of quality of life: A model for the effect of heterogeneous importance weights. *Sociological Methods & Research*, 35, 455–496.
- Hampshire, K. R. & Matthijsse, M. (2010). Can arts projects improve young people's wellbeing? A social capital approach. *Social Science & Medicine*, 71, 708-16.
- Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing fidelity with flexibility and fit: What do we really know about fidelity of implementation in schools?. *Exceptional Children*, 79(2), 181-193.

- Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D. M., & Vandermeer, B. (2017). Grey literature in systematic reviews: A cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Medical Research Methodology*, 17(1), 64.
- Hauser, R., Brown, B. V., & Prosser, W. (Eds.). (1997). *Indicators of children's well-being*. New York: Russell Sage.
- Hawkins, J. D., Catalano, R. F., & Miller, J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance-abuse prevention. *Psychological Bulletin*, 112, 64–105.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2013). *Multilevel and longitudinal modeling with IBM SPSS. Quantitative Methodology Series*. (2<sup>nd</sup> ed.). New York, NY: Routledge.
- Holden, R. R. (2010). Face validity. *Corsini Encyclopedia of Psychology*.
- Holloway, D. L., & LeCompte, M. D. (2001). Becoming somebody! How arts programs support positive identity for middle school girls. *Education and Urban Society*, 33(4), 388-408.
- Howell, E. M., & Yemane, A. (2006). An assessment of evaluation designs: Case studies of 12 large federal evaluations. *American Journal of Evaluation*, 27(2), 219-236.
- Hulett, S. (1997). Using logic models in arts programs for at-risk youth. *Americans for the Arts Monographs*, 1(6). Washington, DC: Americans for the Arts.
- Hur, Y., & Testerman, R. (2012). An index of child well-being at a local level in the U.S.: The case of North Carolina Counties. *Child Indicators Research*, 5, 29–53. DOI 10.1007/s12187-010-9087-x
- IBM Corp. (2017). IBM SPSS Statistics for Mac, Version 25.0.

- Iwai, K. (2002). The contribution of arts education to children's lives. *Prospects*, 32(4), 407-420.
- Kisida, B., Bowen, D. H., & Greene, J. P. (2016). Measuring critical thinking: Results from an art museum field trip experiment. *Journal of Research on Educational Effectiveness*, 9(sup1), 171-187.
- Kochel, K. P., Ladd, G. W., & Rudolph, K. D. (2012). Longitudinal associations among youth depressive symptoms, peer victimization, and low peer acceptance: An interpersonal process perspective. *Child Development*, 83, 637–650. doi:10.1111/j.1467-8624.2011.01722.x
- Laidlaw Foundation (2015, June 15). Retrieved from <http://laidlawfdn.org/about-us/mission-vision-values>
- Lakeshore Arts (2016, March). Retrieved from <http://lakeshorearts.ca>
- Lamb, V. L. & Land, K. C. (2013). Methodologies used in the construction of composite child well-being indices. In A. Ben-Arieh (Ed.), *Handbook of child well-being* (pp. 2739-2755). New York, NY: Springer.
- Lampert, N. (2013). Inquiry and critical thinking in an elementary art program. *Art Education*, 66(6), 6-11.
- Land, K. C., Lamb, V. L., Meadows, S. O., & Taylor, A. (2007). Measuring trends in child well-being: An evidence-based approach. *Social Indicators Research*, 80, 105–132.
- Land, K. C., Lamb, V. L., & Mustillo, S. K. (2001). Child and youth well-being in the United States, 1975-1998: Some findings from a new index. *Social Indicators Research*, 56(3), 241-320.
- Land, K. C., Lamb, V. L., & Zheng, H. (2011). How are the kids doing? How do we know? Recent trends in child and youth well-being in the United States and some international

- comparisons. *Social Indicators Research*, 100, 463–477. DOI 10.1007/s11205-010-9624-5.
- Land, K. C. (Ed.). (2012). *The well-being of America's children: Developing and improving the child and youth well-being index*. New York, NY: Springer.
- Larson, R. W. (2000). Toward a psychology of positive youth development. *American psychologist*, 55(1), 170–183. doi:10.1037/0003-066X.55.1.170
- Larson, R. W., & Angus, R. M. (2011). Adolescents' development of skills for agency in youth programs: Learning to think strategically. *Child Development*, 82(1), 277-294.
- Laslett, P. (1989). *A fresh map of life. The emergence of the third age*. London: Weidenfeld & Nicholson.
- Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance–covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, 39(5), 307-332.
- Lee, B. J. (2014). Mapping domains and indicators of children's well-being. In *Handbook of child well-being* (pp. 2797-2805). Springer Netherlands.
- Lerner, R. M. (2004). *Liberty: Thriving and civic engagement among American youth*. Thousand Oaks, CA: Sage.
- Lerner, R. M. (2006). Editor's introduction: Developmental science, developmental systems, and contemporary theories. In R. M. Lerner, & W. Damon (Eds.), *Theoretical models of human development. Vol. 1 of Handbook of Child Psychology* (6<sup>th</sup> ed., pp. 1-17). Hoboken, NJ: John Wiley & Sons Inc.
- Lerner, R. M., Fisher, C. B., & Weinberg, R. A. (2000). Toward a science for the people: Promoting civil society through the application of developmental science. *Child*

*Development, 71*, 11-20.

Lerner, R. M., Lerner, J. V., Almerigi, J. B., Theokas, C., Phelps, E., . . . von Eye, A. (2005).

Positive youth development, participation in community youth development programs, and community contributions of fifth-grade adolescents: Findings from the first wave of the 4-H study of Positive Youth Development. *The Journal of Early Adolescence, 25*, 17–71. doi:10.1177/0272431604272461

Leventhal, T., & Brooks-Gunn, J. (2000). The neighborhoods they live in: The effect of

neighborhood residence on child and adolescent outcomes. *Psychological Bulletin, 126*, 309–37.

Lima, M. G. (1995). From aesthetics to psychology: Notes on Vygotsky's Psychology of Art.

*Anthropology and Education Quarterly, 26*(4), 410-425.

Lippman, L., Moore, K. A., & McIntosh, H. (2011). Positive indicators of child well-being: A

conceptual framework, measures, and methodological issues. *Applied Research in Quality of Life, 6*, 425–449.

Luftig, R. L. (1994). The schooled mind: Do the arts make a difference? An empirical evaluation

of the Hamilton Fairfield SPECTRA+ Program, 1992-93, *Center for Human Development, Learning, and Teaching*, Miami University, Oxford, Ohio.

Luke, J. J., Stein, J., Foutz, S., & Adams, M. (2007). Research to practice: Testing a tool for

assessing critical thinking in art museum programs. *Journal of Museum Education, 32*(2), 123-135.

Marmot, M., & Wilkinson, R. G. (Eds.) (1999). *Social determinants of health*. Oxford: Oxford

University Press.

Mantie, R. (2008). Getting unstuck: The One World Youth Arts Project, the music education paradigm, and youth without advantage. *Music Education Research*, 10(4), 473-483.

Martin, J., Sugarman, J., & Thompson, J. (2003). *Psychology and the question of agency*. SUNY Press.

Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., . . . Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology*, 41, 733–746.  
doi:10.1037/0012-1649.41.5.733

Masuda, J. R., Creighton, G., Nixon, S., & Frankish, J. (2011). Building capacity for community-based participatory research for health disparities in Canada: The case of “partnerships in community health research”. *Health Promotion Practice*, 12(2), 280-292.

Matarasso, F. (1996a). *Defining values: Evaluating arts programmes, Social Impact of the Arts*. Working Paper 1, Comedia, Stroud.

Matthews, J. (1989). How young children give meaning to drawing. In A. Gilroy & T. Dalley (Eds.), *Pictures at an exhibition: Selected essays on art and art therapy* (pp. 127–142). London: Tavistock/Routledge.

McAuley, C., & Rose, W. (2010). *Child well-being: Understanding children's lives*. London: Jessica Kingsley, Publisher.

McCay, E., Quesnel, S., Langley, J., Beanlands, H., Cooper, L., Blidner, R., Aiello, A., Howes, C., & Bach, K. (2011). A relationship-based intervention to improve social

connectedness in street-involved youth: A pilot study. *Journal of Child and Adolescent Psychiatric Nursing*, 24(4), 208-215.

McCormick, M. P., Cappella, E., O'Connor, E. E., & McClowry, S. G. (2015). Context matters for social- emotional learning: Examining variation in program impact by dimensions of school climate. *American Journal of Community Psychology*, 56(1-2), 101-119.

McCreary Report: Growing up in B.C. (2015). Office of the Provincial Health Officer. Retrieved from <http://www2.gov.bc.ca/assets/gov/health/about-bc-s-health-care-system/office-of-the-provincial-health-officer/reports-publications/special-reports/guibc-2015.pdf>

McGregor, A., Coulthard, S., & Camfield, L. (2016). Measuring what matters: The role of well-being methods in development policy and practice. *Development Progress*, 04, 1-20.

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., . . . Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, 49, 980–989.  
doi:10.1016/j.jaac.2010.05.017

Minkler, M., & Wallerstein, N. (Eds.). (2011). *Community-based participatory research for health: From process to outcomes*. John Wiley & Sons.

Moore, K. A., Brown, B. V., & Scarupa, H. J. (2003). *The uses (and misuses) of social indicators: Implications for public policy*. Washington, DC: Child Trends.

- Moore, K. A., Lippman, L., & Brown, B. (2004). Indicators of child well-being: The promise for positive youth development. *The ANNALS of the American Academy of Political and Social Science*, 591, 125–145.
- Moore, K. A., Murphey, D., & Bandy, T. (2012). Positive child well-being: An index based on data for individual children. *Maternal and Child Health Journal*, 16, 119–128.
- Moore, K. A., Murphey, D., Bandy, T., & Lawner, E. (2014). Indices of child well-being and developmental contexts. In *Handbook of child well-being* (pp. 2807-2822). Springer Netherlands.
- Moore, K. A., Theokas, C., Lippman, L., Bloch, M., Vandivere, S., & O'Hare, W. (2008). A microdata child well-being index: Conceptualization, creation, and findings. *Child Indicators Research*, 1(1), 17-50.
- Moore, K. A., Vandivere, S., Lippman, L., McPhee, C., & Bloch, M. (2007). An index of the condition of children: The ideal and a less-than-ideal US example. *Social Indicators Research*, 84(3), 291–331. doi:10.1007/s11205-007-9120-8.
- Mynaříková, L. (2012). Art-based Program for social and emotional development of children. *US-China Educational Review*, 8, 720-726.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22(4), 287-293.
- Newman, T., Curtis, K., & Stephens, J. (2003). Do community-based arts projects result in social gain? *Community Development Journal*, 38(4), 310-322.
- O'Hare, W. P. (2012). Development of the child indicator movement in the United States. *Child Development Perspectives*, 6(1), 79-84.
- O'Hare, W. P., & Gutierrez, F. (2012). The use of domains in constructing a comprehensive composite index of child well-being. *Child Indicators Research*, 5(4), 609-629.

- O'Hare, W., & Lee, M. (2007). Factors affecting state differences in child well-being. *Baltimore, MD: The Annie E. Casey Foundation.*
- Ontario Trillium Foundation (2018). *Who we are*. Retrieved from <https://otf.ca/who-we-are>.
- Organization for Economic Cooperation and Development Report. (2015). *How's life in Canada?* OECD Better Life Initiatives Report.
- Organization for Economic Cooperation and Development. (2008). *Handbook on constructing composite indicators: Methodology and user guide* (p. 13). Paris: OECD.
- Otte-Trojel, T., & Wong, G. (2016). Going beyond systematic reviews: realist and meta-narrative reviews. *Evidence-Based Health Informatics*, 275.
- Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice*, 15(3), 311-319.
- Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*. (3<sup>rd</sup> ed.). Newbury Park, CA: Sage Publications.
- Patton, M. Q. (2008). *Utilization-focused evaluation*. (4<sup>th</sup> ed.). Thousand Oaks: SAGE Publications Inc.
- Patton, M. Q. (2010). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. Guilford Press.
- Patton, M. Q. (2013). *Utilization-Focused Evaluation (U-FE) Checklist*. Retrieved from [http://www.wmich.edu/sites/default/files/attachments/u350/2014/UFE\\_checklist\\_2013.pdf](http://www.wmich.edu/sites/default/files/attachments/u350/2014/UFE_checklist_2013.pdf)
- Phelps, E., Balsano, A. B., Fay, K., Peltz, J. S., Zimmerman, S. M., . . . Lerner, J. V. (2007). Nuances in early adolescent developmental trajectories of positive and problematic/risk behaviors: Findings from the 4-H study of positive youth development. *Child and*

*Adolescent Psychiatric Clinics of North America*, 16, 473–496.

doi:10.1016/j.chc.2006.11.006

Pitman, W. (1998). *Learning the arts in an age of uncertainty*. Toronto, ON, Canada: Arts Education Council of Ontario.

Pittman, K., & Irby, M. (1997). Promoting investment in life skills for youth: Beyond indicators for survival and problem prevention. *Eurosociological Reports*, 62, 239-246.

Pollard, E. I., & Lee, P. D. (2002). Child well-being: A systematic review of the literature. *Social Indicators Research*, 61, 59–78.

Postman, N. (1982). *The disappearance of childhood*. Delacorte Press.

Prilleltensky, I. (1997). Values, assumptions, and practices: Assessing the moral implications of psychological discourse and action. *American Psychologist*, 52(5), 517.

Quinby, R. K., Fagan, A. A., Hanson, K., Brooke-Weiss, B., Arthur, M. W., Hawkins, J. D. (2008). Installing the Communities That Care prevention system: Implementation progress and fidelity in a randomized controlled trial. *Journal of Community Psychology*, 36, 313–332.

Remer, J. (1996). *Beyond enrichment: Building effective arts partnerships with schools and your community*. American Council for the Arts, One East 53rd Street, New York, NY 10022.

Reynolds, M. W., Nabors, L., & Quinlan, A. (2000). The effectiveness of art therapy: Does it work?. *Art Therapy*, 17(3), 207-213.

Roth, J. L., & Brooks-Gunn, J. (2003). What exactly is a youth development program? Answers from research and practice. *Applied Developmental Science*, 7, 94–111.

doi:10.1207/S1532480XADS0702\_6

- Rowe, C., Watson-Ormond, R., English, L., Rubesin, H., Marshall, A., Linton, K., ... & Eng, E. (2017). Evaluating art therapy to heal the effects of trauma among refugee youth: The Burma art therapy program evaluation. *Health Promotion Practice, 18*(1), 26-33.
- Sampson, R. J., Raudenbush, S.W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918–924.
- Schalock, R. L. (1995). *Outcome-based evaluation*. New York, NY: Plenum Press.
- Schmitz, C., & Parsons, B. A. (1999). Everything you wanted to know about logic models but were afraid to ask. Retrieved May 2017, from <http://insites.org/publications.html#articles>
- Schwartz, S. H., & Melech, G. (2000). National differences in micro and macro worry: Social, economic, and cultural explanations. In E. Diener & E. M. Suh (Eds.), *Culture and subjective well-being*, (pp. 219-256). MIT Press.
- Sidani, S. (2015). *Health intervention research: Understanding research design and methods*. Sage Publications.
- Statistics Canada (2011). *Population census*. Retrieved from <http://www12.statcan.gc.ca/censusrecensement/2011/rt-td/index-eng.cfm>
- Statistics Canada (2013). *The boundaries of youth: Policy horizons in Canada*. Retrieved from <http://www.horizons.gc.ca/eng/content/2-boundaries-youth>
- Statistics Canada (2014). *Suicide rates. Canadian Community Health Survey, Vital Statistics, 2014*. Retrieved from <http://www.statcan.gc.ca/pub/11-631-x/11-631-x2018001-eng.htm#a3>

- Statistics Canada (2017). *Census datasets*. Retrieved from <http://www12.statcan.gc.ca/datasets/Index-eng.cfm?Temporal=2011&Theme=-1&VNAMEE=&GA=53087&S=0>
- Statistics Canada (2018). *A portrait of Canadian youth*. Retrieved from <http://www.statcan.gc.ca/pub/11-631-x/11-631-x2018001-eng.htm#a3>
- Stepping Up Ontario Annual Report (2016). *A strategic framework to help Ontario's youth succeed*. Retrieved from <http://www.children.gov.on.ca/htdocs/English/documents/youthopportunities/steppingup/Steppingup-2015AnnualReport.pdf>
- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1), 99-103.
- Taut, S. M., & Alkin, M. C. (2003). Program staff perceptions of barriers to evaluation implementation. *American Journal of Evaluation*, 24(2), 213-226.
- Taut, S. & Palacios, D. (2016). Intended and unintended interpretations and uses of PISA results: A consequential validity perspective. *Relieve*, 22 (1), 1-16.
- Teague, G. B., Bond, G. R., & Drake, R. E. (1998). Program fidelity in assertive community treatment: Development and use of a measure. *American Journal of Orthopsychiatry*, 68(2), 216.
- The Students Commissions. (2015). *Measuring youth wellbeing in Canada: A synthesis of youth voices*. Report submitted to UNICEF Canada.
- Tiffany, J. S., Exner-Cortens, D., & Eckenrode, J. (2012). A new measure for assessing youth program participation. *Journal of Community Psychology*, 40(3), 277-291.
- Tobler, N. S. (1986). Meta-analysis of 143 adolescent drug prevention programs: Quantitative

outcome results of program participants compared to a control or comparison group.

*Journal of Drug Issues*, 16, 537–567.

Torsheim, T., Currie, C., Boyce, W., Kalnins, I., Overpeck, M., & Haugland, S. (2004). Material deprivation and self-rated health: A multilevel study of adolescents from 22 European and North American Countries. *Social Science and Medicine* 59(1), 1–12.

UNICEF Canada (2016). Retrieved from <https://www.unicef.ca/en/our-work>

UNICEF Office of Research (2013). Child well-being in rich countries: A comparative overview, *Innocenti Report Card 11*, UNICEF Office of Research, Florence.

UNICEF Office of Research (2016). Fairness for children: A league table of inequality in child well-being in rich countries. *Innocenti Report Card 13*, UNICEF Office of Research, Florence.

*United Nation's Convention on the Rights of the Child* (1989). Definition of a child.

Upitis, R. (2011). *Arts education for the development of the whole child*. A report prepared for the Elementary Teachers' Federation of Ontario.

Upitis, R., & Smithrim, K. (2003). *Learning through the arts: National assessment 1999-2002*. Toronto, Ontario, Canada: Royal Conservatory of Music.

Van Lith, T., Fenner, P., & Schofield, M. J. (2009). Toward an understanding of how art making can facilitate mental health recovery. *Australian e-Journal for the Advancement of Mental Health*, 8(2), 183-193.

Vygotsky, L. (1971). *Psychology of art*. Cambridge, MA. MIT Press.

Wallace-DiGarbo, A., & Hill, D. C. (2006). Art as agency: Exploring empowerment of at-risk youth. *Art Therapy*, 23(3), 119-125.

- Wallander, J. L., & Koot, H. M. (2016). Quality of life in children: A critical examination of concepts, approaches, issues, and future directions. *Clinical Psychology Review*, 45, 131-143.
- Waller, D. (2006). Art therapy for children: How it leads to change. *Clinical Child Psychology and Psychiatry*, 11(2), 271–282.
- Whitbourne, S. K., Whitbourne, S. B., & Konnert, C. (2014). *Adult development & aging: Biopsychosocial perspectives*. Canadian Edition. Jon Wiley & Sons, Inc.: Toronto.
- Wilkinson, R. & Pickett, K. (2011). *The spirit level: Why greater equality makes societies stronger*. Tantor Media.
- W. K. Kellogg Foundation (2004). *W. K. Kellogg Foundation logic model development guide*. Battle Creek, MI: W. K. Kellogg Foundation. Retrieved May 2017, from <http://www.wkkf.org/default.aspx?tabid=75&CID=281&NID=61&LanguageID=0>.
- Wright, R., John, L., Alaggia, R., & Sheel, J. (2006). Community-based arts program for youth in low-income communities: A multi-method evaluation. *Child and Adolescent Social Work Journal*, 23(5-6), 635-652.
- Wright, R., John, L., Ellenbogen, S., Offord, D. R., Duku, E. K., & Rowe, W. (2006). Effect of a structured arts program on the socio-emotional functioning of youth from low-income communities: Findings from a Canadian longitudinal study. *Journal of Early Adolescence*, 26(2), 186-205.
- Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: Meta-narrative reviews. *BMC Medicine*, 11(1), 20.

World Health Organization (2010). *Social determinants of health and well-Being among young people. Health Behaviour In School-Aged Children (HBSC) Study: International Report From The 2009/2010 Survey*. Health Policy for Children and Adolescents; No. 6: Regional Office for Europe.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage

*Youth Development Index Results Report* (2013). Commonwealth Youth Programme. Retrieved from  
[http://www.youthdevelopmentindex.org/cms/cmsyouth/\\_images/197918019952385f3219c75.pdf](http://www.youthdevelopmentindex.org/cms/cmsyouth/_images/197918019952385f3219c75.pdf)

*Youth Development Index Results Report* (2016). Commonwealth Youth Programme. Retrieved from  
[http://www.youthdevelopmentindex.org/cms/cmsyouth/\\_images/197918019952385f3219c75.pdf](http://www.youthdevelopmentindex.org/cms/cmsyouth/_images/197918019952385f3219c75.pdf)

Zarrett, N., & Eccles, J. (2006). The passage to adulthood: Challenges of late adolescence. *New Directions for Youth Development*, 111, 13-28.