

THE SATISFACTION WITH LIFE SCALE ADAPTED FOR CHILDREN:
INVESTIGATING THE STRUCTURAL, EXTERNAL, AND SUBSTANTIVE
ASPECTS OF CONSTRUCT VALIDITY

by

Anne Maria Gadermann

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Measurement, Evaluation, & Research Methodology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2009

© Anne Maria Gadermann, 2009

Abstract

Measuring and monitoring children's satisfaction with life is of great significance for improving children's lives. In order to do this, validated measures to assess children's satisfaction with life are necessary. This dissertation describes a program of research for the validation of the Satisfaction with Life Scale adapted for Children (SWLS-C). The introductory chapter provides a theoretical background for subjective well-being and validity/validation research and definitions of key terms. The first manuscript presents psychometric findings on the structural and external aspects of construct validity. A stratified random sample of 1233 students in grades 4 to 7 (48% girls, mean age of 11.7 years) provided data on the SWLS-C and measures of optimism, self-concept, self-efficacy, depression, empathic concern, and perspective taking. The SWLS-C demonstrated a unidimensional factor structure, high internal consistency, and evidence of convergent and discriminant validity. Furthermore, differential item functioning and differential scale functioning analyses indicated that the SWLS-C measures satisfaction with life in the same way for different groups of children. The second manuscript investigated the substantive aspect of construct validity for the SWLS-C by examining the cognitive processes of children when responding to the items. Think-aloud protocol interviews were conducted with 55 students in grades 4 to 7 (58 % girls, mean age of 11.0 years) and content analysis was used to analyze the data. In their responses, children mainly used an 'absolute strategy' (statements indicating the presence/absence of something they consider important for their satisfaction with life) or a 'relative strategy' (statements indicating comparative judgments). The absolute statements primarily referred to social relationships, personal characteristics, time use, and possessions. In the

relative statements, children primarily compared what they have to what (a) they want, (b) they had in the past, (c) other people have, and (d) they feel they need. The results are in line with multiple discrepancies theory (Michalos, 1985) and previous empirical findings. These two studies provide evidence for the meaningfulness of the inferences of the SWLS-C scores. The concluding chapter highlights the contributions of the dissertation, discusses limitations of the presented research, and delineates a future validation program for the SWLS-C.

Table of Contents

Abstract	ii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Acknowledgements	viii
Widmung	ix
Co-authorship statement	x
1. Introduction	1
The Middle Years Development Instrument and the SWLS-C	3
Subjective well-being and satisfaction with life	4
Validity and validation	9
Purpose and structure of the dissertation	16
References	18
2. Investigating validity evidence of the SWLS-C	25
Method	33
Results	40
Discussion	54
References	57
3. Investigating the substantive aspect of construct validity for the SWLS-C: A focus on cognitive processes	65
Method	71
Results	79
Discussion	99
References	107
4. Concluding chapter	113
Novel contributions of the presented research	114

Limitations of the presented studies and future directions.....	116
Implications of the research findings	119
References	121
Appendices.....	123
Appendix A: Behavioural Research Ethics Board certificate of approval, student assent form, and parent consent form for the study in chapter 2	123
Appendix B: Behavioural Research Ethics Board certificate of approval, student assent form, and parent consent form for the study in chapter 3	130

List of Tables

Table 2.1 The SWLS-C.....	36
Table 2.2 Item Response Percentages of the SWLS-C Items.....	40
Table 2.3 Intercorrelations between the Items of the SWLS-C Using the Polychoric Correlation Matrix	41
Table 2.4 Factor Loadings of the Items of the SWLS-C	43
Table 2.6 Intercorrelations between the SWLS-C and Convergent and Discriminant Measures	53
Table 3.1 The SWLS-C.....	73

List of Figures

<i>Figure 2.1.</i> Nonparametric item response functions of the five items of the SWLS-C....	46
<i>Figure 2.2.</i> Conditional reliability of the SWLS-C plotted against the expected total score.	47
<i>Figure 2.3.</i> Differential scale functioning for the SWLS-C.	50
<i>Figure 3.1.</i> Tree diagram of the coding categories.	79
<i>Figure 3.2.</i> Tree diagram of the coding categories for item 1.	85
<i>Figure 3.3.</i> Tree diagram of the coding categories for item 2.	87
<i>Figure 3.4.</i> Tree diagram of the coding categories for item 3.	88
<i>Figure 3.5.</i> Tree diagram of the coding categories for item 4.	89
<i>Figure 3.6.</i> Tree diagram of the coding categories for item 5.	90
<i>Figure 3.7.</i> Summary of the differences and commonalities in strategy and content (sub)categories for the five items of the SWLS-C.	91
<i>Figure 3.8.</i> Frequencies for the five SWLS-C items for strategy and content (sub)categories	93

Acknowledgements

I would like to express my deepest gratitude to my supervisor and mentor Dr. Bruno D. Zumbo for the gentle guidance throughout my doctoral studies, for sharing his time and expertise, for the rich discussions, and most of all for his wonderful support at all times. I am also very grateful to my committee members, Dr. Anita M. Huble and Dr. Kimberly A. Schonert-Reichl, as well as Dr. Susan James for their advice, feedback, mentorship, and support.

In addition, I would like to sincerely thank my friends and my family, my brother Moritz and my parents Monika and Ernst, and most of all Martin for all their love, care, and support throughout the beautiful and at times challenging years.

Widmung

Voller Liebe für meine Eltern und Martinus.

Co-authorship statement

Chapters 2 and 3 of my dissertation have been co-authored, with me as lead author.

Chapter 2 was co-authored by Dr. Kimberly A. Schonert-Reichl and Dr. Bruno D.

Zumbo. My contribution was in the formulation of the research question, the data analyses and interpretation of the results, and the manuscript preparation. Dr. Kimberly A. Schonert-Reichl provided the data and contributed to the interpretation of the results.

Dr. Bruno D. Zumbo contributed to the data analyses and interpretation of the results.

Chapter 3 was co-authored by Dr. Martin Guhn and Bruno D. Zumbo. My contribution was in the formulation of the research question, design of the study, data collection, data analyses and interpretation of the results, and the manuscript preparation. Dr. Martin Guhn and Dr. Bruno Zumbo contributed to the data analyses and the interpretation of the results.

Introduction

In the last three decades, a vast amount of literature has addressed the subjective well-being of adults, in contrast to early psychology where positive psychological states were mainly disregarded and the focus was on the negative aspects of the psychological spectrum (Diener, 1984). The journal ‘Social Indicators Research’, devoted to publishing research on quality of life and subjective well-being, was founded in 1974. During the 1980s, the number of publications listed in Psychological Abstracts using the terms ‘well-being’, ‘happiness’, and ‘life satisfaction’ quintupled (Myers & Diener, 1995) and a search of PsycArticles, PsycInfo, and PsycExtra with the term ‘well-being’ indicated that more than 1700 articles containing this term have been published annually in the last few years on adults. This trend is very much in line with the growing movement of Positive Psychology, which has the aim “to catalyze a change in the focus of psychology from preoccupation only with repairing the worst things in life to also building positive qualities” (Seligman & Csikszentmihalyi, 2000, p. 5). Proponents of this approach have also argued that data on subjective well-being should be collected systematically and in an ongoing, large-scale fashion—in addition to existing social and economic indicators, such as rates of infant mortality and the gross domestic product—to evaluate and compare countries with regard to their inhabitants’ subjective well-being to inform policy makers (e.g., Diener, 2000; Diener, Kesebir, & Lucas, 2008; Diener & Seligman, 2004).

In contrast to research with adults, the topic of subjective well-being has received less attention with regard to children and adolescents (Gilligan & Huebner, 2007; Huebner, 1991) and has become the emphasis of empirical studies only more recently (Huebner & Gilman, 2002). This is reflected in the number of articles published on the

topic. When conducting the same literature search as above, except constraining it to childhood (from birth to age 12), less than ¼ of the number of articles appear as for adults. However, the topic of positive youth development has started receiving more attention, and a shift to focusing on positive developmental outcomes besides problem behavior and its prevention has occurred (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 1999). This is reflected in the positive youth development approach (e.g., Damon, 2004; Lerner et al., 2005) and the developmental assets approach for youth (e.g., Benson, 2003; Lerner, 2003), which emphasize the strength and potential of youth rather than taking a deficit-oriented view.

Yet, there has been a lack of child-oriented indicators (Ben-Arieh et al., 2001), and, as Catalano et al. (1999, Executive summary) noted, “a major obstacle to tracking indicators of positive youth development constructs is the absence of widely accepted measures for this purpose”¹. Similarly, the relative underrepresentation of research on children and adolescents’ subjective well-being and life satisfaction may have, in part, resulted from the situation that there have been relatively few psychometrically sound measures² and those have been developed more recently (Gilligan & Huebner, 2007; Huebner & Diener, 2008; Seligson, Huebner, & Valois, 2005; see Gilman & Huebner, 2000 for a review of life satisfaction measures for adolescents).

The purpose of my dissertation is to investigate the validity of the inferences from the scores of a newly adapted measure assessing satisfaction with life in children,

¹ Today, an increasing interest in the issue is reflected, for example, by the foundation of the ‘International Society for Child Indicators’ in 2006, and the start of the society’s journal, *Child Indicators Research*, in 2008, which publishes research on “measurement and indicators of children’s well-being and their usage within multiple domains and in diverse cultures”.

² In my dissertation the terms ‘measure’, ‘scale’, ‘measurement instrument’, and ‘test’ are used interchangeably, although it is acknowledged that the term test is typically used to refer to educational achievement tests.

namely, the Satisfaction with Life Scale for Children (SWLS-C). In this regard, two studies investigating validity evidence of the SWLS-C with children in grades 4 to 7 are presented in chapters 2 and 3. In order to set up the research purpose of this dissertation, the following sections will provide (i) a description of the project within which the SWLS-C was adapted, (ii) a discussion of the historical roots and definitions of subjective well being and related concepts, and (iii) an overview of the evolving conceptualization of validity and validation and how this has influenced measurement practice.

The Middle Years Development Instrument and the SWLS-C

The Middle Years Development Instrument (MDI) has been developed as part of a research community collaborative project between the Human Early Learning Partnership (HELP) at the University of British Columbia, the Vancouver School Board, and United Way of the Lower Mainland. The MDI is a student self-report survey, which is planned to be administered annually at a population level to grade 4 and grade 7 students in British Columbia, Canada. The aim of this survey is to gain an understanding of students' developmental status in middle childhood (age period between 6 and 12 years) from a developmental assets and strengths-based perspective. The survey integrates a number of existing and newly developed/adapted scales to assess the five domains (i) Social and Emotional Competence and Well-Being, (ii) Connectedness, (iii) School Experiences, (iv) Physical Health and Well-Being, and (v) Time Use.

One construct that the MDI aims to assess within the Social and Emotional Competence and Well-Being domain is satisfaction with life. Accordingly, the newly

adapted *Satisfaction with Life Scale adapted for Children* (SWLS-C) has been included in the MDI. The SWLS-C was adapted from the *Satisfaction with Life Scale* (SWLS, Diener, Emmons, Larsen, & Griffin, 1985)³, a widely used measure to assess global life satisfaction in adults. The adaptation was conducted by Dr. Denise Buote, Angela Jaramillo, and Dr. Kimberly Schonert-Reichl, all experts in the area of children’s socio-emotional development.

Before using the SWLS-C on a large scale basis within the MDI project, it is, however, necessary to examine evidence with regard to the validity of the inferences of the scale scores of the SWLS-C. In regard to this issue, my dissertation presents two studies that investigate validity aspects of the SWLS-C. In order to introduce the reader to some fundamental issues and definitions pertaining to the areas of subjective well-being and validity/validation, the following sections provide (i) a delineation of the constructs of subjective well-being, satisfaction with life, and quality of life, and (ii) a description of the conceptualization of validity/validation that is endorsed in the presented studies.

Subjective well-being and satisfaction with life

Current research on well-being evolved out of two philosophical traditions addressing ‘happiness’, the *eudaimonic* and the *hedonic* view (Ryan & Deci, 2001). Eudaimonia literally translates as ‘favoured by the daimones (near-gods or gods)’, and philosophers that were engaged in this tradition (e.g., Socrates, Plato, and Aristotle) argued that “people should reflect on their lives as a whole, discover what is most

³ For reviews of this scale, please see Pavot and Diener (1993, 2008).

important or valuable (i.e., life's final end or TELOS), and plan and live their lives to achieve that end" (Michalos, 2008, p. 355). Socrates and Plato described the good life as "the possession of *goods* which made possible the best life for a man", and this concept was further adopted by Aristotle, who proposed that happiness is "the accumulation of the greatest *goods* accessible to a man" (Tatarkiewicz, 1976, p. 29, 30). There has been much discussion on what these 'greatest goods' actually are. According to Aristotle, a variety of goods are needed (such as honour, virtue, health, and social relationships), whereas the Stoics argued that there is one supreme good, namely virtue (Tatarkiewicz). In the eudaimonic approach, the measure of happiness was objective, as the criterion according to which happiness was assessed was considered to reflect objective values rather than a subjective judgment (Kashdan, Biswas-Diener, & King, 2008). According to this view, only a person that is leading a life in accordance with certain moral values may be viewed as happy.

In contrast, in hedonism the emphasis is on the subjective experience of pleasure, and happiness is equated with the sum of subjectively experienced pleasures. According to this view, the primary motivation of human beings is to pursue pleasure and avoid pain (McGill, 1967). Specifically, Aristippus proposed that "the goal of life is to experience the maximum amount of pleasure" and that "happiness is the totality of one's hedonic moments" (Ryan & Deci, 2001, pp. 143-144). This concept has been further adopted in different philosophical approaches. In utilitarianism, for example, the concept was extended in that the focus was no longer on individual pleasure, but on the largest degree of pleasure for the greatest number of people, a concept that has been called universalistic hedonism (McGill, 1976).

The debate about these two different approaches to happiness has had a recent renaissance in psychology (Kashdan et al., 2008; Ryan & Deci, 2001; Waterman, 1993, 2008). The two contemporary conceptions of happiness are related but have different foci (Keyes, Shmotkin, & Ryff, 2002). Proponents of the eudaimonic view have focused on concepts such as personal expressiveness, which emphasizes self-realization by fulfilling one's potential (Waterman, 1993, 2008), and psychological well-being, which consists of the six dimensions autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance, and which emphasizes positive functioning (Ryff & Keyes, 1995).

In contrast, hedonic psychology, which has been defined as the research of “what makes experiences and life pleasant and unpleasant” (Kahneman, Diener, & Schwarz, 1999, p. ix), has mainly focused on subjective well-being, which, in a broad conception of hedonism, incorporates individuals' values, emotions, and evaluations (Diener, Sapyta, & Suh, 1998). Specifically, subjective well-being focuses on individuals' self-evaluation “and does not grant complete hegemony to the external judgments of behavioural experts” (Diener et al., pp. 33-34). Subjective well-being is therefore regarded a “particularly democratic scalar” that considers people's values and goals and thus allows for aspects of cultural relativism (Diener & Suh, 2000, p. 4).

With regard to empirical research, the majority of studies have employed measures that (intend to) assess subjective well-being, rather than psychological well-being or personal expressiveness (Pavot in Sirgy et al., 2006). One of the measures in this tradition is the SWLS (Diener et al., 1985), which assesses one aspect of subjective well-being, namely, satisfaction with life. In order to clarify the usage of the terms in this

dissertation, the following section provides definitions of subjective well-being, satisfaction with life, and the related concept of quality of life.

Definitions of subjective well-being, satisfaction with life, and quality of life

Diener (2006) defined subjective well-being as the following:

Subjective well-being refers to all of the various types of evaluations, both positive and negative, that people make of their lives. It includes reflective cognitive evaluations, such as life satisfaction and work satisfaction, interest and engagement, and affective reactions to life events, such as joy and sadness. Thus, subjective well-being is an umbrella term for the different valuations people make regarding their lives, the events happening to them, their bodies and minds, and the circumstances in which they live. (pp. 399-400)

This definition indicates that subjective well-being is a multi-faceted concept and consists of several components: Positive and negative affect, life satisfaction, and domain satisfactions (Diener, Suh, Lucas, & Smith, 1999). Previous research utilizing a multitrait-multimethod approach has shown that the affective components and life satisfaction are related, but discriminable (Lucas, Diener, & Suh, 1996).

As the present discussion focuses on the measurement of the life satisfaction component of subjective well-being, this is defined in the following (for a discussion of the other components, please see Diener, 2006): “Life satisfaction represents a report of how a respondent evaluates or appraises his or her life taken as a whole. It is intended to

represent a broad, reflective appraisal that a person makes of his or her life” (Diener, 2006, p. 401).

It is critical to relate this discussion to the research literature on ‘quality of life’. The term quality of life is frequently used in the health sciences, and in many instances, there is considerable similarity with respect to the conceptualization and measurement of quality of life and subjective well-being. However, it must be noted that the usage of the term quality of life is not consistent. In some definitions, the concept of quality of life emphasizes external and objective aspects of individuals’ lives, such as external circumstances; other definitions emphasize individuals’ subjective experiences (Diener, 2006). The latter approach is reflected in the definition of quality of life that is proposed by the World Health Organization’s Quality of Life assessment (WHOQOL) group (1995):

[Quality of life is defined as] individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns. It is a broad ranging concept, incorporating in a complex way individuals' physical health, psychological state, level of independence, social relationships, personal beliefs and their relationships to salient features of the environment. (p. 1405)

The WHOQOL’s definition of quality of life clearly shows similarities with Diener’s (2006) definition of subjective well-being. Specifically, both definitions

highlight the subjectivity of these concepts, incorporate negative and positive aspects of individuals' lives, and are considered multidimensional. (For a further discussion of the similarities and distinctions between quality of life and subjective well-being, see Camfield & Skevington, 2008.) Given that this dissertation focuses on the validation of a measure assessing satisfaction with life in children, I provide a brief overview of validity and validation and relevant definitions in the following.

Validity and validation

There is still much discussion and controversy about the conceptualizations and definitions of validity and validation⁴. In the following, I will briefly review the evolving conceptualization of validity with a special emphasis on Messick's (1988, 1989, 1995, 1998) unified concept of validity. It needs to be acknowledged that there are contrasting views on validity, such as the one proposed by Borsboom, Mellenbergh, and van Heerden (2004) and Borsboom, Cramer, Kievit, Zand Scholten, and Franic (2009).

The evolving conceptualization of validity

Validity lies at the foundation of measurement and testing, as "...without validity, a test, measure, or observation and any inferences made from it are meaningless" (Hubley & Zumbo, 1996, p. 207). There has been much debate on what validity is and how it can be conceptualized. In the first half of the last century, the criterion-based view of validity was dominant. According to this view, a test or scale could be validated by examining to

⁴ The distinction between validity and validation is that validity refers to a concept and validation to a process (e.g., Zumbo, 1998, 2007, 2009).

what extent test or scale scores were correlated with or predictive of a criterion (Kane, 2001). This view is reflected in the following quote by Anastasi (1950):

To claim that a test measures anything over and above its criterion is pure speculation of the type that is not amenable to verification and hence falls outside the realm of experimental science. To the question, "What does this test measure?", the only defensible answer can thus be that it measures a sample of behavior which in turn may be diagnostic of the criterion or criteria against which the particular test was validated. (p. 67)

The criterion-related view of validity was, however, not the sole view during that time. Several scholars contended that correlational evidence was not sufficient to establish validity. Rather, it was considered necessary to also assess whether an instrument was measuring the relevant content for a given purpose as evaluated by subject matter experts. Accordingly, this notion of validity was labeled content-validity (cf. Sireci, 1998).

In the early 1950s, the *Technical recommendations for psychological tests and diagnostic techniques* (American Psychological Association (APA), 1954) described four types of validity, namely, content, predictive, concurrent⁵, and construct validity. According to Cronbach and Meehl—both key members of the committee working on the technical recommendations—the chief innovation of the report was the explication of the concept of construct validity. According to the report, a test's construct validity was to be "evaluated by investigating what psychological qualities a test measures, i.e., by

⁵ Predictive and concurrent validity are both validity types that can be subsumed under the concept of criterion-oriented validity (cf. Cronbach & Meehl, 1955).

demonstrating that certain explanatory constructs account to some degree for performance on the test” (APA, 1954, p. 14). In a separate publication, Cronbach and Meehl (1955) elaborated on this new concept of construct validity. Therein, they explicate that construct validity does not only need to be investigated empirically, but also theoretically, as the process of construct validation is, in essence, equivalent to validating the theory underlying the test. The focus of construct validity is on a (latent psychological) construct—that is, a postulated attribute of individuals—which is presumed to be reflected in the individuals’ performance/scores on the measure. Cronbach and Meehl further explicate the implications of (the new concept of) construct validity for validation research by means of referring to what they coined the nomological network. The nomological network presents an interlocking system of (statistical and/or deterministic) laws, which describe the relationships between latent (that is, unobservable) psychological constructs and measurable (or observable) variables.

The meaning of the term validity and construct validity has further evolved over the years. This is illustrated in Sireci’s (2009) review of the validity terminology as used in the series of the *Technological recommendations* (later editions of this series were called *Standards for educational and psychological testing*) from 1952 to 1999. In that review, Sireci notes that, in 1974, a shift occurred from viewing the validity types as separate to a unified view of validity. A major proponent of this unified view was Messick (1988, 1995, 1998), who stated:

From the perspective of validity as a unified concept, all educational and psychological measurement should be construct-referenced because construct interpretation undergirds all score-

based inferences—not just those related to interpretative meaningfulness but also the content- and criterion-related inferences specific to applied decisions and actions based on test scores. (Messick, 1988, p. 35)

In this unified view, the importance of investigating and integrating relevant evidence from multiple sources is described as preferable to any validation practice that solely relies on one type of validity. In other words, Messick (1989) defined validity as “an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment” (p. 13). Messick thus emphasizes that it is not the test that is being validated, but the inferences and actions based on the test scores. That is, it needs to be investigated to what degree score meaning and actions apply across different people and contexts. Therefore, validity is seen as a developing concept and validation as an ongoing process. Furthermore, he considers the evaluation of intended and unintended consequences of test use as an aspect of (construct) validity. At the core of Messick’s unified validity framework lies construct validity, which he defined as comprising “the evidence and rationales supporting the trustworthiness of score interpretation in terms of explanatory concepts that account for both test performance and score relationships with other variables” (Messick, 1995, p. 743). He distinguishes between six aspects of construct validity that can be used as validity criteria and that can inform the validation process: (i) the content aspect, which includes evidence of content representativeness and relevance; (ii) the substantive aspect, which highlights the function of substantive theories and process models for (a) identifying participants’

response processes utilized in the measurement situation, and (b) empirically investigating whether participants actually engage in these processes; (iii) the structural aspect, which evaluates the structural fidelity, i.e., whether the scoring structure of the measurement task is in line with the theory of the construct domain; (iv) the generalizability aspect, which investigates to what degree score properties and interpretations can be generalized across individuals, contexts, and tasks; (v) the external aspect, which examines the convergent and discriminant relationships to other variables; and (vi) the consequential aspect, which evaluates the value implications and the intended and unintended consequences of the score interpretation and use.

This unified view suggests that multiple sources of evidence need to be integrated for the investigation of construct validity, but, depending on the measurement task of interest, different aspects of construct validity may be the main foci of investigation.

Similarly, the *Standards for educational and psychological testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME), 1999) state that “the process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations” (p. 9). They list five sources of evidence, namely (i) evidence based on test content, (ii) evidence based on response processes, (iii) evidence based on internal structure, (iv) evidence based on relations to other variables, and (v) evidence based on consequences of testing. These are obviously very much in line with the ones by Messick, except that the generalizability aspect is not listed as an individual source but is subsumed under evidence based on relations to other variables.

After this discussion of the evolving term of validity and the implications for validation practice, it is of interest to review how information on sources of validity evidence is reported in research practice.

Sources of validity evidence reported in research practice

The fundamental importance of validity and validation is widely acknowledged, and the importance of investigating and reporting relevant evidence when using measurement instruments is strongly advocated by influential sources, such as the *Standards for educational and psychological testing* (AERA, APA, & NCME, 1999), the guidelines for statistical methods in psychology journals (Wilkinson & The APA Task Force on Statistical Inference, 1999), and the APA Publication manual (APA, 2001). However, several studies investigating the reporting practices on validity evidence of instrument developers and users indicate that the information provided in published articles and reports is often insufficient and lacking (Cizek, Rosenberg, & Koons, 2008; Hogan & Agnello, 2004; Qualls & Moss, 1996). Qualls and Moss reviewed 622 studies published in APA journals in 1992. Information on validity evidence was only provided for 31.7% of the measurement instruments that were used in these studies; interestingly, evidence for validity was only provided for 15.7% of the new instruments. The majority of the studies reported construct validity evidence; however, it is not detailed which aspect(s) of construct validity. Similarly, reliability information was reported for only 41% of the instruments.

Hogan and Agnello (2004) examined the reporting practices regarding validity evidence using 696 research reports on measurement instruments published in the Directory of Unpublished Experimental Measures by APA, which, as they state, is one of the most widely cited references for information on psychological instruments. The Directory provides information (including, among other things, information on validity) on measures predominantly from psychology, education, and psychology that were published in 36 journals between 1991 and 1995. Hogan and Agnello state that only 52.3% of the entries in the Directory report information on one source of validity for the measures, and 2.3% on two sources. The most frequently reported source of validity evidence was correlations to other variables (i.e., an unspecified other variable, another scale, a specified other variable), which were reported in 85% of the cases, followed by subtest correlations (7%), group contrasts (4%), factor analysis (2%), and other (3%). This is surprisingly restrictive in that evidence for several other aspects of construct validity, such as the content or substantive aspects, are not mentioned at all, showing that much research practice is still conducted in the tradition of the criterion-related model.

Cizek et al. (2008) specifically compared the perspective of reporting validity evidence with 'modern validity theory', which they view in line with the *Standards* and Messick as described above. For their study, Cizek and colleagues used the 16th Mental Measurements Yearbook that reviews 283 educational and psychological tests that are new or substantially revised since 2003. Based on examining different indicators (such as whether validity is portrayed as a unitary concept, or whether it is conceptualized as a characteristic of the inference), they conclude that "a modern conceptualization of validity is not the norm" (p. 404); for example, the unitary perspective was taken in only

2.5% of the reports. Furthermore, they investigated which sources of validity were reported. The most frequently reported source of validity evidence was relationships to other variables. Specifically, criterion-related evidence of validity was used in 67.2% of the cases, followed by the so-called construct validity (58%), which included factor-analytic and convergent and discriminant evidence. Furthermore, for 48.4% of the tests, evidence for content validity was reported. In contrast, evidence based on consequences of testing and evidence for response processes were only reported for 2.5% and for 1.8% of the tests, respectively.

Purpose and structure of the dissertation

The purpose of my dissertation is to investigate the validity of the inferences from the scores of the SWLS-C by investigating several aspects of construct validity (in the language of Messick) or sources of validity evidence (in the language of the *Standards*). Specifically, I am investigating the structural, external, and substantive aspects of construct validity (i.e., evidence based on internal structure, evidence based on relations to other variables, and evidence based on response processes).

This dissertation is written in manuscript-based format following the guidelines by the Faculty of Graduate Studies at the University of British Columbia. The first manuscript, chapter 2, focuses on psychometric analyses that investigate the structural aspect (investigating dimensionality, reliability, and differential item functioning) and the external aspect (investigating convergent and discriminant validity evidence) of construct validity for the SWLS-C. The second manuscript, chapter 3, investigates the substantive aspect of construct validity by examining the response processes of children when

responding to the items of SWLS-C. Finally, contributions and limitations of the presented research as well as future directions will be discussed in the closing chapter.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 51, (Part 2), 1-38.
- American Psychological Association (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- Anastasi, A. (1950). The concept of validity in the interpretation of test scores. *Educational and Psychological Measurement*, 10, 67-78.
- Ben-Arieh, A., Kaufman, N. H., Andrews, A. B., George, R. M., Lee, B. J., & Aber, L. J. (2001). *Measuring and monitoring children's well-being*. Dordrecht, Netherlands: Kluwer Academic Press.
- Benson, P. L. (2003). Developmental assets and asset-building community: Conceptual and empirical foundations. In R. M. Lerner, & P. L. Benson (Eds.), *Developmental assets and asset-building communities: Implications for research, policy, and practice* (pp. 19-43). Norwell, MA: Kluwer Academic Publishers.
- Borsboom, D., Cramer, A. O. J., Kievit, R. A., Zand Scholten, A., & Franic, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 135-170). Charlotte, NC: Information Age Publishing.

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*, 1061-1071.
- Camfield, L., & Skevington, S. M. (2008). On subjective well-being and quality of life. *Journal of Health Psychology, 13*, 764-775.
- Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (1999). *Positive youth development in the United States: Research findings on evaluations of positive youth development programs*. Report to the US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation and National Institute for Child Health and Human Development. Retrieved August, 2009, from <http://aspe.hhs.gov/hsp/PositiveYouthDev99/index.htm>.
- Cizek, G. J., Rosenberg, S. L., & Koons, H. K. (2008). Sources of validity evidence for educational and psychological tests. *Educational & Psychological Measurement, 68*, 397-412.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.
- Damon, W. (2004). What is positive youth development? *The ANNALS of the American Academy of Political and Social Science, 591*, 13-24.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542-575.
- Diener, E. (2000). Subjective well-being: The science of happiness and a proposal for a national index. *American Psychologist, 55*, 34-43.
- Diener, E. (2006). Guidelines for national indicators of subjective well-being and ill-being. *Journal of Happiness Studies, 7*, 397-404.

- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*, 71-75.
- Diener, E., Kesebir, P., & Lucas, R. (2008). Benefits of accounts of well-being—for societies and for psychological science. *Applied Psychology: An International Review, 57*, 37-53.
- Diener, E., Sapyta, J. J., & Suh, E. (1998). Subjective well-being is essential to well-being. *Psychological Inquiry, 9*, 33-37.
- Diener, E., & Seligman, M. E. P. (2004). Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest, 5*, 1-31.
- Diener, E., & Suh, E. (2000). Measuring subjective well-being to compare the quality of life of cultures. In E. Diener, & E. Suh (Eds.), *Culture and subjective well-being* (pp. 13-36). Cambridge, MA: The MIT Press.
- Diener, E., Suh, E., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*, 276-302.
- Gilligan, T. D., & Huebner, E. S. (2007). Initial development and validation of the Multidimensional Students' Life Satisfaction Scale—adolescent version. *Applied Research in Quality of Life, 2*, 1-16.
- Gilman, R., & Huebner, E. S. (2000). Review of life satisfaction measures for adolescents. *Behaviour Change, 17*, 178-83.
- Hogan, T. P., & Agnello, J. (2004). An empirical study of reporting practices concerning measurement validity. *Educational & Psychological Measurement, 64*, 802-812.
- Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *Journal of General Psychology, 123*, 207-215.

- Huebner, E. S. (1991). Initial development of the Students' Life Satisfaction Scale. *School Psychology International, 12*, 231-240.
- Huebner, E. S., & Gilman, R. (2002). An introduction to the Multidimensional Students' Life Satisfaction Scale. *Social Indicators Research, 60*, 115-122.
- Kahneman, D., Diener, E. & Schwarz, N. (Eds.). (1999). *Well-Being: The Foundations of Hedonic Psychology*. New York: The Russell Sage Foundation.
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*, 319-342.
- Kashdan, T. B., Biswas-Diener, R. & King L. A. (2008). Reconsidering happiness: The costs of distinguishing between hedonics and eudaimonia. *Journal of Positive Psychology, 3*, 219-233.
- Keyes, C. L. M., Shmotkin, D., & Ryff, C. D. (2002). Optimizing well-being: The empirical encounter of two traditions. *Journal of Personality & Social Psychology, 82*, 1007-1022.
- Lerner, R. M. (2003). Developmental assets and asset-building communities: A view of the issues. In R. M. Lerner, & P. L. Benson (Eds.), *Developmental assets and asset-building communities: Implications for research, policy, and practice* (pp. 3-18). Norwell, MA: Kluwer Academic Publishers.
- Lerner, R. M., Lerner, J. V., Almerigi, J. B., Theokas, C., Phelps, E., Gestsdottir, S. et al. (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. *Journal of Early Adolescence, 25*, 17-71.

- Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality & Social Psychology, 71*, 616-628.
- McGill, V. J. (1967). *The idea of happiness*. New York: F. A. Praeger.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 12-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Messick, S. (1998). Test validity: A matter of consequences. *Social Indicators Research, 45*, 35-44.
- Michalos, A. C. (2008). Education, happiness and wellbeing. *Social Indicators Research, 87*, 347-366.
- Myers, D. G., & Diener, E. (1995). Who is happy? *Psychological Science, 6*, 10-19.
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction with Life Scale. *Psychological Assessment, 5*, 164-172.
- Pavot, W., & Diener, E. (2008). The Satisfaction with Life Scale and the emerging construct of life satisfaction. *Journal of Positive Psychology, 3*, 137-152.
- Qualls, A. L., & Moss, A. D. (1996). The degree of congruence between test standards and test documentation within journal publications. *Educational & Psychological Measurement, 56*, 209-214.

- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, 52, 141-166.
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality & Social Psychology*, 69, 719-727.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55, 5-14.
- Seligson, J. L., Huebner, E. S., & Valois, R. F. (2005). An investigation of a Brief Life Satisfaction Scale with elementary school children. *Social Indicators Research*, 73, 355-374.
- Sireci, S. G. (1998). The construct of content validity. In B. D. Zumbo (Ed.), *Validity theory and the methods used in validation: Perspectives from the social and behavioral sciences* (pp.83-117). Amsterdam: Kluwer Academic Press.
- Sireci, S. G. (2009). Packing and unpacking sources of validity evidence: History repeats itself again. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 19-37). Charlotte, NC: Information Age Publishing.
- Sirgy, M. J., Michalos, A. C., Ferriss, A. L., Easterlin, R. A., Patrick, D., & Pavot, W. (2006). The quality-of-life (QOL) research movement: Past, present, and future. *Social Indicators Research*, 76, 343-466.
- Tatarkiewicz, W. (1976). *Analysis of happiness*. The Hague, Netherlands: Martinus Nijhoff.

- Waterman, A. S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality & Social Psychology, 64*, 678-91.
- Waterman, A. S. (2008). Reconsidering happiness: A eudaimonist's perspective. *Journal of Positive Psychology, 3*, 234-252.
- WHOQOL Group (1995). The World Health Organization Quality of Life Assessment (WHOQOL): Position paper from the World Health Organization. *Social Science & Medicine, 41*, 1403-1409.
- Wilkinson, L., & The APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.
- Zumbo, B. D. (Ed.). (1998). Validity theory and the methods used in validation: Perspectives from the social and behavioral sciences [Special issue]. *Social Indicators Research, 45*(1-3).
- Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26: Psychometrics* (pp. 45-79). Amsterdam: Elsevier Science B. V.
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 65-82). Charlotte, NC: Information Age Publishing.

Investigating validity evidence of the Satisfaction with Life Scale adapted for Children⁶

Measuring children's subjective well-being is important for a number of reasons. For example, it can increase our understanding of children's subjective well-being and its correlates; it is a critical component of monitoring children's subjective well-being; and it can provide useful information for improving children's subjective well-being (cf. Ben-Arieh & Frones, 2007). The strategy of assessing children's positive developmental outcomes, such as subjective well-being, provides a different focus and a complementary approach to strategies that focus on the assessment of problem behaviors or aspects of pathology of children and youth. Such an approach is in line with the *Positive Psychology* movement, which primarily examines how to foster quality of life and 'what makes life worth living' (Seligman & Csikszentmihalyi, 2000, p. 13), as well as with the positive youth development approach (e.g., Damon, 2004) and the developmental assets approach for youth (e.g., Mannes, Roehlkepartain, & Benson, 2005), which have a strength-oriented perspective. Specifically, the emphasis is not only on the absence of mental or behavioral disorders but also on 'optimal functioning' to depict the whole psychological spectrum.

In contrast to research with adults, the topic of subjective well-being and satisfaction with life has received less attention with regard to children and adolescents (Gullone & Cummins, 1999; Huebner, 1991a). It has been suggested that this situation is,

⁶ A version of this chapter is in press: Gadermann, A. M., Schonert-Reichl, K. A., & Zumbo, B. D. (in press). Investigating validity evidence of the Satisfaction with Life Scale adapted for Children. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*.

at least in part, related to the fact that instruments for assessing children's subjective well-being and satisfaction with life have been developed only relatively recently (see Huebner & Diener, 2008; Huebner, 1991b; Seligson, Huebner, & Valois, 2005).

This paper presents several psychometric validation analyses of a recently developed instrument on which children rate their satisfaction with life; namely, the Satisfaction with Life Scale for Children, which has been adapted from the Satisfaction with Life Scale (SWLS, Diener, Emmons, Larsen, & Griffin, 1985). Before the measure and the psychometric analyses are presented, a brief overview of research pertaining to children and adolescents' satisfaction with life is presented.

Life satisfaction refers to a "...cognitive judgmental evaluation of one's life" (Diener, 1984, p. 550) and is considered to be one component of subjective well-being besides positive and negative affect and domain satisfaction (Diener, Suh, Lucas, Smith, 1999). Life satisfaction has been shown to be an important predictor for a variety of factors in children and adolescents (cf. Huebner, 2004 and Gilman and Huebner, 2003 for reviews). Longitudinal studies with adolescents indicate that global life satisfaction predicts externalizing and internalizing behavior over a 2-year time period (Haranin, Huebner, & Suldo, 2007), and moderates the association between stressful life events and the development of externalizing behavior in the following year (Suldo & Huebner, 2004). Furthermore, it has been shown that global life satisfaction correlates over a 1-year time period with a variety of clinical variables (e.g., social stress, anxiety, and depression) and adaptive variables (relations with parents and interpersonal relations) for adolescents (Huebner, Funk III, & Gilman, 2000) and that adolescents' general life satisfaction predicts relational victimization and prosocial experiences one year later

(Martin, Huebner & Valois, 2008). In addition, cross-sectional studies have shown that low general life satisfaction is associated with adolescents' substance use, such as tobacco, alcohol and cocaine, and with the age of first substance use (Zullig, Valois, Huebner, Oeltmann, & Drane, 2001) as well as with suicide behaviors, suicide ideation, and decreased mental health (Valois, Zullig, Huebner, & Drane, 2004). Therefore, life satisfaction can be seen as an important protective psychological factor for children and adolescents, which is associated with positive growth and development.

There are several scales reviewed by Gilman and Huebner (2000) that assess life satisfaction in adolescents. These instruments are based on three conceptual models of life satisfaction, two being unidimensional (global and general life satisfaction) and one being multidimensional (Gilman & Huebner, 2000; Antaramian, Huebner, & Valois, 2008). The two unidimensional models assume that the items of the measure can be summed up, and that the total score can be used to represent respondents' life satisfaction. However, the unidimensional models differ insofar as that the global model measures life satisfaction with items that are context free (i.e., they do not refer to a specific life domain). Thus, the global model allows respondents to rate their life satisfaction according to their subjective criteria. In contrast, in the general model, items addressing different life domains are summed (thereby assuming that the items included in the instrument cover the important life domains of respondents' life satisfaction) and the overall score then reflects life satisfaction. Instruments based on the multidimensional model also address different life domains; however, in this case, the domains are used to create a profile of life satisfaction by only adding up items within each life domain/subscale.

One of the scales reviewed by Gilman and Huebner (2000) is the SWLS (Diener et al. 1985), which is based on the unidimensional and global conceptual framework. The SWLS is a commonly used measure developed for the assessment of adults' satisfaction with life. Previous studies using this measure have reported strong psychometric properties (Diener et al., 1985; Pavot, Diener, Colvin, & Sandvik, 1991; Pavot & Diener, 2008).

Although most commonly administered to adults, the SWLS has also been used with adolescents in different countries. Specifically, Neto (1993) administered a Portuguese version of the SWLS to an adolescent sample (age range 14-17, mean age of 14.7 years) in Portugal, and Leung and Leung (1992) and Shek (2007) report on Chinese versions of the SWLS⁷ in adolescent samples (age range 11-16 years (no mean age provided) and 11-19 years, mean age of 12.7 years, respectively) in Hong Kong.

In these studies, which used translated versions of the SWLS, the following psychometric properties were reported: The internal consistency ranged between .67 and .84 in the three studies. Furthermore, Shek (2007) reported the stability of the scale over a 1-year time period as .44. The authors reported evidence for the construct validity of the translated SWLS versions for their respective samples, showing that the scale had moderate to large correlations to psychological variables with which it was expected to be related, such as happiness, self-concept, loneliness, hopelessness, and self-esteem. Neto (1993) also reports that the scale was unidimensional according to a principal component analysis. These studies thus provide some evidence for the reliability and construct validity of the scores of the SWLS in samples with adolescents.

⁷ Shek (2007) adapted the response format from the original 7-point to a 6-point Likert-type scale.

We were interested in using the SWLS with a younger age group—namely, children aged 9-14, the so-called developmental periods of middle childhood and early adolescence⁸—as part of a larger project funded by United Way of the Lower Mainland in British Columbia, Canada, aiming at exploring the psychological and social world of children aged 9-14. However, the concern was that children of this age group might have difficulties understanding the SWLS in its original form. Therefore, the SWLS was adapted for children (the SWLS-C) by three researchers working in the area of children’s social-emotional development by changing the wording of the items and the response format to make it more understandable for children.

It is of primary importance to study children’s development during the periods of middle childhood (ranging approximately from age 6 to 10⁹) and early adolescence (ranging approximately from age 11 to 14), as these are critical developmental periods that are marked by biological, cognitive, and social changes and transitions (Eccles, 1999; Carnegie Council on Adolescent Development, 1995). Specifically, during this time children can develop important competencies and self-confidence, which have long-term implications for their later development (Huston & Ripke, 2006). Although this is a time of positive growth for the majority of children (Eccles, 1999), it is also a period of vulnerability or risk, and an increasing number of children experience mental health problems, such as depression or anxiety, during this time, which can have long-term effects for the children (Eccles, Lord, & Buchanan, 1996; Carnegie Council on

⁸In the following when the term children is used, I refer to individuals in the developmental periods of middle childhood and early adolescence.

⁹Slightly different age periods are used to talk about middle childhood, e.g. Collins (1984) refers to the period between 6 and 12 years.

Adolescent Development, 1995). This illustrates the importance of investigating children's well-being and life satisfaction during this developmental period.

It needs to be pointed out that Huebner (1991a) developed the Students' Life Satisfaction Scale (SLSS) for the assessment of children's global life satisfaction, which is "formulated from the work of Diener and his colleagues" (Gilman & Huebner, 1997, p. 231). The SLSS consists of seven items, and its scores have been shown to be reliable, and evidence of construct validity has been provided (e.g., Huebner & Alderman, 1993). However, the advantage of adapting the SWLS for children (and staying close to the original items of the SWLS) is that, if the scales are shown to be measurement equivalent across children and adults, this will facilitate (i) theoretical comparisons between the literature on children and adults when using the SWLS and the SWLS-C, and (ii) research on the change of satisfaction with life over time by using the SWLS-C for children and the SWLS for adults in longitudinal studies.

Purpose of the paper

The purpose of this paper is to examine evidence for the validity of the inferences from the scale scores of the SWLS-C. It is recognized in contemporary validity theory that validity is not the property of a test or scale, but rather of the inferences based on the responses (e.g., Zumbo, 2007a). In the following, when the term validity is used, I refer to the inferences from the scores. According to Messick (1995, p. 743), "construct validity comprises the evidence and rationales supporting the trustworthiness of score interpretation in terms of explanatory concepts that account for both test performance and score relationships with other variables". In the present study, I investigated several

aspects of the construct validity of the SWLS-C. Specifically, I investigated (i) the factor structure and internal consistency of the SWLS-C, (ii) whether the SWLS-C measures satisfaction with life in the same way for different groups of children (with regard to gender, first language learned at home—English (non-ESL) versus other language(s) than English (ESL)—and across different grades) using differential item functioning, (iii) whether there are significant group differences in the SWLS-C score with regard to gender, grade, and first language background (i.e., ESL versus non-ESL), and (iv) whether the SWLS-C relates to other constructs (namely, optimism, self-concept, self-efficacy, depression, empathic concern and perspective taking) as predicted by previous research.

With regard to the factor structure, I hypothesized the scores of the SWLS-C to be unidimensional based on studies that used the SWLS with adolescents (Neto, 1993) and adults (Pavot & Diener, 1993). Regarding group differences on the SWLS-C based on gender, grade, and first language background, I hypothesized—based on previous findings in this area (cf. Gilman & Huebner, 2003)—that these are either not statistically significant or of small effect size (cf. Cohen, 1988). In terms of convergent and discriminant evidence, I hypothesized the SWLS-C to have the following relationships with these constructs. Due to previous research indicating relationships between satisfaction with life in children/adolescents and measures of depression¹⁰ (Huebner & Alderman, 1993; Piko, 2006), optimism (Lucas, Diener, & Suh, 1996 with university students), general self-concept (Dew & Huebner, 1994; Huebner, 1991a), and self-efficacy (Neto, 1993), I expected relationships of medium or large effect size between the

¹⁰ The correlation between life satisfaction and depression was negative, whereas the other ones were positive.

SWLS-C and scales assessing these variables (indicating convergent evidence). In contrast, I hypothesized correlations of small effect size with scales assessing two aspects of empathy, namely empathic concern and perspective taking, thereby indicating evidence for discriminant validity (cf. Grühn, Rebucal, Diehl, Lumley, Labouvie-Vief, 2008 with a sample of diverse age groups). Specifically, I was interested in the pattern of the convergent and discriminant coefficients, in terms of a relative comparison and hypothesized that the convergent coefficients would be statistically significantly larger than the discriminant ones.

Method

Participants

Ethics approval was obtained from the University of British Columbia (see Appendix A) and the school boards of the seven school districts that were contacted. The stratified random sample consisted of 1266 students from 23 schools in these school districts from Vancouver and the Lower Mainland, British Columbia, Canada. Stratification was conducted according to neighborhood-level vulnerability rates for children's development, as reported by the Human Early Learning Partnership (see Kershaw, Irwin, Trafford, & Hertzman, 2005; www.earlylearning.ubc.ca). The vulnerability rates were determined according to the Early Development Instrument (EDI; Janus & Offord, 2007), a measure via which Kindergarten teachers assess their Kindergarten children's developmental status as reflected in their school readiness in five developmental domains¹¹. For the sampling strategy, schools located within school districts in Vancouver and the Lower Mainland were randomly selected and approached, stratified by 'high', 'medium' and 'low' vulnerability rates to have a diverse representation of children in the sample. The sampling strategy was chosen to obtain a representative sample with regard to children's developmental outcomes (at the aggregate neighborhood level). At the same time, the sampling strategy maximized the likelihood of capturing neighborhoods from a representative range of socioeconomic status, as Kershaw et al. (2005) report high correlations between vulnerability rates and socioeconomic status at the neighborhood level. Finally, randomly choosing schools from

¹¹ A child that receives ratings below a specific cut-off score for one or more developmental domains is considered developmentally vulnerable, and neighborhood-level vulnerability rates are determined by calculating the percentage of children within a given neighborhood that are considered as developmentally vulnerable.

school districts throughout Vancouver and the Lower Mainland maximized the representativeness of the sample with regard to certain neighborhood characteristics (e.g., urban and suburban neighborhoods).

Schools were contacted in the seven school districts and 23 schools provided their consent for participation. The participation rate of students was 86%.

Thirty participants did not complete several scales of the questionnaire package, including the SWLS-C, and three participants had missing data on one item of the SWLS-C; these participants were excluded from the analysis to have complete data on all five items of the SWLS-C. Therefore, the sample for the analysis consisted of 1233 students (48% girls) with a mean age of 11.7 years (ranging from 9.2 years to 14.0 years, with a standard deviation of 1.0 year). The students attended grades 4 to 7 (8% grade 4, 21% grade 5, 33% grade 6, and 38% grade 7). With regard to first language learned at home (a proxy for children's immigrant background), students reported up to four languages. Of the 1225 students who answered this item, 55% reported having learned English only as a first language at home, 37% participants reported having learned one (or more) other language(s) than English as their first language (most common were Chinese¹², Punjabi, and Vietnamese), and 8% participants reported having learned English and another language (for the languages besides English, Chinese, Punjabi, and Vietnamese were also most common).

Teachers were asked to provide information about students' ethnicity and whether they had special needs. Of the 1195 students that were rated on their ethnicity, 1111 were rated as having one ethnicity with 48% White, 28% Asian, 10% Indo-Canadian, 4%

¹² The language label 'Chinese' consisted of children responding with Chinese, Cantonese, Mandarin, and Taiwanese.

Filipino, 2% Latin, 2% First Nations, 2% Black, and 4% with another ethnicity. Eighty-one students were rated as having two ethnicities (most of them as White and Asian, and White and First Nations) and 3 students were rated as having three ethnicities. Of the 1203 students that were rated by teachers with regard to special needs, 83% did not have special needs and 17% had special needs.

Measures

Satisfaction with Life. Satisfaction with life was assessed with the SWLS-C. As mentioned in the introduction, this scale was adapted from the SWLS (Diener et al., 1985), a 5-item instrument that assesses global life satisfaction. On the SWLS, respondents are asked to respond on a 7-point Likert-type scale, which ranges from strongly disagree (1) to strongly agree (7). The total scale score can therefore range from 5 to 35. There are no reverse scored items. Previous studies provide evidence supporting the internal consistency, stability, concurrent and predictive validity of the SWLS (Diener et al., 1985; Pavot et al., 1991). Three researchers in the area of children's socio-emotional development adapted the items of the SWLS for children by changing the wording of the item stem and response format to make it more understandable for children. At the same time, the goal was to stay very close to the original version of the SWLS. For example, the item wording of item 4 stayed the same. Furthermore, the number of response options was reduced to 5-point Likert-type scale, a format that is commonly used in research with children (cf. Developmental Studies Center, 2003). The response format for the SWLS-C ranges from disagree a lot (1) to agree a lot (5). The total scale score can therefore range from 5 to 25. In order to get a sense of the reading

level necessary to comprehend the SWLS-C, the Flesch-Kincaid grade level score was computed, which provides information about the reading level of the text based on a grade-school level. This resulted in a score of 1.9, indicating that an average child with grade 2 reading skills can understand the scale. The SWLS-C is provided in Table 2.1. The reliability of the SWLS-C will be discussed in the results section.

Table 2.1

The Satisfaction with Life Scale Adapted for Children.

For each of the following statements, please circle the number that describes you the best. Please read each sentence carefully and answer honestly. Thank you.

	Disagree a Lot	Disagree a Little	Don't Agree or Disagree	Agree a Little	Agree a Lot
1. In most ways my life is close to the way I would want it to be.	1	2	3	4	5
2. The things in my life are excellent.	1	2	3	4	5
3. I am happy with my life.	1	2	3	4	5
4. So far I have gotten the important things I want in life.	1	2	3	4	5
5. If I could live my life over, I would have it the same way.	1	2	3	4	5

Optimism. The optimism subscale of the short form of the Resilience Inventory (Song, 2003) was used. This subscale consists of nine items, of which five are reverse

scored. Song used a 5-point Likert-type response scale ranging from always false to always true, which was adjusted by our research team to be consistent with several other measures that were part of the data collection so that it ranged from not at all like me (1) to always like me (5). Song provided evidence for the reliability (internal consistency and test-retest reliability over a three week-interval) and construct validity of this inventory. In the present sample, the subscale assessing optimism had a Cronbach's coefficient alpha of .79. I also computed ordinal alpha, an estimate of internal consistency that was specifically developed for items with Likert-type response formats, taking into account the ordinal nature of the data (Zumbo, Gadermann, & Zeisser, 2007). Ordinal coefficient alpha was .83.¹³

Self-efficacy. To assess self-efficacy, another subscale (consisting of eight items) from the short form of the Resilience Inventory (Song, 2003) was used. As for the optimism subscale, the response options were adjusted to range from not at all like me (1) to always like me (5). Cronbach's alpha in the present data was .72 and ordinal alpha was .76.

Self-concept. In order to measure general self-concept, the self-concept subscale of the Marsh Self Description Questionnaire for preadolescents (Marsh, 1988) was used. This subscale consists of eight items that are rated on a 5-point Likert-type scale ranging from never (1) to always (5). Evidence for the reliability and validity of this scale has

¹³ For the 5- and 4-point Likert-type response format of the SWLS-C and the other measures, ordinal coefficient alpha is the coefficient of choice but Cronbach's coefficient alpha is also reported due to its familiarity to most researchers.

been provided by Marsh (1988, 1990). In this sample, Cronbach's coefficient alpha was .87 and ordinal coefficient alpha was .90.

Depression. Depression was assessed with one subscale of the Seattle Personality Questionnaire (Rains, 2003). This subscale consists of 11 items that are answered on a 4-point Likert-type scale ranging from not at all (1) to always (4). Rains (2003) and Greenberg and Lengua (1995) provide evidence for the reliability and construct validity of this instrument. In the present study, Cronbach's coefficient alpha was .85 and ordinal coefficient alpha was .89.

Empathy. Two aspects of empathy were assessed with two subscales of the Interpersonal Reactivity Index (Davis, 1980), namely empathic concern and perspective taking. The subscales consist of seven items each that are answered on a 5-point Likert-type scale with the response options not at all like me (1) to always like me (5). Evidence for the reliability and construct validity of this measure has been provided by Davis (1980, 1983). Cronbach's coefficient alpha was .85 for the empathic concern subscale and .79 for the perspective taking subscale, and ordinal coefficient alpha was .87 and .82, respectively.

Procedure

The present data were collected as part of a larger research project that investigated after school activities of students in grade 4 to 7 in Canada. During the first visit in the classrooms, research assistants explained the purpose and provided details about the study. Permission slips/parent consent forms (see Appendix A) were handed

out and the children were asked to bring these to their parents and return them signed (either yes or no) to the classroom. Translated versions of the parent consent form were also available. As an incentive for the children to return the parent consent form, they were given an eraser upon returning the form. During the second visit of the research assistants, the data collection took place. Children whose parents agreed to have them participate in the study were asked for their assent (see Appendix A for the student assent form). Students that provided their assent could participate in the study. The data collection took place in the general classrooms and was conducted and observed by the research assistants. After the data collection, students were given a small gift, and a pizza party for each class was organized.

Results

Distributions and intercorrelations of the items of the SWLS-C

The distributions of the children's responses to the five items of the SWLS-C are provided in Table 2.2. All items were negatively skewed (ranging from -1.29 to -.40) as commonly found in research on satisfaction with life in North America (e.g., Johnson & Krueger, 2006), indicating that most respondents were satisfied rather than dissatisfied with their lives.

Table 2.2

Item Response Percentages of the SWLS-C Items

SWLS-C	Disagree a Lot	Disagree a Little	Don't Agree or Disagree	Agree a Little	Agree a Lot
Item 1	5.1%	9.8%	16.3%	40.1%	28.7%
Item 2	4.5%	9.9%	18.7%	35.6%	31.3%
Item 3	2.4%	5.4%	12.0%	30.5%	49.7%
Item 4	3.8%	8.7%	17.0%	34.8%	35.7%
Item 5	12.6%	14.9%	19.4%	26.9%	26.2%

Furthermore, the inter-item correlations of the SWLS-C were investigated. The scale has 5 response options, and is therefore considered as ordinal. Thus, I used the polychoric correlation matrix, which is recommended for ordinal data. This procedure

provides an estimate of the correlation between two unobserved continuous variables, given observed ordinal data. (The assumption is that the observed values are due to an unobserved underlying continuous distribution; Flora & Curran, 2004). As shown in Table 2.3, the magnitude of the intercorrelations of the SWLS-C ranged between .56 and .75, indicating that all items are highly related to one another.

Table 2.3

Intercorrelations between the Items of the SWLS-C Using the Polychoric Correlation Matrix

	1	2	3	4	5
Item 1	-				
Item 2	.69	-			
Item 3	.69	.75	-		
Item 4	.56	.57	.60	-	
Item 5	.61	.61	.65	.56	-

Factor analysis

In the following, the results of the factor analysis are reported. As the data are considered ordinal, I used the polychoric correlation matrix (with the software PRELIS) in the factor analysis using MINRES as the estimation method. As Flora and Curran (2004) remind us, the assumptions when using the polychoric correlation are about the

hypothesized underlying item response distributions, which are not testable from data. But their simulation study showed that, if there is moderate nonnormality of the latent response distributions, the estimation of the polychoric correlation is robust. In order to aid the determination of the number of factors, I first ran a principal components analysis on the polychoric correlation matrix. The first eigenvalue was 3.6 and explained 71% of the variance; the second eigenvalue was 0.48 explaining 9.6% of the variance. The fact that there was only one eigenvalue larger than 1 and the large ratio between the first and second eigenvalue indicates that the scale is unidimensional. A parallel analysis (cf. Russell, 2002) also identified one factor (the second random factor had an eigenvalue of 1.05, which is higher than the second eigenvalue in this dataset of 0.48). Having determined the number of factors to retain based on the components analysis, I then turned to the factor analysis. Fitting a one-factor model, I investigated the residuals between the observed and reproduced correlations as another way of assessing statistical model fit. The absolute values of the residuals ranged between .0007 and .034. According to McDonald (1985), residuals should be below .10 for a good model fit; therefore, this also supported the assumption of unidimensionality of the SWLS-C in the present data. The factor loadings of the five items of the SWLS-C, as determined by the factor analysis, were all high (equal to or above .70), as shown in Table 2.4.

Table 2.4

Factor Loadings of the Items of the SWLS-C

Item	Factor 1
Item 1	.81
Item 2	.84
Item 3	.87
Item 4	.70
Item 5	.76

As the results of the principal components and factor analysis indicate unidimensionality of the SWLS-C, the internal consistency was computed on the basis of the scores of the five items. Cronbach's coefficient alpha was .86 and ordinal coefficient alpha was .90.

In addition, I calculated the corrected item-total correlations of the SWLS-C using the polyserial correlation matrix, and also the ordinal alphas if item deleted (see Table 2.5). The corrected item-total correlations ranged between .63 and .77; and the ordinal alphas if item deleted ranged between .86 and .89. This indicates that, in the present sample, all items related highly to the corrected total scale and that deleting any of the items would slightly decrease the reliability of the full scale.

Table 2.5

Corrected Item-Total Correlations Using the Polyserial Correlation and Ordinal Alpha if Item Deleted of the SWLS-C Items

	Corrected item-total correlation	Ordinal alpha if item deleted
Item 1	.72	.87
Item 2	.74	.86
Item 3	.77	.86
Item 4	.63	.89
Item 5	.69	.88

Graphical representation of the items using nonparametric item response theory

For a graphical representation of the item response functions for the five items, nonparametric item response theory (NIRT) was used. The nonparametric approach was chosen because of its flexibility and adaptive nature to the data (Ramsay, 1997).

Specifically, in NIRT, the form of the item response function (IRF; also called item characteristic curve), which relates the probability of an item response to corresponding values on the latent variable, is determined by the data. Thus, NIRT uses the existing data optimally whereas, in parametric IRT, the IRF is a pre-determined function of a model, such as a logistic function or normal ogive (Ramsay, 1997; Zumbo, Gelin & Hubley, 2002). That is, the nonparametric IRF is allowed to take any form, as opposed to the IRF in parametric IRT.

As Zumbo et al. (2002) state, Jim Ramsay's NIRT, which is implemented in the software TestGraf, presents a useful method for nonparametric item response modeling. Ramsay's (2000) approach uses a class of nonparametric regression methods that partitions the continuum of variation into intervals, within which the likelihood of an item response is estimated. Figure 2.1 shows the nonparametric IRFs of the five items with regard to their expected scores. The x-axis represents the expected scores along the latent variable, and the y-axis represents the likely item response at a given level of the latent variable. On top of each graph the percentiles of the score distributions are provided, indicated by dashed vertical lines. This information allows the reader to take into account the distribution of the scores when interpreting the IRFs. The small solid vertical lines along the IRFs provide the 95% pointwise confidence limits of the IRFs. Given the short length of the scale, I also investigated the rest scores (the total scale score minus the item under investigation) and found that the plots were very similar; hence, the commonly used expected scores are reported. The IRFs show that all five items performed well. Specifically, the IRFs start in the lower left corner and increase monotonically to the upper right corner, with relatively steep slopes indicating that the items are discriminating well along the entire continuum of the latent variable among students. Few students scored 10 or less (i.e., 5%), which is reflected in the wider confidence intervals at the lower end of the score continuum.

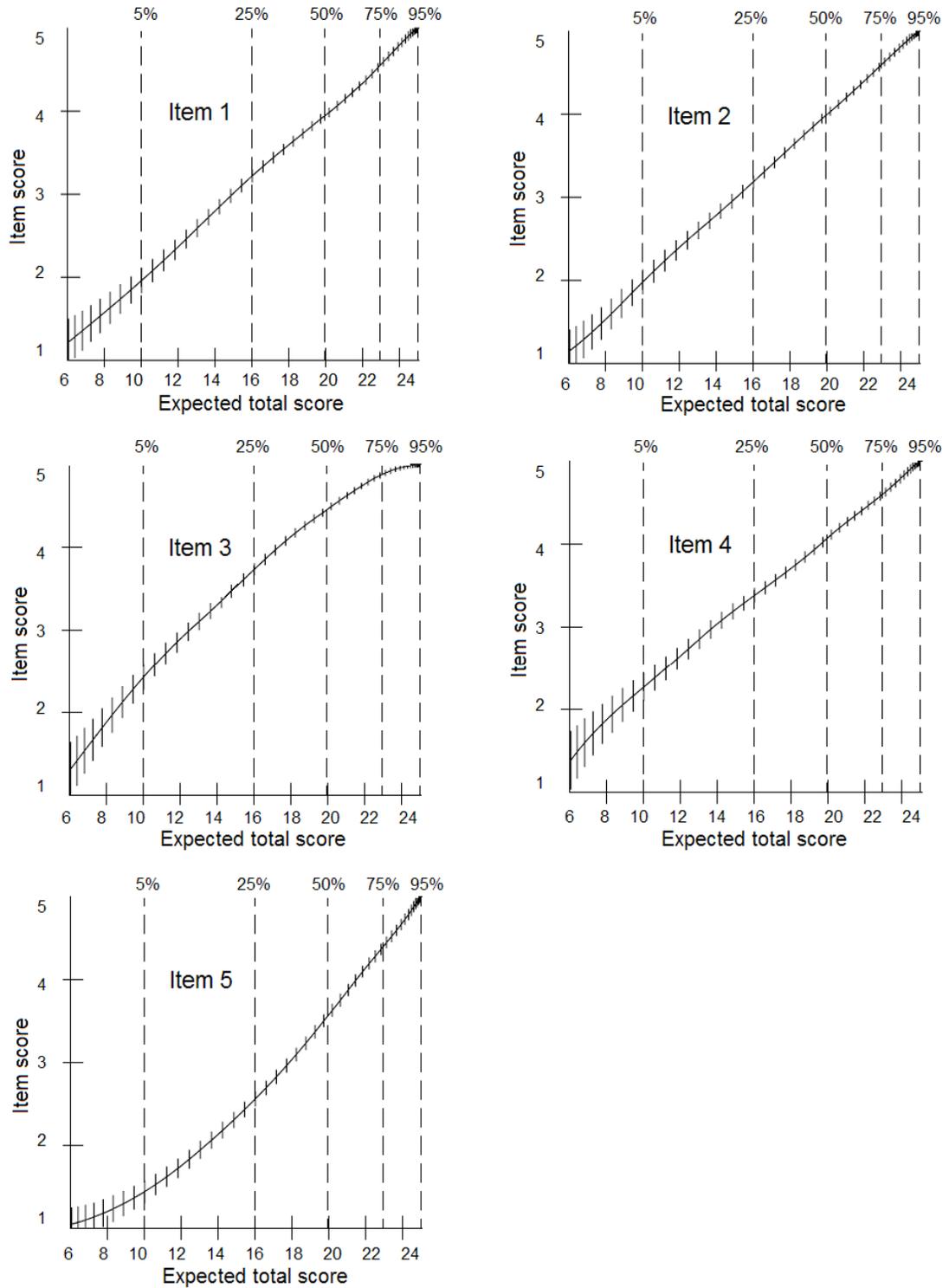


Figure 2.1. Nonparametric item response functions of the five items of the SWLS-C.

In addition, I investigated the measurement precision of the SWLS-C across the continuum of the scale by estimating the conditional reliability function¹⁴ of the SWLS-C with NIRT.

Figure 2.2 shows that the reliability varied between approximately .75 and .87, indicating that the SWLS-C measures most precisely for students who scored between 20 and 24 on the total score. As mentioned above, few students scored 10 or less; so this section of the graph is not well-specified and should not be considered as accurate.

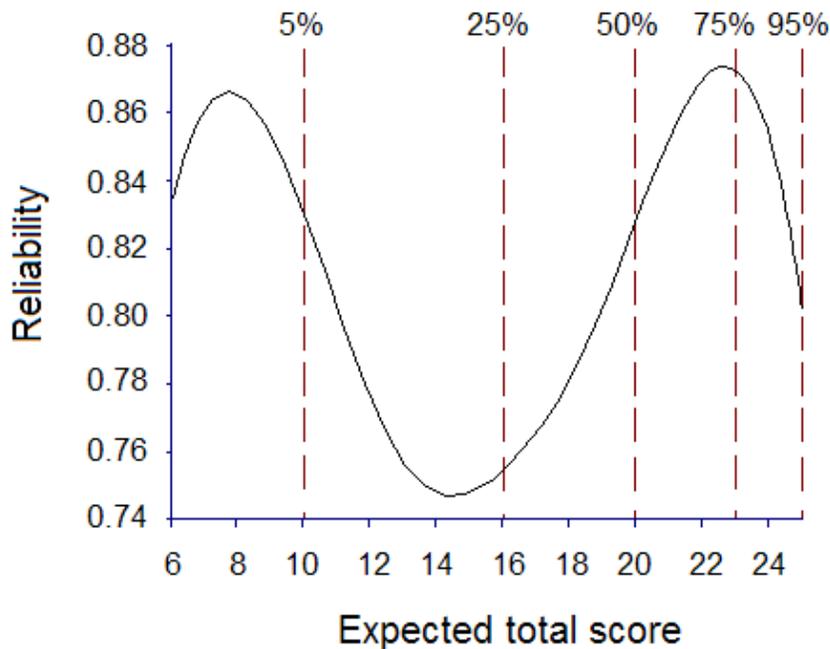


Figure 2.2. Conditional reliability of the SWLS-C plotted against the expected total score.

¹⁴ It should be noted that the conditional standard error of measurement conveys similar information; generally, where the reliability is the highest the standard error of measurement is the lowest.

Differential item and scale functioning using ordinal logistic regression

The next step was to investigate whether the SWLS-C measures satisfaction with life in the same way for different groups of children, namely with regard to gender (boys versus girls), first language learned at home (ESL versus non-ESL¹⁵), and different grades (grades 4 to 7). Thereby, the question of measurement invariance was addressed.

Differential item functioning (DIF) occurs “when examinees from different groups show differing probabilities of success on (or endorsing) the item after matching on the underlying ability that the item is intended to measure” (Zumbo, 2007b, p. 12). However, the presence of DIF does not necessarily reflect the existence of bias. Rather, in the presence of DIF, one then needs to examine whether this reflects item impact (i.e., the groups’ differing probabilities of endorsing an item result from true differences with regard to the construct that is measured by the item) or item bias (i.e., the groups’ differing probabilities of endorsing an item result from some construct-irrelevant characteristic of the item or the assessment situation) (Zumbo, 1999). To investigate the sources of DIF, one can, for example, ask subject matter experts whether they see the source of DIF as item bias or item impact. There are several statistical techniques available to investigate DIF. I used ordinal logistic regression because it can be utilized for the analysis of ordinal item response formats and in the context of moderate-to-small-scale testing (with 500 or less participants per group and less than 50 items in the scale, Witarsa, 2003), which is the case in the present study. As the matching criterion, the corrected total scale (that is, the total scale score minus the item score under investigation) was used.

¹⁵ Children who reported that they first learned English and another language at home (8%) were excluded from the analysis.

None of the items showed statistically significant uniform or non-uniform DIF with regard to gender, ESL versus non-ESL, and across different grades. Furthermore, following the effect size criteria by Jodoin and Gierl (2001), none of the items displayed substantial uniform or non-uniform DIF with regard to gender, ESL versus non-ESL, and across different grades.

Besides investigating DIF, one can also investigate differential scale functioning, a technique first described in Guhn, Gadermann, and Zumbo (2007). This examines whether small DIF of several items can add up to a substantial effect at the scale level. Differential scale functioning was investigated with regard to gender, ESL versus non-ESL, and across grades. There was no differential scale functioning across any of these groups, i.e., the graphs for the respective groups in the analysis were nearly identical. Figure 2.3 provides an illustration of this with regard to differential scale functioning for gender. The x-axis represents the SWLS-C total score¹⁶ and the y-axis represents the average composite model predicted scale score—i.e., the predicted scale score is obtained for each item and then summed to a scale-level predicted score. These curves are similar in purpose to test characteristic curves of item response theory. The graphs of boys and girls are nearly identical; it is therefore concluded that there is neither DIF nor differential scale functioning present for boys and girls on the SWLS-C.

¹⁶ For the analysis of each item, corrected item-total scale was used.

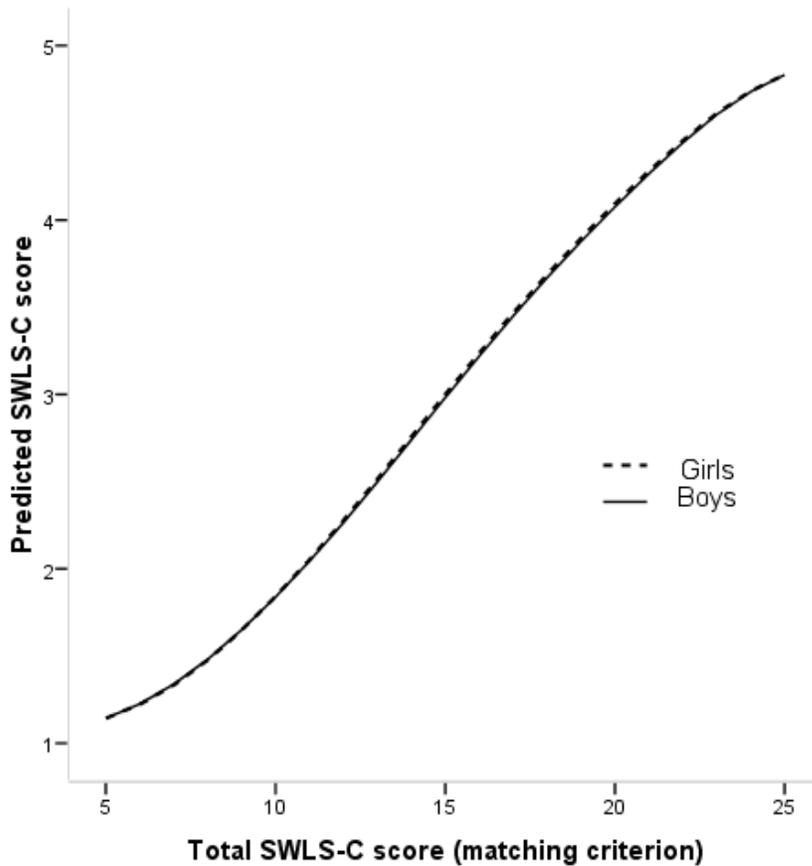


Figure 2.3. Differential scale functioning for the SWLS-C.

Demographic differences on the SWLS-C

In order to test for main effects and interactions on the SWLS-C with regard to these groups (i.e., gender, ESL versus non-ESL, and grades), a 2x2x4 factorial ANOVA was conducted. First, the assumptions of equal variances and normality were tested. The Levene's test of equality of error variances was non-significant $F(15, 1114) = 1.62, p = .06$. There were very few outliers (all of which were real values), i.e., 10 out of the 1130 cases had absolute values of the standardized or studentized residuals greater than 3; given the large sample size, all data were retained for the analysis. There was no statistically significant main effect for gender. The results indicate a statistically

significant main effect for ESL versus non-ESL ($F(1, 1114) = 6.37, p = .01, \eta_p^2 = .006$) with non-ESL students reporting higher life satisfaction than ESL students. Furthermore, there was a statistically significant main effect for grades ($F(3, 1114) = 8.13, p < .0001, \eta_p^2 = .02$). A posthoc test, Tukey's HSD, showed that the significant differences were between grades 7 and 4 and grades 7 and 5. That is, on average, students in grade 7 reported less life satisfaction than students in grades 4 and 5 (there was actually a trend across the four grades that with increasing grades/age students reported less life satisfaction). In terms of the interactions, only the interaction grade-by-gender was statistically significant ($F(3, 1114) = 3.2, p = .02, \eta_p^2 = .009$) indicating that girls reported slightly higher life satisfaction in grades 4, 5, and 7, yet, the opposite was the case for grade 6. However, it has to be pointed out that the effect sizes for these effects were all small (cf. Cohen, 1988).

Concurrent and discriminant validity evidence for the SWLS-C

The discriminant and convergent zero-order correlations between the SWLS-C and the other scales are provided in Table 2.6. In terms of missing data, participants had to have responded to at least 80% of the items of a scale, and then the average was taken. Treatment of missing data for the correlations then was listwise deletion. This resulted in a sample of 1203 students for the analysis. Listwise deletion allowed us to make comparisons of correlations based on the same data.

As predicted, the SWLS-C showed statistically significant correlations to optimism ($r = .65, p < .001$), self-concept ($r = .57, p < .001$), self-efficacy ($r = .52, p < .001$) and depression ($r = -.46, p < .001$) with large and medium effect sizes (Cohen,

1992¹⁷). In addition, the SWLS-C showed statistically significant correlations to empathic concern and perspective taking ($r = .27, p < .001$; $r = .29, p < .001$, respectively) with small (but close to medium) effect sizes.

The convergent correlations were all larger in terms of absolute values than the discriminant ones. The differences between the four convergent to each of the two discriminant correlations were tested with a t-test of the difference between dependent correlations. In all cases, as hypothesized, the two correlations for the discriminant variables were statistically significantly smaller than the convergent ones. The eight t-statistics ranged from 4.89 to 13.73 (with 1200 degrees of freedom).

¹⁷ Cohen suggested a rule of thumb for interpreting effect sizes according to which correlations of .10 are considered to be of small, .30 to be of medium, and .50 to be of large effect size.

Table 2.6

Intercorrelations between the SWLS-C and Convergent and Discriminant Measures

	SWLS-C	Optimism	Self- concept	Self- efficacy	Depression	Empathic concern	Persp. taking
SWLS-C	-						
Optimism	.65**	-					
Self-concept	.57**	.60**	-				
Self-efficacy	.52**	.56**	.73**	-			
Depression	-.46**	-.61**	-.38**	-.29**	-		
Empathic concern	.27**	.25**	.40**	.47**	-.03	-	
Perspective taking	.29**	.29**	.46**	.55**	-.07**	.66**	-

** p < .01.

Discussion

This study introduced the SWLS-C, a scale based on the SWLS, to assess satisfaction with life in children and provided information with regard to the psychometric properties of the SWLS-C. The presented results provide some evidence for the validity of the inferences of the SWLS-C scores for children in grades 4 to 7. Specifically, the results indicate that the SWLS-C (i) has a unidimensional factor structure, (ii) has high internal consistency, (iii) performs in the same way for different groups of children (with regard to gender, first language learned at home, and across grades) as investigated with a DIF and a differential scale functioning analysis, (iv) has statistically non-significant or small associations with demographic variables, and (v) has relations to other scales as expected based on previous research. Specifically, children who reported high satisfaction with life tended to report a positive self-concept, scored high in self-efficacy and optimism, and scored low in depression. As was hypothesized, the correlation of the SWLS-C to these measures was of medium or large effect size, which is in accordance with previous research (e.g., Lucas et al., 1996; Huebner, 1991a; Neto, 1993; Piko, 2006). Furthermore, the correlations of the SWLS-C to aspects of empathy (empathic concern and perspective taking) were of small (but close to medium) effect size (cf. Grühn et al., 2008). The results also support the stated hypotheses in terms of relative magnitude of association. The correlations of the SWLS-C to the hypothesized convergent measures were all larger than the ones to the hypothesized discriminant measures and this difference was statistically significant; this finding is therefore interpreted as evidence for convergent and discriminant validity.

With regard to the relationship between children's scores on the SWLS-C and demographic variables, these findings are in line with previous research that has shown that these relationships are 'modest at best' (Gilman & Huebner, 2003, p. 196). Consistent with past

research, I did not find any gender differences (e.g., Seligson, Huebner, & Valois, 2005; Dew & Huebner, 1994). Although several studies have indicated that children and adolescents' life satisfaction is not related to age/grade (e.g., Dew & Huebner, 1994; Seligson et al., 2005), other studies have shown a statistically significant negative relationship between life satisfaction and increased age/grade (Martin et al., 2008; Haranin et al., 2007), which is in line with the presented finding (of small effect size). Furthermore, there was a statistically significant difference for ESL versus non-ESL children (a proxy for immigrant background), indicating that non-ESL children reported higher life satisfaction (again, the effect had a small effect size).

Overall, the results provide preliminary validity evidence that indicates the appropriateness for using the SWLS-C to assess satisfaction with life in children aged 9-14. This study was a first step in the validation of the SWLS-C. However, further validation research is necessary, because validation is an ongoing process (Huble & Zumbo, 1996). For example, it should be investigated to what extent the presented findings can be replicated and generalized to other contexts and populations (e.g., children of different cultural background or age) (AERA, APA, & NCME, 1999). Furthermore, the findings are limited as only self-report measures were used; it would be of interest to utilize multi-method assessments, for example, by also including teacher- and parent-reports. In addition, future research should investigate the substantive aspect of construct validity (Messick, 1995), with regard to identifying the cognitive processes that children employ when answering the items of the SWLS-C, for example, using think-aloud protocols (cf. Ericsson & Simon, 1980).

As was suggested in the introduction, the SWLS-C can potentially be used to make direct comparisons between research findings with children and with adults (using the SWLS) as well as in longitudinal studies using these measures. However, before that it is attempted, the

measurement invariance of these scales needs to be investigated, which is another future direction for research using the SWLS-C.

Notwithstanding these limitations, the presented results provide evidence for the validity of the SWLS-C scores for children in grades 4 to 7. Provided that future validation research supports the presented psychometric findings, the SWLS-C can be used to complement other measures of mental health for research purposes (e.g., to further investigate the construct and correlates of satisfaction with life in children), as well as for more applied purposes, such as program evaluation or informing social policy.

References

- AERA, APA, & NCME (1999). Validity. *Standards for educational and psychological testing*. (Chap. 1, pp. 9-24). Washington, DC: AERA, APA, & NCME.
- Antaramian, S. P., Huebner, E. S., & Valois, R. F. (2008). Adolescent life satisfaction. *Applied Psychology: An International Review*, *57*, 112-126.
- Ben-Arieh, A., & Frones, I. (2007). Indicators of children's well-being: Concepts, indices and usage. *Social Indicators Research*, *80*, 1-4.
- Carnegie Council on Adolescent Development (1995). *Great transitions: Preparing adolescents for a new century (Executive Summary)*. New York: Carnegie. Retrieved from http://www.carnegie.org/sub/pubs/reports/great_transitions/gr_exec.html#intro
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.
- Collins, W. A. (Ed.) (1984). *Development during middle childhood: The years from six to twelve*. Washington, DC: National Academy Press.
- Damon, W. (2004). What is positive youth development? *The ANNALS of the American Academy of Political and Social Science*, *591*, 13-24.
- Davis, M. A. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113-126.
- Developmental Studies Center (2003). Retrieved November, 23, 2008, from http://www.devstu.org/pdfs/cdp/DSC_ElemSch_scales.pdf

- Dew, T., & Huebner, E. S. (1994). Adolescents' perceived quality of life: An exploratory investigation. *Journal of School Psychology, 32*, 185-199.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542-575.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*, 71-75.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*, 276-302.
- Eccles, J. S. (1999). The development of children ages 6 to 14. *Future of children, 9*, 30-44.
- Eccles, J. S., Lord, S., & Buchanan, C. M. (1996). School transitions in early adolescence: What are we doing to our young people? In J. L. Graber, J. Brooks-Gunn, & A. C. Petersen (Eds.), *Transitions through adolescence: Interpersonal domains and context* (pp. 251-84). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466-491.
- Gilman, R., & Huebner, E. S. (1997). Children's reports of their life satisfaction: Convergence across raters, time and response formats. *School Psychology International, 18*, 229-243.
- Gilman, R., & Huebner, E. S. (2000). Review of life satisfaction measures for adolescents. *Behaviour Change, 17*, 178-83.
- Gilman, R., & Huebner, S. (2003). A review of life satisfaction research with children and adolescents. *School Psychology Quarterly, 18*, 192-205.

- Greenberg, M., & Lengua, L. (1995). *Scale construction of the Seattle Personality Inventory. (Fast Track Project Technical Report)*. Seattle, Washington: University of Washington.
- Grühn, D., Rebucal, K., Diehl, M., Lumley, M., & Labouvie-Vief, G. (2008) Empathy across the adult lifespan: Longitudinal and experience-sampling findings. *Emotion, 8*, 753-765.
- Guhn, M., Gadermann, A. M., & Zumbo, B. D. (2007). Does the EDI measure school readiness in the same way across different groups of children? *Early Education and Development, 18*, 453-472.
- Gullone, E., & Cummins, R. A. (1999). The comprehensive quality of life scale: A psychometric evaluation with an adolescent sample. *Behaviour Change, 16*, 127-139.
- Haranin, E. C., Huebner, E. S., & Suldo, S. M. (2007). Predictive and incremental validity of global and domain-based adolescent life satisfaction reports. *Journal of Psychoeducational Assessment, 25*, 127-138.
- Hubley, A. M., & Zumbo, B. D. (1996). A dialectic on validity: Where we have been and where we are going. *The Journal of General Psychology, 123*, 207-215.
- Huebner, E. S. (1991a). Initial development of the Students' Life Satisfaction Scale. *School Psychology International, 12*, 231-240.
- Huebner, E. S. (1991b). Further validation of the Students' Life Satisfaction Scale: The independence of satisfaction and affect ratings. *Journal of Psychoeducational Assessment, 9*, 363-368.
- Huebner, E. S. (2004). Research on assessment of life satisfaction of children and adolescents. *Social Indicators Research, 66*, 3-33.

- Huebner, E. S., & Alderman, G. L. (1993). Convergent and discriminant validation of a children's life satisfaction scale: Its relationship to self- and teacher-reported psychological problems and school functioning. *Social Indicators Research, 30*, 71-82.
- Huebner, E. S., & Diener, C. (2008). Research on life satisfaction of children and youth. In M. Eid, & R. J. Larsen (Eds.), *The Science of Subjective Well-Being* (pp. 376-392). New York: Guilford Press.
- Huebner, E. S., Funk III, B. A., & Gilman, R. (2000). Cross-sectional and longitudinal psychosocial correlates of adolescent life satisfaction reports. *Canadian Journal of School Psychology, 16*, 53-64.
- Huston, A. C., & Ripke, M. N. (2006). Middle childhood: Contexts of development. In A. C. Huston, & M. N. Ripke (Eds.), *Developmental Contexts in Middle Childhood: Bridges to Adolescence and Adulthood* (pp. 1-22). New York: Cambridge University Press.
- Janus, M., & Offord, D. (2007). Development and psychometric properties of the Early Development Instrument (EDI): A measure of children's school readiness. *Canadian Journal of Behavioural Science, 39*, 1-22.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error rate and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329-349.
- Johnson, W., & Krueger, R. F. (2006). How money buys happiness: Genetic and environmental processes linking finances and life satisfaction. *Journal of Personality and Social Psychology, 90*, 680-691.

- Kershaw, P., Irwin, L., Trafford, K., & Hertzman, C. (2005). *The British Columbia atlas of child development* (1st ed., Vol. 40). Vancouver, BC: Human Early Learning Partnership, Western Geographical Press.
- Leung, J., & Leung, K. (1992). Life satisfaction, self-concept, and relationship with parents in adolescence. *Journal of Youth and Adolescence, 21*, 653-665.
- Lucas, R. E., Diener, E., & Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality & Social Psychology, 71*, 616-628.
- Mannes, M., Roehlkepartain, E. C., & Benson, P. L. (2005). Unleashing the power of community to strengthen the well-being of children, youth, and families: An asset-building approach. *Child Welfare, 84*, 233-250.
- Marsh, H. W. (1988). *Self-Description Questionnaire: A theoretical and empirical basis for the measurement of multiple dimensions of preadolescent self-concept: A test manual and a research monograph*. San Antonio, Texas: The Psychological Corporation.
- Marsh, H. W. (1990). A multidimensional, hierarchical self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2*, 77-172.
- Martin, K., Huebner, E. S., & Valois, R. F. (2008). Does life satisfaction predict victimization experiences in adolescence? *Psychology in the Schools, 45*, 705-714.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, N.J.: Lawrence Erlbaum.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

- Miles, J., & Shevlin, M. (2004). *Applying regression and correlation: A guide for students and researchers*. London: Sage Publications.
- Neto, F. (1993). The Satisfaction with Life Scale: Psychometric properties in an adolescent sample. *Journal of Youth and Adolescence*, 22, 125-134.
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction with Life Scale. *Psychological Assessment*, 5, 164-172.
- Pavot, W., & Diener, E. (2008). The Satisfaction with Life Scale and the emerging construct of life satisfaction. *Journal of Positive Psychology*, 3, 137-152.
- Pavot, W., Diener, E., Colvin, C. R., & Sandvik, E. (1991). Further validation of the Satisfaction with Life Scale: Evidence for the cross-method convergence of well-being measures. *Journal of Personality Assessment*, 57, 149-161.
- Piko, B. F. (2006). Satisfaction with life, psychosocial health and materialism among Hungarian youth. *Journal of Health Psychology*, 11, 827-831.
- Rains, C. (2003). *Seattle Personality Questionnaire-Original*. (Fast Track Project Technical Report). Retrieved September, 2007, from: <http://www.fasttrackproject.org/>
- Ramsay, J. O. (1997). A functional approach to modeling test data. In W. J. van der Linden, & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 381-394). New York: Springer Verlag.
- Ramsay J.O. (2000). *TESTGRAF: A program for the graphical analysis of multiple choice test and questionnaire data*. McGill University. Retrieved November 2007, from <http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html>

- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, 28, 1629-1646.
- Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist*, 55, 5-14.
- Seligson, J. L., Huebner, E. S., & Valois, R. F. (2005). An investigation of a Brief Life Satisfaction Scale with elementary school children. *Social Indicators Research*, 73, 355-374.
- Shek, D. T. L. (2007). A longitudinal study of perceived parental psychological control and psychological well-being in Chinese adolescents in Hong Kong. *Journal of Clinical Psychology*, 63, 1-22.
- Song, M. (2003). *Two studies on the Resilience Inventory (RI): Toward the goal of creating a culturally sensitive measure of adolescence resilience*. Unpublished doctoral dissertation, Harvard University, MA.
- Suldo, S. M., & Huebner, E. S. (2004). Does life satisfaction moderate the effects of stressful life events on psychopathological behavior during adolescence? *School Psychology Quarterly*, 19, 93-105.
- Valois, R. F., Zullig, K. J., Huebner, E. S., & Drane, J. W. (2004). Life satisfaction and suicide among high school adolescents. *Social Indicators Research*, 66, 81-105.
- Witarsa, P. M. (2003). *Nonparametric item response modeling for identifying differential item functioning in the moderate-to-small-scale testing context*. Unpublished doctoral dissertation, University of British Columbia, B.C.

- Zullig, K. J., Valois, R. F., Huebner, E. S., Oeltmann, J. E., & Drane, W. (2001). The relationship between life satisfaction and selected substance abuse behaviors among public high school adolescents. *Journal of Adolescent Health, 29*, 279-288.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.
- Zumbo, B. D. (2007a). Validity: Foundational issues and statistical methodology. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of Statistics, Vol. 26: Psychometrics* (pp. 45-79). Elsevier Science B. V.: The Netherlands.
- Zumbo, B. D. (2007b). Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*, 223-233.
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21-29.
- Zumbo, B. D., Gelin, M. N., & Hubley, A. M. (2002). The construction and use of psychological tests and measures. In the Psychology theme of the *Encyclopedia of Life Support Systems* (EOLSS), Oxford, UK: Eolss Publishers.

Investigating the substantive aspect of construct validity for the Satisfaction with Life Scale adapted for Children: A focus on cognitive processes¹⁸

Measuring and monitoring children's well-being has received increasing attention and interest over the last decade (Ben-Arieh, 2006; Ben-Arieh & Goerge, 2001). One of the reasons for this is the "movement toward accountability-based public policy that requires increasing amounts of data to provide more accurate measures of the conditions children face and the outcomes various programs achieve" (Ben-Arieh, 2005, p. 573). Specifically, measuring and monitoring children's well-being are important to gain a better understanding of, and enhanced knowledge about, their well-being and to inform and evaluate policies and programs with the aim of improving children's well-being (e.g., Ben-Arieh & Goerge, 2006; Ben-Arieh et al., 2001; Frones, 2007).

With regard to what is measured, there has been a shift away from early indicators that focused on measuring survival or negative facets of children's lives to an approach that is more holistic by also measuring assets and positive aspects of children's lives. Furthermore, indicators assessing children's 'well-becoming' (predicting transition to, and well-being in, adulthood) have been supplemented by indicators assessing current well-being during childhood (Ben-Arieh, 2006; Ben-Arieh & Goerge, 2001).

One approach that combines the focus of emphasizing the positive aspects of individuals' lives with the focus on current well-being is the field of subjective well-being research. Subjective well-being is considered to consist of positive and negative affect, life satisfaction,

¹⁸ A version of this chapter has been submitted for publication: Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2009). Investigating the substantive aspect of construct validity for the Satisfaction with Life Scale adapted for Children: A focus on cognitive processes.

and domain satisfaction (e.g., Diener, Suh, Lucas, Smith, 1999). One of the most commonly used instruments to assess satisfaction with life in adults is the Satisfaction with Life Scale (Diener, Emmons, Larsen, & Griffin, 1985). This scale has been adapted by subject matter experts for children in grades 4 to 7 and its psychometric properties have been shown to be favorable with a sample of children in grades 4 to 7 (Gadermann, Schonert-Reichl, & Zumbo, in press). However, in order to develop and/or validate an instrument, it is recommended to use experiential experts (i.e., members of the target population) to investigate the cognitive processes that respondents use to answer questions (Collins, 2003; Willis, 2005). This is especially important for measures developed for children, as the conceptualization of the adult test developers can potentially be quite different from the one of children. Therefore, in the present study I used think-aloud protocols, a cognitive interviewing technique, with children for evaluating the items of the SWLS-C. This technique has shown to be useful to investigate the cognitive processes of children in previous studies (e.g., Cremeens, Eiser, & Blades, 2006a; Fox, Houston, & Pittner, 1983; Lodge, Harte, & Tripp, 1998; Lodge, Tripp, & Harte, 2000; Rebok et al., 2001). This technique was used to investigate the cognitive processes of children when answering the items of the SWLS-C, in order to explore how the children arrived at their specific response. The investigation of cognitive processes during a measurement task is one way to evaluate the substantive aspect of construct validity (Messick, 1995). In the following, I will first provide a brief overview of the importance of cognitive interviewing techniques for the validation of self-report measures before describing the study.

The importance of validating self-report measures using cognitive interviewing

Self-report measures, such as questionnaires and surveys, are commonly used in the social sciences to collect data on psychological constructs, such as subjective well-being. The information from such questionnaires and surveys is used for a variety of reasons, for example, to evaluate intervention programs, to describe societal conditions, and to inform public policy. Accordingly, self-report measures can have far-reaching consequences. However, the data collected with such measures are obviously only as meaningful as the questions that are asked and the responses that participants provide (Schwarz, 1999). Therefore, the thorough development and ongoing validation of questionnaires and surveys is of special relevance. In this regard, it is of interest to investigate the substantive aspect of construct validity (Messick, 1995). Specifically, it is of interest to investigate how and why respondents arrive at their answers and how this is influenced by the characteristics of the respondent and the questionnaire (and their potential interactions). In other words, one needs to ask the question: What are the underlying cognitive processes that result in respondents providing responses to self-report questions? In the last three decades, this topic has become of increasing interest for researchers in areas such as psychology and survey methodology. In the 1980s, the Cognitive Aspects of Survey Methodology (CASM) initiative started, an interdisciplinary movement with the aim to improve the quality of self-report data and “to bridge the communication gaps between survey research and the cognitive and social sciences, and to initiate CASM research that would benefit survey applications as well as basic cognitive research” (Sirken & Schechter, 1999, p. 1). CASM research investigates the cognitive processes that underlie self-reports in order to understand how these processes function. CASM research can thereby influence questionnaire design (e.g., by

suggesting how to redesign a questionnaire if the items do not perform/function as expected) as well as stimulate basic research on cognition (Sirken & Schechter).

This is in line with contemporary views of measurement validity, in that cognitive processes or models are investigated in the validation process to support the inference one makes from the scale scores. As Messick (1989) stated, “validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness of inferences* and *actions* based on test scores or other modes of assessment” (p. 13). In Messick’s unified view of validity, construct validity lies at the core and “comprises the evidence and rationales supporting the trustworthiness of score interpretation in terms of explanatory concepts that account for both test performance and score relationships with other variables” (Messick, 1995, p. 743). As mentioned above, one aspect of construct validity is the substantive aspect, which highlights the importance of theories and process modeling in examining the processes that are involved in the measurement task, and which can be investigated using different approaches such as cognitive interviewing or modeling response times. Evidence based on response processes is also listed as one of the five sources of validity in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME), 1999). The research question ‘What does a score on a self-report measure provided by a participant mean?’ is also very much in line with what has been described as a strong form of construct validity, which “should provide an *explanation* for the test scores, in the sense of the theory having explanatory power for the observed variation in test scores” (Zumbo, 2009, p. 69; see also Zumbo, 2007).

Although this illustrates the importance of investigating the substantive aspect of

construct validity, this aspect is often not investigated in the development or evaluation of measures. It is worth noting that much of the validation research is about correlations with other variables and hence is not explanatory. For example, a study by Cizek, Rosenberg, and Koons (2008) investigated (among other things) the types of validity evidence reported in the current edition of the *Mental Measurements Yearbook*. The authors reported that response processes were only investigated for 1.8% of the measures, whereas criterion-related (correlational) validity evidence was provided for 67.2% of the measures. None of the personality/psychological measures or social measures reported on response processes as sources of validity evidence, whereas it was reported in 5.9% of the cases of the developmental measures, 4.0% of the behavioural measures, and 3.7% of the achievement measures. Similarly, in Cremeens, Eiser, and Blades' (2006b) review of health-related measures, including quality of life measures, for children aged 3 to 8 years, cognitive processes are not reported at all as a source of validity evidence. (It should be noted, however, that children were consulted during the process of item development for 40% of the measures, e.g., through interviews, and that some form of pilot testing with children was conducted for 47% of the measures.)

In line with this, McColl, Meadows, and Barofsky (2003) state that cognitive techniques have rarely been applied to well-being or quality of life research, although in recent years there has been a development in this direction as indicated by the formation of the 'International Study Group on Cognitive Aspects of Quality of Life Research' (Barofsky, Meadows, & McColl, 2003). Nonetheless, with regard to investigating the cognitive processes underlying self-reports of children, there are few studies that employed think-aloud protocols with children in the area of quality of life and subjective well-being (Cremeens et al., 2006a; see also Riley, 2004, for the area of health). It is noteworthy that children and adolescents are often not included in the

evaluation of measures, given that several studies have indicated that having children as subject matter experts adds a critical component to the development and evaluation of measures for children and adolescents (e.g., Cremeens et al., 2006a; Rebok et al, 2001; Schilling et al., 2007; Stewart, Lynn, & Mishel, 2005).

One of the few studies using think-aloud protocols with children in the area of quality of life and well-being was conducted by Cremeens et al. (2006a). In their study, children aged 5-9 were asked to think aloud while responding to the TedQL, a generic quality of life measure. The authors report that children used several strategies for responding to the items, namely (i) social comparisons; (ii) stable character references; (iii) concrete examples; (iv) other reasons; and (v) no reason given. The type of strategy utilized was related to the age of children and type of item (ability and social items). Specifically, older children were more likely to use the social comparison and concrete examples strategies than younger children, whereas younger children were more likely to provide no reason than older children. Furthermore, concrete examples was the most frequently used strategy and there was no statistically significant difference in the use of this strategy by type of item. In contrast, the social comparison strategy was used more frequently for ability than social items, whereas the stable character references strategy was used more frequently for social than ability items.

Similar to Cremeens et al.'s (2006a) study, I was interested in investigating the cognitive processes of children when responding to the items of the SWLS-C and whether response strategies used would be associated with demographic characteristics of the children.

Method

This section is structured into the following parts: (i) Sample; (ii) measure; (iii) think-aloud protocols; (iv) procedure; (v) development of coding categories; and (vi) quantitative analysis.

Sample

The study was conducted in one elementary school in the Lower Mainland of British Columbia. This sample was completely unrelated to the sample in chapter 2. Ethics approval was obtained from the University of British Columbia (see Appendix B) and the school board of the district of that school. The school is located in an urban, low income environment. The median family income (ca. CAN\$ 30,000) in the neighbourhood surrounding the school is approximately one standard deviation (CAN\$ 12,000) below the median family income of the entire province of BC (Can\$ 43,000).

Seventy percent of parents of the children who returned the signed parental consent form provided consent; all of these students provided their assent (see Appendix B for the parent consent and student assent forms). These 61 students in grades 4 to 7 provided think-aloud protocols. Because 6 of them had strong problems with the language due to English as a second language, these were excluded from the analysis; therefore, the total sample size for the analysis was 55. The students were from six classrooms: Two grade 4/5 classrooms, one grade 5/6 classroom, and three grade 6/7 classrooms. Fifty-eight percent of the students were girls. The mean age was 11.0 years (with a standard deviation of 1.2) ranging from 8.8 to 12.8 years. In terms of the grades, 16% of the children were in grade 4, 24% in grade 5, 29% in grade 6, and 31% in grade 7. The school has 350 students who speak more than 25 languages. In this sample,

children reported having learned 20 different languages as their first language at home. Specifically, 31% of the children reported having learned English only, 33% reported having learned another language than English, and 36% reported having learned English and another language at home¹⁹. With regard to the children who had first learned another language than English only at home, the most frequently learned languages were Farsi, Chinese²⁰, and Korean. With regard to children who had learned English and another language, the most frequently learned languages were Chinese, Punjabi, Spanish, French, and Korean.

With regard to how difficult it is for them to read and write in English, 53% of the children reported it as being very easy, 38% as easy, and 9% as hard. Two of the participants asked the interviewer to read out the items for them as they had reading problems.

Measure

The Satisfaction with Life Scale adapted for Children (SWLS-C). The SWLS-C was adapted from the Satisfaction with Life Scale (SWLS; Diener et al., 1985), a commonly used measure to assess satisfaction with life in adults. The SWLS was adapted for children by three subject matter experts in the area of socio-emotional development of children. The SWLS-C consists of five items addressing the respondents' life satisfaction with a 5-point response scale ranging from 'disagree a lot' to 'agree a lot' (see Table 3.1). Gadermann et al. (in press) provided psychometric evidence for the construct validity of the SWLS-C in a sample of 1233 students in grades 4 to 7. Specifically, the results indicated that the scale was unidimensional, had a high reliability, and measured life satisfaction in the same way across different groups of children

¹⁹ Two of these children reported having learned English and two other languages at home.

²⁰ The language label Chinese includes Mandarin, Cantonese, and Taiwanese.

(namely, across gender, first language learned at home, and different grades) at the item and scale level as investigated by differential item and scale functioning analyses in that sample. Furthermore, the SWLS-C showed relationships to convergent and discriminant measures as was expected based on previous research.

Table 3.1

The Satisfaction with Life Scale for Children

For each of the following statements, please circle the number that describes you the best. Please read each sentence carefully and answer honestly. Thank you.

	Disagree a Lot	Disagree a Little	Don't Agree or Disagree	Agree a Little	Agree a Lot
1. In most ways my life is close to the way I would want it to be.	1	2	3	4	5
2. The things in my life are excellent.	1	2	3	4	5
3. I am happy with my life.	1	2	3	4	5
4. So far I have gotten the important things I want in life.	1	2	3	4	5
5. If I could live my life over, I would have it the same way.	1	2	3	4	5

Think-aloud protocols

According to Messick (1995), there are six aspects of construct validity, one of which is the substantive aspect. The substantive aspect of construct validity highlights the importance of identifying and modeling the processes that respondents employ in completing assessment tasks.

Evidence for this construct validity aspect can be provided from different sources, and one is the think-aloud protocol (Ericsson & Simon, 1980). In a think-aloud protocol, respondents are typically given the instruction to think aloud while completing a questionnaire, which is called concurrent verbalization. Also, respondents may be asked to describe previous cognitive processes (e.g., right after having finished a task) and this procedure is called retrospective verbalization (Ericsson & Simon, 1980). In both of these think-aloud procedures, the researcher hardly interjects. A related approach is verbal probing, where respondents are probed for specific information by the interviewer; that is, the interviewer utilizes specific verbal probes, such as asking the respondent to reformulate an item, or to define some of the key terms in their own words (Willis, DeMaio, & Harris-Kojetin, 1999). Oftentimes, researchers use a combination of the think-aloud and verbal probing approaches. In the present study, a combination of the concurrent verbalization with verbal probing was used.

Procedure

During the first visit in the classrooms, I or the teacher explained the purpose and provided details about the study. Permission slips/parent consent forms were handed out and the children were asked to bring these to their parents and return them signed (either yes or no). Translated versions of the parent consent form were not available. As an incentive to return the parent consent form, students were given a pencil upon returning the form. During my second visit in the classrooms, the data collection took place. During class time, individual students were asked to come to a quiet room for the think-aloud protocol. The think-aloud protocols were audiotaped. Three practice items were used to familiarize the children with the task of thinking-aloud. Specifically, the first practice item was verbally presented to the children (adapted from

Creameens et al., 2006a): “When you are answering the items, I would like you to say out loud all the things that come into your head when you are choosing your answer. For example, I am answering a question about whether I am good at tidying my bedroom...Now what do I think? I don’t like to tidy my bedroom, but I do tidy it when my mother tells me to...and I make sure that all my things are put away...so I think I am good at tidying my bedroom, and I point here (i.e., on the high end of the rating scale). Now we are going to answer some more questions, and I want you to remember to talk aloud, and say what you are thinking as you answer.” The children were then asked to respond themselves to this item. Then they were asked to respond to two more practice items “In general, I like to eat vegetables.” and “I enjoy reading books.” For each item, the children were asked to ‘think-aloud’ while they were considering their responses. If a child was silent for more than 10 seconds, s/he was given up to two prompts, such as “Remember to say out loud all the things that come into your head” and “What are you thinking and saying to yourself now?” (Creameens et al., 2006a, p. 85).

After the three practice items were completed, the children were asked to respond to the SWLS-C while thinking aloud. A subsample of 23 of the children was asked, after the think-aloud protocol, what they thought about the items and what they thought about giving these items to other children their age. On average, the session took about ½ hour per child. At the end of the interview, the children were given two erasers.

Development of coding categories

All interviews were transcribed by a professional transcriptionist, and then checked for accuracy by the first author. Content analysis was used for deriving a coding scheme of the transcripts (Berg, 2004) in order to decipher and interpret the data (Böhm, 2004). Content

analysis is “the systematic, objective, quantitative analysis of message characteristics” (Neuendorf, 2002; p. 1).

The children’s responses were coded for each of the five items of the SWLS-C separately, using the software Atlas.ti 5.0. First, open coding was used for a wide inquiry into the data. After the open coding, codes were assigned to categories. The categories were developed to (i) reflect the research purpose, (ii) be exhaustive and mutually exclusive, (iii) be grounded conceptually in the theoretical quality of life research literature, and (iv) be grounded empirically in the data (see Dey, 1993; Holsti, 1969). Accordingly, the coding combined an inductive and deductive approach, with a larger reliance on the inductive approach.

Themes were chosen as the unit of analysis for the coding. Themes in its simplest form can be a “simple sentence, a string of words with a subject and a predicate” (Berg, 2004, p. 273). Children’s responses were coded according to themes, which were then assigned to the accordant category. Generally, one primary theme was coded for each response to a particular item. Children frequently provided more than one theme in response to a single item (without one being primary). In those cases, the different themes were coded into separate categories. Eventually, for a category to be kept in the overall final coding scheme, a category had to occur at least three times in any of the five items of the SWLS-C (cf. Berg, 2004).

The development of the coding categories was guided by three general research purposes. The first purpose was to investigate how the children understand, interpret, and respond to the items of the SWLS-C. Specifically, I was interested in the strategies that children employed to respond to these items. A second purpose was to identify the content the children talked about; in other words, the aim was to investigate on which content children focused when using a certain strategy for making their life satisfaction judgments. A third research purpose was to find out

whether children use positive or negative statements in their responses. According to these purposes, the process of developing the coding categories was informed by the following specific questions:

- (i) What strategies do the children employ when responding to the item?
- (ii) What are the general content topics that come up for the children when responding to the items?
- (iii) Is the valence of these content topics typically positive (i.e., presence of something positive or absence of something negative) or negative (i.e., presence of something negative or absence of something positive)?

Furthermore, it was of interest to examine whether children had any difficulties in terms of understanding the SWLS-C items and/or the response format.

After preliminary categories were developed, the most prominent categories were further elaborated, and a coding scheme was developed with main categories and sub-categories. Based on the coding scheme, I went through the transcripts again to check the codes and recode the data. After that, a second rater, who is a researcher in the area of child development, coded the data based on the coding scheme, and the inter-rater reliability was computed. In case there was a difference in coding, the two raters discussed the code and came to an agreement.

Quantitative analysis

The frequencies of the codes were then transferred into SPSS 17.0 for further analysis. Specifically, the frequencies of the use of different categories were investigated. Furthermore, it

was of interest to examine whether children used certain strategies significantly more often than another for the respective items. This was investigated by calculating the McNemar test, a paired test of equality of proportions, for each item separately. Moreover, it was examined whether there were differences in the use of strategies depending on the children's grade. In order to investigate this, the sample was divided into two groups: A younger group (with the children in grades 4 and 5) and an older group (with children in grades 6 and 7). The McNemar test was then run separately for the two grade categories.

In addition, in order to detect potential demographic differences with regard to the use of the strategies, Poisson or binary logistic regression analyses (depending on whether the response variable was a count or binary variable, respectively) were run with the factors gender, grade, and first language background.

Finally, one additional set of analyses was conducted to examine whether the valence (positive or negative) of children's responses was correlated with their scores on the individual SWLS-C items.

Results

Definitions of categories and subcategories

All of the participants' responses to the five items were coded into four levels of categories, according to specific coding definitions. In the following, a definition of each category is provided. The coding categories are represented by the tree diagram in Figure 3.1.

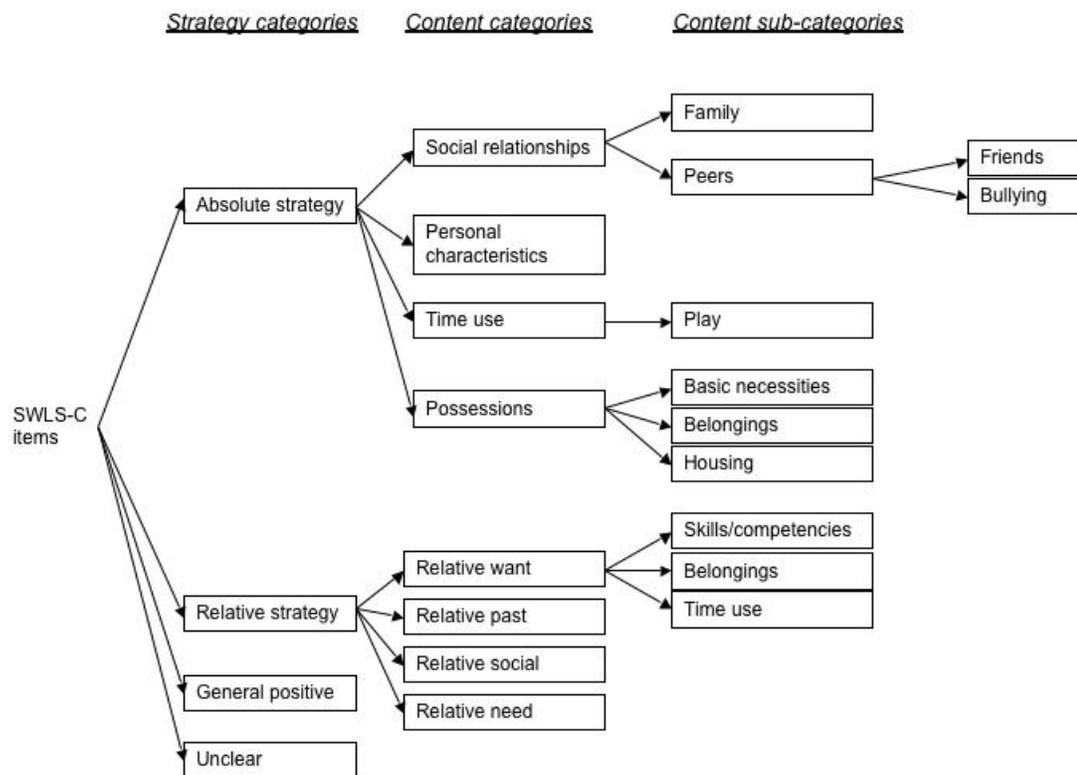


Figure 3.1. Tree diagram of the coding categories.

Strategy categories. As a first step, for each response, it was examined whether the participant used an *absolute* (A) or a *relative* (R) strategy in her/his response thought process. That is, it was examined whether a participant used absolute or relative statements while responding to an item. In this context, an absolute statement indicated the presence or absence of

something that was apparently important with regard to a participant's judgment of her/his satisfaction with life (e.g., "I agree a lot with that cause I have very nice parents and a really nice sister."). A relative statement, on the other hand, included a comparative statement with regard to the presence or absence of something that was important with regard to a participant's judgment of her/his satisfaction with life (e.g., "I wish I would get *better* grades."; "I want to have *more* friends."). In addition to these two categories, a category labeled *General positive* was defined to capture all statements that did not include an absolute or relative strategy, but that included a general, positive statement (e.g., "Because it's fun, and I just like it."; "Well, it's not boring, it's kind of fun.") Finally, any responses that could not be coded into any of these three categories were assigned to a category labeled *unclear* (e.g., "It's because mostly sometimes it happens and sometimes it doesn't."; "I don't exactly know what I want in life"; "It's just a life. I just live on a daily basis or something.").

Content categories. In a second step, two clusters of *content categories* were developed. The first cluster of content categories was assigned to the absolute strategy category, and the second cluster of content categories to the relative strategy category. For the absolute strategy category, four content categories were defined: (A1) *Social relationships*; (A2) *Time use*; (A3) *Personal characteristics*; and (A4) *Possessions*. Similarly, for the relative strategy, the following four categories were defined: (R1) *Relative social (social comparisons)*, (R2) *Relative to one's wants*; (R3) *Relative to one's past*; and (R4) *Relative to one's needs*.

For the four content categories under the absolute strategy category (see Figure 3.1), the following definitions were developed. (A1) *Social relationships*: Each response that used an absolute statement referring to a social relationship was coded in this category. In order to

capture the diversity of social relationships that were mentioned, this category was subdivided into two subcategories, according to whether the statement referred to *Family* members, such as parents, siblings, or grandparents (e.g., “I am happy with my life because I have a really caring family.... And like my mom is always there for me whenever I’m sad or there’s something that I really want to tell her.” or “My dad gets mad at me for no reason. And he swears a lot.... My dad keeps getting mad because he was on drugs and stuff.”); or *Peers*. The category *Peers* was further subdivided into the subcategories *Friends* (“My friends, they’re very supportive of me and they’re wonderful.”) and *Bullying* (“Because people make fun of me and call me names. Like Big Apple because they think I’m fat. They bully me a lot, like start punching me and kicking me.”). (A2) *Time use*: Any statement referring to an activity was coded into this category (e.g., “I like going shopping on Saturdays.”). The subcategory *Play* was created, which included statements referring to games and play activities (e.g., “Because most of the time I like to play.”). (A3) *Personal characteristics*: This category included statements referring to personal characteristics, competences, skills, and likes (e.g., “Cause I am not doing that good in school and stuff.”; “I’ve got these good talents in singing and a lot of knowledge, too.”). (A4) *Possessions*: Any references to personal belongings, possessions, or access (or lack of) material things were coded into this category, according to the following subcategories: *Basic necessities* (this subcategory included things that fulfill basic needs, e.g., “We’re not living in poverty and that’s also important. And I have shelter and all the other basics, like, water.”); *Belongings* (this category included material things, such as computer games, as well as pets; e.g., “Since I have lots of Lego.”; “Because I have a cat.”); and *Housing* (statements in this category did not indicate the presence of shelter as a necessity, but referred to the quality of the housing situation; e.g., “Because we have a nice house.”).

The four content categories within the relative strategy category (see Figure 3.1) were coded according to the following definitions: (R1) *Relative social*: This category included statements that made social comparisons (e.g., “I have a younger brother, so my parents like him more. And they care for him more.”; “We should realize how it will be to be like other people with struggles in other parts of the world.”). (R2) *Relative to one’s wants*: Any statement indicating a comparison between what a child currently has and what s/he wants was coded in this category. This category was subdivided into three subcategories: *Time use* (e.g., “Because I’ve always wanted to start karate and right now I’m starting my monthly karate.”); *Belongings* (e.g., “We live in an apartment, but I want to live in a house.”; “I agree a little cause I want to get a Wii and a bigger house and a car.”); and *Skills/competencies* (e.g., “I want better grades..., but I don’t usually get good grades, but they’re okay.”; “Because I always wanted to be a good drawer and now I’m a really good drawer.”). (R3) *Relative to one’s past*: If a child made a comparative reference to her/his past, the response was coded into this category (e.g., “Back at the old place, they [my cousins] teased me, but here they don’t tease me.”; “Because I have the things I wanted to happen. Oh, like, one night, when we were in Afghanistan, so when there was a fight, there was a war with Taliban, so we wanted a good life. So I wished that we could go to another country or somewhere else. Or maybe Canada or America. First we went to Uzbekistan, then we lived there for 8 or 9 years and then we came here.”). (R4) *Relative to one’s needs*: Any statements that made a comparison between a child’s status quo and his/her (stated) needs was included in this category (e.g., “I don’t have the best life, I don’t think. But I have one that suits me. I have everything I need right now”).

Category frequency counts

In the method section, it was described that the criterion for creating (and keeping) a category was that at least three participants used statements referring to this category in response to an item. It must be noted that categories that occurred three or more times for one item (e.g., item 1) were then also used for the coding of the other items (e.g., 2-5). This led to the scenario that, for some of the items, a category was used by less than three participants. In those cases, the codes for these statements were assigned to the next higher level category (e.g., if statements in reference to the *peers* category occurred only twice in response to item 2, these statements were counted towards the higher level category *social relationships*).

In Figure 3.1, it should be noted that all categories that occurred for at least one item are shown. However, when I report the results for individual items and present them graphically in the respective items' tree diagrams (Figures 3.2 to 3.6), only those categories that were used at least three times for that particular item are reported and are shown as individual boxes in the diagram. In those cases, any categories with less than three codes were collapsed into a general category labeled as *other*, and the count of statements within this general category was simply counted towards the higher level category. For example, if, on item 3, only one participant referred to *basic necessities* and one participant referred to *housing* (in the tree diagram, see: *absolute* (strategy category) → *possessions* (content category) → *belongings* and *housing* (subcategory)), these two codes were collapsed, represented in a box labeled *other*, and counted towards the next higher level category, *possessions*. It must also be noted that there were a few responses, which were coded and assigned to a category, even if the codes did not fit into any of the category's subcategories—because they occurred less than three times on all five items. For example, a child referred to the relationship with her/his teacher. This reference was, naturally,

coded under the category *social relationships*, but the code could not be assigned to any of the subcategories of *social relationships* (that is, *family* or *peers*). In the frequencies reported in the result section, as well as in the graphic representations of the results, such codes also appear under the generic *other* category or subcategory. This procedure allowed us to maintain the highest level of detail in the descriptive reporting of the data, and allowed us to conduct the statistical analyses of the category frequencies at a methodologically adequate level.

As mentioned above, the objective was to develop categories that reflect the research purpose, and that are exhaustive and mutually exclusive. These requirements were met except for being exhaustive. Specifically, there was one category entitled *Unclear*; however, as Holsti (1969) points out “even the most carefully designed study is likely to fall short of completely satisfying this requirement” (p. 99).

The inter-rater reliability using Cohen’s Kappa was .84. According to Mayring (2004), Kappa coefficients of .70 or larger are considered to be sufficient.

Item 1: In most ways my life is close to the way I would want it to be. For item 1, the absolute strategy was used 46 times and the relative strategy 33 times. Furthermore, 3 responses were coded in the general positive category, and three responses were coded in the unclear category. The frequencies of the strategy and content (sub)categories for item 1 are illustrated in Figure 3.2. As mentioned above, children frequently used more than one strategy and/or content category in response to one item²¹. Therefore, it is also reported how many children used the relative and absolute strategies and/or content categories: Of the 55 children, 7 children used both the relative and absolute strategy in responding to item 1, 20 children used only the absolute

²¹ As a result, the sum of all frequencies of the categories is not equal to the sample size of 55.

strategy, and 22 children used only the relative strategy in order to respond to the item. Of the children who used the absolute strategy, 11 used one content category, 14 mentioned two content categories, 1 child mentioned three, and 1 child mentioned four different content categories. Of the children who used the relative strategy, 25 mentioned one content category and 4 mentioned two content categories. Seven children had some difficulty responding to the item as they found the item somewhat difficult to understand. With regard to the valence of the statements, 70% of the statements were positive, 22% negative, and 8% were mixed.

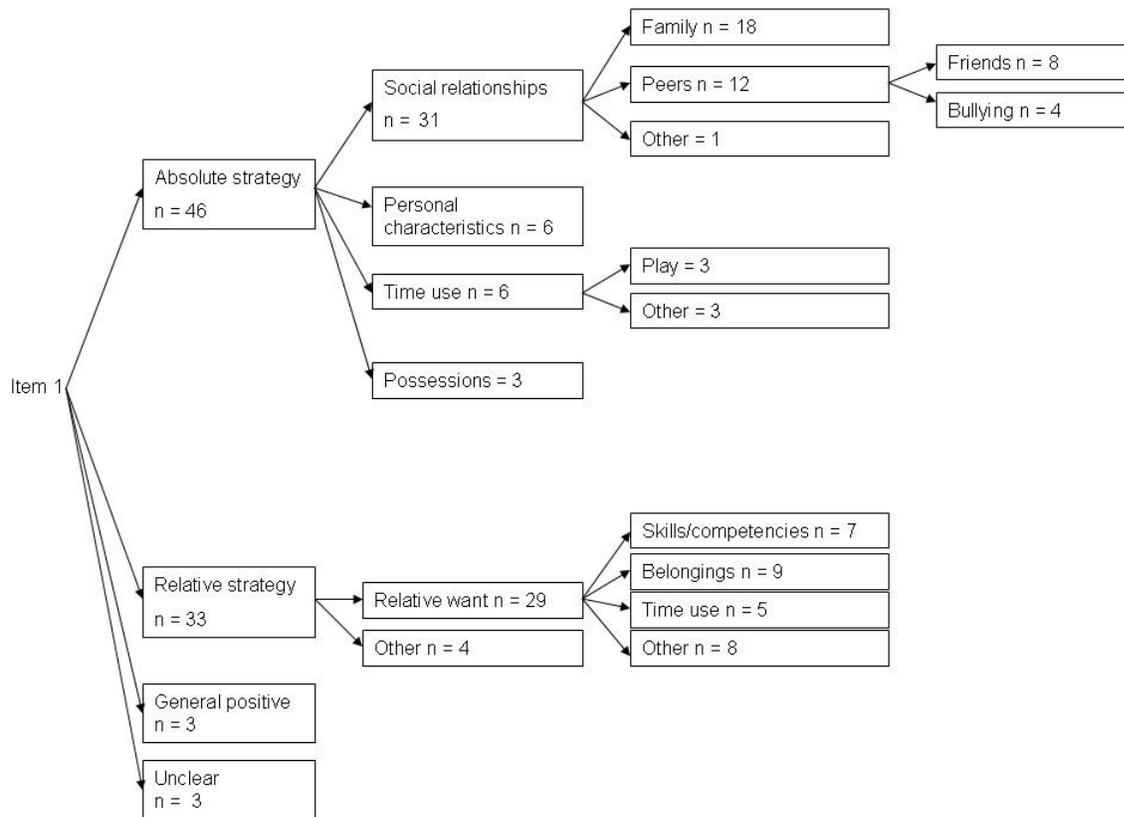


Figure 3.2. Tree diagram of the coding categories for item 1.

Item 2: The things in my life are excellent. For item 2, the absolute strategy was used 69 times and the relative strategy was used 15 times. Furthermore, three responses were coded as unclear. The frequencies of the strategy and content (sub)categories for item 2 are illustrated in Figure 3.3. Two children used both the relative and absolute strategy, 38 used only the absolute strategy, and 12 children used only the relative strategy. Among those who used the absolute strategy, 20 mentioned one content category, 14 mentioned two content categories, 3 mentioned three content categories, and 3 mentioned four content categories. For the relative strategy, 13 children mentioned one content category, and 1 child mentioned two. With regard to item understanding, 1 child asked to what the word “things” was referring, and 1 child commented that he felt the wording was grammatically incorrect. With respect to the valence of the statements, they were predominantly positive (72%), and only relatively few were negative (21%) or mixed (7%).

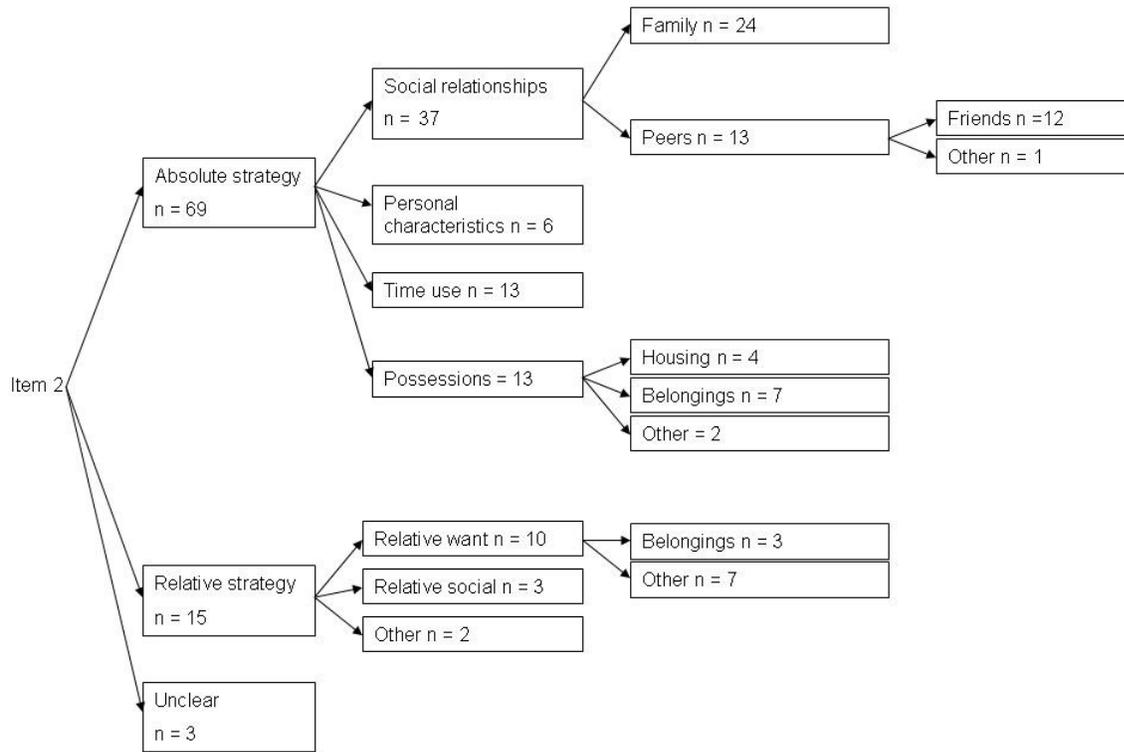


Figure 3.3. Tree diagram of the coding categories for item 2.

Item 3: I am happy with my life. For item 3, the absolute strategy was used 40 times and the relative strategy was used 13 times. Furthermore, 10 responses were coded in the general positive category, five responses were coded in the unclear category. The frequencies of the strategy and content (sub)categories for item 3 are illustrated in Figure 3.4. One child did not provide any explanation for his response. One child used both the absolute and relative strategies, 26 used only the absolute strategy, and 12 used only the relative strategy. For the absolute strategy, 19 children mentioned one content category, 5 children mentioned two content categories, 1 mentioned three, and 2 mentioned four content categories. For the relative strategy, all mentioned one content category. Two children commented that the item was similar to the previous ones and one child had problems with the response format to respond to the item. The

valence of the statements was mostly positive (80%) with a few negative (11%) and mixed (9%) ones.

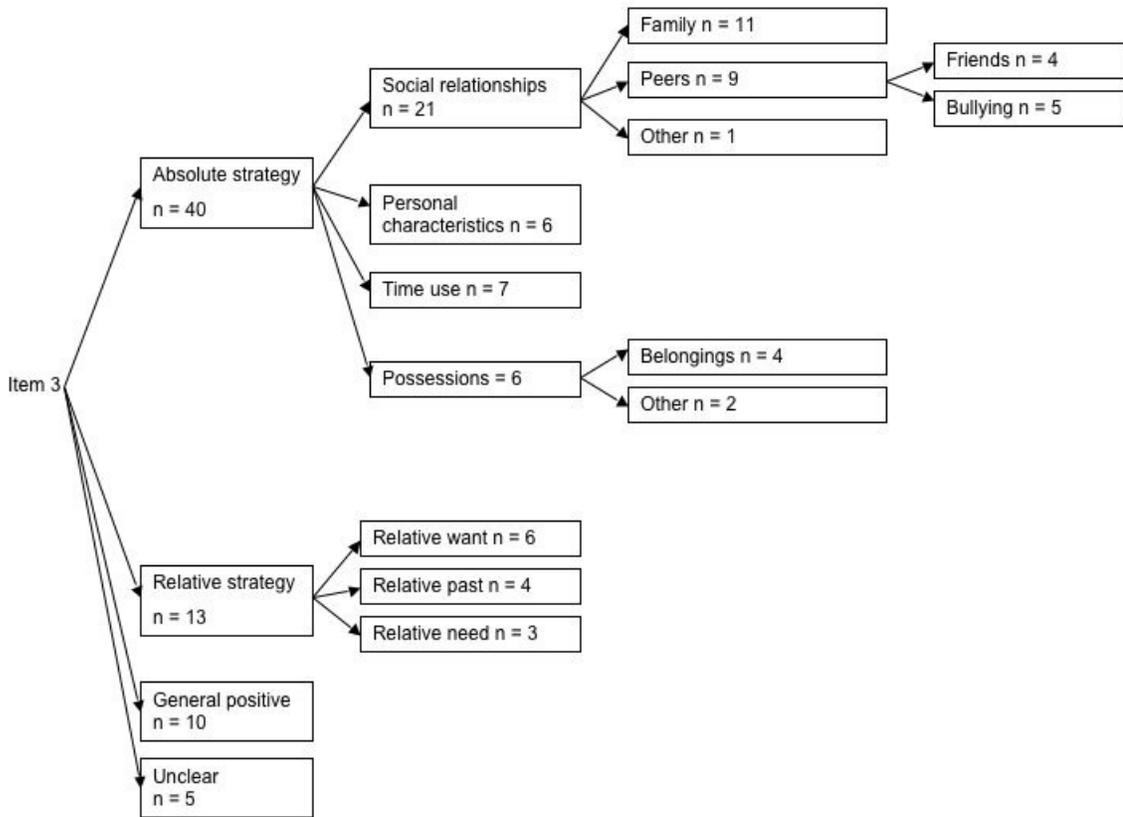


Figure 3.4. Tree diagram of the coding categories for item 3.

Item 4: So far I have gotten the important things I want in life. For item 4, the absolute strategy was used 40 times and the relative strategy was used 47 times. Furthermore, three responses were coded as unclear. The frequencies of the strategy and content (sub)categories for item 4 are illustrated in Figure 3.5. Eleven children used both the relative and absolute strategy, 12 only the absolute strategy, and 29 only the relative strategy. For the absolute strategy, 13 children mentioned one content category, 6 mentioned two content categories, 1 mentioned three

content categories, and 3 mentioned four content categories. For the relative strategy, 34 children mentioned one content category, 5 mentioned two content categories, and 1 child mentioned three content categories. One child had problems responding to the item. With respect to the valence of the statements, they were predominantly positive (75%), and only relatively few were negative (15%) or mixed (7%).

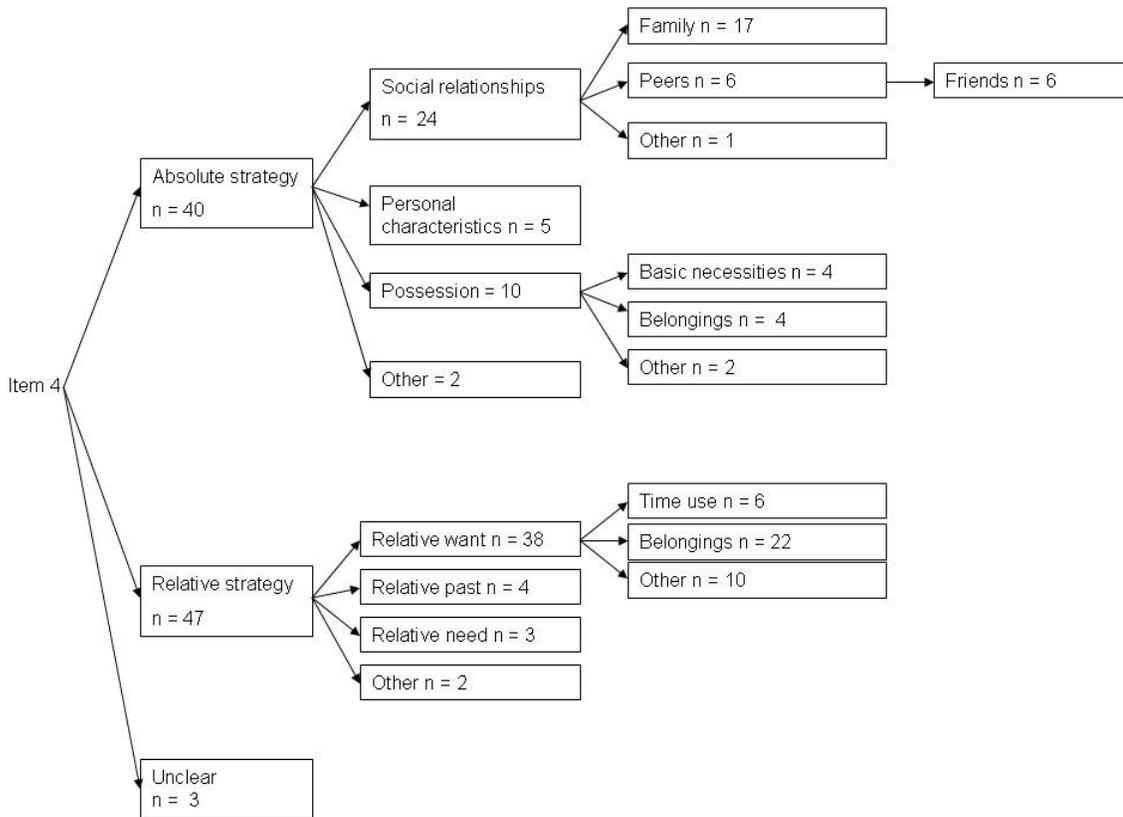


Figure 3.5. Tree diagram of the coding categories for item 4.

Item 5: If I could live my life over, I would have it the same way. For item 5, the absolute strategy was used 13 times and the relative strategy was used 34 times. Furthermore, nine responses were coded in the general positive and 10 responses in the unclear category. The

frequencies of the strategy and content (sub)categories for item 5 are illustrated in Figure 3.6. Three children used both the relative and absolute strategy, 7 only the absolute strategy, and 26 only the relative strategy. For the absolute strategy, 7 children mentioned one content category, and 3 mentioned two content categories. For the relative strategy, 24 children mentioned one content category, and 5 mentioned two content categories. Furthermore, 5 children had problems responding to the item. With respect to the valence of the statements, 52% were positive, 40% were negative, and 8% were mixed.

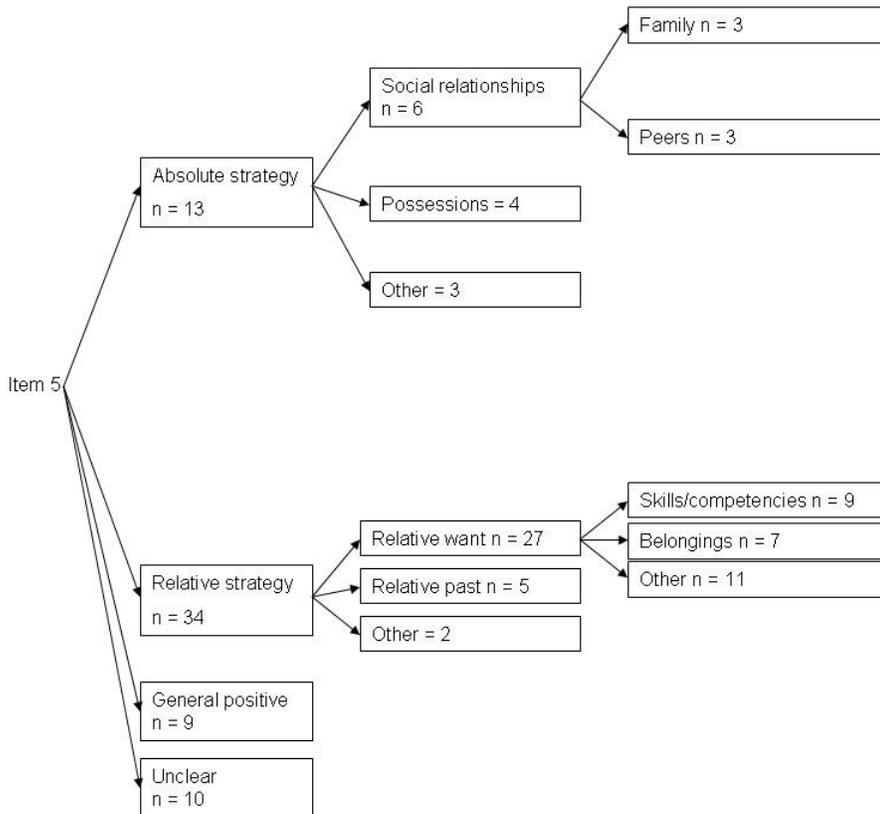


Figure 3.6. Tree diagram of the coding categories for item 5.

Comparison of items

There are several ways to look at the patterns of the findings. One way is to compare whether the tree diagrams—that is, the occurrence of categories and subcategories—are similar or different across items. This information is summarized graphically in Figure 3.7.

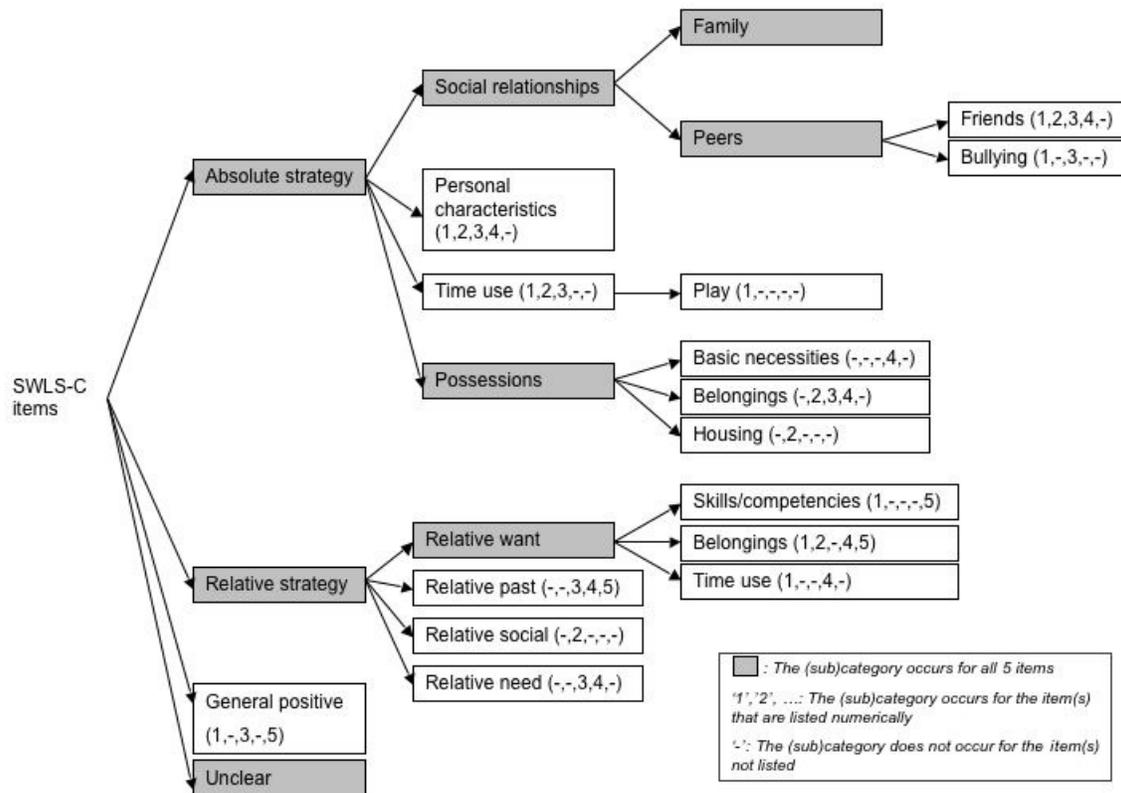


Figure 3.7. Summary of the differences and commonalities in strategy and content (sub)categories for the five items of the SWLS-C.

As can be seen, for all items, children used the absolute and relative strategies. Furthermore, the category ‘unclear’ was assigned to responses for all five items. However, comments assigned to the category ‘general positive’ only occurred (three or more times) for items 1, 3, and 5. Among the content categories, ‘social relationships’, ‘possessions’, and ‘relative want’ occur for all five items. The only subcategories that occur for all five items are

the social relationship subcategories ‘peers’ and ‘family’. All other (sub)categories occurred for a subset or only one of the items. For example, the content category ‘personal characteristics’ occurred for items 1 to 4, whereas the content category ‘relative social’ only occurred for item 2.

A further way to explore the patterns of results is to visualize the frequencies with which the different strategies, categories, and subcategories were used across the items. Figure 3.8 presents the frequencies for the absolute and relative strategies, the content categories, and the most frequently used subcategories.

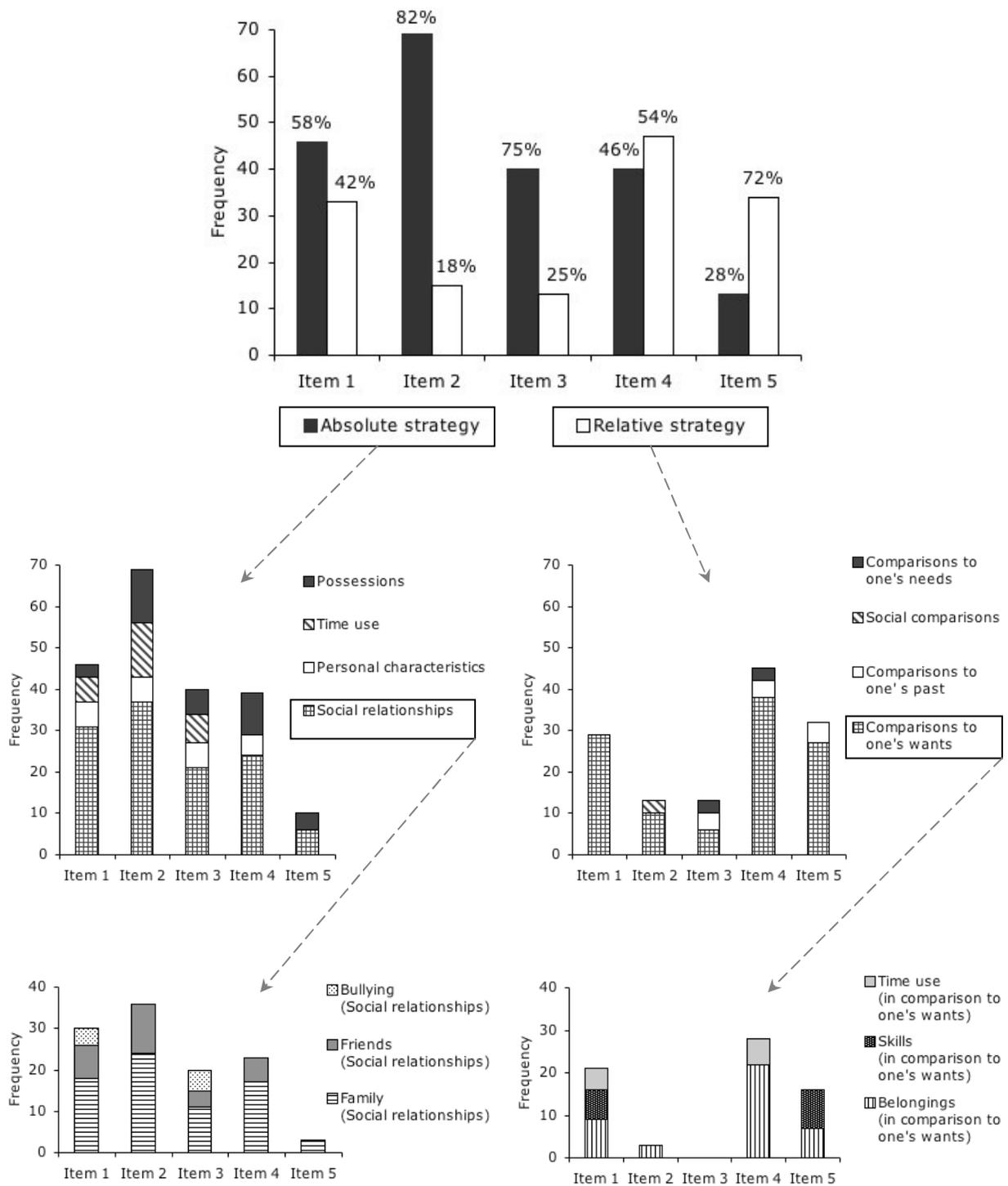


Figure 3.8. Frequencies for the five SWLS-C items for *strategy* and *content (sub)categories*

Frequencies for the five SWLS-C items for *absolute* versus *relative strategies* (top panel), the content categories for absolute strategies (middle panel, left) and relative strategies (middle panel, right), and the subcategories for *social relationships* (bottom panel, left) and for *comparisons to one's wants* (bottom panel, right).

In the top panel of Figure 3.8, it can be seen how often children used absolute versus relative/comparative strategies, for each item. In the middle panel of the figure, it can be seen how often the different content categories of the absolute strategy (left) and the relative strategy (right) were used. As can be seen, within each of the five items, the content category ‘social relationships’ was used most frequently in the absolute strategy. In the relative strategy, the content category ‘comparison to one’s wants’ was used most frequently for each of the five items. In the bottom panel of the figure, the subcategories for the ‘social relationship’ content category (left) and the subcategories for the ‘comparison to one’s wants’ content category (right) are displayed, showing how often each of the respective subcategories occurred. (Note: If the total numbers of the bars in the middle and lower figures do not correspond to their respectively corresponding bars in the figure(s) one level above, it is because the codes that fell under ‘other’ are left out of these figures.)

Figure 3.8 illustrates several interesting patterns. First of all, the absolute strategy is used more frequently for items 1, 2, and 3, but the relative strategy is used more frequently for items 4 and 5. The difference in the use of strategies is most pronounced for items 2, 3, and 5. With regard to the content categories for the absolute strategy, ‘social relationships’ were mentioned most frequently in the children’s responses for each of the five items. The content category ‘time use’ only occurred (three or more times) in items 1, 2, and 3; which are the three items that do not make a reference to a time frame (Item 4: ‘*So far*, I have gotten ...’; Item 5: ‘If I *could live my live over*, I would ...’). The content category ‘possessions’ occurred most frequently for items 2 and 4, both of which contain the word ‘things’ in it (Item 2: ‘The *things* in my life are excellent.’; Item 4: ‘So

far, I have gotten the important *things* I want in life.’). For the relative strategy, the content category ‘comparisons to one’s wants’ occurred most frequently for all five items. Comparisons to one’s past were made for items 3, 4, and 5—with items 4 and 5 being the two items that make an explicit reference to a time frame.

Within the content category ‘social relationships’ (bottom panel, left), it can be seen that children most frequently made references to their ‘family’ in their responses, and that ‘friends’ were mentioned with the second-highest frequency. The subcategory ‘bullying’ solely occurred for items 1 and 3. With respect to the content category ‘comparisons to one’s wants’, it can be seen that ‘belongings’ were most frequently mentioned by children in response to item 4, which makes reference to the past (‘*So far, ...*’), and to ‘things’.

Comparison of the use of strategies within items

The use of the relative versus the absolute strategy was compared overall and separately for the two grade groups for each item. For the comparison, each child received a binary code (0 or 1) depending on whether s/he used the absolute or relative strategy or not, and then the McNemar test, a paired test of equality proportions, was calculated. With regard to item 1, there were no statistically significant differences overall or between the grade groups²². The results for item 2 indicate that there was a statistically significant difference in the use of the absolute versus the relative strategy overall, with the absolute strategy being used more often ($\chi^2(1) = 12.50$; $p = .0001$; OR = 3.0). This difference was only statistically significant in the younger grade group; i.e., the

²² According to Figure 3.8, it appears as if there should be a statistically significant difference in favour of the absolute strategy. Please note that in contrast to the data in Figure 3.8, the data for the McNemar test were recoded into binary data, which explains why the test was statistically non-significant.

younger children used the absolute strategy statistically significantly more often than the relative strategy (exact significance $p = .001$; OR = 6.0). With regard to item 3, there was a statistically significant difference in the overall use of the absolute versus the relative strategy, with the absolute one being used more frequently ($\chi^2(1) = 4.45$; $p = .04$; OR = 2.1), but there were no statistically significant differences within the grade groups. With regard to item 4, there was a statistically significant difference in the overall use of the strategies, with the relative one being used more frequently ($\chi^2(1) = 7.23$; $p = .007$; OR = 2.7); this difference was only statistically significant within the older grade group ($p = .04$; OR = 2.9). With regard to item 5, there was a statistically significant difference in the overall use of the strategies, with the relative one being used more frequently ($\chi^2(1) = 9.82$; $p = .002$; OR = 3.6); this difference was statistically significant only within the older grade group ($p = .0001$; OR = 9.2).

Relationship to demographic variables

In the next step, it was investigated whether there are statistically significant differences with regard to demographic variables when using the absolute or relative strategies. Therefore, Poisson or binary logistic regression analyses (depending on whether the data were counts or binary) were run with the factors of gender, grade, and first language background. The results indicate that, for the relative strategy, there was only one statistically significant result, namely for gender on item 3. Specifically, girls used the relative strategy significantly more often than boys (Wald $\chi^2(1) = 4.24$; $p = .04$; OR = 5.5). With regard to the absolute strategy, there was also only one statistically significant result. Specifically, girls used the absolute strategy significantly more often

than boys on item 4 (Wald $\chi^2(1) = 6.67$; $p = .01$; expected rate for girls = .97; expected rate for boys = .35).

Correlations between valence of the responses and SWLS-C item scores

In order to calculate the correlations between the valence of children's responses and the SWLS-C scores, the positive statements were coded as +1, the negative statements as -1, and the neutral statements as 0, and a sum score was calculated for each item. This score was then correlated with the children's respective item scores. For all five items, statistically significant ($p \leq .001$), positive Spearman rank correlations were found (the SWLS-C item mean and standard deviations (SD) are provided in parentheses): Item 1: $r = .56$ (mean = 4.1; SD = .92); Item 2: $r = .43$ (mean = 4.2; SD = .83); Item 3: $r = .44$ (mean = 4.6; SD = .63); Item 4: $r = .48$ (mean = 4.4; SD = .87); Item 5: $r = .66$ (mean = 4.0; SD = 1.28).

Feedback on the items

Of the 23 children who were asked for their feedback on the items, 22 responded that they thought it is important to give these items to children and that they enjoyed responding to them. Specifically, several children said that it was a good way to find out how children their age are feeling, for example: "So you can know how they're feeling in life and - like how they're feeling with their families, friends, teachers and stuff like that."; "Because it's easier then to understand how at our age people think. And what's happening at home and their life, if they're stressed out or not."; "I think you should know what's going on in their heads, because a lot of kids have problems. And they don't

talk about it. So, you need to know this stuff.”; “It really helps them just [to] get their feelings out. Instead of holding all their feelings inside.”

Furthermore, several children said that this would be a good way to get information that would be important to help children, for example: “Because if you wanted to change something and if most people say it, then you could change it.”; “So then people can help us more.”

In addition, several children mentioned that they enjoyed answering the items, for example: “It’s good, because I never even thought about these questions before in my life.”; “Because you’re asking them what they like the most. And what they do or they don’t like the most. So they’re encouraging.”; “Because then you can think of your life a bit. And see that maybe you made a mistake in your life and then you said it in here, realizing that you did make a mistake, so that you can fix the mistake over in your life if it ever happens again.”

One child also mentioned that it would be good to give this scale to older students in high school “because... they have too much homework. They’re stressing out and stuff. They have lots of problems in their life.”

Discussion

The purpose of the present study was to investigate the cognitive processes of children when responding to the items of the SWLS-C to provide evidence for the substantive aspect of construct validity. This study showed that children used two main strategies to answer the items on life satisfaction, namely an absolute strategy and a relative or comparative strategy. In the former, children referred to the presence or absence of something that was of relevance for their satisfaction with life. In the latter, children made comparisons of their current state to what they want, what others have, what they had in the past, and what they need to rate their life satisfaction. The presented findings are in line with the *multiple discrepancies theory* (MDT; Michalos, 1985) in several regards. MDT makes several propositions about the processes used by individuals to make judgments on their life satisfaction and domain satisfaction. The first proposition of MDT postulates that reported net satisfaction is a function of perceived discrepancies between what an individual currently has compared to (i) what s/he wants ('self-want'), (ii) what relevant others have ('self-others'), (iii) the best s/he has had in the past ('self-best past'), (iv) what s/he expected to have 3 years ago at this point in life ('self-progress'), (v) what s/he expects to have after 5 years ('self-future'), (vi) what s/he deserves ('self-deserves'), and (vii) what s/he needs ('self-needs').

The MDT also proposes that the discrepancy between what an individual currently has and what s/he wants is a mediating variable between the other discrepancies and life satisfaction (Michalos, 1985, pp. 347-348)²³. Even though the mediation could not be tested with the present data, it is of interest to note that the children used the self-want comparison with the highest frequency. This finding suggests that children assign a

²³ For the other propositions, please see Michalos (1985).

particular importance to the self-want category in their judgment of life satisfaction. Furthermore, the presented findings show parallels to findings from previous studies that tested the MDT with university students (Michalos, 1985; 1991). In particular, Michalos (1985) tested how successfully the MDT could be used to predict/explain life satisfaction in a Canadian undergraduate student sample. In that study, the discrepancies that were most salient with regard to predicting/explaining variance in the students' life satisfaction ratings were—in order—self-want, self-others, self-needs, self-best past, self-deserved, self-progress, and self-future. Similarly, in a study that investigated the relative importance of the discrepancies with regard to the prediction of life satisfaction in a large sample of undergraduates from 39 countries, the self-wants and the self-others discrepancies had the largest impact (Michalos, 1991). The findings of the relative strategy show that children in grades 4 to 7 use some of the same discrepancies to make evaluations of their satisfaction with life when responding to the items of the SWLS-C. Particularly, the four discrepancies that were used by the children are the ones that were most successfully predicting life satisfaction in those previous studies, namely the self-want, self-past, self-need, and self-other discrepancies (ordered according to frequency of occurrence in children's responses). It needs to be pointed out that the self-past discrepancy was used differently by the children than it is conceptualized in MDT. In MDT it is the discrepancy between what one currently has and the best one has had in the past. In contrast, the children were mostly making comparisons with the past, where their lives or a specific occurrence in the past was considered to be negative, and they were commenting on the improvement in their lives since then.

Furthermore, the presented findings show some similarities with Cremeens et al.'s (2006a) findings, regardless of the fact that the items of the measure in Cremeens et al.'s study and the SWLS-C are quite different. The items of the TedQL used by Cremeens et al. are quite specific (addressing abilities, such as children's reading ability, or social aspects, such as having friends at school), whereas the items of the SWLS-C are more general (pertaining to overall evaluations of children's lives). Also, the children in the study by Cremeens et al. were younger than the ones in this study (mean age of 7.1 versus 11.0 years). These differences notwithstanding, there is some overlap in the strategies that children used in responding to the respective measures. Specifically, Cremeens et al. report that children used social comparisons for answering the items, which was also found for item 2. Furthermore, they report that children used stable character references, which in the present study was coded under the absolute strategy and the content category personal characteristics and was used for items 1 to 4. In addition, Cremeens et al. report on children using concrete examples as a strategy, which was also present in children's responses to the SWLS-C, but which was not coded as a strategy in itself, as it occurred within the different strategies when children used concrete examples for illustrative purposes. Lastly, they report on other reasons or no reason given, which is similar to the strategies termed general positive strategy and unclear strategy in the presented study.

In regard to the demographic variables of gender, first language background, and grade, the regression analyses with the relative and absolute strategies as dependent variables did not show any systematic patterns across the five items of the SWLS-C, but

it would be of interest to investigate these relationships in future studies with a larger sample size, and a larger range of age/grades.

In a separate set of analyses, it was examined whether children's use of the absolute or relative strategy was associated with the (wording of the) items. The results indicate that the relative strategy was used more frequently than the absolute one when children responded to the two items that make reference to the past (items 4—'So far, ...'—and item 5—"If I could live my life over, ...'), whereas the absolute strategy was used more frequently for the two items that make reference to the present (items 2 and 3). (There were no statistically significant differences in the use of strategies for item 1). When looking at the response strategies children used within the respective grade groups of younger (grade 4 and 5) and older (grade 6 and 7) children, it was found that older children used the relative strategy significantly more often than the absolute strategy for items 4 and 5, whereas there was no such difference for the younger children. For item 2, the younger children were more likely to use the absolute than the relative strategy, whereas there was no difference for the older children. It goes beyond the scope of this study to speculate about the reasons for this. It might be the case that the reference to the past in items 4 and 5 is more likely to elicit a relative strategy rather than an absolute strategy in children, and particularly for older children. It would be of interest to examine in future studies with a larger sample and age range whether age-related cognitive development is associated with specific response strategies in response to the SWLS-C items. In fact, an age-related pattern could be expected based on developmental theories that propose that children's understanding of self becomes increasingly specific during middle childhood because their cognitive skills become more complex (Stone &

Lemanek, 1990; De Civita et al., 1990), and because their self-descriptions become more comparative (Bee, 1989).

With regard to the content topics that came up for the children when responding to the items, the content category of the absolute strategy that was used most frequently was 'social relationships', which was used by the children for all five items, with social relationships with family members being especially prominent. This indicates the importance of social relationships, especially with family members, for children's life satisfaction, which is in line with previous empirical research (Huebner, Suldo, Smith, & McKnight, 2004). Huebner (1991) reports that the strongest association between global life satisfaction and domain satisfaction ratings was with the domain family, but the relationship to the domain peers was also significant for children in grades 5 to 7. Similarly, Man (1991) found that parent orientation had a stronger relationship to life satisfaction than peer orientation with adolescents. In the present study, children often mentioned parental support during the think-aloud procedure. Young, Miller, Norton, and Hill (1995) also report that perceived parental support was positively correlated with adolescents' ratings of life satisfaction.

For the relative strategy, the content category that was used most frequently was 'comparisons to one's want', which was utilized by the children in responding to all five items (this was also the discrepancy with the highest success rate in the test of MDT). Within the self-want category, children most frequently referred to belongings. Similarly, 'possessions' was a content category of the absolute strategy that was frequently used. The school in which the study was conducted is located in a neighborhood with relatively low socio-economic status, and several children said that their families do not have

enough money to buy them certain things. At the same time, many children were also commenting on the things they (or their family) owned and which they considered important. Furthermore, several children were weighing the things they owned with the ones they would have liked to have had, typically arriving at a positive judgment of their life satisfaction. Empirical findings on the relationship between socio-economic status and life satisfaction for children and adolescents have been ambiguous, with some studies reporting a statistically significant association of moderate effect size (e.g., Dew & Huebner, 1994), and other studies reporting a statistically non-significant correlation of negligible effect size (e.g., Huebner, 1991). It would be of interest to investigate whether children from a different socio-economic background also mention belongings or possessions as frequently in think-aloud protocols.

With regard to the valence of the children's statements, children predominantly talked about positive experiences and aspects of their lives (i.e., the presence of something positive or the absence of something negative). In fact, 70% of the statements were of positive valence across the items. In contrast, 22% of the statements were of negative valence (i.e., the presence of something negative or the absence of something positive) and 8% were of mixed valence (both positive and negative). This suggests that (most) children are predominantly thinking about positive experiences and aspects of their lives when making judgments on their life satisfaction. In addition, the valence of children's responses to the items was positively related to the respective item scores. Also, the item mean scores were all equal to or above 4.0, indicating that the children in this sample, on average, rated their life satisfaction as positive. These findings are in line with previous empirical research that has shown that most children and adolescents rated

their life satisfaction positively (e.g., Gadermann et al., in press; Greenspoon & Saklofske, 1997; Huebner & Alderman, 1993; Huebner, Drane, & Valois, 2000).

The majority of the children did not have any difficulties with the item content or the response format of the SWLS-C. However, several children found it difficult to respond to items 1 and 5 (7 and 5 children, respectively). These children were slightly younger than the overall sample (mean age of 10.1 years) and, with the exception of 2 children, all were bilingual. I am hesitant to recommend any changes to the item wording based on the children's feedback as this was quite diverse with regard to the response format and item wording. Furthermore, these items were performing well in previous pilot studies with focus groups of children and in a psychometric analysis with a larger sample (Gadermann et al., in press). However, it is recommended that future studies validating the SWLS-C should have a special focus on these two items.

In a previous study, the SWLS-C showed favourable psychometric properties in terms of reliability, factor structure, differential item and scale functioning, and correlations to convergent and discriminant measures (Gadermann et al., in press). The aim of the present study was to add to the validity evidence by evaluating the substantive aspect of construct validity by investigating the cognitive processes of children when responding to the items of the SWLS-C. This aspect of construct validity is often not investigated (Cizek et al., 2008), although it is one of the six aspects of construct validity proposed by Messick (1995) as well as one of the five sources of evidence proposed by the *Standards for Educational and Psychological Testing* (APA, AERA, & NCME, 1999). This study provides critical evidence with regard to the substantive aspect of validity, as it provides insights into the strategies that children used to respond to the

SWLS-C, and as these strategies are, in turn, congruent with theoretical considerations pertaining to the construct of subjective well-being/quality of life. In other words, the findings illustrate that children's item responses were governed by strategies that are meaningful and reflect ideas in the subjective well-being/quality of life literature. Specifically, the strategies and content of the children's responses were theoretically in line with MDT and also converge with previous empirical findings in the subjective well-being/quality of life literature with children and adolescents, as described above. Additionally, the results indicate that the majority of the children did not have any difficulties in understanding the items. From a practical, applied perspective, it is also important to highlight the finding that the children enjoyed responding to the SWLS-C and thought it was important to ask children their age these questions.

Messick (1995) stated that "validity is an evolving property and validation a continuing process" (p. 741). In future studies, it would be of interest to investigate and compare the cognitive processes that are employed by children and adolescents of different age groups, with different socio-economic background, and of diverse ethno-cultural background when responding to the items of the SWLS-C (i.e., to investigate the generalizability aspect of construct validity). Furthermore, it will be important for future research to monitor for which purposes the SWLS-C is administered and to critically investigate the intended and unintended consequences of the use and interpretation of the SWLS-C scores with regard to these purposes (i.e., to investigate the consequential aspect of construct validity).

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Barofsky, I., Meadows, K., & McColl, E. (2003). Cognitive aspects of survey methodology and quality of life assessment: Summary of meeting. *Quality of Life Research, 12*, 281-282.
- Bee, H. (1989). *The developing child* (5th ed.). New York: Harper & Row Publishers.
- Ben-Arieh, A. (2005). Where are the children? Children's role in measuring and monitoring their well-being. *Social Indicators Research, 74*, 573-596.
- Ben-Arieh, A. (2006). Is the study of the "State of our children" changing? Re-visiting after 5 years. *Children and Youth Services Review, 28*, 799-811.
- Ben-Arieh, A., & Goerge, R. M. (2001). Beyond the numbers: How do we monitor the state of our children? *Children and Youth Services Review, 23*, 603-631.
- Ben-Arieh, A., & Goerge, R. M. (Eds.) (2006). *Indicators of children's well-being: Understanding their role, usage and policy influence*. Dordrecht, The Netherlands: Springer Academic Publishers.
- Ben-Arieh, A., Kaufman, H. N., Andrews, B. A., Goerge, R. M., Lee, B. J., & Aber, J. L. (2001). *Measuring and monitoring children's well-being*. Dordrecht, The Netherlands: Kluwer Academic Press.
- Berg, B. L. (2004). *Qualitative Research Methods for the Social Sciences* (5th edition). Boston: Pearson Education.

- Böhm, A. (2004). Theoretical coding: Text analysis in grounded theory. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 270-275). London: Sage Publications.
- Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*, 397-412.
- Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12*, 229-238.
- Cremeens, J., Eiser, C., & Blades, M. (2006a). A qualitative investigation of school-aged children's answers to items from a generic quality of life measure. *Child: Care, Health, & Development, 33*, 83-89.
- Cremeens, J., Eiser, C., & Blades, M. (2006b). Characteristics of health-related self-report measures for children aged three to eight years: A review of the literature. *Quality of Life Research, 15*, 739-754.
- De Civita, M., Regier, D., Alamgir, A. H., Anis, A. H., Fitzgerald, M. J., & Marra, C. A. (2005). Evaluating health-related quality-of-life studies in paediatric populations. Some conceptual, methodological and developmental considerations and recent applications. *Pharmacoeconomics, 23*, 659-685.
- Dew, T., & Huebner, E. S. (1994). Adolescents' perceived quality of life: An exploratory investigation. *Journal of School Psychology, 32*, 185-199.
- Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London: Routledge.

- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment, 49*, 71-75.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*, 276-302.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.
- Fox, J. E., Houston, B. K., & Pittner, M. S. (1983). Trait anxiety and children's cognitive behaviors in an evaluative situation. *Cognitive Therapy and Research, 7*, 149-154.
- Frones, I. (2007). Theorizing indicators. *Social Indicators Research, 83*, 5-23.
- Gadermann, A. M., Schonert-Reichl, K. A., & Zumbo, B. D. (in press). Investigating validity evidence of the Satisfaction with Life Scale adapted for Children. *Social Indicators Research*.
- Greenspoon, P. J., & Saklofske, D. H. (1997). Validity and reliability of the Multidimensional Students' Life Satisfaction Scale with Canadian children. *Journal of Psychoeducational Assessment, 15*, 138-155.
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley Publishing Company.
- Huebner, E. S. (1991). Correlates of life satisfaction in children. *School Psychology Quarterly, 6*, 103-111.
- Huebner, E. S., & Alderman, G. L. (1993). Convergent and discriminant validation of a children's life satisfaction scale: Its relationship to self- and teacher-reported psychological problems and school functioning. *Social Indicators Research, 30*, 71-82.

- Huebner, E. S., Drane, W., & Valois, R. F. (2000). Levels and demographic correlates of adolescent life satisfaction reports. *School Psychology International, 21*, 281-292.
- Huebner, E. S., Suldo, S. M., Smith, L. C., & McKnight, C. G. (2004). Life satisfaction in children and youth: Empirical foundations and implications for school psychologists. *Psychology in the Schools, 41*, 81-93.
- Lodge, J., Harte, D. K., & Tripp, G. (1998). Children's self-talk under conditions of mild anxiety. *Journal of Anxiety Disorders, 12*, 153-176.
- Lodge, J., Tripp, G., & Harte, D. K. (2000). Think-aloud, thought-listing, and video-mediated recall procedures in the assessment of children's self-talk. *Cognitive Therapy and Research, 24*, 399-418.
- Man, P. (1991). The influence of peers and parents on youth life satisfaction in Hong Kong. *Social Indicators Research, 24*, 347-365.
- Mayring, P. (2004). Qualitative content analysis. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 266-269). London: Sage Publications.
- McColl, E., Meadows, K., & Barofsky, I. (2003). Cognitive aspects of survey methodology and quality of life assessment. *Quality of Life Research, 12*, 217-218.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 12-103). New York: Macmillan.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

- Michalos, A. C. (1985). Multiple discrepancies theory. *Social Indicators Research, 16*, 347-413.
- Michalos, A. C. (1991). *Global report on student well-being: Life satisfaction and happiness* (Vol. 1). New York: Springer.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.
- Rebok, G., Riley, A., Forrest, C., Starfield, B., Green, B., Robertson, J., et al. (2001). Elementary school-aged children's reports of their health: A cognitive interviewing study. *Quality of Life Research, 10*, 59-70.
- Riley, A. W. (2004). Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics, 4*, 371-376.
- Schilling, L. S., Dixon, J. K., Knafl, K. A., Grey, M., Ives, B., & Lynn, M. R. (2007). Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. *Nursing Research, 56*, 361-366.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.
- Sirken, M. G., & Schechter, S. (1999). Interdisciplinary survey methods research. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and Survey Research* (pp. 1-10). New York: Wiley.
- Stewart, J. L., Lynn, M. R., & Mishel, M. H. (2005). Evaluating content validity for children's self-report instruments using children as content experts. *Nursing Research, 54*, 414-418.

- Stone, W. L., & Lemanek, K. L. (1990). Developmental issues in children's self-reports. In A. M. La Greca (Ed.), *Through the eyes of the child: Obtaining self-reports from children and adolescents*. Boston: Allyn & Bacon
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.
- Willis, G. B., DeMaio, T. J., & Harris-Kojetin, B. (1999). Is the Bandwagon headed to the methodological promised land? Evaluating the validity of cognitive interviewing techniques. In M. G. Sirken, D. J. Herrmann, S. Schechter, N. Schwarz, J. M. Tanur, & R. Tourangeau (Eds.), *Cognition and Survey Research* (pp. 133-153). New York: Wiley.
- Young, M. H., Miller, B. C., Norton, M. C., & Hill, E. J. (1995). The effect of parental supportive behaviors on life satisfaction of adolescent offspring. *Journal of Marriage & Family*, 57, 813-822.
- Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26: Psychometrics* (pp. 45-79). Amsterdam: Elsevier Science B. V.
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.) *The concept of validity: Revisions, new directions and applications* (pp. 65-82). Charlotte, NC: Information Age Publishing.

Concluding chapter

The purpose of this dissertation was to investigate several aspects of construct validity of the Satisfaction with Life Scale adapted for Children (SWLS-C). This was addressed in two studies. The first study investigated the structural and external aspects of construct validity with regard to the SWLS-C by examining the dimensionality, reliability, and differential item functioning of the SWLS-C, and by examining the relationships between the SWLS-C and other variables in a sample of children in grades 4 to 7. The second study addressed the substantive aspect of construct validity of the SWLS-C by investigating the cognitive processes of an independent sample of children in grades 4 to 7 when responding to the items of the SWLS-C.

Messick (1995) highlights six aspects of construct validity as “general validity criteria or standards for all educational and psychological measurement” (p. 741). Similarly, the *Standards for educational and psychological testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME), 1999) provide five sources of validity evidence that can be used in evaluating the score interpretations of measurement instruments²⁴.

Out of the six aspects of construct validity (in the language of Messick) or the five sources of validity evidence (in the language of the *Standards*), I addressed three in the presented studies. As validation is an ongoing process, the two studies of my dissertation can thus be seen as part of a larger validation research program for the SWLS-C. In this concluding chapter, I will highlight what the two presented studies contributed to this

²⁴ The six aspects of construct validity (Messick, 1995) and the five sources of validity evidence (AERA, APA, & NCME, 1999) are described in the introduction.

validation research program and will delineate how future validation research can complement these studies.

Novel contributions of the presented research

The two validation studies of this dissertation provide examples of the important intersection of methodology/measurement and life satisfaction research with regard to children. The contribution of this dissertation is twofold:

(i) This dissertation contributes to measurement and validation practice in general by providing two examples of systematic validation research. Study 1 parallels previous validation research for measures of children's well-being, as the study validates the inferences of the SWLS-C by investigating its dimensionality and reliability, as well as by examining its relationships to other measures of interest (e.g., Huebner, 1991a; Neto, 1993). In addition, study 1 investigated measurement invariance at the item and scale level across different subgroups; a component of validation research that is often overlooked (cf. Cizek, Rosenberg, & Koons, 2008). With regard to making a contribution to the field of measurement practice, study 2 needs to be especially highlighted. As I described in the introduction, much validation research has mainly focused on correlational evidence, and cognitive processes have rarely been investigated (Cizek et al., 2008). Study 2 investigated the cognitive processes of children when responding to the SWLS-C and provides important insights into children's thought processes. Hence it becomes an exemplar for other work that aims at investigating this aspect of construct validity and, in so doing, it pushes forward the paradigm that has been advocated by the *Standards*, by Messick, and others.

(ii) This dissertation contributes to the measurement of children’s satisfaction with life in particular by systematically investigating the structural, external, and substantive aspects of construct validity of the SWLS-C. Specifically, the presented studies provide evidence that one can meaningfully interpret the scores derived from children’s responses to the SWLS-C. This is in line with Messick’s (1995) approach, which views “validation of inferences” as the “scientific inquiry into score meaning” (p. 741), as well as Zumbo’s (2007; 2009) explanatory-focused approach to validation/validity. Specifically, based on the findings of the two studies, one can infer that children’s responses were not ‘random check mark behavior’, but that they were based on an interpretable process, which relates to existing theoretical ideas (i.e., multiple discrepancies theory; Michalos, 1985) and empirical findings in the literature on life satisfaction (and subjective well-being/quality of life; e.g., Huebner, 1991b, 1991c; Neto, 1993; Piko, 2006). Accordingly, one can argue that the variation that is observed in the SWLS-C scale scores is, in fact, attributable to the construct the SWLS-C is intended to measure.

These contributions are critical for the further use of the SWLS-C in research. At the same time, it needs to be reiterated that validity is an evolving context and validation an ongoing process—a process that Zumbo (2009) likened to the process of building and rebuilding the ship at sea. The findings from the presented studies may thus not be generalized to other contexts without scrutinizing to what extent and for which contexts such generalizations are defensible (see also Messick, 1995). This issue is further addressed in the following section.

Limitations of the presented studies and future directions

The presented studies provide systematic evaluations of several aspects of construct validity. As mentioned above, Messick (1995) described six aspects of construct validity that can be used to evaluate measurement instruments. My dissertation addressed three of these aspects (the substantive, structural, and external aspects of construct validity) and it would be crucial for future research to investigate the other three, namely the content, generalizability, and consequential aspects of construct validity. Although all six aspects of construct validity are important to be evaluated, depending on the measurement instrument under investigation, different aspects may need to be prioritized in the validation process.

The content aspect of construct validity, for example, is of special importance with regard to measures that assess different domains of satisfaction with life, because such measures include a certain number of domains chosen from the large number of potential domains (Diener, Lucas, Schimmack, & Helliwell, 2009). An investigation of the content aspect of construct validity would, in that case, investigate the content relevance and representativeness of the chosen domains, for example, by interviewing subject matter experts and/or members of the target population. In contrast, the SWLS-C assesses global life satisfaction and asks for overall evaluations of children's lives, and does not focus on specific life domains. Arguably, there is thus no sampling issue (with regard to life satisfaction domains) at hand; therefore, the investigation of the content aspect of construct validity for an overall life satisfaction measure might not be as crucial (Diener et al., 2009). This argument notwithstanding, it would be of interest for future studies to investigate whether children's responses on the SWLS-C reflect construct-

irrelevant variance (e.g., the inclusion of other constructs, such as social desirability). Future studies that examine the content aspect of construct validity for the SWLS-C would therefore need to test whether this is the case or not, for example, by conducting focus groups or interviews with children.

Furthermore, it is especially important that in future studies researchers investigate the generalizability and consequential aspects of construct validity. The two studies of my dissertation were conducted in the same geographical region and, although the sample composition of both samples was diverse in terms of children's ethno-cultural background, it would be of great interest to validate the SWLS-C with samples from different cultural groups, in different regions/countries, and with different age groups in order to investigate to what degree the score properties and interpretations generalize to these groups of children in different contexts. In addition, another important line of research would be to compare the findings with the SWLS-C with children to the findings with the SWLS with adults. Specifically, one advantage of the adaptation of the SWLS for the use with children was that this will potentially allow one to make direct comparisons between satisfaction with life scores of adults and children, as well as to investigate change in satisfaction with life in longitudinal studies using the SWLS-C for children and the SWLS for adults. However, such comparisons presuppose that the underlying conceptual structure of satisfaction with life is invariant across children and adults, i.e., that the two measures are psychometrically equivalent. If that is not the case, the comparison might be confounded and one might not compare the same construct. Therefore, for future studies it would be of great interest to investigate measurement invariance across children and adults, using the SWLS-C and the SWLS, respectively,

with psychometric methods such as differential item functioning or multi-group factor analysis.

With regard to the consequential aspect, future studies need to investigate the intended and unintended consequences of the score interpretation and use of the SWLS-C. Specifically, it needs to be evaluated whether potential negative consequences of the interpretation and use of the SWLS-C are due to aspects of scale invalidity, such as construct-irrelevant variance (Messick, 1989). For example, it has been argued that, after more validation work, life satisfaction measures may be useful in different clinical contexts for children and adolescents, in combination with measures assessing psychopathology, to provide a more comprehensive assessment (Gilman & Huebner, 2000; Gilman & Huebner, 2003). In fact, Diener et al. (2009) have proposed that regular school check-ups of children's subjective well-being and mental health might be useful, in combination with check-ups on physical health, in terms of identifying (groups of) children with low subjective well-being and/or mental health problems, and for using that information to inform the provision of support and services and the implementation of (intervention) programs for those children. It needs to be emphasized that Diener et al. (2009) explicitly base their proposal on the prerequisite that "valid tools and interventions are available" (p.142). In other words, if a measure such as the SWLS-C were to be used for such purpose, it would first have to be investigated whether the scores of the measure can be utilized for screening or diagnostic purposes. Second, potential negative consequences of the use and interpretation of the SWLS-C due to scale invalidity (e.g., construct-irrelevant variance due to social desirability) would need to be investigated. For example, there might be negative consequences for children that are

misidentified/mislabeled, which may cause unwarranted stress to the children and/or their families. Likewise, a measure's lack of sensitivity might lead to a scenario where children that might benefit from interventions will not receive such support.

These examples illustrate that a comprehensive, theoretically guided validation framework is critical for any measure, but especially for measures that are (potentially) administered at a large scale and may (potentially) have far-reaching consequences. Therefore, a close, bi-directional communication process between those that conduct research, administer the measure, interpret and communicate results, and use the results for decision making is necessary.

Implications of the research findings

A systematic, ongoing validation process that takes into account the six aspects of construct validity may be considered ideal from a validity perspective, but it must be acknowledged that such a process not only takes time and resources, but may, in practice, not be completely feasible. This opens up the question of "At what stage can one begin using a measure?" or "How much and what validity evidence is needed?". A deliberate response to this issue has been offered by Zumbo (2009), who notes "one can start (cautiously) using [a] measure as one gains a deeper understanding and explanation, but that the stakes for the measurement use should guide this judgment" (p. 75). In other words, it may be suggested that the stakes involved in a measurement usage should only increase according to the extent to which associated validity evidence accumulates. The presented studies provide initial evidence for the meaningfulness of a number of inferences that are based on children's responses to the SWLS-C. The findings from this

dissertation indicate that the SWLS-C may be appropriately used for research purposes with children in middle childhood and in contexts similar to the one described in the presented study, as the SWLS-C scores were unidimensional for a representative sample of children, showed convergent and discriminant validity in theoretically predicted ways with other measures (e.g., of optimism), and measured satisfaction with life in the same way for different groups of children. In addition, the findings indicate that children from diverse (language) backgrounds use strategies and content categories in their responses to the SWLS-C that are in line with Michalos' (1985) MDT as well as with previous research. On the other hand, the SWLS-C cannot (yet) be recommended for diagnostics or for classification purposes of individual children at this stage, because there is no validity evidence available with regard to such application. Future studies should be taking the validation research program to the next stage as delineated in this discussion. The presented findings and the theoretical arguments for a comprehensive validation program pertaining to the SWLS-C make a useful, practicable contribution in this regard. More importantly, it is hoped that the presented research contributes to a process that culminates in research as well as the translation of research into practice that fosters the motivating cause of this research, namely the goal to benefit the well-being of children and youth.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement, 68*, 397-412.
- Diener, E., Lucas, R. E., Schimmack, U., & Helliwell, J. F. (2009). *Well-being for public policy*. New York: Oxford University Press.
- Gilman, R., & Huebner, E. S. (2000). Review of life satisfaction measures for adolescents. *Behaviour Change, 17*, 178-183.
- Gilman, R., & Huebner, E. S. (2003). A review of life satisfaction research with children and adolescents. *School Psychology Quarterly, 18*, 192-205.
- Huebner, E. S. (1991a). Further validation of the Students' Life Satisfaction Scale: The independence of satisfaction and affect ratings. *Journal of Psychoeducational Assessment, 9*, 363-368.
- Huebner, E. S. (1991b). Initial development of the Students' Life Satisfaction Scale. *School Psychology International, 12*, 231-240.
- Huebner, E. S. (1991c). Correlates of life satisfaction in children. *School Psychology Quarterly, 6*, 103-111.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 12-103). New York: Macmillan.

- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.
- Michalos, A. C. (1985). Multiple discrepancies theory. *Social Indicators Research, 16*, 347-413.
- Neto, F. (1993). The Satisfaction with Life Scale: Psychometric properties in an adolescent sample. *Journal of Youth and Adolescence, 22*, 125-134.
- Piko, B. F. (2006). Satisfaction with life, psychosocial health and materialism among Hungarian youth. *Journal of Health Psychology, 11*, 827-831.
- Zumbo, B. D. (2007). Validity: Foundational issues and statistical methodology. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of statistics, Vol. 26: Psychometrics* (pp. 45-79). Amsterdam: Elsevier Science B. V.
- Zumbo, B. D. (2009). Validity as contextualized and pragmatic explanation, and its implications for validation practice. In R. W. Lissitz (Ed.) *The concept of validity: Revisions, new directions and applications* (pp. 65-82). Charlotte, NC: Information Age Publishing.

Appendices

Appendix A: Behavioural Research Ethics Board certificate of approval, student assent form, and parent consent form for the study in chapter 2



The University of British Columbia
Office of Research Services and Administration
Behavioural Research Ethics Board

Certificate of Approval

PRINCIPAL INVESTIGATOR Schonert-Reichl, K.A.		DEPARTMENT Educ & Couns Psych & Spec Educ	NUMBER B05-0921
INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT Elementary School ,			
CO-INVESTIGATORS: Buote, Denise, Counselling Psychology; Jaramillo, Angela, Educational Studies			
SPONSORING AGENCIES United Way of the Lower Mainland			
TITLE: What Do Kids Do When They Are Not In School? The Experiences of Canadian Children during Out-of-School Time			
APPROVAL DATE DEC 15 2005	TERM (YEARS) 1	DOCUMENTS INCLUDED IN THIS APPROVAL: Nov. 14, 2005, Consent forms / Assent form / Questionnaires / Sept. 12, 2005, Contact letter / Cover letter	
<p>CERTIFICATION:</p> <p>The protocol describing the above-named project has been reviewed by the Committee and the procedures were found to be acceptable on ethical grounds for research involving human subjects.</p> <div style="text-align: center;">  <p><i>Approved on behalf of the Behavioural Research Ethics Board by one of the following:</i> Dr. Peter Suedfeld, Chair, Dr. Susan Rowley, Associate Chair</p> </div> <p>This Certificate of Approval is valid for the above term provided there is no change in the experimental procedures</p>			



THE UNIVERSITY OF BRITISH COLUMBIA



Student Assent Form

January, 2006

**Department of Educational and Counselling
Psychology, and Special Education**

Faculty of Education

2125 Main Mall

Vancouver, B.C., V6T 1Z4

Dear Participating Student,

The purpose of this form is to give you the information you need in order to decide whether or not you want to be in our research study entitled **“What do kids do when they are not in School? The Experiences of Canadian Children during Out-of-School Time.”**

Purpose

The purpose of our study is to find out how intermediate grade children spend their time outside of school. This study is the first step undertaken by Dr. Kim Schonert-Reichl with the support of the United Way of the Lower Mainland in order that Lower Mainland communities have information about how children spend their out of school time during middle childhood.

Study Procedures

There are two parts to the questionnaires. In the first part, you will be asked to report on how you spend your out of school time during a typical week and how satisfied you are with your out of school time. This will be done in the classroom for five consecutive days and will take approximately ten minutes per day. You will do this via a questionnaire which will be placed in a sealed envelope so that all answers are confidential. In the second part of our questionnaires, you will be asked to provide information on your feelings about yourself, your classroom, and your relationships with peers, parents and other adults. Completion of these questionnaires will take approximately two class periods of 45-60 minutes each. **THIS IS NOT A TEST.** There are no right or wrong answers. We simply want to know where children are during out of school time and the nature of the activities in which they engage during their out of school time as well as how children understand themselves and others. In addition, information relating to

school attendance, and school achievement (marks) will be collected from your school records and from the BC Ministry of Education (Foundation Skills Assessment). Teachers will also be asked to complete a checklist assessing various dimensions of your social, emotional, academic and physical well-being.

Confidentiality

Remember no one at school or in your community (not even your parents/guardians, teacher, or school principal) will ever see your answers (they will be confidential). We will keep your answers in locked cabinets at UBC. No names will be used when the information is studied. In this way, the information that you give us will be kept private. The only people who will see these materials are research assistants who have been trained in ways to protect confidentiality.

It is your choice whether or not you want to take part of this study. If you change your mind at any time during the study, you tell us that you don't want to participate and there will be no consequences. If you choose not to participate, it will not affect your marks. We will be happy to answer any questions you have before signing or later. Please show that you have read this form by signing your name on the line below. If you want a copy of this form, please ask us.

Thank you for your help!

Date: _____

Name (Please print): _____

Signature: _____



THE UNIVERSITY OF BRITISH COLUMBIA

Parent Consent Form



January, 2006

**Department of Educational and
Counselling Psychology, and
Special Education**

Dear Parent/Guardian:

Faculty of Education
2125 Main Mall
Vancouver, BC, Canada V6T 1Z4

We are writing to request permission for your son/daughter to participate in an important new research project that we are conducting at his/her school. The project is entitled **“What do kids do when they are not in School? The Experiences of Canadian Children During Out-of-School Time.”** and is taking place in several districts in the Lower Mainland.

Purpose

The purpose of our study is to find out how intermediate grade children spend their time outside of school. This study is the first step undertaken by Dr. Kim Schonert-Reichl with the support of the United Way of the Lower Mainland in order that Lower Mainland communities have information about how children spend their out of school time during middle childhood.

Study Procedures

There are two parts to the questionnaires. In the first part, students will be asked to report on how they spent their out of school time during a typical week and their level of satisfaction with their out of school time. This will be done in the classroom for five consecutive days and will take approximately ten minutes per day. Children will do this via a questionnaire which will be placed in a sealed envelope by the student so that all answers are confidential. In the second part of our questionnaires, children will be asked to provide information on their feelings about themselves, their classroom, and their relationships with peers, parents and other adults. Completion of these questionnaires will take approximately two class periods of 45-60 minutes each. In our project, we are not, in any sense “testing” the children. We simply want to know where children are during out of school time and the nature of the activities in which they

engage during their out of school time as well as how children understand themselves and others. In addition, information relating to school attendance, and school achievement (marks) will be collected from students' school records and from the BC Ministry of Education (Foundation Skills Assessment). Teachers will also be asked to complete a checklist assessing various dimensions of each child's social, emotional, academic and physical well-being. We have found that children genuinely enjoy these questionnaires, and are eager and happy to participate in helping us better understand Canadian children. As these questionnaires will be administered during class time, any child who does not have permission to participate will work on an activity that is related to their regular program in the classroom.

Confidentiality

All of your child's answers on all questionnaires will be completely confidential and will not be available to teachers, parents, or other school personnel. No specific child will be referred to by name or identified in any way in the report of the results. Children's names will be removed from any questionnaires and be replaced with a code number. All information will be kept in a locked file cabinet in Dr. Schonert-Reichl's research office at UBC.

Contact

If you have any questions about this research project, please do not hesitate to call us at 604-822-2215 or e-mail me at: kimberly.schonert-reichl@ubc.ca. You can also contact Denise Buote at 604-671-1441 or e-mail her at dbuote@shaw.ca. If you have any concerns about your child's treatment as a research participant, you may contact the Research Subject Information Line in the UBC Office of Research Services at 604-822-8598. Participation in this study is entirely voluntary and you or your child may refuse to participate or withdraw from the study at any time, even after signing this consent form. Refusing to participate or withdrawal will not jeopardize your child's standing at his/her school in any way.

Please keep a copy of this consent form for your own records.

Sincerely,

Kim Schonert-Reichl, Ph.D.
Principal Investigator
Associate Professor
University of British Columbia

Department of Educational and Counselling Psychology, and Special
Education
Faculty of Education, 2125 Main Mall Vancouver, B.C. V6T 2E8
Phone: 604-822-2215 Fax: 604-822-3302
Email: kimberly.schonert-reichl@ubc.ca

Denise Buote, Doctoral Candidate

Project Coordinator
Phone: (604) 671-1441,
E-mail: dbuote@shaw.ca

PARENT CONSENT FORM: STUDENT PARTICIPATION

Study Title:

“What do kids do when they are not in school? The experiences of Canadian children During out-of-school time”

Principal Investigator:

Kimberly A. Schonert-Reichl, Ph.D.
University of British Columbia
Department of Educational and Counselling Psychology, and Special Education
Phone: (604) 822-2215, e-mail: kimberly.schonert-reichl@ubc.ca

(KEEP THIS PORTION FOR YOUR RECORDS)

PARENT CONSENT FORM: STUDENT PARTICIPATION

I have read and understand the attached letter regarding the study entitled “**What do kids do when they are not in school? The Experiences of Canadian Children During Out-of-School Time.**” I have also kept copies of both the letter describing the study and this permission slip.

- Yes, my son/daughter has my permission to participate.
- No, my son/daughter does not have my permission to participate.

Parent's Signature _____

Son or Daughter's Name _____

Date _____



(DETACH HERE AND RETURN TO SCHOOL)

PARENT CONSENT FORM: STUDENT PARTICIPATION

I have read and understand the attached letter regarding the study entitled “**What do kids do when they are not in School? The Experiences of Candian Children During Out-of-School Time.**” I have also kept copies of both the letter describing the study and this permission slip.

- Yes, my son/daughter has my permission to participate.
- No, my son/daughter does not have my permission to participate.

Parent's Signature _____

Son's or Daughter's Name _____

Date _____

Appendix B: Behavioural Research Ethics Board certificate of approval, student assent form, and parent consent form for the study in chapter 3



The University of British Columbia
Office of Research Services
Behavioural Research Ethics Board
Suite 102, 6190 Agronomy Road, Vancouver, B.C. V6T 1Z3

CERTIFICATE OF APPROVAL - FULL BOARD

PRINCIPAL INVESTIGATOR: Bruno Zumbo	INSTITUTION / DEPARTMENT: UBC/Education/Educational & Counselling Psychology, and Special Education	UBC BREB NUMBER: H08-00530
INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT:		
Institution UBC Other locations where the research will be conducted: Morley Elementary School		Site Vancouver (excludes UBC Hospital)
CO-INVESTIGATOR(S): Anne Gadermann		
SPONSORING AGENCIES: UBC Faculty of Education		
PROJECT TITLE: Validating the Satisfaction with Life Scale for Children		
REB MEETING DATE: March 27, 2008	CERTIFICATE EXPIRY DATE: March 27, 2009	
DOCUMENTS INCLUDED IN THIS APPROVAL:		DATE APPROVED: May 12, 2008
Document Name	Version	Date
Consent Forms: Parent consent form	2	April 24, 2008
Assent Forms: Student assent form	2	April 24, 2008
Questionnaire, Questionnaire Cover Letter, Tests: Questionnaires: SWLS-C and SWLS	Version 1	March 1, 2008
Letter of Initial Contact: Letter of initial contact to teachers	Version 1	March 1, 2008
The application for ethical review and the document(s) listed above have been reviewed and the procedures were found to be acceptable on ethical grounds for research involving human subjects.		
<p style="text-align: center;">Approval is issued on behalf of the Behavioural Research Ethics Board and signed electronically by one of the following:</p> <p style="text-align: center;"> Dr. M. Judith Lynam, Chair Dr. Ken Craig, Chair Dr. Jim Rupert, Associate Chair Dr. Laurie Ford, Associate Chair Dr. Daniel Salhani, Associate Chair Dr. Anita Ho, Associate Chair </p>		



THE UNIVERSITY OF BRITISH COLUMBIA

Student Assent Form

September 2008

**Department of Educational and Counselling
Psychology, and Special Education
Faculty of Education**

2125 Main Mall
Vancouver, B.C., V6T 1Z4

Dear Student,

The purpose of this form is to give you the information you need in order to decide whether or not you want to be in our research study entitled, **“Validating the Satisfaction with Life Scale adapted for Children”**.

Purpose

The purpose of our study is to find out how children your age respond to the items of the ‘Satisfaction with Life Scale adapted for Children’, a measure that is designed to assess life satisfaction of children. Specifically, we are interested in the thinking processes of children your age while answering the items of the Satisfaction with Life Scale adapted for Children.

Study Procedures

If you decide to be a part of this study, you will be asked to respond to the five items of the Satisfaction with Life Scale adapted for Children. While answering each item, you will be asked to tell us what you are thinking, a so called ‘think-aloud interview’. Completion of this interview will take no more than 15 minutes.

The think-aloud interview will be audiotaped. The Satisfaction with Life Scale adapted for Children and the think-aloud interview are NOT TESTS. There are no right or wrong answers. We simply want to know how children your age respond to the items of this scale.

Confidentiality

Remember no one at school or in your community (not even your parents/guardians, teacher, or school principal) will ever see your answers (they will be confidential). We will keep your answers from the questionnaire, the tapes and the transcriptions in locked cabinets at UBC. No names will be used when the information is studied. In this way, the information that you give us will be kept private. The only people who will

see these materials are the researchers who have been trained in ways to protect confidentiality.

Potential Risks: A potential risk of this research is that it might disrupt the class. However, by working with the classroom teacher and determining days/classes that are best suitable for data collection in the classrooms we will strive to minimize this risk.

Potential Benefits: The results of this study will help us in finding out about how children of this age respond to the Satisfaction with Life Scale adapted for Children. It will provide us with important information that can be used for revising the scale so that it can be utilized in future research for children of this age. Information obtained from this survey will assist researchers and educators who wish to learn about children's life satisfaction and find ways to promote positive development in all children.

It is YOUR CHOICE whether or not you want to take part in this study. If you change your mind at any time during the study, you may stop answering the scale and there will be no consequences (nothing will happen to you). If you choose not to participate, it will not affect your marks.

We will be happy to answer any questions you have before signing this form now or later. Please show that you have read this form and agree to participate by signing your name on the line below. If you want a copy of this form, please ask us.

I have read and understand the attached letter regarding the study entitled "**Validating the Satisfaction with Life Scale adapted for Children**".

Date: _____

Name (Please print): _____

Thank you for your help!



THE UNIVERSITY OF BRITISH COLUMBIA

Parent Consent Form

September 2008

**Department of Educational and
Counselling Psychology, and
Special Education**

Faculty of Education

2125 Main Mall

Vancouver, BC, Canada V6T 1Z4

Dear Parent/Guardian:

We are writing to request permission for your son/daughter to participate in a research project that researchers at the University of British Columbia are conducting at your child's elementary school. The project is entitled, "**Validating the Satisfaction with Life Scale adapted for Children**", and is taking place at Morley Elementary School in Burnaby. This research study is concerned with investigating how children understand and answer the items of the Satisfaction with Life Scale adapted for Children, a measure that is designed to assess life satisfaction of children in grades 4 to 7. Listed below are several aspects of this project that you need to know.

Purpose

The purpose of our study is to find out how children (grades 4 to 7) respond to the items of the Satisfaction with Life Scale adapted for Children. Specifically, we are interested in children's thought processes while answering the items of this scale. This study is being undertaken by the principal investigator Dr. Bruno Zumbo and his co-investigator Anne Gadermann. This research will be part of the PhD dissertation of co-investigator Anne Gadermann.

It is hoped that the information obtained from this research will help inform us whether the Satisfaction with Life Scale adapted for Children is a useful measure for children of this age group so that it can be utilized in future educational programs to evaluate children's satisfaction with life.

Study Procedures

Students who participate in this study will be asked to fill out a scale designed to assess information about their satisfaction with life. The scale will be administered individually to students in a quiet room in their school. While responding to each item, the children are asked to tell the

researcher about what they are thinking. Students will also have the opportunity to ask questions about the items. This session will be about 15 minutes in length. The session will be audiotaped and then later transcribed so that we can gather all the important information that the children have to tell us about the scale. There are no known risks to participating in the study. The researchers will not provide any form of counseling during the session. If participating children appear to be stressed when participating in this research, they will be referred to a school counselor.

In our project, we are not “testing” the children. We simply want to find out about the thought processes of children when answering the items of the Satisfaction with Life Scale adapted for Children.

Confidentiality

All of your child's answers will be completely confidential and will not be available to teachers, parents, or other school personnel. No specific child will be referred to by name or identified in any way in the report of the results. Children's names will be removed from any transcriptions. All transcriptions, tapes and information will be kept in a locked file cabinet in Dr. Zumbo's research office at UBC. The transcriptions of the sessions will not be available to teachers, parents, or other school personnel. All information obtained from children will be combined at a group level that will not allow one to identify individual student responses.

Potential Risks: A potential risk of this research is that it might disrupt the class. However, by working with the classroom teacher and determining days/classes that are best suitable for data collection in the classrooms we will strive to minimize this risk.

Potential Benefits: The results of this study will help us in finding out about how children of this age respond to the Satisfaction with Life Scale adapted for Children. It will provide us with important information that can be used for revising the scale so that it can be utilized in future research for children of this age. Information obtained from this survey will assist researchers and educators who wish to learn about children's life satisfaction and find ways to promote positive development in all children.

Contact

If you have any questions about this research project, please do not hesitate to call us at 604-822-1931 or e-mail me at: bruno.zumbo@ubc.ca. If you have any concerns about your child's treatment as a research participant, you may contact the Research Subject Information Line in the

UBC Office of Research Services at 604-822-8598. Participation in this study is entirely voluntary and you or your child may refuse to participate or withdraw from the study at any time, even after signing this consent form. Also, we always respect a child's wishes as to whether he or she wants to participate. Refusing to participate or withdrawal will not jeopardize your child's education in any way. Please keep a copy of this consent form for your own records.

Sincerely,

Bruno D. Zumbo, Ph.D.

Principal Investigator
Faculty of Education
University of British Columbia

Anne Gadermann, Ph.D. candidate

Co-Investigator
Faculty of Education
University of British Columbia

PARENT CONSENT FORM: STUDENT PARTICIPATION

Study Title:

"Validating the Satisfaction with Life Scale adapted for Children"

Principal Investigator:

Bruno D. Zumbo, Ph.D.

University of British Columbia

Department of Educational and Counselling Psychology, and Special Education

Phone: (604) 822 1931, e-mail: bruno.zumbo@ubc.ca

**(KEEP THIS PORTION FOR YOUR RECORDS)
PARENT CONSENT FORM: STUDENT PARTICIPATION**

I have read and understand the attached letter regarding the study entitled "**Validating the Satisfaction with Life Scale adapted for Children**". I have also kept copies of both the letter describing the study and this permission slip.

- Yes, my son/daughter has my permission to participate.
- No, my son/daughter does not have my permission to participate.

Parent's Signature _____

Son or Daughter's Name _____

Date _____



**(DETACH HERE AND RETURN TO SCHOOL)
PARENT CONSENT FORM: STUDENT PARTICIPATION**

I have read and understand the attached letter regarding the study entitled "**Validating the Satisfaction with Life Scale adapted for Children**". I have also kept copies of both the letter describing the study and this permission slip.

- Yes, my son/daughter has my permission to participate.
- No, my son/daughter does not have my permission to participate.

Parent's Signature _____

Son's or Daughter's Name _____

Date _____