The Contribution of Domain Satisfaction to Life Satisfaction: Convergent Validity of Importance Measures

by

Christopher Zou

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Graduate Department of Psychology
University of Toronto

© Copyright by Christopher Zou 2010
Abstract

Global life satisfaction is the evaluation of one’s life among important domains. Therefore, importance ratings should improve the relation between domain satisfaction and life satisfaction. However, this is not the case because studies have shown that importance ratings do not improve the model. This study examined the validity of importance and satisfaction measures using a multi-method approach. 316 participants were recruited in friendship pairs to fill out a series of questionnaires about themselves and their friend. Based on the self-informant agreement scores, there is some validity in importance and the satisfaction ratings. The importance measures for some domains also predicted people’s behavioural patterns. Indirect importance ratings were also extracted using regression and correlational analysis. In conclusion, there was also evidence for convergent validity of the direct and indirect importance measures. Therefore, people do think about the important domains to some extent when making a global life satisfaction judgment.
I owe my deepest gratitude to my supervisor, Dr. Ulrich Schimmack, for giving me the opportunity to embark on this amazing journey. He has not only helped me develop and mature as an academic but also as a person. Although it has only been one short year, I have learned more from this year than in any other year of my life. I could not have hoped for anything more from this experience and I look forward to continue working with him in the upcoming years.
# Table of Contents

1 Introduction ........................................................................................................................................... 1  
1.1 The Definition of Well-Being .................................................................................................................. 2  
1.1.1 Well-Being is a Subjective Evaluation of One’s Life ........................................................................ 2  
1.1.2 Cognitive Measures of Well-being ...................................................................................................... 3  
1.1.3 Problems of Global Judgments ........................................................................................................... 4  
1.2 Evaluation of Life Domains ................................................................................................................... 4  
1.2.1 Problems of Domain Satisfaction Judgments ...................................................................................... 5  
1.3 Improper Statistical Procedures ............................................................................................................ 7  
1.3.1 Issues with the Multiplicative Approach ............................................................................................ 7  
1.3.2 Regression Analysis and Domain Importance .................................................................................... 9  
1.3.3 Inappropriate Selection of Domains .................................................................................................. 10  
1.3.4 Characteristics of Importance Ratings .............................................................................................. 11  
1.4 Alternative Approaches to Examining Validity .................................................................................... 14  
1.4.1 Informant Ratings .............................................................................................................................. 14  
1.4.2 Validation through Behaviour ............................................................................................................ 15  
2 Method .................................................................................................................................................... 17  
2.1 Participants ............................................................................................................................................ 17  
2.2 Materials ................................................................................................................................................ 17  
2.3 Procedure .............................................................................................................................................. 19  
3 Results ...................................................................................................................................................... 20  
3.1 Analysis across Individuals .................................................................................................................... 20  
3.1.1 Characteristics of Satisfaction Ratings ............................................................................................... 20  
3.1.2 Characteristics of Importance Ratings .............................................................................................. 20  
3.1.3 Self-Informant Agreement .................................................................................................................. 20
3.1.4 Behavioural Measures ........................................................................................... 21

3.2 Analysis at the Aggregate Level ....................................................................................... 23
  3.2.1 Multi-Method Examination of Importance Measures ........................................... 23
  3.2.2 Factor Analysis...................................................................................................... 25

4 Discussion................................................................................................................................. 27

5 References ................................................................................................................................ 31
List of Tables

Table 1. Self-Informant Agreement ................................................................. 35

Table 2. Importance Ratings and Behaviour .................................................. 36

Table 3. Direct and Indirect Ratings of Importance ...................................... 37
1 Introduction

Happiness is an important, if not the ultimate goal, of people’s lives. Therefore, it should not be surprising to see that interest in studying happiness has been an ongoing endeavour ever since the 19th century (Schimmack, 2009). However, because the term “happiness” is commonly used in our everyday lives to mean many different things, researchers prefer to use the term “well-being” to avoid the colloquial nature of the word “happiness” (Diener, 2000). Most of the research on well-being looks for determinants of high well-being, in hopes of providing people with the tools to live a fulfilling life. However, before one can determinants of well-being, a valid measure of well-being is required (Schimmack, 2009). Being able to accurately quantify and measure well-being has been an ongoing concern for policy makers and ordinary citizens alike (Diener, Lucas, Schimmack & Helliwell, 2009). Policy makers can gauge how citizens’ of a nation are doing by analyzing the trends in their well-being and indicate where they need to make improvements.

In a similar vein, although many people are unaware of the process, people naturally quantify well-being and to some extent make decisions based on the net costs and gains to well-being. For example, a person may choose to watch a boring movie with a friend because the well-being gained from being with the friend outweighs the costs to their well-being associated with watching a boring movie. On the other hand, if the movie is extremely terrible, or if the friend is not all that important, then the cost to well-being begins to outweigh the benefits and the person would choose to not watch the boring movie with the friend. In addition, many of the purchases people make are based on the idea that the possession of the item increases their well-being beyond the costs associated with losing money, in hopes that the purchase will provide a net gain in well-being.

Although lay people do quantify well-being, scientists have been trying to create more reliable and valid quantitative measures of well-being. Accurate measurement is essential in many fields. For example, while people can sense the temperature of their surroundings, scientists created thermometers to measure temperature more precisely. There exist numerous scales designed to measure well-being, and new measures of well-being are still being published to measure well-being more accurately. However before one can accurately quantify well-being, a clear definition of the construct needs to be first laid out in order to have a clear idea of what one is trying to
measure. However, since there is no universally accepted definition of the term “well-being”, I selected a definition that I assume to be the best description of the common notion of happiness and well-being.

1.1 The Definition of Well-Being

1.1.1 Well-Being is a Subjective Evaluation of One’s Life

Well-being is commonly measured using objective indicators of well-being, such as economic indicators of well-being, which include Gross Domestic Product (GDP) and income per capita. These indicators are used based on the assumption that these economic indicators are a good reflection of one’s true well-being because money allows for the opportunities to realize one’s preferences. While life satisfaction and happiness are positively correlated with money, the strength of this correlation suggests that money is an imperfect indicator of well-being (Diener & Biswas-Diener, 2002). One reason is that there are many aspects of people’s lives that influence well-being independent of wealth (Diener et al., 2009).

In addition to economic indicators of well-being, well-being is also commonly measured using social indicators of well-being, such as life expectancy, crime rates and literacy rates. The use of these social indicators is based on the assumption that each of these factors contributes to increasing a nation’s well-being. However, deciding who should determine which indicators to include, which domains should be included and how they should be integrated in assessing a nation’s well-being is difficult (Diener et al., 2009). Recently, the president of France created a Commission on the measurement of economic and social progress in 2008 to address the problems of relying exclusively on GDP as an indicator of well-being and asked the commission to develop additional indicators of citizen’s well-being.

However, one major problem of objective indicators of well-being is that well-being is intrinsically a subjective construct. Objective indicators of well-being are based on the assumption that there are certain criteria that each person uses to judge their well-being regardless of their own preferences. On the other hand, subjective indicators of well-being take an individual’s preferences into account when making a judgment about one’s well-being (Schimmack, 2009). However, people’s preferences do matter because each person has different preferences for what makes them happy; some people feel enjoyment when watching wrestling,
while others feel extreme displeasure from watching the same sport. Therefore, the definition of well-being is a subjective one. Well-being is defined as the actual realization of subjective preferences (Schimmack, 2009). In order for researchers to explicitly emphasize that they are using a subjective definition of well-being, the term “subjective well-being” (SWB) is commonly used (Diener, 1984). It is important to note that I am not suggesting for the eradication of objective measures of well-being in favour of subjective measures, but as Diener et al. (2009) suggest, subjective measures of well-being can work in complement with objective measures to gain a better measure of one’s true well-being.

1.1.2 Cognitive Measures of Well-being

Since the definition of well-being is a subjective one, an important feature of a well-being measure is to allow for individuals to base their judgment of their lives on their own set of criteria that they deem to be important, rather than on some external criteria determined to be important by the researcher (Diener, 1984). If well-being is defined as the actualization of one’s preferences, then the definition we are following is a cognitive one. For example, Cantril (1965) created a well-being measure to assess people’s evaluation of their lives. His Self-Anchoring Striving Scale asked people to indicate where they fall on a ladder between the “best possible life” and the “worst possible life.” Cantril’s scale allows for individuals to anchor themselves based on their perspective, rather than being assigned a set of objective criteria to follow. The survey has even been included in past Gallup polls as a measure of well-being (Gallup, 2009).

The most popular well-being measure used in studies is the Satisfaction with Life Scale (SWLS) developed by Ed Diener and his colleagues (Diener, Emmons, Larsen & Griffin, 1985). They consider life satisfaction to be a cognitive judgement of one’s life. Similar to Cantril’s scale, the life satisfaction scale allows individuals to judge their lives based on their own personal set of standards without imposing any external standards. However unlike Cantril’s scale, they included multiple items in their measure, since single item scales have lower reliability (Diener et al., 1985). The SWLS asks participants to indicate their agreement with 5 items on a scale from 1 (strongly disagree) to 7 (strongly agree). For example, one item asks participants to indicate their agreement to the question, “In most ways my life is close to my ideal.” Note that the statement does not provide any standards to which participants must follow, which is the hallmark of all SWB measures (Diener et al., 1985).
1.1.3 Problems of Global Judgments

An assumption in the well-being literature is that when respondents make a global evaluation of their lives, they think about all the aspects of their lives that are important to them and appropriately weight each domain by their importance to come up with a global LS score (Schimmack, Diener & Oishi, 2002). With global judgments, an individual’s subjective preferences are taken into account so that each person can assign their own weight for each domain. However, this measure relies on the assumption that people are able to appropriately apply these weighting procedures.

Schwarz and Strack (1999) expressed some concerns to the use of global well-being measures. First, they argued that people do not think about all the important aspects of their lives when making a global judgment, but rather the selection of domains used in their judgment is context dependent. Since people prefer to use the least amount of cognitive resources in their thought process, they often use mental shortcuts when making global LS judgments. For example, if respondents were asked about their job satisfaction prior to assessing their general life satisfaction, they would be more likely to use job satisfaction when making their global LS judgment regardless of its importance because the information is temporarily accessible in the respondents’ minds.

Second, global judgments of well-being are also susceptible to mood effects. For example, people may report higher well-being on sunny days than on rainy days because people are in a better mood on a sunny day. However, since weather is generally not an important domain in people’s lives (Schimmack et al., 2002; Schimmack & Oishi, 2005), weather should not influence people’s global judgments of well-being. Furthermore, since mood is unstable, life satisfaction judgments based on mood would also be too unstable for the life satisfaction measure to be reliable. Due to these problems associated with global judgments of well-being, some researchers have used other methods to assess well-being to alleviate these issues, such as assessing people’s satisfaction with specific life domains.

1.2 Evaluation of Life Domains

Note that Schwarz and Strack’s (1999) criticisms of well-being measures apply only to global judgments of well-being. Since the global evaluation of one’s life is complex, respondents are
prone to errors during the process of making the global judgment. However, the evaluation of a specific life domain is less complex, so there are fewer opportunities for mistakes to occur when judging domain satisfaction. Additionally, there is some evidence that domain satisfaction judgments may be more valid because they produce higher self-informant agreement (Schneider & Schimmack, 2010).

There are also some practical reasons to use domain satisfaction judgments over the global satisfaction judgments. For example, it can be informative to measure individuals’ satisfaction with particular life domains because general satisfaction does not tell us anything about what contributes to its rises and falls (Campbell, Converse & Rodgers, 1976). For example, domain satisfaction (DS) can be used to examine where the greatest cause of dissatisfaction is coming from, so that resources can be allocated to improving that domain to increase general well-being of the population.

1.2.1 Problems of Domain Satisfaction Judgments

A problem of domain satisfaction judgments is that it is unclear how this information should be integrated to obtain a global measure of well-being. A basic assumption is that more important domains should contribute more heavily to well-being. For example, if a person values their intimate relationship, any fluctuations in their relationship satisfaction should influence changes in their overall LS. However, if the relationship is not all that important to them, then relationship satisfaction should have a minimal influence on LS.

Since the definition of well-being takes into account people’s subjective preferences, deriving one’s global life satisfaction from a simple average of domain satisfactions ignores the subjectivity of well-being. Therefore, in order for domain satisfaction ratings to be used as measures of well-being, it is necessary to incorporate information about the importance of domains in the computation of a global well-being indicator.

One solution to the problem of unweighted averages of domain satisfaction scores is to ask respondents to rate the importance of life domains and to use these importance ratings to create a weighted average of domain satisfaction. Weighted domain satisfaction scores should be more valid indicators of well-being because they take people’s preferences into account. However, many studies have found that a simple average of DS correlates as highly with global judgments
of LS as averages that are weighted by importance ratings (Andrews & Withey, 1976; Campbell et al., 1976; Trauer & Mackinnon, 2001; Wu & Yao, 2006a; Wu, Chen, & Tsai, 2009). This contradicts the assumption that global judgments and weighted domain satisfaction scores are valid indicators of well-being.

There are several explanations for this finding. First, it is possible that importance ratings are invalid. As a result, using them as weights does not increase prediction of a validation criterion. Second, researchers may have used questionable statistical methods when they used importance ratings as weights of domain satisfaction. In this case, importance ratings would be valid and would correlate more highly with global measures of life-satisfaction when appropriate procedures are used. Finally, it is possible that global life-satisfaction judgments of well-being are an invalid validation criterion. If people do not make these judgments by reflecting about satisfaction in important life domains, these ratings do not provide information about the importance of life-domains for well-being. As a result, weighting domain satisfaction ratings by importance ratings would not improve prediction of life-satisfaction judgments even if importance ratings are valid.

It is important to note that a few studies have found some evidence for the validity of importance ratings. Schimmack et al. (2002) asked respondents about the sources that they used to make global life-satisfaction judgments. The results showed that source reports moderated the relation between domain satisfaction and life-satisfaction judgements. Specifically, when a domain is used as a source for one’s life satisfaction judgment, the correlation between the domain satisfaction and overall life satisfaction was higher. Furthermore, the sources they reported were related to the importance of the domain. For instance, people who used family relationships as a source report also reported family relationships to be more important than people who did not use family relationships as a source.

In another study, Schneider and Schimmack (2010) found that important domains (e.g., health, family) contributed significantly to self-ratings and informant ratings of life-satisfaction, whereas unimportant domains (e.g., weather) did not. Although these findings suggest that importance ratings have some validity, they did not examine whether weighted averages increase prediction of life-satisfaction judgments over a simple unweighted average.
1.3 Improper Statistical Procedures

There may be some psychometric issues as to why incorporating importance ratings often provides no benefit in improving the LS-DS relation. First, since many of the studies used the multiplicative approach to incorporate importance with DS, the use of multiplicative scores may pose a problem because multiplicative scores are sensitive to the unit of measurement. Second, assumptions involving regression analysis may be violated so that the weights derived from regressing the LS scores on DS scores are inaccurate representations of importance. Third, there may be some issues surrounding the selection of domains to include in the DS-LS model.

1.3.1 Issues with the Multiplicative Approach

One of the most common methods to incorporate importance ratings is the multiplicative approach. Individuals are asked to provide importance ratings and satisfaction ratings for each domain, typically on a Likert scale from one (not important at all) to seven (extremely important). Then the two scores are multiplied and summed together to provide an overall weighted score of life satisfaction. While this is the most common method of weighting, empirical studies have continuously shown that weighting by multiplicative scores is often no better and in some instances worse than unweighted satisfaction scores (Trauer & Mackinnon, 2001; Wu & Yao, 2006a; Wu, Chen, & Tsai, 2009).

Trauer and Mackinnon (2001) argued that multiplicative scores are difficult to interpret because it is difficult to discern whether the multiplied score was obtained due to a high satisfaction rating with low importance or by a low satisfaction rating with high importance. For example, if an individual indicates that a domain has an importance of 7 and a satisfaction of 1, his or her overall multiplicative score is 7 (1×7) but another individual can obtain the same multiplicative score by indicating an importance of 1 with a satisfaction of 7. It would seem dubious to conclude that these two scores derived from two different situations represent the same “true” level of well-being. Furthermore, they argued that multiplicative scores have unfavourable psychometric properties. Specifically, correlations between multiplicative scores and another third variable have been shown to vary depending on the scale of the importance and the satisfaction ratings. Even changing the Likert scale from 1-7 to 0-6 can result in different outcomes.
Although the sensitivity of multiplicative scores to units of measurement poses challenges, it can be useful if proper units of measurement are used. As Hsieh (2004) argued, the question is not whether one should use importance ratings, but how importance ratings should be incorporated into overall well-being scores. Hsieh (2004) attempted to improve the correlation between the weighted DS and LS by using a weighting scheme where the weighted scores are divided by the sum of the importance scores. Dividing scores by the sum of importance ratings reduces individual differences in the use of a rating scale. As a result, importance scores reflect the importance of one domain relative to other domains. Although this approach is more sensible, Hsieh (2004) found that it did not improve prediction of life-satisfaction ratings over a simple unweighted average.

Hsieh (2004) also examined whether a ranking task produces more valid results. The advantage of rankings is that it forces participants to make decisions about the relative importance of domains. Rankings have been shown to be superior to ratings in the assessment of importance values (Oishi et al., 2005; Oishi, Schimmack, Diener, & Suh, 1998). Consistent with this finding, Hsieh (2004) observed slight improvements in the correlation between the weighted DS and LS when rankings were used to weight domains by importance.

Some researchers have also argued that domain satisfaction scores inherently include an importance component, so that weighting the domain satisfaction scores by their importance is statistically redundant (Trauer & Mackinnon, 2002; Wu, 2008). A simple method to address whether or not domain importance and satisfaction judgments tap into the same construct is to examine the correlation between two scores. If the correlation coefficient is high, then measuring the two separately would be redundant. Cummins, McCabe, Romeo, and Gullone (1994) found correlations between domain importance and DS to range from -.03 to .33. Wu and Yao (2006a) also found correlations ranging from -0.03 to 0.45 in their study. Similarly, Russell, Hubley, Palepu and Zumbo (2006) reported correlations between -.14 to .30 in their sample. From these limited number of studies, it seems improbable that importance and satisfaction measures tap into the same construct, since their observed correlations are never high.

In a similar vein, some authors have suggested that importance of a domain affects the variability of its domain satisfaction score. In other words, if a person reports an extreme domain satisfaction score, then this would imply that the domain is important to the individual. This line
of argument stems from Locke’s range-of-affect hypothesis, which proposes that people’s affective evaluation depends on the discrepancy between what they want and what they have, interacting with the importance of what they want (Wu & Yao, 2006b). If what they want is important, then it can produce greater variability between extreme dissatisfaction and extreme satisfaction. On the other hand, if it is unimportant, then the range is restricted so that the evaluation lies somewhere close to neutral. For example, if school is viewed as an important domain, then success or failure in that domain should produce great feelings of satisfaction or dissatisfaction. Conversely, if school is unimportant, then the same episode should produce close to neutral evaluations of the event. However, there has been very little empirical support for this hypothesis in the well-being research (Russell & Hubley, 2009).

1.3.2 Regression Analysis and Domain Importance

Importance ratings are needed if researchers are interested in optimal measurement of individuals’ well-being. However, sometimes researchers are more interested in the average importance of domains in a given population. Policy makers are also more likely to focus on the importance of domains for populations than for individuals. The focus on populations provides an alternative approach to weight domains in the prediction of life-satisfaction. Rather than trying to measure importance of domains directly, one can infer domain importance from the regression weights in a multiple regression equation in which life satisfaction judgments are regressed on a set of domain satisfactions. Regression analysis finds the optimal weights to maximize prediction of the criterion. Given a set of assumptions, these regression weights can be interpreted as the average importance of domains.

Based on this logic, Campbell et al. (1976) examined whether these regression coefficients were related to importance ratings. They extracted unstandardized regression coefficients by regressing 12 domain satisfaction scores on global satisfaction. He also obtained direct importance ratings by asking respondents to rate how important each domain was to them on a scale ranging from “Not at all important” to “Extremely important.” They found that the indirect importance measures derived from the regression equation and the direct importance ratings correlated moderately, \( r = .41 \). However, they also noted some marked discrepancies between the indirect and direct importance scores for certain domains (i.e., health, money). They suggested that direct importance ratings are susceptible to socially desirable responses, so indirect
importance ratings provide a more accurate representation of reality. However, they provided no empirical support for this claim.

Furthermore, there are a few problems for using the regression approach to derive indirect importance ratings. One problem for this approach is that it only provides unbiased results when all domains that respondents considered are included as predictors. Regression equations with missing domains may produce biased results. Moreover, it is difficult to predict how omitted domains influence results because the effect depends on the correlation among predictor variables. However, since the scope of one’s life is wide and varies from person to person, it is virtually impossible to assess all the important domains of people’s lives. Since not all important life domains are ever assessed, the use of multiple regression to derive importance weights is inappropriate.

Second, it is also important to make sure that two or more predictors do not highly correlate with one another. In the presence of multicollinearity, the beta weights derived from multiple regression may be inaccurate. Since multiple regression extracts the unique variance of each predictor in predicting the outcome variable, multicollinearity reduces the amount of variance available to predict the outcome and this results in large standard errors for the beta weights derived from the regression equation.

1.3.3 Inappropriate Selection of Domains

One crucial point to note is that many of these studies have focused on the important domains. For example, Trauer and Mackinnon (2001) noted that most Quality of Life measures cover domains that are already deemed to be important by the general public and that unimportant domains are rarely ever included. Similarly, Andrews and Withey’s (1975) identified people’s general concerns through structured interviews, surveys and previous research to develop a large questionnaire based around their general findings, and Campbell et al. (1976) selected a set of domains that they believed would be relevant to most of the population. For instance, 11 of the 12 domains listed by Campbell et al. (1976) were indicated to be greatly important by 35% or more respondents. Similarly, 6 of the 7 domains listed by Cummins et al. (1994) were indicated to be greatly important by 40% or more respondents and 1% or less judged them as being unimportant.
One problem of this approach is that it makes it more difficult to demonstrate the validity of importance ratings (Russell & Hubley, 2009). For example, Schimmack, Diener and Oishi (2002) found an average importance of 5.26 (on a 1 to 6 scale) for Family Relationships and an average importance of 5.10 for Health. Given the small difference in importance, it is extremely difficult to falsify the null-hypothesis that both domains are equally important. However, the authors found much larger differences in average importance of Family Relationships and Weather (Mean, difference = 1.22). Given this large difference, it is much easier to examine the validity of importance ratings. Thus, it can be useful to include a few domains that are expected to be unimportant in studies that examine the validity of importance ratings. Using this approach, Schimmack and colleagues were able to demonstrate that unimportant domains are indeed weaker predictors of life-satisfaction judgments (Schimmack et al., 2002; Schimmack & Oishi, 2005; Schneider & Schimmack, 2010).

When all domains have similar importance, measurement error can make it very difficult to test validity because sampling error can overpower valid differences in importance. As a result, researchers would need to use large samples or measures with very high reliability. However, most existing studies have used small samples and single item measures with moderate reliability. As a result, the results of these studies are inconclusive. The fact that a simple average predicts life-satisfaction judgments as well as weighted scores may simply reflect the fact that most domains were approximately equally important for most participants.

In addition to the psychometric issues surrounding the use of importance ratings, there are other possible reasons as to why virtually no improvements are observed in the weighted DS over the unweighted DS scores. The most commonly proposed explanation is that the importance ratings are invalid, so weighting the DS by importance would be detrimental in obtaining the overall LS. However, an alternate explanation that has been overlooked is that life-satisfaction judgments may be less valid than commonly assumed in the well-being literature.

1.3.4 Characteristics of Importance Ratings

1.3.4.1 Issues with Reliability

Some authors have expressed concern that importance ratings often have low reliability, which would also threaten the validity of such ratings because reliability is necessary but not sufficient
for validity. Campbell et al. (1976) observed that the test-retest reliability of importance ratings were lower than that of domain satisfaction, global satisfaction and happiness ratings. However, the difference was rather small (i.e., about one decimal difference). They suggested that the low reliability scores could be due to some measurement error, or because people often change their mind about the importance of each domain, or, most likely, to a combination of both. Marsh (1986) reported a test-retest reliability \((r = .57)\) for the single item measures of importance, which is similar to test-retest reliability scores for other single item measures \((r = .60)\) (Schimmack, Wagner, Krause, & Schupp, 2010).

Rice, Gentile & McFarlin (1991) examined whether increasing reliability produced better results. They collected seven self-report measures of importance, which included four rating methods, two ranking methods, and a point-distribution method, and found minimal improvement in reliability using multiple items. They found moderate correlation among the individual measures, but the average importance of the seven methods was highly reliable. Nevertheless, importance did not moderate the relationship between satisfaction with job aspects and overall job satisfaction.

Moreover, the seven measures produced very similar rank orderings of the average importance of job aspects. This finding is consistent with findings in the well-being literature, where importance ratings produce similar rank orders of life domains across studies (Schimmack et al., 2002). Health is the most, if not one of the most, important domains in virtually all studies across different populations (Schimmack et al., 2002; Hsieh, 2004; Wu & Yao, 2006). Overall, these findings suggest that low reliability does not explain why importance ratings do not improve prediction of global satisfaction measures.

1.3.4.2 Issues with Validity

As noted earlier, many authors have proposed that importance ratings are invalid. Some authors have expressed concern that importance ratings are plagued with a host of biases. Hofstede (1980), for example, believed that importance ratings were influenced by acquiescence bias, which is the tendency to judge all domains as being important. Mastekaasa (1984) also found some evidence for acquiescence bias in his study. However, there is still variation in importance scores across the domains, so the problem of acquiescence bias can be alleviated by ipsatizing the scores so that an individual’s importance rating in one domain is expressed as a deviation
from their own mean score across items. A superior alternative is to obtain ranking data that force individuals to make judgments of the relative importance (Oishi et al., 1998). However, even ranking data produce only slight improvements (Hsieh, 2004).

There have also been some concerns that importance ratings may be susceptible to social norms, so respondents may be completing importance measures in socially desirable ways. For example, a student may indicate that school is important to them because they do not want to be seen as a lazy person. People may also downplay a domain’s importance due to the values and norms of their culture even though that domain may be personally important to them. For example, a person may indicate that money is not important to them because our culture tends to emphasize the idea that materialism is a bad quality to possess. However, since very little research has been done to examine the influence of these biases on importance ratings, support for these hypotheses is lacking (Russell & Hubley, 2009).

Blood (1971) examined the usefulness of using importance weighting for specific aspects of the job to derive overall job satisfaction. He proposed that importance ratings do not improve the relation because importance ratings are invalid. He used a multi-method approach whereby he attained importance ratings using different methods. First, he extracted beta weights by regressing satisfaction scores with certain job aspects on overall job satisfaction scores. Second, he calculated the correlation between satisfaction scores with each job aspect to overall job satisfaction scores. Lastly, he also asked participants to rank the five job aspects by importance. He expected the regression weights, the correlation coefficients and the direct importance ratings to all produce the same ranking of importance for specific job aspects. Contrary to his expectations, he found that none of the results followed the same order. Based on the results, he concluded that importance ratings are invalid because they do not relate to how much a specific job domain contributes to overall job satisfaction. While Blood believed the fault lied with the validity of importance measures, it would be equally plausible to argue that measures of job satisfaction may be invalid. In other words, people may not necessary think about the importance aspects of their job when making an evaluation about their overall job satisfaction, which would invalidate job satisfaction measures. However, this possibility was neglected throughout his paper.
In conclusion, existing research has failed to demonstrate clearly whether importance ratings are invalid or not. It is possible that weighting domain satisfaction by importance does not increase prediction life satisfaction judgments because importance ratings are invalid. However, it is equally likely that low validity of life-satisfaction judgments is responsible for this finding. To address this problem, it is necessary to obtain additional measures of importance and to examine the validity of importance ratings and life satisfaction judgments from a multi-method perspective (Campbell & Fiske, 1959).

1.4 Alternative Approaches to Examining Validity

All of the previous studies on relating importance and DS-LS have used self-report ratings to assess all the measures. A major problem with this approach is that mono-method studies (i.e., studies that rely on self-ratings for all measures) can produce biased results because method variance produces spurious correlations among measures (Campbell & Fiske, 1959). The correlation could, for example, arise from people’s tendency to respond in socially desirable ways.

Campbell and Fiske (1959) popularized the idea of using multiple methods to separate valid relationships between measures from spurious relations due to method variance (Schimmack, 2010). The present study uses the logic of a multi-method approach. However, rather than examining convergent validity in terms of correlations among individuals, the focus is on convergent validity of the importance of domains. Past studies show very little evidence for convergent validity between direct importance ratings and indirect measures of importance in regression analysis with global life-satisfaction judgments as criterion. To examine which of the two approaches produces invalid results, it is necessary to examine whether one of these approaches shows convergent validity with other measures of importance.

1.4.1 Informant Ratings

Since there is no objective measure of importance, it is impossible to use objective validation criteria. However, it is possible to examine validity of measures by examining convergent validity with other measures, while none of the measures is considered a gold standard against which other measures are being evaluated (Campbell & Fiske, 1959). Well-being researchers often use informant ratings as an alternative measure to self-reports of well-being (Schneider &
Schimmack, 2009). A high correlation between the self and informant rating would suggest a minimal effect of self-rating biases. However, a low correlation between the two methods would suggest that there are some problems with the validity of the measure, possibly due to rating biases.

In research on self-esteem, Marsh (1986) obtained informant reports of importance ratings for specific domains of self-esteem. The average self-informant agreement for importance measures was $r = .26$, which is comparable to self-informant agreement scores often observed for satisfaction measures ($r = .26$ to $r = .40$) (Schneider, Schimmack, Petrician & Walker, 2009; Schneider & Schimmack, 2009). Thus, there is some evidence for validity of importance measures. However, Marsh (1986) observed great discrepancies in self-informant agreement scores between the different domains. Therefore, it is possible that the validity of importance ratings may be domain specific.

Although many researchers have suggested that importance measures are invalid, studies often do not use a multi-method approach to examine validity. This is the first study to examine the validity of importance measures in the context of well-being research using a multi-method approach. Informant ratings for the importance measures and the satisfaction measures were collected to properly examine the validity of the measures.

1.4.2 Validation through Behaviour

A similar idea to importance is the contingency of self-worth, proposed by Crocker and Wolfe (2001). Contingency of self-worth comprises of the domains in which people stake their self-esteem on. In other words, a domain’s contingency of self-worth represent how much of one’s self-esteem is contingent on doing well in that domain. Crocker and Wolfe (2001) suggested that these contingencies of self-worth can also have an influence on people’s behaviour. Specifically, people should choose situations that provide them with an opportunity to experience positive events relevant to their contingencies of self-worth. Using Crocker and Wolfe’s (2001) example, if an individual’s self-esteem is highly contingent on academic achievement, then they should be more likely to apply to graduate school than equally accomplished people whose self-esteem is not contingent on that domain. This proposition was supported by Crocker, Luhtanen, Cooper and Bouvrette (2001) who found that the contingencies of self-worth predicted how students spent their time during their freshmen year in college. For example, women whose self-esteem
was based on their appearance were more likely to indicate that they were planning to join a sorority in their freshmen year and were more likely to actually join a sorority. Consequently, the selections of situations also in turn affected how they allocated their time. For example, women whose self-esteem was highly contingent on appearance spent more time partying, shopping, and grooming. Students who rated academic achievement as being highly contingent on their self-esteem spent more time studying and less time partying. Therefore, these contingencies of self-worth are reflected through people’s behavioural choice and how they allocate their time.

Using these ideas, one can apply the same logic to importance ratings. The use of behavioural measures to reveal preferences is a pervasive line of thought in economic theories of well-being (Diener et al., 2009). For instance, if school is important to an individual, then they should spend more time studying, when compared to someone who does not find school to be all that important. Thus, people's behavioural choices can be used as another validation criterion for the importance measures. If the importance measures do bear some relation with the behavioural measures, then it further provides evidence of convergent validity of importance ratings.
2 Method

2.1 Participants

158 friendship pairs (N = 316) were recruited for this study. Age of the participants ranged from 16 to 35 years with a mean of 18.9 years (SD = 1.56 years). 158 participants were Asian, 45 participants were white, 37 participants had a mixed ethnic background, 38 participants were African American and 38 participants were of another ethnicity or unknown. The sample consisted of 71.2% females and 28.8% males. Participants were recruited from the PSY100 pool at the University of Toronto Mississauga campus. In order to enrol in the study, participants were required to bring in a friend to also take part in the study. Participants were given partial course credit for their participation. The average length of the friendship was 24.7 months (SD = 39.1 months).

2.2 Materials

Global Life Satisfaction. Global Life Satisfaction was assessed with the three-item version of the Satisfaction With Life Scale (SWLS) (Diener et al., 1985), where participants were asked to answer questions about their life satisfaction in the past six months on a 7 point Likert scale, where 7 indicates “strongly agree” and 1 indicates “strongly disagree.” An example of one item on the scale is, “The conditions of my life are excellent,” and participants are asked to indicate their level of agreement with that statement.

Domain Satisfaction Scale. Domain satisfaction was assessed using the same questionnaire that was used in Schimmack et al.’s (2002) previous study. Participants were asked to rate their satisfaction with a series of domains on a 5-point scale, from 1 = rock bottom cannot get worse to 5 = top cannot get better in the past six months. Satisfaction with academic performance, family relationships, romantic relationship, health and housing were assessed. These domains were shown to be frequently used when making life satisfaction judgments (Schimmack et al., 2002). Satisfaction with appearance and religion were also included. These two domains were identified as domains that people frequently based their self-esteem on and because self-esteem is the best predictor of life satisfaction, it is possible for these two domains to influence LS judgements (Crocker & Wolfe, 2001). Satisfaction with their current financial situation was also included since previous studies have found that financial satisfaction does influence global LS to
some extent (Diener & Biswas-Diener, 2002). Weather and commute were included as unimportant domains.

Direct Importance Ratings. Participants were asked to indicate how important each domain was to them on a scale of 1 (not at all important) to 7 (most important).

Importance Pie Chart. As was stated previously, it is important to properly obtain importance ratings so that they can be weighted appropriately. As Hseih (2004) argued, it is important to obtain ratio level measurements for each of the domain importance items in order to appropriately weight each domain by importance. Thus a pie chart was created to obtain ratio level measurements of importance to see if ratio level measurement produces different results. Participants were asked to indicate how much they feel each domain is important to them by creating a pie chart that pits each domain against one another, so that the larger the slice they assign to one domain, the more important that domain is to them. They were not required to assign a slice to each domain, so if they feel that a domain is not at all important to them, they are allowed to assign a value of “0”. However, by assigning a larger slice to one domain, they were left with less room to assign to the other domains. Therefore, participants were required to rank each domain in order to complete the pie chart. However by using the pie chart, this allows for the possibility of ties, so that multiple domains can be given the same weightings, given that all the slices sum to the entire circle. Furthermore, this method is also superior to the ranking method because one can see how far the distance is between each of the domain importance. The ranking method would not be able distinguish if one domain is predominantly more important than all the other domains or if one domain is slightly more important than the other domains, but the pie chart provides this information.

Frequency of Behaviours. Participants were asked to indicate how many times they engage in certain behaviours related to a specific domain. For example, in the health domain, participants were asked how many times per week they engage in unhealthy and healthy behaviours, such as consuming fast food, exercising, and brushing their teeth.

Allocation of Time. As another behavioural measure, participants were also asked to indicate how much time they allocate for each domain on average each week. In particular, they were asked how many hours per week they spend on an activity that relates to the specific domain. For
example, in the academic domain, participants were asked to indicate how many hours per week they spend studying.

Informant Ratings. For each of the above measures, the participants were also asked to fill out each of the questionnaires for their friend.

Demographic questionnaire. Participants were asked about themselves and their relationship with their friend. They were asked about their age, sex, income, highest level of education, ethnic background and at what age they came to Canada. Then, the participants were also asked about the length and quality of their friendship with the friend they have brought in to the study.

2.3 Procedure

Participants were asked beforehand to bring in a friend who will also be participating in the study. Participants first came into the laboratory to fill in a consent form. Then they were asked to separate into two rooms and were asked to fill out a series of questionnaires. Participants first made self-ratings for all questions and then made informant ratings with their friend as the target. After they filled out the questionnaires, they were debriefed and thanked for their participation.
3 Results

3.1 Analysis across Individuals

3.1.1 Characteristics of Satisfaction Ratings

Many participants were unsure of the definition of autonomy, so the autonomy domain was excluded from subsequent analysis. The mean general life satisfaction score of the sample was 4.86 (SD=1.16). The mean domain satisfaction scores ranged from 3.66 (SD=1.52) for academics to 5.65 (SD=1.22) for friendships (see Table 1). Cronbach’s alpha (an estimate of internal consistency) was .84 for the three-item life satisfaction scale. Scores from the three-item life satisfaction scale and the single-item life satisfaction scale had a high positive correlation, \( r(316) = .71 \).

3.1.2 Characteristics of Importance Ratings

The mean domain importance scores for the direct ratings of importance ranged from 3.88 (SD = 1.57) for weather and 3.88 (SD=1.53) for commute to 6.54 (SD =0 .88) for family (see Table 2). As expected, importance scores for weather and commute (M = 3.88, SD = 1.25) were significantly lower than the importance scores for other domains (M = 5.73, SD = 0.63).

The mean domain importance percentages for the pie chart ranged from 1.95% (SD = 2.54%) for commute to 24.1% (SD = 12.4%) for family (see Table 2). Once again, average importance percentages for weather and commute (M = 2.00%, SD = 2.12%) were significantly lower than importance percentages for other domains (M = 10.7%, SD = 0.472%).

I then correlated the direct importance ratings with the pie chart percentages for each of the 11 domains. As expected, both ratings were positively correlated, average \( r(314) = .59 \). This finding shows a reliability of .6 for single item measures of importance, which is consistent with previous studies and reliability of single-item measures in general.

3.1.3 Self-Informant Agreement

Consistent with previous studies, there was evidence for moderate self-informant agreement in domain satisfaction judgments (range from \( r = .07 \) for friendships to \( r = .39 \) for romantic relationships; average \( r = .182 \)) and life-satisfaction judgments (while self-informant agreement...
for the satisfaction scores was $r = .215$. In addition, this study showed evidence for self-informant agreement in importance ratings, although agreement varied considerably across domains ranging from $r = .02$ for friendships to $r = .55$ for religion (see Table 1). This finding suggests that there is no simply answer to the question of the validity of importance ratings. Whereas, importance ratings of religion have rather high validity, other domains showed no evidence of convergent validity. This could be one explanation why weighing by importance ratings does not improve the validity of aggregated domain satisfaction measures.

3.1.4 Behavioural Measures

3.1.4.1 Allocation of Time

Since each participant used a different reference point for time, the allocation of time scores were standardized for each individual, to represent the relative time people spent in each domain. Using regression analysis, I examined whether there was any relation between domain importance and how much time people devote to those domains. A summary of the findings is presented in Table 2. Health was not assessed in this section of the questionnaire. Importance scores for family, romantic relationships, friendships, appearance and religion significantly predicted how much time people spent doing activities related to the respective domain. Conversely, importance scores for academics, housing, weather and commute did not predict the amount of time people spent on activities related to these domains. The domains in which the importance scores did not predict behaviour also tended to have lower self-informant agreement scores. This finding is consistent with the previous analyses in suggesting that the validity of importance ratings is domain specific.

3.1.4.2 Frequency of Behaviours

Once again using regression analysis, I examined if domain importance can predict how often people engage in certain behaviours. The unstandardized regression coefficients were calculated over the standardized regression coefficients because it is more easily interpretable in this context. In general, importance with the domains of appearance, family, romantic relationships, friendships, academics, religion and weather were related to behaviour. Importance with appearance scores significantly predicted how often people looked at themselves in the mirror, $b = 1.56$ times/day, how often people fix their makeup or hair, $b = 0.561$ times/day, how often people attend parties, $b = .253$ times/month, and how often people weigh themselves, $b = 1.19$
times/month. Importance with family scores significantly predicted how often people socialize with their family, $b = 3.57$ hours/week. Importance with romantic relationship scores significantly predicted how often people spend time with their romantic partners, $b = 4.32$ hours/week, and how often they buy gifts for their romantic partners, $b = 1.69$ times/year. Importance with friendship scores significantly predicted how often people spend time with their friends, $b = 4.42$ hours/week. Importance with academic scores significantly predicted how much time people spend studying, $b = 2.35$ hours/week. Importance with religion significantly predicted how often people attended religious ceremonies, $b = .314$ times/week, how often they pray, $b = 2.52$ times/week, and how often they consume alcohol, $\beta = -.411$ times/month. Importance with weather significantly predicted how often people checked the weather, $b = .467$ times/week, and how often they complained about the weather, $b = .489$ times/month.

Conversely, domains of health, money and commute did not significantly predict people’s behaviours. Importance with health scores did not significantly predict how long people slept, unstandardized coefficient $b = -.585$ hours/night, how often people smoke, $b = -.443$ times/week, or consume alcohol, $b = -.157$ times/month, how often they cook a healthy meal, $b = .429$ times/week, how often they go to the doctors, $b = .041$ times/year, and how often they exercised, $b = .258$ hours/week. While importance with health scores significantly predicted how often people brushed their teeth, $b = -.075$ times/day, this relation was negative, which is contrary to what one would expect. Importance with commuting did not significantly predict how often people check the traffic conditions, $b = .134$ times/week, and also did not significantly predict how often they complained about the traffic, $b = -.024$ times/week. Importance with housing did not significantly predict how often people refurnish their house, $b = 208$ times/year, and also did not significantly predict how often people decorate their house, $b = 1.15$ times/year. Importance with money did not significantly predict how often people gambled, $b = .830$ times/year, how often they searched for jobs, $b = -.250$ times/year, or how often they worked, $b = -.509$ hours/week. Overall, it seems as though importance with money and commute do not predict any expected behavioural patterns in addition to having low self-informant agreement. This further suggests that the validity of importance ratings may be domain specific.
3.2 Analysis at the Aggregate Level

3.2.1 Multi-Method Examination of Importance Measures

An average domain importance was calculated for the direct importance ratings (see Table 3). For the pie chart, each individual’s domain percentage was divided by the sum of the individual’s domain percentages because not every participant had domain percentages add up to 100%. Then each domain percentage was averaged across individuals to examine average importance ratings for each domain based on the pie chart (see Table 3). The same procedure was used for informant ratings to obtain average importance ratings of domains for the direct ratings and the pie chart (see Table 3).

Indirect importance ratings were acquired by regressing self-ratings of life satisfaction on the eleven self-ratings of domain satisfaction scores to obtain unstandardized regression coefficients for the eleven domains. The eleven domains explained a significant portion of the variance in general life satisfaction scores, $R^2 = .47$. The unstandardized regression coefficients ranged from $b = -.10$ for religion to $b = .216$ for appearance (see Table 3). This procedure was repeated with informant ratings of life satisfaction and domain satisfaction. I also regressed informant ratings on self-ratings and vice versa.

I used unstandardized coefficients because standardized regression coefficients are dependent on the variability in scores of the independent and dependent variable, which would pose a problem for domains in which there is not much variability in the scores. For example, there would be low variability in scores for the health domain because most people are healthy and are satisfied with their health (Schimmack et al., 2002). Therefore, using a standardized regression coefficient would be problematic because it would downplay the role of health when making global life satisfaction judgements simply due to the fact that most people are healthy.

Regression coefficients are most meaningful when the predictor variables are uncorrelated. In general correlations among domain satisfaction ratings were low, but some correlations were more substantial, (e.g., housing satisfaction and financial satisfaction correlate at $r = .41$; see Table 1 for details). Therefore, I also computed simple correlations between each domain satisfaction scores and general life satisfaction scores. Simple correlations can reveal that two overlapping domains are important. Both approaches yielded similar information about the
average importance of domains, $r = .78$. The correlation between the self-ratings of domain satisfaction and self-ratings of global life satisfaction ranged from $r = .146$ for religion to $r = .472$ for appearance (see Table 3).

As expected, a rank-order correlation of the average domain importance between self-reports of the direct rating scores and self-reports the pie chart values was found to be highly correlated $\rho(11) = .86$. This provides further evidence that the pie chart and the direct ratings of importance provide similar information about the relative importance of domains.

A rank-order correlation of the average domain importance between self-reports of the direct rating scores and informant-reports of the direct ratings scores was very high, $\rho(11) = .94$. A rank-order correlation of the average domain importance between self-reports of the pie chart scores and informant-reports of the pie chart scores was also high, $\rho(11) = .78$. The high correlations suggest that informant ratings converge with the self ratings of importance.

A rank-order correlation of the average domain importance scores from self-reports of the direct ratings and the unstandardized regression weights derived from self-reports of life satisfaction on self-reports of domain satisfaction was moderate, $\rho(11) = .41$. A rank-order correlation of the average domain importance scores from self-reports of the pie chart and the regression weights derived from self-reports of life satisfaction on self-reports of domain satisfaction was also moderate, $\rho(11) = .21$. Since regression makes a set of assumptions that are violated in this study, I also examined the correlations between direct importance ratings and the correlations of domain satisfaction ratings with life satisfaction ratings. The results were practically identical, direct ratings $\rho(11) = .40$, pie chart $\rho(11) = .20$.

I repeated these analyses with the regression coefficients based on informant data. A rank-order correlation of the average domain importance scores from self-reports of the direct ratings and the unstandardized regression weights derived from informant ratings of life satisfaction on informant ratings of domain satisfaction was moderate, $\rho(11) = .247$. Substituting the pie chart percentages for direct ratings resulted in a correlation close to zero, $\rho(11) = -.045$. I also examined the correlations between informant ratings of life satisfaction and informant ratings of domain satisfaction. The use of correlations improved the results, direct ratings $\rho(11) = .47$, pie chart $\rho(11) = .20$. 

I repeated the analyses across methods. Cross-method analysis reduces the amount of shared method variance between the domain satisfaction ratings and the global life satisfaction ratings. A rank-order correlation of the average domain importance scores from self-ratings of the direct ratings and the unstandardized regression weights derived from self-ratings of life satisfaction on informant ratings of domain satisfaction was low, ρ(11) = .128. Substituting the pie chart for the direct ratings also produces a similar correlation, ρ(11) = .118. I also examined the correlations between self-ratings of life satisfaction and informant ratings of domain satisfaction. Once again, the use of correlations yielded better results, direct ratings ρ(11) = .32, pie chart ρ(11) = .26.

A rank-order correlation of the average domain importance scores from the direct self-ratings and the unstandardized regression weights derived from regressing informant ratings of life satisfaction on self-ratings of domain satisfaction was low, ρ(11) = .126. Repeating the analysis with the pie chart improved the correlation, ρ(11) = .337. I also examined the correlations between informant reports of domain satisfaction and self-reports of life satisfaction. The correlations once again produced improved results, direct ratings ρ(11) = .41, pie chart ρ(11) = .59.

The average correlation between the 12 different methods was moderate, ρ = .27. This shows some evidence of convergent validity for importance measures across the different methods. While people do think about the important domains when making a global evaluation of their lives, the relation is not very strong. In addition, correlational analyses often seem to perform better than regression. Since regression makes a series of assumptions that are violated, correlations may be more appropriate for this analysis.

### 3.2.2 Factor Analysis

The previous analysis showed some moderate consistency across methods. To examine whether the different methods reflect a common factor of domain importance, I subjected them to a factor analysis. As some measures used very similar methods, I included only measures that used somewhat different methods. I used correlation coefficients rather than regression coefficients because correlations showed somewhat higher convergent validity. I used direct ratings rather than pie chart percentages because direct ratings seemed to be at least as valid as pie chart percentages. I performed a factor analysis on the direct self and informant ratings and the four correlations of domain satisfaction with life satisfaction. I used principal component analysis to
determine the amount of explained variance in the six measures. The principal component had an Eigenvalue of 3.91. The second principal component had an Eigenvalue of 1.13. This finding suggests that the six measures mostly capture one common attribute. The first principal component explained 65% of the variance. This finding shows that the six measures provide some consistent information about domain importance.
4 Discussion

The present study is the first to examine the validity of life satisfaction measures and importance measures in the same context. Theoretically, valid life-satisfaction measures should be more strongly related to important life domains than to unimportant life domains. To demonstrate the validity of life-satisfaction measures it is necessary to have valid information about the importance of life domains. This study is the first study that examined the validity of measures of domain importance using a multi-method approach. First, I found relatively weak self-informant agreement in ratings of domain importance. Moreover, convergent validity varied across domains with some domains showing no convergent validity (e.g., money) and others showing fairly high convergent validity (e.g., importance of religion).

While previous studies were quick to blame the importance measures for the failure of the multiplicative approach, this study shows that importance scores do have some validity, although the validity of importance measures may be domain specific. Examining the validity of importance measures across domains, it seems as though importance ratings of money, housing and commute are not valid. These domains were found to have low self-informant agreement scores and also did not relate to people’s behavioural patterns. It is also possible that the low convergence for the domains of money, housing and commute may be due to the lack of observable behaviour available to illustrate their importance. For example, it is difficult to predict any behaviour that illustrates that a person values his or her house. This would result in low self-informant agreement scores and low utility in predicting behavioural patterns from housing importance.

It is also unclear whether the low self-informant agreement are due to the invalidity of self-reports of importance or to the informant-reports of importance or to a mix of both. Furthermore, the domains in which self-informant agreement was the highest also had the highest variation in the importance scores. This is an important point because it is easier to find a correlation between two measures if there is more variation in the scores in the two measures. Therefore, it is possible that self-informant agreement scores are not an appropriate method to examining validity of measures in this study because many of the domains have low variability in importance scores, which reduces the chance of finding a correlation between the self and informant ratings. Thus, measures that assess these constructs should be further examined for
their validity. However, since there is some evidence for validity of importance ratings in other domains, the use of importance ratings as weights should still improve the prediction of general life satisfaction from domain satisfaction.

It is also important to note that indirect importance ratings derived from regressing domain satisfaction on life satisfaction did correlate (ρ = .41) with direct importance ratings. This correlation is similar to that found in Campbell et al.’s (1976) study, which also found a correlation of .41 between the indirect and direct ratings of importance. Furthermore, the importance scores derived from the 12 different methods did show some convergence using rank-order correlations and factor analysis. This suggests that people do somewhat think about the important domains when making global life satisfaction judgments, but the relation is not very strong.

Some researchers have proposed that indirect importance ratings derived from regression and correlational analysis do not represent importance, but a different construct altogether termed determinance. While determinance and importance are related constructs, they do not necessarily always coincide. To use Myers and Alpert’s (1977) example, most people indicate that the safety of a car is important to them, yet the safety features of a car often do not determine whether they will purchase the car since most cars meet the minimum safety requirements. This may explain why there are striking differences between the direct and indirect importance ratings in some domains, such as health. Campbell et al. (1976) also noted that health satisfaction ranked highly when measured directly for its importance, but made one of the least contributions to life satisfaction. While health is rated to be highly important across multiple studies, as long as people meet the minimum requirement for health, the influence of health satisfaction on life satisfaction is minimal. This would illustrate that importance and determinance may represent different, yet related constructs.

While I have included two unimportant domains (weather and commute) to increase the effect size, the inclusion of unimportant domains still did not reveal any utility in including the importance measures. This may be because these unimportant domains were still judged to be somewhat important, since the average importance scores for these domains were around 4. In addition, while there is more variation across domains in the importance scores with the inclusion of the unimportant domains, there is not much variation across domains in the
satisfaction scores, which also reduces power to find significant effects. Therefore, it would be helpful to include domains that are unimportant while also including domains that people often find to be unsatisfactory in order to maximize the chances of finding significance.

Although weather and commute were rated to be generally unimportant to people, these domains can still make a contribution to predicting overall life satisfaction due to top-down effects of life satisfaction on domain satisfaction. For example, some people may have a general tendency to indicate that they are satisfied and this general tendency would apply to all domains regardless of their importance. Therefore, unimportant domains are also influenced by this bias, which would mean that these domains still do hold some information about a person’s overall evaluation of his or her life, thus contributing to the prediction of one’s global life satisfaction. However, these unimportant domain scores measure hold information about people’s response biases, so they are not a valid reflection of people's true well-being. One way of alleviating this issue is to use structural equation modeling, to tease apart the top down effects from the bottom up effects to acquire more accurate weights for each domain.

Another problem with this study is that in order to increase sample size, there was no requirement for friendship length to participate in the study. Schneider et al. (2010) suggest that in order to maximize the validity of informant ratings, informants should have known the target for at least three years. Thus, lower self-informant agreement in domain satisfaction may have arisen since most of the friendship pairs in the study did not know each other for longer than three years. While this study did have a larger sample size than most other studies, the sample size is still not large enough to detect small effects. The next step is to merge the data from this sample with data from previous studies to drastically increase the sample size.

Based on results from past studies, researchers have concluded that unweighted domain satisfaction scores perform nearly as well as weighted domain satisfaction scores. This may be true for most situations where the sample size is often too small to observe any noticeable difference in incorporating importance ratings into the LS-DS model. However, if these findings are to be used to inform public policies where millions of people are involved, researchers need to be careful in applying findings from their small sample to a much larger population. While the effects of importance may not matter in the context of a small sample size, in a larger sample, importance ratings may play a crucial role in increasing the accuracy of well-being measures.
Overall, what is clear from this study is that it would be incorrect to assume that importance measures are invalid. Though the validity of importance ratings for some domains is questionable, importance ratings do have some validity for certain domains. Furthermore, it would also be incorrect to conclude that importance is unimportant when predicting general life satisfaction from domain satisfaction. As was shown in the study, people do think about the important domains to some extent when evaluating their overall lives. However, the relation is not as strong as one would expect. Hence, future studies need to more closely examine the potential reasons for this unexpected trend to ensure that the measures researchers are using are valid. If people do not think about the important aspects of their lives when they make a global evaluation of their lives, then researchers have been using a measure that is not measuring what they intended to measure. The invalidation of the life satisfaction measure would not only invalidate the measure itself, but would also invalidate results from hundreds of studies that have used this measure to assess well-being. In conclusion, people do assess important aspects of their lives to some extent when making an overall evaluation of their lives, but the relation is not as strong as one would expect. Therefore, further research is required to understand why this relation does not strongly conform to people’s expectations.
5 References


Table 1

Self-informant agreement for domain satisfaction and domain importance.

<table>
<thead>
<tr>
<th></th>
<th>US</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
<th>LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS</td>
<td>1.00</td>
<td>0.58</td>
<td>0.48</td>
<td>0.74</td>
<td>0.68</td>
<td>0.75</td>
<td>0.81</td>
<td>0.80</td>
<td>0.68</td>
<td>0.75</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td>1.00</td>
<td>0.66</td>
<td>0.66</td>
<td>0.70</td>
<td>0.75</td>
<td>0.78</td>
<td>0.77</td>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.83</td>
<td>0.85</td>
<td>0.87</td>
<td>0.87</td>
<td>0.85</td>
<td>0.83</td>
<td>0.85</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.91</td>
<td>0.88</td>
<td>0.88</td>
<td>0.91</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.91</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.91</td>
<td>0.91</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.91</td>
</tr>
<tr>
<td>LS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: Correlations below the diagonal represent correlations between satisfaction measures. Correlations above the diagonal represent correlations between importance measures, except for LS and iLS, which correlated importance measures with general life satisfaction.
Table 2

Assessing the relation between domain importance and behaviour.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Regression (β)</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family</td>
<td>.241</td>
<td>.057</td>
</tr>
<tr>
<td>Romantic</td>
<td>.326</td>
<td>.034</td>
</tr>
<tr>
<td>Academic</td>
<td>.107</td>
<td>.059</td>
</tr>
<tr>
<td>Appearance</td>
<td>.102</td>
<td>.027</td>
</tr>
<tr>
<td>Housing</td>
<td>.001</td>
<td>.015</td>
</tr>
<tr>
<td>Friendships</td>
<td>.183</td>
<td>.046</td>
</tr>
<tr>
<td>Money</td>
<td>.022</td>
<td>.031</td>
</tr>
<tr>
<td>Religion</td>
<td>.103</td>
<td>.015</td>
</tr>
<tr>
<td>Commute</td>
<td>.026</td>
<td>.018</td>
</tr>
<tr>
<td>Weather</td>
<td>.012</td>
<td>.011</td>
</tr>
</tbody>
</table>
### Table 3

Comparing direct and indirect importance ratings including informant reports

<table>
<thead>
<tr>
<th>Self Report</th>
<th>Informant Report</th>
<th>Indirect Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Direct Measure</td>
<td>Direct Measure</td>
</tr>
<tr>
<td>Family</td>
<td>6.54 (1)</td>
<td>6.03 (1)</td>
</tr>
<tr>
<td></td>
<td>24.1 (1)</td>
<td>21.8 (1)</td>
</tr>
<tr>
<td></td>
<td>5.82 (3)</td>
<td>9.2 (4)</td>
</tr>
<tr>
<td></td>
<td>5.92 (2)</td>
<td>7.39 (5)</td>
</tr>
<tr>
<td></td>
<td>6.11 (4)</td>
<td>11.7 (3)</td>
</tr>
<tr>
<td></td>
<td>5.16 (6)</td>
<td>9.24 (2)</td>
</tr>
<tr>
<td>Romantic</td>
<td>5.34 (7)</td>
<td>9.62 (5)</td>
</tr>
<tr>
<td></td>
<td>6.33 (2)</td>
<td>9.64 (4)</td>
</tr>
<tr>
<td>Health</td>
<td>5.51 (5)</td>
<td>7.65 (7)</td>
</tr>
<tr>
<td></td>
<td>0.146 (11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.71 (9)</td>
<td>7.88 (6)</td>
</tr>
<tr>
<td></td>
<td>5.42 (6)</td>
<td>4.0 (9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Religion</td>
<td>4.71 (9)</td>
<td>7.88 (6)</td>
</tr>
<tr>
<td></td>
<td>5.42 (6)</td>
<td>4.0 (9)</td>
</tr>
<tr>
<td>Housing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather</td>
<td>3.88 (10)</td>
<td>2.06 (10)</td>
</tr>
<tr>
<td>Commute</td>
<td>3.88 (10)</td>
<td>1.95 (11)</td>
</tr>
</tbody>
</table>

Note: Rankings are provided in brackets.