In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Lara B. Russell
Name of Author (please print)
23/04/2004
Date (dd/mm/yyyy)

Title of Thesis: IS IT WORTH THE WEIGHT? REVISITING WEIGHTED AND UNWEIGHTED SCORES WITH A QUALITY OF LIFE MEASURE

Degree: Master of Arts Year: 2004
Department of Educational and Counselling Psychology & Special Education
The University of British Columbia
Vancouver, BC Canada
Abstract

Subjective assessments of importance have been used as a weighting factor in measurement in a number of areas of research including quality of life, self esteem and job satisfaction. Despite the powerful intuitive appeal of this practice, conceptual and psychometric concerns with importance weighting have been raised, and research using weighted scores has produced mixed results. The advantages of importance weighting have therefore not been clearly established. The present study revisits importance weighting using data collected with the Injection Drug User Quality of Life Scale (IDUQOL). Weighted and unweighted IDUQOL scores from a subset of 241 participants from the Vancouver Injection Drug User Study (VIDUS) were correlated with measures of convergent and discriminant validity and a large number of criterion variables including drug use, stability of housing, involvement in drug treatment, and hospitalization. The contribution of importance ratings to scores on a global measure of life satisfaction was calculated using regression analysis. To determine whether importance ratings contributed significantly to the weighted IDUQOL total scores, analysis of variance was employed. Overall results of these analyses suggest that incorporating importance does not enhance the measurement of quality of life for this sample. However, the mean of satisfaction ratings for all important domains correlated significantly higher than the mean of satisfaction ratings for all unimportant domains with measures of convergent validity. It appears that the impact of importance depends at least in part on how it is measured and used. Further research may uncover methods for incorporating subjective importance that do increase the sensitivity of the IDUQOL and other quality of life measures.
# Table of Contents

Abstract ................................................................................................................................. ii  
Table of Contents ................................................................................................................... iii  
List of Tables .......................................................................................................................... v  
Acknowledgements ............................................................................................................... vi  
Importance Weighting ........................................................................................................... 1  
Issues Around Importance Weighting .................................................................................. 2  
  Reliability .......................................................................................................................... 2  
  Score Interpretation ......................................................................................................... 3  
  Are Importance Scores and Domain Ratings Measuring the Same Thing? .................... 4  
  Are Weighted Scores Effective? ...................................................................................... 5  
  Criterion Measures Used to Evaluate Weighted Scores ................................................. 7  
Present Research .................................................................................................................. 9  
Research Questions ............................................................................................................ 10  
Methodology ....................................................................................................................... 10  
  Sample .............................................................................................................................. 10  
  Measures .......................................................................................................................... 11  
  Procedures ....................................................................................................................... 13  
Results .................................................................................................................................. 14  
  Reliability ........................................................................................................................ 14  
  Importance/Satisfaction Correlations .......................................................................... 15  
  Correlation of IDUQOL Scores with Measures of Convergent and Discriminant Validity ............................................................................................................. 15  
  Correlation of IDUQOL Scores with Criterion Measures ............................................ 16  
  External and Internal Tests of the Contribution of Importance .................................... 18  
  Important Versus Non-important Domains ..................................................................... 20  
Summary and Discussion .................................................................................................... 20  
References ......................................................................................................................... 25  
Author's Note ....................................................................................................................... 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tables</td>
<td>31</td>
</tr>
<tr>
<td>Abstract</td>
<td>39</td>
</tr>
</tbody>
</table>
List of Tables

Table 1: Hsieh's Solution for Difficult-to-Interpret Multiplicative Scores .................. 31
Table 2: Criteria Used in the Literature on Weighting ............................................. 32
Table 3: Test-Retest Reliability Coefficients for Importance and Satisfaction Scores
and Importance/Satisfaction Correlations ........................................................... 33
Table 4: Correlations of Weighted and Unweighted Total IDUQOL Scores With
Convergent and Discriminant Measures ............................................................ 34
Table 5: Correlations of Weighted and Unweighted Total IDUQOL Scores
With Criterion Measures ....................................................................................... 35
Table 6: Correlations of Selected Domain Weighted and Unweighted IDUQOL
Scores With Selected Criterion Variables ........................................................ 36
Table 7: Frequency of Domains Selected as Important ............................................. 37
Table 8: Correlations of MLimp and MUimp With Convergent and Discriminant
Measures ............................................................................................................... 38
Acknowledgements

A big thank you is due to my supervisor, Dr. Anita Hubley, who not only shared her expertise, guidance, enthusiasm, and insight with me, but also helped me discover what I really want to be when I grow up. I would also like to thank my committee members, Dr. Bruno Zumbo and Dr. Anita Palepu, for their assistance and encouragement.

A project of this size could never be completed without significant behind-the-scenes support. I want to express my gratitude to my wonderful husband, who never lost faith in my ability to see this through, even during the times when I forgot to have faith in myself.
Importance Weighting

The concept of using subjective importance ratings as a weighting factor in measurement has a long and tenacious history in psychology and the social sciences. From James (1950) through to Coopersmith (1967, as cited in Marsh & Hattie, 1996), Rosenberg (1979) and Harter (1990, 1999), subjective ratings of importance have been incorporated into numerous models of self esteem. In the quality of life literature, importance ratings have been seen as contributing, and even essential, to effective assessment (Carr & Higginson, 2001; Gill & Feinstein, 1994). In the area of job satisfaction, importance weighting has been discussed by numerous authors including Vroom (1964), Hofstede (1980) and Staples and Higgins (1998). Although the findings regarding the effectiveness of importance weighting have been mixed, the intuitive appeal of the practice is clearly powerful; even those authors who find no support for importance weights in their research are reluctant to dismiss the practice completely. Trauer and Mackinnon (2001), despite finding fault with many aspects of the most common approach to importance weighting, concluded that importance itself remains a valid area of study and suggested that the problem is one of methodology, not concept. Hsieh (in press) also stated that the issue should not be whether, but rather how, to weight. Clearly, the question of whether to use importance as a weighting factor in measurement has not yet been resolved. Marsh (1986) argued that, although there was little empirical support for importance weighting, the idea has "too much intuitive appeal to be completely rejected, and so further examination of the issue is needed" (p. 1233).
The most common approach to importance weighting is to ask individuals to provide a Domain Rating for various aspects of the construct of interest\(^1\). The importance of each of these domains, as judged by the respondent, is rated as well. A multiplicative domain score is obtained by calculating the product of the importance rating and Domain Rating for each domain. A total score, which represents the measurement of the overall construct, is generally obtained by summing these weighted domain scores. The main assumption underlying the use of weighted scores is that they will increase the sensitivity of an instrument compared to the use of Domain Ratings alone. A review of the research on importance weighting, however, reveals a number of concerns that have been raised in connection with this practice.

**Issues Around Importance Weighting**

*Reliability*

Several authors have reported that importance scores appear to have low internal reliability and temporal stability, especially when compared to the corresponding Domain Ratings. Campbell, Converse, and Rodgers (1976) noted this issue in their large-scale study of quality of life in the United States. Marsh (1986) reported an average test-retest reliability at one month of .57 for his single-item importance scores, but .70 for single-item Domain Ratings. Farmer, Jarvis, and Berent (2001) found that test-retest reliability (at over one month) ranged from .57 to .80 for importance scores and from .71 to .83 for Domain Ratings. Alpha coefficients for domain importance ranged from .65 to .82, compared to .70 to .82 for Domain Ratings. Another study reported alpha coefficients only .65 for domain importance, but .73 for

---

\(^1\) The term 'satisfaction' is used in the measurement of job satisfaction, quality of life and well-being. In the area of self esteem, it is more common to use the term 'self concept' or 'competence' rather than satisfaction. For the sake of clarity, 'Domain Rating' will be used to refer to specific primary ratings of construct domains. This term is distinct from 'domain rating', which does not refer to specific scores but rather to domain scores in a more general sense. The term 'domain' will be used to refer to the components, also referred to in the literature as 'areas' and 'facets', that make up a larger construct such quality of life or self esteem.
Domain Ratings (Cummins, McCabe, Romeo, & Gullone, 1994). While the determination of acceptable minimum reliability coefficients is perhaps somewhat a matter of individual standards, it should be noted that these reliability coefficients for importance did not, at the low end, meet the minimum level of .70 recommended by Nunnally (1978) for research purposes. This raises the question of the effect of multiplying less reliable scores (e.g., importance) by more reliable ones (e.g., Domain Ratings), a question that has not been extensively explored in the measurement literature (Trauer & Mackinnon, 2001).

However, it is worth noting that not all research supports the finding of lower reliability of importance scores. Ebbeck and Stuart (1993) reported alpha coefficients of .88 for importance and .87 for Domain Ratings in a study of self esteem. Kalleberg (1977) noted alphas of .68 to .85 for domain importance and .68 to .87 for Domain Ratings in his research on job satisfaction. Therefore, while importance ratings may be less reliable than their corresponding Domain Ratings in some instances, this finding is not consistent. In addition, because many studies do not report reliability coefficients for either importance scores or Domain Ratings, the overall frequency with which a difference occurs is difficult to judge.

**Score Interpretation**

Under certain conditions, weighted scores calculated through the multiplicative method may be difficult to interpret. Specifically, a high Domain Rating combined with a low importance rating may yield a domain score that is similar, or even identical, to one resulting from a low Domain Rating combined with high importance. For example, an individual who assigns an importance of 5 and a Domain Rating of 7 to a domain on a quality of life measure will obtain the same multiplicative score for that domain as someone assigning an importance of 7 and a Domain Rating of 5. Yet, as Trauer and Mackinnon (2001) asked, "are we to
Hsieh (in press) resolved this problem by dividing each individual's weighted domain scores by the sum of his or her importance ratings, essentially creating ipsatised scores (see Table 1). Marsh (1986) proposed an additional step. Standardising Domain Ratings results in scores that are positive if they fall above the mean and negative if they fall below the mean. The resulting weighted multiplicative scores will also be negative or positive, thus making it possible to distinguish between scores obtained from high and low Domain Ratings.

Are Importance Scores and Domain Ratings Measuring the Same Thing?

Some authors have argued that individuals' Domain Ratings already incorporate assessments of importance, making separate importance ratings redundant (Quinn & Mangione, 1973; Rice, Gentile, & McFarlin, 1991; Trauer & Mackinnon, 2001). Most often, the domains included in measures of quality of life, job satisfaction and self-esteem are selected by the instrument's developers precisely because they are likely to be important to the vast majority of people. As a result, most domains will be rated as at least somewhat important by most respondents (Harter, 1999; Harter & Whitesell, 2001; Quinn & Mangione, 1973; Trauer & Mackinnon, 2001). Other authors are not so sure that this is the case. Hsieh (2003; in press) pointed to the less-than-perfect correlation between scores on global and domain-based measures of life satisfaction as an indication that the latter do not necessarily incorporate all the possible domains that comprise overall life satisfaction. It might even be argued that the very exhaustiveness of some measures increases the likelihood that they will include at least a few domains that are not important to some people.

This debate raises the question of whether Domain Ratings and importance scores are tapping the same construct. One way to assess this possibility is to look at the correlations
between the two sets of scores. Positive correlations between Domain Ratings and importance scores have been reported across several studies but, overall, these are not high enough to suggest that the two sets of scores are measuring the same construct. For example, Marsh (1986) found a mean correlation of .35 between Domain Ratings and importance across various domains of self-esteem, with 10 out of 12 of these correlations lower than .47 (the highest was .86). In another study of self-esteem, Farmer et al. (2001) reported correlations between Domain Ratings and importance scores ranging from .37 to .74 across domains. Cummins et al. (1994) found correlations between Domain Ratings and domain importance ratings ranging from -.08 to .33. Turning to a different method of judging the distinct nature of importance, Kalleberg (1977) used factor analysis to show that importance and satisfaction items on the job satisfaction scale used in his study loaded onto two separate factors. However, most studies on weighting have ignored this issue, and have neither reported the correlations between Domain Ratings and importance scores nor otherwise explored the relationship between these scores.

**Are Weighted Scores Effective?**

Perhaps the most important question regarding weighted scores concerns their effectiveness. Do weighted scores improve the measurement of constructs such as quality of life, self esteem, and job satisfaction?

The answer in the literature on quality of life has been almost uniformly negative. Andrews and Withey (1976) and Campbell et al. (1976) found that weighted scores performed no better than unweighted scores in predicting overall life satisfaction. Cummins et al. (1994) assessed the use of the weighted scores with the Comprehensive Quality of Life Scale (ComQol) and found that importance provided no advantage over simple Domain Ratings. Trauer and Mackinnon (2001), using unpublished data collected with the ComQol, showed that
total weighted quality of life scores were far more highly correlated with the mean of the Domain Ratings than with mean importance. Only Hsieh (in press; 2003) has argued in favour of using weighted scores, yet even his research failed to provide much empirical support for this. In his study, the correlations of weighted and unweighted life satisfaction scores with a global measure were almost identical. Finally, Mastekaasa (1984), investigating the use of importance ratings with data on both quality of life and job satisfaction, reported that, in the case of quality of life, there was no evidence to support the use of weighted scores.

Findings in the area of self esteem have been somewhat more mixed. Ebbeck and Stuart (1993), Kaplan (1980), and Rosenberg (1979) all reported evidence to support the use of some form of importance weighting. Other authors have found otherwise, notably Marsh (1986; 1994) and Marsh and Sonstroem (1995). For example, Marsh (1986), in a study conducted with 930 youth, found that weighted Domain Ratings correlated .48 (for a sum of single-item measures) and .51 (for a scale based on multi-item domain ratings) with a global measure of self esteem. The unweighted Domain Ratings, in contrast, correlated .59 (for the summed single-items) and .67 (for the scale) with the global measure. Using standardised Domain Ratings and ipsatised importance scores merely succeeded in eliminating the gap between the correlations of the weighted scores and unweighted scores with the global measure. Hoge and McCarthy (1984) also attempted various transformations of Domain Ratings and importance scores, but found that all of these were less effective than using simple unweighted Domain Ratings. While Farmer et al. (2001) did find that importance contributed significantly to overall self esteem in 375 college students, the effect size of this contribution was very small (0.7%), especially compared to the contribution of the Domain Ratings, which accounted for 53% of the variance in global self esteem.
Research in job satisfaction has also produced mixed results. Kalleberg (1977) reported a small, but nevertheless significant, effect of importance on overall job satisfaction. Mastekaasa (1984) found that the impact of individual domains on overall job satisfaction was linearly related to the importance assigned to those domains. Vroom (1964) also found evidence suggesting an effect from the interaction of domain importance and Domain Ratings on overall levels of job satisfaction. However, other evidence suggests that incorporating importance does not necessarily improve the measurement of job satisfaction. This was the finding in a number of studies including those of Ewen (1967), Staples and Higgins (1998), Smith Mikes and Hulin (1968), Quinn and Mangione (1973), and Rice et al. (1991). One intriguing finding in the research of Rice et al. is that, although domain importance did not contribute to the prediction of overall job satisfaction, it was a significant predictor of domain satisfaction. This points to the interesting possibility of exploring the effectiveness of weighted scores at the domain level.

Although the use of importance weights been explored fairly extensively in previous research, the results of this research have been so mixed that it is not yet possible to arrive at a definitive recommendation whether to weight or not to weight. While some authors have reported finding support for the use of importance scores, evidence from other research suggests that the effectiveness of weighting is doubtful. Yet even authors who are generally unsupportive of weighted scores are reluctant to abandon the concept entirely, and have recommended further research (Marsh, 1986; Trauer & Mackinnon, 2001).

Criterion Measures Used to Evaluate Weighted Scores

The vast majority of studies on importance weighting have assessed the effectiveness of weighted scores by comparing these to scores from a global or facet-free measure of the
construct of interest (see Table 2). Very few studies have used other types of assessment
criteria, and very few authors have explained the selection of global measures as their main
criterion. Hoge and McCarthy (1984) noted in one study that the total scores on a domain-based
measure of self-esteem correlated only .45 with scores on a global measure of self-esteem, and
concluded that domain-based and global scores cannot necessarily be equated. Although it is
certainly not unreasonable to assume that scores from global and domain-based measures of the
same construct should overlap to some extent, it has been argued that the two types of measures
do not necessarily reflect identical constructs (see Bollen & Lennox, 1991; Diener, 1984; and
Heady, Veenhoven, & Wearing, 1991, for examples of the discussion around this issue).
Therefore, it could be argued that global measures may not be the best measure against which
to assess the performance of weighted scores. At the very least, the superiority of this approach
has not been clearly established. This was pointed out by Quinn and Mangione (1973), who
also noted that very few studies have used more than a single global measure as the criterion
for evaluation. Research on importance weighting that incorporates a larger number and range
of criterion variables is needed to fill this gap in the literature.

Although correlating weighted and unweighted scores with a global measure has been
the most frequent approach to assessing importance weighting, some authors have instead used
analysis of variance and regression analysis to test the effectiveness of weighting. Here again,
the dependent variable is generally the score from a global measure of the construct of interest,
with the Domain Ratings, importance scores and the interaction of Domain ratings and
importance as the independent variables (e.g., Ebbeck & Stuart, 1993; Marsh, 1986;
Mastekaasa, 1984; Quinn & Mangione, 1973). This model raises the same question as the
correlational analyses mentioned above; while it can be assumed that global and domain-based
measures of the same construct will overlap to some extent, it cannot be necessarily assumed that the global measure is the best criterion against which to assess weighted scores. It would be valuable to identify a way to assess the relative contribution of importance scores, Domain Ratings and the interaction of these two that does not rely on scores from a global measure.

Present Research

The present study looks at importance weighting using data collected with the Injection Drug User Quality of Life scale (IDUQOL; Brogly, Mercier, Bruneau, Palepu, & Franco, 2003) in order to (a) examine a number of issues surrounding importance ratings that have been raised in the literature, and (b) determine whether the continued use of importance ratings in the IDUQOL seems advisable. This research brings together different methods, such as correlational analyses and regression analyses, as well as various criteria for assessing scores including two measures of convergent validity, one measure of discriminant validity, and a large number of criterion variables. Although multiple methods have been used to explore importance weighting in some previous research (e.g., Quinn and Mangione, 1973; Marsh, 1986; 1994), this has not yet been undertaken with quality of life data. This research also extends the typical approach to importance weighting by including a much larger number of criterion variables than any previous study. In addition, the impact of importance on various criterion measures and variables is systematically investigated at both the domain and overall level. As in several previous studies, the contribution of importance ratings to scores from a global measure of life satisfaction was assessed, but, in a more unique step, the contribution of importance to the total weighted IDUQOL scores themselves was measured as well.
Research Questions

This study addresses the following overarching question: Does the use of importance scores improve the validity of inferences made using the IDUQOL scale? The answer is based on the responses to four more specific questions that arise from the literature on importance weighting.

Question 1: Do the data from the IDUQOL reflect some of the problems found in other research on importance weights? Specifically, will (a) lower reliability estimates be found for importance scores compared to satisfaction scores?, and (b) high correlations be found between importance and satisfaction scores?

Question 2: Do IDUQOL scores, weighted by importance, correlate more highly with measures of convergent validity and less highly with measures of discriminant validity than unweighted scores?

Question 3: Do IDUQOL scores, weighted by importance, correlate more highly with criterion variables than unweighted scores?

Question 4: Do the IDUQOL data provide evidence of a unique contribution of importance to overall ratings of quality of life in the present sample?

Methodology

Sample

Participants were drawn from a cohort of individuals participating in an existing longitudinal study of the incidence of HIV among injection drug users in Vancouver, Canada (Vancouver Injection Drug User Study, VIDUS). Two hundred and fifty individuals participated in the present study. Data from three participants had to be discarded because they were deemed, at the time of collection, to be too impaired or too tired to focus on the research
tasks. An additional six participants had to be removed from the analysis of weighted and unweighted scores due to missing data, resulting in a final sample of 241. Of these, 151 (62.7%) were male and 90 (37.3%) were female. The majority (84.6%) had completed high school.

To obtain retest data, the first 50 participants were invited to return for a second session within 6-8 days. Participants were paid $10 CDN for each session of the study.

Measures

Injection Drug User Quality of Life Scale (IDUQOL). The IDUQOL was developed as a culturally sensitive measure of quality of life for injection drug users. It incorporates 21 life domains, many of which (e.g., Drugs, Drug Treatment, Harm Reduction and Neighbourhood Safety) were included because of their particular relevance to the social and physical reality of injection drug users. Each of the life domains is represented on a 5 by 5 inch card, with the name of the domain printed on the front along with a simple picture. Graphic representation of the life domains is intended to make the instrument more accessible to individuals who have low literacy skills or do not speak English as a first language.

In the first step of administration of the IDUQOL, the interviewer shows the respondent the 21 life domain cards, reading out and discussing the description of each domain. The participant selects those domains that he/she deems important to his/her quality of life, and the remaining cards are set aside. The cards representing important domains are laid out and the participant is given 3 small plastic chips, similar to poker chips, for each of the selected cards. The total number of chips can therefore range from 0 (no life domains are important) to 63 (all life domains are important). The participant then distributes the chips across the cards to indicate the level of importance of each life domain, with more chips indicating greater
importance. Each card selected as important must be marked with at least one chip, but aside from this single requirement the chips can be distributed freely. If, during the process of distributing the chips, a participant decides that a domain is not important after all, the card is removed, along with 3 chips, and the remaining chips are redistributed according to the importance of the remaining life domains.

Next, the participant provides a satisfaction rating each domain, using a 7-point Likert-type scale anchored by 1 (very dissatisfied) to 7 (very satisfied) and illustrated with seven stylised frowning and smiling faces. Visual representation is included as a guide for respondents with limited English or literacy skills.

Calculation of the weighted domain scores is based on a multiplicative model. The importance rating (number of chips) of each domain is divided by the total number of chips used by that participant and then multiplied by the domain's importance rating. This produces a domain score. Finally, all domain scores are summed to obtain an overall quality of life score ranging from 1 (lowest) to 7 (highest).

*Marlowe-Crowne Social Desirability Scale (MC X2).* Strahan and Gerbasi's 1972 MC X2 is a 10 item version of the Marlowe-Crowne Social Desirability Scale (MC SDS). The measure uses a true/false response format. It provides an estimate of socially desirable responding as a potential source of measurement error and functions as a measure of discriminant validity in this study. The short form is made up of items 2, 4, 6, 12, 14, 20, 21, 24, 28 and 30 from the original scale, and has been found to correlate .80 or higher with the full-length MC SDS (Strahan & Gerbasi, 1972). The alpha coefficient for the MC X2 obtained with the present sample was .62. One-week test-retest reliability was .64.

---

This rating system is quite unique. In the vast majority of studies, importance is rated on a Likert-type scale, and participants are required to assign at least a minimum importance rating to all domains.
Satisfaction With Life Scale (SWLS). Diener’s Satisfaction With Life Scale (Diener, Emmons, Larsen, & Griffin, 1985) is a 5 item measure of global life satisfaction that is rated 1 (strongly disagree) to 7 (strongly agree). The alpha coefficient obtained with the present sample was .85, whereas the one-week test-retest reliability was .69.

Rosenberg Self-Esteem Scale (RSES). The Rosenberg Self Esteem Scale (Rosenberg, 1965) is a 10 item measure of self-esteem containing both positively and negatively worded items which are commonly rated on a 4 point Likert-type scale ranging from 1 (strongly disagree) to 4 (strongly agree). Negatively worded items are reverse scored so that higher total scores represent greater self esteem. The alpha coefficient for the RSES in the present sample was .82, and the one-week test-retest reliability was .31.

Demographic Information. Demographic information was obtained from the VIDUS databank on the stability of participants’ housing, their sex trade involvement, type and frequency of drug use, whether participants were lending or borrowing needles for injecting drugs, involvement in a methadone program or drug treatment program, recent overdoses, and recent hospitalisation or emergency room visits. All variables were measured and coded dichotomously.

Procedures

The data were collected on an individual basis by one of three VIDUS staff in sessions of approximately 25-30 minutes. The IDUQOL was administered first, following the procedure outlined in the "Measures" section, with an additional step to allow for the collection of unweighted satisfaction scores. Once the chips and the cards representing important life domains had been removed, the cards representing those life domains previously designated as unimportant to the participant's quality of life were laid out. The participant provided a
satisfaction rating for each domain using the same 7-point scale described previously. The total unweighted score was computed by summing the satisfaction scores for all domains, using both the satisfaction scores from the weighted section of the administration and those obtained for 'unimportant' domains, and dividing the result by 21 (the total number of domains).

The next step was the administration of the MC X2, followed by the SWLS and finally the RSES. All participants completed the measures in the same order and within a single session. Retest sessions followed the same format and involved the same tasks as the initial session.

Results

Reliability

In order to calculate reliability coefficients for the IDUQOL data, it was necessary to make some adjustments to the data. As noted above, the IDUQOL allows respondents to designate domains as unimportant, which removes them completely from the calculation of the total weighted score. As a result, participants' weighted scores were based on varying numbers of domains. While this presented no difficulties for the correlation and regression analyses, the calculation of estimates of reliability required that the same number of domains be used for each participant. It was therefore decided to assign a value of 0 to those domains that were designated unimportant as it was felt that this is what was implied by declaring a domain unimportant. Although this did not change the total weighted scores (as the resulting multiplicative scores were all zero), the calculation of reliability estimates could now be based on a 21 item scale.³

³ It should be noted that the step of assigning a value of 0 to unimportant domains was only undertaken for the purposes of calculating reliability estimates and the correlations between importance and satisfaction scores. For all other analyses involving importance scores, the number of cases used was determined by whether the relevant domain(s) was (were) considered important by each participant.
The internal reliability estimate for the importance ratings, using Cronbach's alpha, was .65 whereas the reliability estimate for the satisfaction ratings was .88. The reliability findings from the IDUQOL are therefore consistent with other reports in the literature on weighting that suggest importance scores tend to be less reliable than satisfaction ratings.

One-week test-retest reliability estimates were based on the subsample of 50 participants who completed the IDUQOL twice. The correlation of the composite importance scores across the two sessions was .69, with correlations for individual domain scores across the two sessions ranging from .24 to .82. The correlation of the composite satisfaction scores over the two sessions was .78, with domain score correlations across the two sessions ranging from .32 to .67. Test-retest correlations for all domains and the composite score are listed in Table 3. Overall, satisfaction scores appear to exhibit greater stability than importance scores with this sample.

Importance/Satisfaction Correlations

Correlations between importance and satisfaction scores ranged from -.003 to .31. Only two of the 22 correlations were significant at p<.01, and two were significant at p<.05. Table 3 provides the importance/satisfaction correlations for each domain and the composite scores.

Correlation of IDUQOL Scores with Measures of Convergent and Discriminant Validity

Table 4 shows the correlations of weighted and unweighted total IDUQOL scores with the SWLS, the RSES, and the MC X2. Both the weighted and unweighted scores were positively and significantly correlated with all of the measures but, as expected, the convergent measures (SWLS, RSES) were more highly correlated with the IDUQOL total scores than was the discriminant measure (MC X2). Steiger's (1980) Z tests of the difference between the weighted and unweighted IDUQOL correlations with each of the convergent and discriminant
measures are also shown in Table 4. For example, the first row of Table 4 shows that the weighted IDUQOL total scores correlated .60 with the SWLS, while the unweighted total IDUQOL scores correlated .59 with the SWLS. The non-significant Steiger's Z test of differences between these two correlations with the SWLS (Z=0.44) indicates that the difference between them was not statistically significant. In fact, all the Z tests in this table showed statistically nonsignificant results. This finding is perhaps not surprising given the correlation of .90 between the weighted and unweighted IDUQOL total scores.

**Correlation of IDUQOL Scores with Criterion Measures**

Table 5 shows the correlations of the weighted and unweighted IDUQOL total scores with the criterion variables as well as the results of Steiger's Z test for each pair of correlations. Of the statistically significant correlations, most were in the expected direction. Weighted and unweighted IDUQOL total scores were statistically significantly and negatively correlated with unstable housing, sex trade involvement, borrowing and lending needles, and daily use of heroin and speed. The unweighted IDUQOL scores were statistically significantly negatively correlated with daily use of speed and overdoses in the past six months, but weighted scores were not. The weighted and unweighted IDUQOL scores did not correlate significantly with any of the other variables. There were no statistically significant differences between the correlations of weighted and unweighted IDUQOL total scores with any of the criterion variables.

Some authors have suggested that using standardised satisfaction scores may improve the correlation of weighted scores with other measures (e.g., Marsh, 1986). Total IDUQOL scores were calculated using standardised satisfaction ratings and correlated with the MC X2, SWLS and RSES. The resulting correlations were identical to those based on unstandardised
weighted scores, and tests of the differences in correlations were not significant. Thus, in the case of the IDUQOL, standardised satisfaction ratings did not improve the performance of the weighted scores.

Total IDUQOL scores are based on a wide range of domains that encompass social, physical and emotional realms, and therefore might not necessarily correlate with the more specific criterion variables. To explore this possibility, analyses were carried out at the domain level. Table 6 shows the correlations of selected weighted and unweighted IDUQOL domain scores, based only on those individuals who selected that domain as important, with corresponding criterion variables. Tests of the differences between the correlations are also included.

Once again, the statistically significant correlations are all in the expected direction, with unstable housing negatively correlated with weighted and unweighted scores in the domain of Housing, involvement in a methadone program or drug treatment positively correlated with Drug Treatment, and scores on the RSES positively correlated with the domain of Feeling Good About Yourself. Unweighted Health domain scores were significantly and negatively correlated with hospitalization and involvement in the sex trade, but weighted scores were not. Only two of these domains showed a significant difference in the correlation of weighted and unweighted scores with the criterion variable. For both the Housing domain and the Feeling Good About Yourself domain, unweighted scores correlated significantly better than the weighted scores with the criterion variables. The vast majority of the differences between weighted and unweighted scores were not significant, and provide no evidence for the uniqueness of weighted scores.
As noted earlier, many satisfaction measures include domains that will be selected as being important by most people. This can have the effect of reducing the variability in importance scores, which then reduces the effectiveness of analyses. The pattern of selecting most domains as important is repeated with the IDUQOL. Table 7 shows the frequencies of domain selection and the percentage of individuals selecting each number of domains as important. The minimum number of domains selected was three, chosen by only 0.8% of the sample. In contrast, close to 11% assigned importance to all 21 domains. Approximately 85% of the sample indicated that 10 or more domains were important to them, and about half selected 15 or more domains.

To determine if the limited variation in the number of domains selected was affecting analyses, correlations were computed for the weighted and unweighted IDUQOL total scores with the MC X2, the SWLS and the RSES based only on those participants who selected 15 or fewer domains (N=128), or 11 or fewer domains (N=52), as important. Although the correlations with the measures of discriminant and convergent validity remained significant, none of the Steiger's Z tests resulted in significant values. Comparing the correlations of weighted and unweighted IDUQOL total scores with selected criterion variables for those individuals who selected 15 or fewer domains as important resulted in only one significant difference. Overall, limiting the analyses to only those individuals who did not select most domains as important did not increase the differences in the correlation of weighted and unweighted scores with the criterion variables.

External and Internal Tests of the Contribution of Importance

In an assessment of weighted scores against an external criterion, the contribution of importance ratings to scores on a global measure was measured using a series of hierarchical
regressions with scores on the SWLS as the dependent variable. These analyses were based on a design presented by Marsh (1986). The influence of each domain was tested by entering the domain satisfaction score in the first step, the domain importance score in the second step, and finally the interaction of satisfaction and importance. The contribution of the components of each step was assessed using the $R^2$ change as a guide. A final analysis saw all domain satisfaction scores entered in the first step, all domain importance scores in the second step, and finally all satisfaction $\times$ importance interactions (essentially the weighted scores) in the final step. The results showed that the contribution of satisfaction was significant for all 21 domains, but neither importance nor the satisfaction $\times$ importance interaction resulted in a significant change in $R^2$ in any of the analyses.

Because it is unclear if an external global measure is the best criterion against which to assess the effectiveness of importance weighting, a second set of analyses was carried out using analysis of variance to look at the contribution of domain satisfaction, domain importance and weighted domain scores to the weighted IDUQOL total score. By looking at the effect of importance on the weighted IDUQOL total scores themselves, this internal test was aimed at circumventing the need for an external criterion. The results showed that domain importance was significant at $p<.05$ in one domain (Drug Treatment), and at $p<.01$ in two domains (Family and Money). The interaction of importance and satisfaction was significant ($F(1,167)= 7.14$, $p<.01$) only for the domain of Family. In contrast, domain satisfaction failed to reach significance at $p<.05$ in only 3 out of 21 domains. Clearly, the overall effect of satisfaction on the total weighted scores was greater than that of either importance or satisfaction weighted by importance.

4 Before conducting each analysis, the relevant weighted domain score was subtracted from the weighted IDUQOL total score. This adjusted the total score for the effect of that weighted domain score.
Important Versus Non-important Domains

A unique aspect of importance weighting on the IDUQOL lies in the fact that participants can choose to completely eliminate domains from the calculation of their weighted total score by designating them as 'unimportant'. Although this does not appear to have increased the sensitivity of the resulting weighted scores, several analyses were conducted in which importance was treated as a dichotomous variable. That is, two scores were calculated for each participant: the mean of satisfaction in all important domains (MImp), and the mean of satisfaction in all unimportant domains (MUimp). Participants who had selected all domains as important were removed, resulting in a subsample of 215. The two sets of scores were then correlated with the MC X2, SWLS and the RSES. These correlations and the corresponding test of the differences in the correlations can be found in Table 8. All correlations were still statistically significant and in the expected direction. However, while the difference in correlations with the MC X2 was still not statistically significant, the differences in correlations with the SWLS and the RSES were statistically significant, with larger correlations found with the mean of important domains.

Summary and Discussion

The present research explored the use of weighted scores in the IDUQOL using a number of different strategies and criteria for evaluating their effectiveness. In keeping with the findings from most previous studies involving importance weighting, the overall results do not support the use of importance scores as they are currently measured and applied in the IDUQOL.

The correlations of total weighted and unweighted IDUQOL scores with the MC X2, SWLS and RSES did not differ significantly, as shown by Steiger's Z test of differences in
dependent correlations. Using standardised satisfaction scores did not increase the effectiveness of weighted scores. The correlations of the weighted and unweighted IDUQOL total scores with 13 different criterion variables did not differ significantly. Overall, these findings provide no evidence to support the continued use of importance weighting.

Arguing that the IDUQOL total score is comprised of scores from a fairly broad range of domains that might not be expected to correlate significantly with very specific criteria, analyses were also conducted at the domain level. Once again, however, there were almost no statistically significant differences in the correlations of weighted and unweighted domain scores with selected criterion variables. Taken over the 11 selected variables, the results appear to indicate that weighting by importance does not result in scores that differ significantly, from unweighted scores, in their correlations to these variables.

These findings support evidence from previous research that weighted scores do not enhance the measurement of a construct, particularly in the prediction of criterion variables and scores from a global measure (e.g., Campbell et al., 1976; Hoge & McCarthy, 1984; Marsh 1986, 1994; Quinn & Mangione, 1973; Rice, et al., 1991; Smith Mikes & Hulin, 1968). Analysis at the domain level also failed to improve correlations with criterion variables.

The internal and test-retest reliability estimates showed that importance scores were less reliable than importance scores for this sample. The alpha coefficient for the satisfaction scores was quite high, and although the test-retest reliability was somewhat lower, it was still within acceptable ranges. In contrast, importance scores exhibited neither high internal consistency nor high temporal stability, similar to the findings in the research of Campbell et al. (1976), Cummins et al. (1994) and Marsh (1986), among others. This has worrisome implications for the validity of inferences made from importance scores, and raises concerns about the effect of
multiplying less reliable scores (importance) with more reliable ones (satisfaction) in the calculation of weighted scores.

The correlations of importance and satisfaction scores within domains, which ranged from -.003 to .31, were lower than those reported in research by Farmer et al. (2001) and Marsh (1986), but very similar to those reported by Cummins et al. (1994). Only three of the correlations were significant at p<.05. This suggests that ratings of importance and ratings of satisfaction are measuring different constructs.

In contrast to these low correlations between domain importance and domain satisfaction, the correlation between the weighted IDUQOL total score and the unweighted IDUQOL total score .90, suggesting a significant amount of overlap between the two and raising questions about the contribution of importance to overall scores. In an attempt to address these questions, regression analyses were carried out with the scores from the SWLS as the dependent variable. According to the model described by Marsh (1986), the incorporation of importance in the IDUQOL would be supported if the satisfaction rating and the satisfaction x importance product each contributed significantly to the prediction of the dependent variable. In the case of the IDUQOL data, satisfaction contributed significantly to scores on the SWLS, but none of the importance x satisfaction products resulted in a significant increase in $R^2$. Turning to the analyses of the weighted scores themselves, ANOVAs at the domain level showed that the contribution of domain importance to the weighted IDUQOL total score was significant in only three domains, whereas the weighted domain score (the product of satisfaction and importance) was significant in only one case. In contrast, domain satisfaction contributed significantly to the weighted IDUQOL total score for 18 of the 21 domains. These analyses did not use a separate global measure such as the SWLS, but rather relied on the
weighted scores themselves as the criteria against which the effectiveness of weighting was assessed. Nonetheless, no evidence was found for an unique contribution from importance ratings.

Returning to the main research question underlying the present study, it would appear that incorporating importance ratings does not improve the validity of inferences made from the IDUQOL and provides no advantages over the use of scores based on satisfaction scores alone. On the whole, the multiplicative model used did not improve correlations with either criterion variables or measures of convergent and discriminant validity, compared to the use of satisfaction scores alone. In addition, importance was not a significant contributor to the weighted quality of life scores. This finding has a number of implications for the administration and scoring of the IDUQOL. Eliminating the measuring of importance would naturally decrease administration time. In addition, the specific procedure used for the IDUQOL, involving the distribution of small plastic chips based on the number of important domains, was found to contribute to measurement error due to occasional misplaced and miscounted chips during administration. Although the "hands on" aspects of measuring importance were popular with participants, removing this step would eliminate one source of measurement error for the IDUQOL.

There are several limitations to the study that may have impacted the results. One is that all of the criterion variables used here were measured dichotomously. Not only did this represent a limitation on the sensitivity of the measurement of these variables, but the resulting limitation in the variability of scores may have reduced the effectiveness of the statistical analyses. A second, and perhaps more important reason, is that although this study employed a significantly larger number of criterion variables than any previous research on weighting, it
has not answered the question of whether these were the best criteria to have used to draw out the effects of weighting. As noted above, this is a major question that has not yet been adequately addressed in the literature on importance weighting, and one that certainly deserves exploration.

Perhaps the most intriguing finding of this study can be found in the comparison of correlations from mean satisfaction for important domains (MImp) and mean satisfaction for unimportant domains (MUimp) with the SWLS and RSES. The correlations of the two sets of scores differed significantly, with MImp more highly positively correlated with these two measures of convergent validity. This suggests that importance can have an impact on the relationship between IDUQOL scores and criterion measures, and that it was the multiplicative model that failed to capture this impact. Further research might reveal a more effective method of using assessments of importance that does, in fact, increase the sensitivity of the scores from IDUQOL and other instruments.

Finally, even if it is found that the selection of different criterion variables and the application of different methods of weighting do not lead to improvements in measurement, should it be assumed that importance has no utility? Trauer and McKinnon (2001) noted that importance may "have explanatory value in its own right" (p. 583), whereas Harter and Whitesell (2001) suggested importance may provide valuable information for the development of interventions. This is an intriguing possibility that has not been explored and yet another reason to say that it is too soon to dismiss importance as a relevant variable in measurement.
References


Hsieh, C.-M. (in press). To weight or not to weight: The role of domain importance in quality of life measurement. *Social Indicators Research, 00*, 1-12.


Author's Note

This research was supported by an operating grant from the Canadian Institutes of Health Research (CIHR) to Dr. Anita Palepu and Dr. Anita Hubley.
Table 1
*Hsieh's Solution for Difficult-to-Interpret Multiplicative Scores*

<table>
<thead>
<tr>
<th>Domain 1 score</th>
<th>Person A</th>
<th>Person B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Importance</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Domain score</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Domain 2 score</th>
<th>Person A</th>
<th>Person B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Importance</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Domain score</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

Sum of all importance scores 10 4

Weighted total score

| Multiplicative weighting | 20  20 |
| Hsieh's ipsatised importance weighting | 2  5 |

Source: Hsieh (in press), p. 8
Note. In the multiplicative model, the total score is obtained by summing the domain scores. For the ipsatised model, the domain scores are divided by the sum of all importance scores before they are summed to produce the total score.
Table 2
Criteria Used in the Literature on Weighting

<table>
<thead>
<tr>
<th>Area of Literature</th>
<th>Criteria used to assess importance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of Life</td>
<td></td>
</tr>
<tr>
<td>Andrews &amp; Withey (1976)</td>
<td>Overall well-being</td>
</tr>
<tr>
<td>Campbell et al. (1976)</td>
<td>Overall life satisfaction</td>
</tr>
<tr>
<td>Cummins et al. (1994)</td>
<td>Weighted quality of life</td>
</tr>
<tr>
<td>Hsieh (2003; in press)</td>
<td>Overall life satisfaction</td>
</tr>
<tr>
<td>Trauer &amp; Mackinnon (2001)</td>
<td>Weighted quality of life</td>
</tr>
<tr>
<td>Self Esteem</td>
<td></td>
</tr>
<tr>
<td>Ebbeck &amp; Stuart (1993)</td>
<td>Global self esteem</td>
</tr>
<tr>
<td>Farmer et al. (2001)</td>
<td>Global self esteem</td>
</tr>
<tr>
<td>Kaplan (1980)</td>
<td>Self-derogation</td>
</tr>
<tr>
<td>Marsh (1986)</td>
<td>General self esteem</td>
</tr>
<tr>
<td>Marsh (1994)</td>
<td>General self esteem, global physical self esteem</td>
</tr>
<tr>
<td>Marsh &amp; Sonstroem (1995)</td>
<td>Global self esteem, physical self esteem, physical activity</td>
</tr>
<tr>
<td>Job Satisfaction</td>
<td></td>
</tr>
<tr>
<td>Ewen (1967)</td>
<td>Overall job satisfaction</td>
</tr>
<tr>
<td>Kalleberg (1977)</td>
<td>Overall job satisfaction</td>
</tr>
<tr>
<td>Mastekaasa (1984)</td>
<td>Overall job satisfaction</td>
</tr>
<tr>
<td>Quinn &amp; Mangione (1973)</td>
<td>Overall job satisfaction, job related tension, voluntary termination, negative mental health</td>
</tr>
<tr>
<td>Rice et al. (1991)</td>
<td>Overall job/satisfaction</td>
</tr>
<tr>
<td>Smith Mikes &amp; Hulin (1968)</td>
<td>Voluntary termination</td>
</tr>
<tr>
<td>Staples &amp; Higgins (1998)</td>
<td>Overall job satisfaction</td>
</tr>
<tr>
<td>Vroom (1964)</td>
<td>Overall job satisfaction</td>
</tr>
<tr>
<td>Domain</td>
<td>One-week test-retest: Importance</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Being Useful</td>
<td>.61**</td>
</tr>
<tr>
<td>Drugs</td>
<td>.39**</td>
</tr>
<tr>
<td>Drug Treatment</td>
<td>.41**</td>
</tr>
<tr>
<td>Education</td>
<td>.54**</td>
</tr>
<tr>
<td>Family</td>
<td>.73**</td>
</tr>
<tr>
<td>Feeling Good</td>
<td>.45*</td>
</tr>
<tr>
<td>Friends</td>
<td>.34*</td>
</tr>
<tr>
<td>Harm Reduction</td>
<td>.50**</td>
</tr>
<tr>
<td>Health</td>
<td>.41**</td>
</tr>
<tr>
<td>Health Care</td>
<td>.37</td>
</tr>
<tr>
<td>Housing</td>
<td>.32**</td>
</tr>
<tr>
<td>Independence</td>
<td>.57**</td>
</tr>
<tr>
<td>Leisure Activities</td>
<td>.30*</td>
</tr>
<tr>
<td>Money</td>
<td>.54**</td>
</tr>
<tr>
<td>Neighbourhood Safety</td>
<td>.42**</td>
</tr>
<tr>
<td>Partner(s)</td>
<td>.54**</td>
</tr>
<tr>
<td>Community Resources</td>
<td>.27</td>
</tr>
<tr>
<td>Sex</td>
<td>.82**</td>
</tr>
<tr>
<td>Spirituality</td>
<td>.73**</td>
</tr>
<tr>
<td>Transportation</td>
<td>.40**</td>
</tr>
<tr>
<td>Treatment by Others</td>
<td>.24</td>
</tr>
<tr>
<td>Composite</td>
<td>.69**</td>
</tr>
</tbody>
</table>

* p < .05, ** p < .01; N = 241
Table 4
Correlations of Weighted and Unweighted Total IDUQOL Scores With Convergent and Discriminant Measures

<table>
<thead>
<tr>
<th></th>
<th>Weighted IDUQOL total score</th>
<th>Unweighted IDUQOL total score</th>
<th>Steiger’s Z test of differences in correlations (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergent Measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction With Life Scale</td>
<td>.60**</td>
<td>.59**</td>
<td>0.44</td>
</tr>
<tr>
<td>Rosenberg Self Esteem Scale</td>
<td>.51**</td>
<td>.54**</td>
<td>-1.15</td>
</tr>
<tr>
<td>Discriminant Measure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marlowe-Crowne Social Desirability Scale (X2)</td>
<td>.33**</td>
<td>.35**</td>
<td>-0.63</td>
</tr>
</tbody>
</table>

** p < .01; N = 241

Note. Steiger’s Z test of differences in dependent correlations tests the significance of the difference in the correlations of measures from two non-independent samples with a single third variable.
Table 5  
Correlations of Weighted and Unweighted Total IDUQOL Scores With Criterion Measures

<table>
<thead>
<tr>
<th>Criterion Variable</th>
<th>Weighted IDUQOL total score</th>
<th>Unweighted IDUQOL total score</th>
<th>Steiger's Z test of differences in correlations (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing (stable=0, unstable=1)</td>
<td>-.13*</td>
<td>-.16*</td>
<td>0.94</td>
</tr>
<tr>
<td>Engaged in sex trade (no=0/yes=1)</td>
<td>-.18**</td>
<td>-.17**</td>
<td>0.11</td>
</tr>
<tr>
<td>Currently borrowing needles (no=0/yes=1)</td>
<td>-.15*</td>
<td>-.19*</td>
<td>1.33</td>
</tr>
<tr>
<td>Currently lending needles (no=0/yes=1)</td>
<td>-.23**</td>
<td>-.25**</td>
<td>0.68</td>
</tr>
<tr>
<td>At least once daily use of heroin (no=0/yes=1)</td>
<td>-.25**</td>
<td>-.26**</td>
<td>0.36</td>
</tr>
<tr>
<td>At least once daily use of cocaine (no=0/yes=1)</td>
<td>-.13</td>
<td>-.14</td>
<td>1.25</td>
</tr>
<tr>
<td>At least once daily use of speed (no=0/yes=1)</td>
<td>-.13*</td>
<td>-.14*</td>
<td>0.21</td>
</tr>
<tr>
<td>At least once daily use of crack (no=0/yes=1)</td>
<td>-.12</td>
<td>-.12</td>
<td>0.28</td>
</tr>
<tr>
<td>Currently on methadone treatment (no=0/yes=1)</td>
<td>.08</td>
<td>.07</td>
<td>0.35</td>
</tr>
<tr>
<td>Overdose in last 6 months (no=0/yes=1)</td>
<td>-.12</td>
<td>-.14*</td>
<td>0.85</td>
</tr>
<tr>
<td>Drug treatment program in last 6 months (no=0/yes=1)</td>
<td>.02</td>
<td>.01</td>
<td>0.10</td>
</tr>
<tr>
<td>Visited ER in last 6 months (no=0/yes=1)</td>
<td>-.06</td>
<td>-.06</td>
<td>-0.10</td>
</tr>
<tr>
<td>Hospitalized in last 6 months (no=0/yes=1)</td>
<td>-.07</td>
<td>-.06</td>
<td>-0.45</td>
</tr>
</tbody>
</table>

* p < .05, ** p < .01; N = 241,

Note. Steiger's Z test of differences in dependent correlations tests the significance of the difference in the correlations of measures from two non-independent samples with a single third variable
Table 6
Correlations of Selected Domain Weighted and Unweighted IDUQOL Scores With Selected Criterion Variables

<table>
<thead>
<tr>
<th>Criterion Variable</th>
<th>Weighted IDUQOL domain score</th>
<th>Unweighted IDUQOL domain score</th>
<th>Steiger’s Z test of differences in correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing (^a) (stable=0, unstable=1)</td>
<td>-0.17*</td>
<td>-0.30**</td>
<td>2.17*</td>
</tr>
<tr>
<td>At least once daily use of heroin (^b) (no=0/yes=1)</td>
<td>-0.01</td>
<td>-0.06</td>
<td>0.55</td>
</tr>
<tr>
<td>At least once daily use of cocaine (^b) (no=0/yes=1)</td>
<td>0.16</td>
<td>0.01</td>
<td>1.72</td>
</tr>
<tr>
<td>At least once daily use of speed (^b) (no=0/yes=1)</td>
<td>-0.08</td>
<td>-0.10</td>
<td>0.23</td>
</tr>
<tr>
<td>At least once daily use of crack (^b) (no=0/yes=1)</td>
<td>0.07</td>
<td>0.02</td>
<td>0.57</td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currently on methadone treatment (^c) (no=0/yes=1)</td>
<td>0.19*</td>
<td>0.22*</td>
<td>-0.49</td>
</tr>
<tr>
<td>Drug treatment program in last 6 months (^c) (no=0/yes=1)</td>
<td>0.19*</td>
<td>0.25**</td>
<td>-0.79</td>
</tr>
<tr>
<td>Health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Currently on methadone treatment (^d) (no=0/yes=1)</td>
<td>0.04</td>
<td>0.05</td>
<td>-0.08</td>
</tr>
<tr>
<td>Drug treatment program in last 6 months (^d) (no=0/yes=1)</td>
<td>0.04</td>
<td>0.01</td>
<td>0.47</td>
</tr>
<tr>
<td>Visited ER in last 6 months (^d) (no=0/yes=1)</td>
<td>-0.09</td>
<td>-0.13</td>
<td>0.60</td>
</tr>
<tr>
<td>Hospitalized in last 6 months (^d) (no=0/yes=1)</td>
<td>-0.13</td>
<td>-0.15*</td>
<td>0.33</td>
</tr>
<tr>
<td>Health Care</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visited ER in last 6 months (^e) (no=0/yes=1)</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.66</td>
</tr>
<tr>
<td>Hospitalized in last 6 months (^e) (no=0/yes=1)</td>
<td>-0.02</td>
<td>0.01</td>
<td>-0.55</td>
</tr>
<tr>
<td>Feeling Good About Yourself</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenberg SES (^a)</td>
<td>0.35**</td>
<td>0.59**</td>
<td>-4.80**</td>
</tr>
<tr>
<td>Engaged in sex trade (^a) (no=0/yes=1)</td>
<td>-0.13</td>
<td>-0.16*</td>
<td>0.51</td>
</tr>
<tr>
<td>Feeling Good About Yourself</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engaged in sex trade (^f) (no=0/yes=1)</td>
<td>-0.01</td>
<td>-0.124</td>
<td>1.63</td>
</tr>
</tbody>
</table>

\(^*\) \(p < .05\), \(^**\) \(p < .01\); \(^a\) \(N = 209\), \(^b\) \(N = 146\), \(^c\) \(N = 150\), \(^d\) \(N = 221\), \(^e\) \(N = 211\), \(^f\) \(N = 177\)

Note. Steiger’s Z test of differences in dependent correlations tests the significance of the difference in the correlations of measures from two non-independent samples with a single third variable.
Table 7  
*Frequency of Domains Selected as Important*

<table>
<thead>
<tr>
<th>Number of domains</th>
<th>Percent</th>
<th>Cumulative percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>4</td>
<td>0.8</td>
<td>1.7</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
<td>3.3</td>
</tr>
<tr>
<td>6</td>
<td>1.2</td>
<td>4.6</td>
</tr>
<tr>
<td>7</td>
<td>2.1</td>
<td>6.6</td>
</tr>
<tr>
<td>8</td>
<td>1.7</td>
<td>8.3</td>
</tr>
<tr>
<td>9</td>
<td>1.7</td>
<td>10.0</td>
</tr>
<tr>
<td>10</td>
<td>5.4</td>
<td>15.4</td>
</tr>
<tr>
<td>11</td>
<td>6.2</td>
<td>21.6</td>
</tr>
<tr>
<td>12</td>
<td>6.2</td>
<td>27.8</td>
</tr>
<tr>
<td>13</td>
<td>5.0</td>
<td>32.8</td>
</tr>
<tr>
<td>14</td>
<td>8.3</td>
<td>41.1</td>
</tr>
<tr>
<td>15</td>
<td>12.0</td>
<td>53.1</td>
</tr>
<tr>
<td>16</td>
<td>7.9</td>
<td>61.0</td>
</tr>
<tr>
<td>17</td>
<td>6.2</td>
<td>67.2</td>
</tr>
<tr>
<td>18</td>
<td>9.1</td>
<td>76.3</td>
</tr>
<tr>
<td>19</td>
<td>7.1</td>
<td>83.4</td>
</tr>
<tr>
<td>20</td>
<td>5.8</td>
<td>89.2</td>
</tr>
<tr>
<td>21</td>
<td>10.8</td>
<td>100.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>100.0</strong></td>
<td></td>
</tr>
</tbody>
</table>

*N = 241*
### Table 8
**Correlations of MImp and MUimp With Convergent and Discriminant Measures**

<table>
<thead>
<tr>
<th></th>
<th>MImp</th>
<th>MUimp</th>
<th>Steiger's Z test of differences in correlations (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Convergent Measures</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction With Life</td>
<td>.62**</td>
<td>.31**</td>
<td>5.20**</td>
</tr>
<tr>
<td>Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenberg Self</td>
<td>.50**</td>
<td>.33**</td>
<td>2.66**</td>
</tr>
<tr>
<td>Esteem Scale</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Discriminant Measure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marlowe-Crowne</td>
<td>.33**</td>
<td>.24**</td>
<td>1.33</td>
</tr>
<tr>
<td>Social Desirability</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scale (X2)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**p <.01; N = 215**

Note. MImp = mean of satisfaction ratings for all important domains; MUimp = mean of satisfaction ratings for all unimportant domains. Steiger's Z test of differences in dependent correlations tests the significance of the difference in the correlations of measures from two non-independent samples with a single third variable.
Appendix

Literature Review

The concept of using importance as a weighting factor in the measurement of psychological and sociological processes has a long and remarkably tenacious history. An early reference to this practice in relation to self concept is found in the writing of psychologist William James. In an oft-cited passage, first published in 1890, James (1950) said that

I, who for the time have staked my all on being a psychologist, am mortified if others know much more about psychology than I. But I am contented to wallow in the grossest ignorance of Greek. My deficiencies there give me no sense of personal humiliation at all. Had I 'pretensions' to be a linguist, it would have been just the reverse. (p. 310).

James proposed the following formula: self-esteem= success/pretensions, and suggested that self-esteem can be increased by either through reducing the denominator (pretensions) or increasing the numerator (success). James' concept of a relationship between ability in a specific area, the centrality of that area to a person's sense of worth, and overall self esteem has made its way, with occasional variations and changes, into several theories of self-concept, including those of Coopersmith (1967, as cited in Marsh, 1996), Wylie (1974, as cited in Marsh, 1996), Rosenberg (1979), and Harter (1985, 1990, 1999).

The idea of the significance of subjective valuation is also found in other areas of psychology and the social sciences, although the terms 'pretensions' and 'desires' have generally been replaced with 'importance'. In quality of life research, the effectiveness of weighted scores was explored in two large-scale, well-known studies by Andrews and Withey (1976) and Campbell et al. (1976). Although these studies did not provide any support for the use of
importance weighting, the appeal of using importance in quality of life measurement has remained strong. For example, one review of quality of life measures in the medical literature cited importance ratings as a criteria for assessment, considering them a desirable feature that increased an instrument's rating (Gill & Feinstein, 1994). The authors of another article stated that a "true assessment of quality of life can only be achieved using weights for individual patients" (Carr & Higginson, 2001). Trauer and Mackinnon (2001), despite finding fault with the most common approach to importance weighting, concluded that importance itself remains a valid area of study and suggested that the problem lies in the methodology, not the concept. Hsieh (in press) stated that the issue should not be whether, but rather how, to weight.

Importance weighting has also been incorporated into the measurement of job satisfaction. Vroom (1964) used the term 'valence' to indicate "affective orientations towards particular outcomes" (p. 15) in his discussion of job satisfaction, and provided a number of methods for including the measurement of valence in the calculation of overall job satisfaction. Other authors who have explored weighted scores in job satisfaction, with varying results, include Quinn and Mangione (1973), Mastekaasa (1984), and Staples and Higgins (1998).

The main assumption underlying the use of importance as a weighting factor is that an area that is more important to a person will have a larger influence, either positive or negative, on his overall level of the construct of interest. For example, a person might feel that physical appearance is a very important factor in his self esteem, but that athletic ability, in contrast, is not. The theory of importance suggests that the level of this person's satisfaction with his physical appearance will have a greater impact on his overall self esteem than will his satisfaction with athletic ability. That is, high satisfaction with physical appearance will increase his self esteem more than a similar level of satisfaction with athletic ability, simply
because the former matters more. Similarly, dissatisfaction with his appearance will deflate his self esteem more than will a sense of athletic incompetence. In measurement terms, this implies that incorporating subjective assessments of importance should improve the measurement of the construct under investigation.

Rather than using a ratio calculation such as the one proposed by James (1950), importance is now usually used to calculate what is generally termed a multiplicative score. The most common approach is to ask individuals to rate both the importance of, and their satisfaction with, various domains or areas of the construct being measured. The two ratings, importance and satisfaction, are multiplied to arrive at a total domain score, and then all domain scores are summed to arrive at an overall weighted score for each respondent. Another approach to importance ratings, sometimes referred to as a discrepancy or subtractive model, has also been used in both self esteem and job satisfaction research. In this model, importance scores are incorporated through subtraction rather than multiplication (see Tables 1 and 2).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Importance</th>
<th>Satisfaction</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>5</td>
<td>25</td>
</tr>
</tbody>
</table>

The intuitive appeal of importance weights remains powerful, yet some of the research on their effectiveness in improving measurement has been discouraging, raising various empirical concerns with these ratings. Three areas in which importance ratings have been explored fairly extensively, and which will be discussed here, are quality of life, self esteem, and job satisfaction. Although the focus of this review is on weighting, specifically subjective weighting,
it is important to understand some of the general theory in each of these three areas in order to make sense of the debate surrounding the utility of importance ratings. Therefore, this general theory will be addressed, followed by a discussion of the main concerns that have been raised with regards to the use of importance ratings. Next, a summary of the evidence for and against importance weighting will be provided. First, however, it is necessary to provide a brief discussion about the terminology that will be used throughout this review.

**Terminology**

The terminology used to discuss importance weighting is inconsistent across authors and areas of research. Some explanation of these differences and the way different terms will be used here will be helpful in preventing confusion.

The terms 'domain', 'area', 'life area', and 'facet' are all used in the literature and will be used here to refer to the components that make up a larger construct such as life satisfaction or self esteem. The concept of 'importance' has been referred to in the literature as 'values', 'salience' and 'desires', to name just a few terms. Although there are likely subtle differences in the meanings of these terms, this review focuses specifically on the issue of subjective evaluations that imply importance, and therefore 'importance' will be used to indicate any of these assessments. The term 'satisfaction' is naturally used in the measurement of job satisfaction, and frequently in quality of life and well-being research. In the area of self esteem, it is more common to use the term 'self concept' or 'competence' rather than satisfaction. Nevertheless, while 'satisfaction', 'competence' and 'self concept' are clearly not identical concepts, for the purposes of discussing importance weighting they function in a similar manner, representing subjective evaluations of the domains that comprise the actual construct of interest. To minimize the use of awkward terms such as 'satisfaction/self-concept' or 'satisfaction/competence',
'satisfaction' will be used to represent these evaluations when discussing more than one area of research at once. If the focus is on just one area, such as for example self esteem, the term that is most frequently used in the relevant literature will be applied.

The multiplicative model used in much of the research involving weighting is also sometimes described as an 'importance-weighted model' or an 'interactive model'. These terms all refer to the basic method (with minor variations) of multiplying satisfaction scores with importance scores. The terms 'discrepancy model' and 'subtractive model' are both used to refer to a different type of model in which importance ratings are subtracted from satisfaction scores. 'Unweighted scores' are generally obtained by simply summing domain satisfaction scores; this is sometimes referred to as an 'additive model'.

Research on quality of life overlaps a great deal with research on subjective well-being; although these are not identical constructs (George, 1991), the findings from subjective well-being will be used to inform the discussion of quality of life. The terms 'quality of life' and 'life satisfaction' will also be used interchangeably. Research on self esteem and self concept overlaps in a similar manner, and both will be addressed together.

The literature in the areas of quality of life, self esteem and job satisfaction provides insight into several concepts that are relevant to a more general discussion of importance weighting. Therefore, some key points in each of these areas will be highlighted before moving on to the discussion of issues in weighting.

Quality of Life

One important debate in the measurement of subjective well-being is the distinction between bottom-up and top-down theories. A bottom-up approach presumes that, when asked to give a rating of their general well-being, people perform a mental tally of the various domains
that make up their lives before arriving at the final number. Thus, the bottom-up approach views overall well-being as a sum of many parts. In contrast, a top-down model views people as possessing general traits that predispose them to see their lives in a positive or negative light. It is this overall disposition that, in turn, influences their assessment of specific life areas (Diener, 1984). In this approach, satisfaction in different domains essentially reflects global satisfaction. These two views are reflected in the two main types of quality of life measures. Some instruments, referred to as global measures, echo the top-down theory. They are composed of one or more items that are quite broad in scope and address overall life satisfaction without referring to specific life areas. Diener's Satisfaction With Life Scale (Diener et al., 1985) is one well-known example of this type of instrument. The other type of quality of life measure, often referred to as a domain-based measure, reflects a bottom-up theory of quality of life. Respondents are asked to rate various life domains and an overall score is obtained by summing the domain scores. Life domains that might be included in such an instrument are family, work, leisure activities, and income. One example of this type of measure is the Injection Drug User Quality of Life scale (IDUQOL, Brogley et al., 2003).

Importance ratings are generally used with domain-based measures of quality of life. Each domain is given an importance rating as well as a satisfaction rating, and these are then combined into a domain score. Domain scores are in turn combined to provide a score that reflects overall quality of life. The multiplicative model described above is the most common approach to incorporating importance ratings into the measurement of quality of life (Trauer & Mackinnon, 2001).
Self Concept/Self Esteem

As with quality of life, there are multiple theories of self-esteem and self concept. Some authors have suggested that self-esteem is a unidimensional construct that is best reflected by a single score. Although measures based on the unidimensional model may ask respondents to provide self-concept ratings in different domains, it is only the overall score, calculated by summing the various items, that is considered significant. Other authors have argued for a more multidimensional view of self-concept in which the scores in the separate domains each convey significant information (Harter, 1986). Harter, for example, used factor analysis to identify five separate domains that figure in the self esteem of children. Rosenberg (1979) has argued for both sides of the debate. He suggested that self-concept is made up of various components that are unequally important to an individual's self-esteem. At the same time, people also have an overall sense of self. According to Rosenberg, these are two distinct, but related, constructs: "although the specific and the general obviously overlap, they are also far from identical" (p. 21). Hoge and McCarthy (1984) also argued that global and facet-based measures may be tapping somewhat different constructs.

In the measurement of self-esteem, much like in quality of life, instruments may be global or facet-based, with importance incorporated into facet-based measures. Marsh (1994) identified a number of approaches to incorporating importance ratings, including an interactive model, actual-ideal discrepancies weighted by importance, actual-importance discrepancies, and profile similarity indices. The interactive (multiplicative) model and actual-importance discrepancies will be discussed in more detail here because they are the most commonly explored approaches in the area of self esteem.
In the interactive model, the contribution of each facet to the total score varies based on the rating of that facet's importance. The most common approach is to use individual importance ratings, although other possibilities include weighting an individual's facet rating by the mean of the sample's (or a subgroup's) importance scores for that facet (Marsh, 1986). Marsh suggested using analysis of variance to test the interactive model. For each facet, the analysis should find a main effect for self-concept, and a significant self-concept/importance interaction, each of which should contribute "significantly, positively, and independently to the prediction of esteem" (p. 1226). According to Marsh, although the presence of a main effect for importance does not support the model, it need not disconfirm it either. Correlational analyses may also be used to assess the effect of importance. To support the multiplicative model, correlations between weighted scores and measures of overall self-esteem should be higher than those between unweighted scores and overall self-esteem (Marsh, 1994).

Actual-importance discrepancy scores also incorporate importance ratings, though not through multiplication. Instead, the rating of perceived competence in each facet is subtracted from the importance of that facet, resulting in a score that reflects the discrepancy between the actual and the desired. The smaller the size of the discrepancies (i.e., the closer the discrepancy scores are to zero), the greater a person's self-esteem (Harter, 1986). In this model, overall self-esteem is inversely related to the extent to which a person's self-concept in each facet is exceeded by the importance of that facet. For example, someone who does not feel confident of her academic ability, but attaches little importance to this area, will experience higher self-esteem than someone who believes she is academically challenged but dreams of obtaining a doctoral degree.
According to Marsh (1986), testing the discrepancy model through analysis of variance should reveal main effects for both facet self concept and facet importance, with the former contributing positively and the latter contributing negatively to the prediction of self esteem. In terms of correlations, facet self concept should contribute positively, but importance should contribute negatively, to overall self esteem.

Job Satisfaction

Studies of the role of importance in job satisfaction have, for the most part, compared an unweighted, additive model, in which domain satisfaction scores are simply summed, with a multiplicative model of importance weighted scores, although discrepancy scores have also been addressed (Vroom, 1964). Job satisfaction measures include both domain-based and global instruments.

Issues Surrounding Importance Ratings

The use of importance ratings has been challenged on a number of psychometric, theoretical, and methodological grounds. These include concerns about the reliability of importance scores, as well as the validity of inferences made from these. Other concerns relate to the theories underlying importance ratings and the models used to incorporate importance into measurement.

The Reliability of Importance Ratings Is Low

One concern is that the reliability of importance ratings tends to be lower than ideal, and is often lower than that of corresponding satisfaction scores. For example, in their large-scale study of American quality of life, Campbell et al. (1976) found that the test-retest reliability of importance ratings was lower than that of domain satisfaction ratings and did "not compare very favorably even with the single-item global reports of happiness or life satisfaction. Either people
change their minds with disconcerting frequency about how important domains are to them, the measurement contains some uncommon error, or, as is most likely, some combination of the two is occurring" (p. 87-88). Other evidence from the quality of life literature suggests that the internal consistency of importance ratings may also be lower than that of satisfaction ratings. For example, in the development of one domain-based quality of life scale, the Cronbach's alpha for satisfaction items was .73 while the alpha for importance items was .65 (Cummins et al., 1994).

This pattern of lower reliability of importance ratings has also been found in studies of self esteem. Marsh (1986) reported that his single-item importance ratings exhibited lower test-retest reliability (r=.57) than either single-item (r=.70) or multi-item (r=.87) self concept ratings when measured twice within a one-month period. In a sample of 375 college students, Farmer et al. (2001) found alphas of .65 to .82 for importance scales and .70 to .82 for self-concept scales. In the same study, test-retest coefficients, measured at over 1 month, ranged from .57 to .80 for importance scales and .71 to .83 for self concept scales.

Whether the reliability estimates for importance ratings reported in these studies are problematic is perhaps a matter of interpretation, but it is worth noting that Nunnally (1978) recommends alphas of at least .70 for research. Measured against this standard, some of the reliability coefficients cited here for importance scores may be cause for concern, or at least caution.

Not all researchers have reported large differences in the reliability of importance and satisfaction scores. In one study of self esteem, for example, researchers found an alpha coefficient of .88 for a subscale measuring importance, which compared favourably with the reliability of the domain competence subscale at .87 (Ebbeck & Stuart, 1993). Some of the research in job satisfaction also suggests that importance scores are not always less reliable than
satisfaction ratings. Kalleberg (1977) reported Cronbach's alphas of .68 to .85 for importance ratings and similar alphas of .68 to .87 for satisfaction ratings with a sample of close to 1,500 individuals. Hofstede (1980), looking at the stability of importance and satisfaction ratings by country, noted coefficients of .39 for overall satisfaction scores but coefficients of .49 to .95 for importance scores. Altogether, the research suggests that importance ratings may be somewhat less reliable than importance ratings, both in terms of internal consistency and temporal stability, but this finding is inconsistent. A serious limitation to the generalizability of this statement is the fact that the majority of studies in which weighting was used do not report reliability coefficients for either satisfaction or importance ratings.

One possible explanation for lower reliability coefficients for importance scores may be the frequent use of single-item measures, a point raised by Marsh (1986). It does appear to be the case that almost all research using importance ratings, whether in quality of life, self esteem, or job satisfaction, has relied on single item-measures of domain importance. However, it should be noted that many studies that use single-item measures of domain importance also use single-item measures of domain satisfaction. While this does not negate concerns about low reliability coefficients for importance ratings, the same concern may also apply to satisfaction scores in these cases.

The authors of one study on job satisfaction collected importance information with seven different measures in order to compare various methods of reporting importance information (Rice et al., 1991). The measures included four rating methods, two ranking methods and one point distribution method. As well as being evaluated individually, the scores from the seven methods were combined into a composite importance measure for each domain. The resulting in alphas ranged from .81 to .92 across twelve domains. While the composite measure did perform
slightly better than any single method of measuring importance, the authors suggested that the difference between the composite measure and the top three individual methods (direct rating, simple ranking and point distribution) was not large enough to warrant the additional time required to administer multiple importance measures. In the end, they recommended the use of direct ratings on a Likert-type scale. Nevertheless, this study does present one possible solution for those who feel that the reliability estimates obtained with single-item measures of importance are unacceptably low.

A different concern about the reliability of weighted scores arises when several sources of error are combined through the multiplication of importance and satisfaction scores. Classical test theory states that summing numerous items increases the reliability estimate of the overall measure, but little research has been carried out on the effect of other functions such as multiplication. Trauer and Mackinnon (2001) expressed the concern that combining satisfaction and importance ratings through multiplication introduces error from both sets of measures, thereby decreasing, rather than increasing, the reliability of the total score. This issue was also raised by Smith Mikes and Hulin (1968), who stated that, even when measures of importance have high reliability, the sum of domain satisfaction by importance products will have lower reliability than the sum of domain satisfaction ratings alone. As a result, they argued, weighted scores will never correlate as highly as unweighted scores with any criterion. This argument is not necessarily borne out by the research however, as some authors have reported higher (albeit generally not significantly higher) correlations between weighted scores and a criterion measure, compared to unweighted scores (e.g.; Hsieh, in press; Marsh, 1986).
Validity

A number of questions around importance ratings relate to the validity of inferences made from these scores. One question is whether satisfaction ratings already include assessments of importance. Related to this is the question of whether high positive correlations between satisfaction and importance ratings are a sign that the two constructs are identical, or at least closely related. Finally, there is the possibility that bias, selectivity, and lack of insight influence the measurement of importance.

*Satisfaction ratings already include importance.* Trauer and Mackinnon (2001) suggested that importance ratings are an unnecessary component in quality of life measures because importance is already inherent in the selection of domains for inclusion on an instrument. They argued that most measures cover an exhaustive range of life domains selected on the basis of interviews, focus groups, reviews of the literature (notably Andrews & Withey, 1976, and Campbell et al., 1976), existing measures, and expert knowledge. In other words, life areas are included in a measure precisely because they are important to the vast majority of people, and there is little to be gained from soliciting additional importance ratings.

Other researchers in other areas have made similar suggestions. Harter (1999) and Harter and Whitesell (2001) noted that about 75 to 80% of respondents in their studies of self esteem among children and adolescents rated all domains on the measures as important. Quinn and Mangione (1973) pointed out that they selected job domains for inclusion in their study precisely because these were believed to be important to most people. They note that instead of capturing the true subjective importance of these domains, this may have forced participants to make artificially fine distinctions between areas of similar importance.
Hsieh (2003) was less convinced that domain-based measures include only important areas, pointing to the less-than-perfect correlations between scores on domain-based quality of life measures and those that ask for a single, global score. He suggested that the reason for this discrepancy may lie in the fact that the domains do not constitute a complete list of all possible important life areas for all people. Even if it is true that most quality of life measures are based on a comprehensive list of important domains, it does not necessarily follow that there is no hierarchy of importance among these domains, or that there is nothing to be gained from drawing this out (Hsieh, in press). In fact, it could be argued that the very exhaustiveness of some measures makes it more likely that at least some domains will be more or less important to some individuals.

Trauer and Mackinnon (2001) proposed that satisfaction scores that fall at either the extreme positive or extreme negative end of the rating scale already indicate that the domain has some importance to the respondent. Quinn and Mangione (1973) suggested that most people mentally assess the importance of a work domain before responding to satisfaction ratings, thereby incorporating importance into satisfaction ratings. They also argued that extreme satisfaction ratings imply importance, pointing to research indicating that those domains rated as most important were also those that showed the greatest variability in satisfaction scores, whereas domains rated as less important tended to be associated with a more limited range of satisfaction scores. Rice et al. (1991) argued that "because facet importance is implicitly reflected in each facet-satisfaction score, it is conceptually and statistically redundant to consider facet importance as a moderator of the relationship between facet satisfaction and overall job satisfaction" (p. 32).
Mastekaasa (1984), for his part, rejected the idea that importance ratings are incorporated into satisfaction ratings, at least for the quality of life component of his study of weighted scores in job and life satisfaction. He argued that, if importance is included in domain satisfaction scores, the effect of these domains on overall life satisfaction would be more or less constant. The data from his study, however, showed significant differences in the effect of different domain scores on overall life satisfaction.

*Measures of satisfaction and importance may tap the same construct.* Do importance scores measure a distinct construct? One large-scale study of subjective well-being found a trend of positive correlations between importance and satisfaction ratings (Campbell et al., 1976). One consequence of such a trend would be that, as the correlations reach high levels, it becomes necessary to question whether satisfaction and importance ratings are measuring the same construct. Unfortunately, it is difficult to assess evidence regarding this possibility through a review of the literature because correlations between importance and satisfaction scores are not routinely reported in research on self esteem, job satisfaction and quality of life. One exception is a study of self esteem among 930 college students, high school students, and athletes (Marsh, 1986), in which correlations between domain importance and domain self-concept ranged from .07 to .86, with 10 out of 12 correlations below .47 and a mean correlation of .35. Although these correlations were all positive, they were not so high as to suggest identical constructs. Farmer et al. (2001) reported domain self concept and importance correlations of .37 to .74, which begin to fall into the moderately high range but still do not suggest the constructs are the same. In the area of quality of life, Cummins et al. (1994) found correlations between domain importance and domain satisfaction ranging from -.08 to .33. Moreover, factor analysis from one study of job
satisfaction indicated that, for that sample, ratings of importance and satisfaction were tapping different constructs (Kalleberg, 1977).

Assessing the research as a whole, it would appear that, while there may be some overlap in the constructs of satisfaction and importance, they are not identical or even necessarily very highly related constructs.

Bias, insight, and selectivity. A number of authors have proposed various sources of response bias in importance scores. Hofstede (1980) suggested that ratings of work domain importance are susceptible to acquiescence bias, or the tendency to rate all areas as important. He argued that this is particularly the case for respondents who have lower social status or educational levels than the researchers administering the measures, or for respondents from more collectivist cultures. Mastekaasa (1984) found some weak evidence for such a bias in both his quality of life and job satisfaction data, and suggested ipsatizing scores to counteract this effect.

Some researchers have also found that importance scores in certain domains are particularly vulnerable to the influence of social and cultural norms. Specifically, direct assessments of the importance of material domains for a sample of Canadian and American adults were found to be much lower than indirect assessments (Atkinson & Murray, 1980). Campbell et al. (1976) found that money was rated directly as relatively unimportant, but indirectly as very important, by their American sample.

Part of the reason for this discrepancy between direct and indirect measures of importance may be related to insight. Mastekaasa (1984) pointed to research suggesting that the more abstract a concept, the harder it is to evaluate. He noted that, in his own research, weighting with importance scores was effective in the area of job satisfaction, but not in the area of life satisfaction. He proposed that the reason for this might be the fact that job satisfaction is more
concrete and quantifiable than life satisfaction, thus making it easier to judge the relative influence that various domains might have on overall job satisfaction.

Campbell et al. (1976) suggested that importance ratings in quality of life are influenced by both actual and ideal situations, and that individuals adjust to less-than-ideal life situations by reducing the importance that they assign to domains in which they are not satisfied. Campbell et al. used findings from the domain of marriage as an example to illustrate this phenomenon. They noted that widowed women rated marriage as less important than did women who were still married. The authors pointed out that it is unlikely that women who do not value marriage are widowed more frequently than others; therefore some other process was most likely responsible for the difference in importance assigned to this domain. Women in the study who were divorced also tended to rate marriage as somewhat, but not excessively, less important than did married women for the first 5 years post-divorce but, after this point, their importance ratings dropped noticeably. Campbell et al. proposed that this drop in importance was the result of an adjustment to single life among those women who did not have an opportunity to remarry (although they also recognised that those women who valued marriage less were also less likely to seek remarriage).

In the area of self-esteem, Rosenberg (1979) suggested that individuals will enhance their self-esteem by assigning greater importance to those domains in which they feel successful and less importance to those areas in which they feel they are lacking. He also acknowledged that the relationship is not unidirectional, and that people will try to excel in areas that they deem important.

Other researchers in self esteem have embraced this idea, which has been called the discounting hypothesis (Harter, 1986) or the selectivity hypothesis Marsh (1986). Harter (1986;
1990; 1999) and Harter and Whitesell, 2001) have cited evidence suggesting that children and adolescents with high self esteem are more likely to assign low importance to areas in which they feel less competent, while their peers who have low self esteem assign more importance to areas of lower competence. For example, children with high overall self esteem gave their least competent domain a mean importance score of 2.6 on a 4 point scale in which higher scores were indicative of greater importance. Children with medium and low self esteem assigned mean importance ratings of 3.3 and 3.6, respectively, to their least competent domain (Harter, 1986).

Other researchers have also suggested that, theoretically at least, the selectivity hypothesis makes sense: "Logically, individuals would be expected to embrace a system of values that is most likely to facilitate the development of positive self-esteem, such as valuing domains of personal competence and devaluing domains of personal incompetence" (Ebbeck & Stuart, 1993, p. 254).

Campbell et al. (1976) treated the issue of adjusting importance ratings to fit life circumstances as a threat to the validity of inferences made from these ratings, but other authors have clearly presented a different perspective where adjusting assessments of domain importance to life circumstances or personal deficiencies may reflect a legitimate aspect of the relationship between satisfaction and domain importance.

The implications for the validity of inferences made from weighted scores raised by questions of bias, insight, and whether importance and satisfaction are distinct concepts are clearly important. They are also complex, and have not yet been fully resolved.

*Multiplicative Scores May Be Difficult to Interpret*

Trauer and Mackinnon (2001) pointed out that multiplicative scores can be difficult to interpret under certain conditions. Identical scores may be obtained by multiplying a high domain satisfaction score with a low importance score, or by multiplying a low domain...
satisfaction score with a high importance score. As an example of this problem, imagine that someone has a satisfaction score of 5 in the area of family and rates the importance of this domain as 7. The total domain score, calculated under the multiplicative model, would be 5 x 7 = 35. Another individual might give family a satisfaction rating of 7 but an importance rating of 5, resulting in the same total domain score of 35. As Trauer and Mackinnon (2001) asked, "are we to conclude that these quite different situations represent the same 'true' level of QoL?" (p. 580). Marsh (1986) noted this same issue in using the multiplicative model to measure self esteem. He presents the hypothetical example of two individuals, one with high self concept in a domain that is rated as unimportant, and another who has low self concept in a domain that is rated as very important. According to the theory underlying the use of importance ratings, in the case of the first individual, the domain should have a small, though positive, influence on overall self esteem. In the case of the second individual, the domain should have a strong, but negative, effect on overall self esteem. Instead, both sets of ratings will produce similar domain scores, thereby clouding their different meanings.

One possible solution to this problem is to divide the weighted domain scores by the sum of all importance scores (Hsieh, in press). Through this method, the problem is resolved, as Hsieh demonstrates with the following example (see Table 3).

Under this alternative weighting scheme, higher importance scores combined with lower satisfaction scores result in a lower overall score (2) than low importance combined with higher satisfaction (4). Not only does this make it possible to distinguish between the scores, but it also ensures that the weighted total scores more clearly reflect the level of satisfaction (whether high or low) with more important domains.
Table 3
Hsieh's Solution for Difficult-to-Interpret Multiplicative Scores

<table>
<thead>
<tr>
<th></th>
<th>Person A</th>
<th>Person B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1 score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Importance</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Domain score</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td><strong>Domain 2 score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Importance</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Domain score</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td><strong>Sum of all importance scores</strong></td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td><strong>Weighted total score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiplicative weighting</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Hsieh's ipsatised importance weighting</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

Source: Hsieh (in press), p. 8

Note. In the multiplicative model, the total score is obtained by summing the domain scores. For the ipsatised model, the domain scores are divided by the sum of all importance scores before they are summed to produce the total score.

Another method for dealing with these similar importance-by-satisfaction products is to transform the satisfaction ratings into standardised scores, such as z scores. As a result, domain scores obtained from multiplying satisfaction and importance ratings will be positive when satisfaction scores are above the mean and negative when the satisfaction scores are below the mean (Marsh, 1986).

Finally, Hsieh (in press) has suggested that using importance rankings, rather than ratings, might also be effective in resolving this problem. In the same study that looked at the alternative weighting scheme mentioned above, Hsieh found that using domain importance rankings as a weighting factor in a multiplicative model improved correlations between a single-item global measure of life satisfaction and domain-based scores from .39 (for unweighted scores) to between .41 and .46 for various rank weighted methods. Although Hsieh pointed out
that these results are preliminary, they are intriguing. Indirect theoretical, if not empirical support for using a ranking system is provided by Atkinson and Murray (1980), who suggested that when people are forced to rate two domains that are of similar importance to them, it may be relative value, not absolute value, that forms the basis of their decision.

Nevertheless, not everyone who has applied domain importance rankings has obtained promising results. In a study of job satisfaction with a sample of 163 workers, Ewen (1967) found that, although the job component ranked as most important by respondents correlated significantly higher with two global measures of job satisfaction than did the least important component, overall, both unweighted scores and those obtained from multiplying satisfaction and importance ratings correlated more highly with criterion measures. A study by Blood (1971) failed to find any support for using importance rankings. Rice et al. (1991) also found that rankings of importance performed no better than ratings in predicting overall satisfaction. Finally, a job satisfaction study with a sample of 3,804 workers found that using importance rankings did not improve correlations with an overall measure compared to a simple additive model (Staples & Higgins, 1998).

The Level of Measurement of Importance Data is Inadequate

Another potential issue with the use of importance weighting that has been only infrequently addressed in the literature relates to levels of measurement. Trauer and Mackinnon (2001) pointed out that multiplication requires ratio level measurements if the results are to be meaningful, yet this condition is not met by most quality of life, self esteem, or job satisfaction data. A defining characteristic of ratio scales is a true zero point, or a point where no amount of the attribute being measured is present. Another characteristic of ratio scales is that the ratio of any two scores does not vary when multiplied by a constant (Pedhazur & Schmelkin, 1991). Yet
how plausible is it that a scale measuring importance will possess either of these characteristics? Hsieh (2003) questioned whether giving a domain an importance rating of zero truly implies that it has no impact on life satisfaction. He also suggested that it is not clear that changes in importance ratings follow a linear pattern. That is, "how much more weight should be given for each one point increase in a five-point (or seven-point) importance rating scale"? (Hsieh, in press). Or, as Mastekaasa (1984) pointed out, it cannot be assumed that a domain importance rating of 2 means that the domain is considered twice as important as a domain given a rating of 1. The choice of scale anchors may also be critical. Trauer and Mackinnon (2001) used hypothetical data to show that multiplicative scores are, in fact, quite susceptible to even slight changes in the scaling of the component satisfaction and importance scores. Discrepancy models, which only assume interval scaling, are somewhat less prone to these problems, yet even the assumption of interval data may be stretching the characteristics of most importance scales (Quinn & Mangione, 1973).

Not all authors agree with Trauer and Mackinnon (2001) that the limitation imposed by the level of measurement implied in a Likert-style rating scale is insurmountable. Mastekaasa (1984) explored alternative methods for calculating weighted scores that may circumvent the need for ratio level data. In a study comparing the use of weighted scores in quality of life and job satisfaction, he proposed that an additive model can be expressed as:

\[ J = a + \sum b_i F_i + e \]

whereas the typical multiplicative model is expressed by:

\[ J = a + \sum d_i F_i I_i + e \]

where \( J \) is overall job satisfaction, \( F \) is domain satisfaction, \( I \) is domain importance, \( a \), \( b_i \) and \( d_i \) represent regression coefficients and \( e \) represents error (p. 150). This formula for the
multiplicative model presupposes that importance is measured on a ratio scale, an assumption which is unlikely to be met. However, Mastekaasa argued that substituting the following formula ensures that the effect of domain satisfaction will increase linearly as importance increases, without requiring ratio data:

\[ J = a + \sum b_j F_j + \sum c_j I_j + \sum d_j F_I + e \]

where \( J \) is overall job satisfaction, \( F \) is domain satisfaction, \( I \) is domain importance, \( a, b_j, c_j \) and \( d_j \) represent regression coefficients, and \( e \) represents error (p. 151). Therefore, the need for ratio-level data may be avoided through a slight adjustment to the formula for the multiplicative model.

*Multiplicative Scores May Not Reflect the Underlying Theory*

The claim has been made that domain scores weighted under the multiplicative model may not properly reflect some of the assumptions of the underlying interactive hypothesis. Marsh (1986) presented the hypothetical scenario of one person who has low self concept in areas deemed unimportant but high self concept in areas considered important, and compares this situation to that of someone who has varying levels of self concept but deems all areas important (see Table 4).

<table>
<thead>
<tr>
<th></th>
<th>Person C</th>
<th>Person D</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Domain 1 score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Importance</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td><strong>Domain 2 score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Importance</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>
According to the interactive hypothesis, the first individual should have higher overall self-esteem, because those areas where self concept is lower are also those that matter less. However, it is the person who rates both domains as more important who has a higher overall score, despite low satisfaction in one of those domains. Marsh's proposed solution is to ipsatise importance ratings, for example by dividing each of an individual's domain importance ratings by the sum of all her importance ratings. This is the same solution proposed by Hsieh (in press) to deal with problematic importance-by-satisfaction products, and may provide a parsimonious solution to two problems at once.

It is evident from the literature that there has been much discussion about the concerns that have been raised regarding importance weighting. While some researchers have suggested that problems with reliability, validity, theory and methodology pose significant barriers to the effectiveness of importance ratings, others are more sanguine, and have proposed various solutions. It is difficult to say if any of the concerns discussed here are insurmountable; perhaps that is a matter, in many cases, of judgement. Clearly, however, any decision to incorporate importance ratings into measurement should be informed by the debate surrounding their potential limitations.

*Do Importance Weights Work?*

Perhaps the most important question to raise in connection with importance weighting is whether this practice can improve the measurement of a construct. In other words, does importance weighting work? The answer to this question depends in part on the criteria used for assessment. Although there are a number of potential criteria, the best place to begin this discussion is with the existing research. Virtually all studies to date have evaluated the effectiveness of weighted and unweighted scores by comparing their relationships to scores on a
global measure of the construct of interest, whether quality of life, self esteem, or job satisfaction. Other criteria, such as more objective outcome measures, have been used in a small number of studies. The use of global measures as the primary criterion is not without problems, but a summary of the relevant research is nevertheless an essential first step towards answering the question of whether importance weighting 'works'.

Quality of Life

Importance ratings have been used in numerous studies of subjective quality of life, but the evidence for their effectiveness in this area is slim. In some cases, reliance appears to be placed primarily on the face validity of importance ratings, and no statistical or psychometric justification is provided for their inclusion in a study (e.g., Carr & Higginson, 2001; Gill & Feinstein, 1994). The only study in the area of quality of life in which importance has been successfully applied is Hsieh's (2003, in press) research using importance ranking. His study found that using importance ranking as a weighting factor resulted in slightly higher correlations with a single-item global measure, compared to using unweighted domain satisfaction scores. Nevertheless, the same study found that ranking did not lead to improved correlations with two multi-item global measures of life satisfaction. When domain importance was rated, instead of ranked, the correlations of the weighted scores with the single-item and multi-item global measures were lower than, or at most identical to, those of the unweighted scores.

Other researchers have also found no support for the use of importance ratings in the measurement of quality of life. Andrews and Withey (1976) found only weak correlations ranging from .15 to .30, between importance ratings and global quality of life scores, and decided that a simple additive model provided the best fit for their data. Despite being drawn to the idea of importance as an influencing factor, Campbell et al. (1976) concluded after testing
various models that assessments of importance did not improve the prediction of overall well-being in their study. During the development of the Comprehensive Quality of Life Scale (Cummins et al., 1994), importance ratings were considered necessary on theoretical grounds, yet subsequent psychometric evaluations found no benefit to their inclusion. When respondents were divided into two groups representing low and high subjective quality of life, it was the domain satisfaction scores, not the importance scores, which discriminated between them. Trauer and Mackinnon (2001) used unpublished data collected with the ComQol to show that weighted total quality of life scores correlated .97 with the mean of the satisfaction scores, but only .29 with the mean of the importance scores.

The alternative weighting scheme proposed by Hsieh (in press) as a way of rendering scores easier to interpret (see "Multiplicative scores may be difficult to interpret", above) may solve the problem of identical importance-by-satisfaction products, but appears to do nothing to improve the correlation weighted scores with an overall measure of self esteem. In a study of 100 adults, Hsieh (in press) found that a simple sum of domain satisfaction scores correlated .39 with a single-item global measure of life satisfaction, whereas scores calculated from his alternative weighting scheme correlated .38, and a weighted score based on the standard multiplicative model correlated .40.

Finally, Mastekaasa (1984) found no support for the use of multiplicative scores in quality of life measurement, despite his use of various techniques to circumvent bias and issues related to levels of measurement.

*Self Esteem*

Research on importance weighting in the area of self-esteem has produced slightly more mixed results than the research on quality of life. Beginning with Rosenberg (1979), several
authors have found that domain importance may enhance the measurement of self esteem. Using data from 2,300 American adults, Rosenberg (1979) established that adults with higher incomes also had higher self esteem, but that the strength of this relationship depended on the importance attached to money. Similar results were found for social status, leading Rosenberg to posit a theory of "psychological centrality" (p. 144), or essentially the idea that the more important a domain to an individual, the greater its impact.

Kaplan (1980) hypothesised that negative self attitudes are caused by, among other things, the perception that one does not possess valued behaviours. To test this, he had several thousand American high school students indicate which of 12 domains or patterns of behaviour they believed accurately described them, as well as how important they felt each of these domains or behaviours to be. At the same time he administered a measure of self-derogation, which was re-administered twice more at 1 year intervals. Results showed that after 1 year, those students who had stated that an item was both self descriptive and important scored significantly lower on the measure of self-derogation than those who felt that an item was important but not descriptive of them. This relationship was found for all 12 domains. At the same time, five domains were related to lower self derogation scores when they were rated as self descriptive but not important, and some domains, if rated as self descriptive, appeared to have a positive impact on self-concept regardless of importance. This suggests that the impact of importance was irregular across domains.

One study looked at the relationship between self esteem, perceived competence in sports, and the importance of football ability for 100 male football players between the ages of 11 and 14 (Ebbeck & Stuart, 1993). Multiple regression analysis showed that two items
measuring the importance of being a good football player predicted 4% of overall self esteem. However, perceived competence in sports contributed far more (44%) to this prediction.

The discrepancy model has been used in the measurement of self esteem, notably by Harter (e.g., 1986, 1990) and Marsh (e.g., 1986). In support of the model, Harter (1986) cited evidence from a study of 90 fifth and sixth grade students whose discrepancy scores across five domains correlated -.76 with overall self worth. In another study with 90 students in grades five to seven, divided into three groups representing children with high, medium and low self esteem (Harter, 1986), the discrepancy score for the high self esteem group was -.27, while the medium self esteem group had a discrepancy score of -.62. The group with low self esteem had the largest discrepancy score: -1.20. Harter (1986) states that "there are converging findings attesting to the role of importance, and the overall pattern suggests that this type of discrepancy model accounts for certain processes involved in the self-worth judgements of children" (p. 156).

Others are less convinced of the effectiveness of importance ratings. Marsh and Hattie (1996) argued that Harter, in her various studies, never specified whether her discrepancy scores were any more effective than summed domain self concept scores for predicting global self esteem. Marsh (1986, 1994) and Marsh and Sonstroem (1995) tested both the discrepancy and multiplicative models in several studies and found little support for either model. In a study of self esteem with 930 youth and adults, Marsh (1986) looked at correlations between the participants' discrepancy scores and two types of domain competence measures. One measured each domain with a single item, while the other used multiple items to create a domain scale. The total discrepancy score based on the single-item measures correlated .33 with overall self esteem, while the discrepancy score based on the scale measures correlated .44 with overall self esteem.

5 Recall that with discrepancy scores, higher discrepancies are assumed to be indicative of lower self esteem.
The multiplicative model performed slightly better, with a correlation of .48 between the single-item self concept measure and overall self esteem and a correlation of .51 between the scale self concept measure and overall self esteem. However, neither model incorporating importance did as well as the unweighted model. The unweighted domain score based on the single-item domain self concept measures correlated .59 with overall self esteem, while the correlation coefficient for the unweighted scale score with overall self esteem was .67. Marsh then tried standardising the domain competence scores and calculating ipsatised importance scores. After standardisation, the correlation of the unweighted scores with overall self esteem rose to .69, while the standardised and ipsative weighted scores correlated .71 with overall self esteem. Although the weighted scores performed slightly better in this last case, the correlation with overall self esteem obtained was still very close to that obtained by using unweighted scores. Marsh also pointed out that the domain importance ratings did not correlate highly with overall self esteem (r=.13 for the sum of all importance ratings, significance not reported) compared to the satisfaction ratings (r=.59 for the sum of the single item measures, r=.67 for the sum of the scale measures).

Marsh has reported similar findings from his other research. In a study of 395 high school students, Marsh (1994) found that, while the mean of unweighted domain self concept scores correlated .79 with global physical self concept and .76 with overall self esteem, the mean untransformed scores weighted through the multiplicative model correlated only .67 and .61, respectively. Using standardised competence scores and ipsatised importance scores once again increased the correlation coefficients for the weighted scores, but only to the level of unweighted scores (r=.79 for global physical self concept and r=.76 for overall self esteem). Discrepancy scores correlated .70 with both global physical self concept and overall self esteem. Overall, the
variance explained by the importance ratings was 0.9% whereas the interaction of importance and self concept contributed 0.8%; both values were not significant. Marsh argued that, once again, even the most successful method of incorporating individual importance ratings did not prove more effective than simple summed domain self concept ratings. Marsh and Sonstroem (1995) also found little support for either the multiplicative or the discrepancy model in a study of 216 adult dancers. However, that research included a measure of exercise activity as a criterion variable along with the more familiar global measures. Although a simple additive model which ignored importance ratings was as highly correlated as various weighted model with measures of overall self esteem and global physical self esteem, the authors found some evidence that importance ratings contributed to the prediction of exercise activity. They tentatively hypothesised that people may be more likely to engage in a behaviour if they think it is important, and pointed out the possible usefulness of "appropriate importance ratings as additional predictors of external behavior" (p. 103). They also noted that much more research is needed in this area.

Other researchers have failed to find much support for the use of importance weighting in the measurement of overall self esteem. Farmer et al. (2001), conducted a study with 375 college students and found that while both domain self-ratings and domain importance accounted for significant variance in self esteem, the effect size of the latter was quite small. Most of the variance (53%) in global self esteem scores could be attributed to domain self concept ratings, whereas the interaction of self concept and importance contributed only an additional 0.7%. Hoge and McCarthy (1984), meanwhile, found that weighting (using individual importance ratings) was less effective than using unweighted scores, even when a number of transformations were attempted with the aim of strengthening the weighted model.
Job Satisfaction

It is perhaps in the area of research on job satisfaction that using importance scores has produced the most mixed results.

Ewen (1967) compared weighted and unweighted job domain scores for three samples of workers, assessed importance using both ratings and rankings. Results showed that, for that sample, the scores weighted by importance ratings correlated about as highly as unweighted scores with the global measures. Importance ranking was found to be considerably less effective than rating. Staples and Higgins’ (1998) study of 3,804 workers determined that, while a simple additive model explained 55% of the variance in overall job satisfaction scores, a multiplicative model accounted for significantly less at 47%. Nevertheless, the authors note that the $R^2$ resulting from the two methods is close enough that it cannot be said that the multiplicative model performed any worse than the additive one. Another study Smith Mikes and Hulin (1968) included voluntary job termination as a criterion, in addition to a measure of overall job satisfaction. In this case, importance did not improve the prediction of external behaviour; scores based on domain satisfaction alone correlated -.46 with job termination, but those weighted by importance correlated only -.33. Quinn and Mangione (1973) also used voluntary job termination as a criterion in their study of 1533 workers. Additional criteria included job related tension, mental health, and overall job satisfaction. Unweighted domain satisfaction scores correlated .46 with overall satisfaction, .37 with job related tension, .29 with voluntary job termination, and .20 with negative mental health. In contrast, weighting domains by importance resulted in a correlation of .34 with overall satisfaction, .28 with job related tension, .23 with voluntary job termination and .10 with negative mental health.\(^6\)

\(^6\) Significance and signs are not reported for this study and the authors simply state that "all significant correlations were in the predicted direction" (Quinn & Mangione, 1973, p. 11).
While Rice et al. (1991) agreed that domain importance did not improve the correlation of domain scores with overall job satisfaction, they suggested, based on their study conducted with 97 employed college students, that importance did affect the relationship between facet (domain) satisfaction and 'facet description', or "affect-free perceptions about the experiences associated with individual job facets" (p. 31). The authors hypothesised that facet descriptions and facet importance would both contribute to facet satisfaction. Analysis showed that for nine of the 12 facets investigated, the interaction of facet importance with facet description produced a significant increase in $R^2$, lending support to the hypothesis. While facet descriptions are perhaps qualitatively different from the domain satisfaction or self concept ratings usually seen in domain importance research, Rice et al. did point the way to some intriguing ways in which weighted scores could be used to investigate the impact of importance at the domain level.

A small number of studies have reported more favourable findings regarding the effectiveness of importance weighting in measuring job satisfaction. Some of these findings were identified by Vroom (1964), who discussed both the multiplicative and discrepancy models in relation to job satisfaction. Kalleberg (1977) also suggested that importance scores influence overall job satisfaction, specifically when the effect of positive correlations between domain importance and job rewards (i.e., subjective assessments of the extent to which certain positive aspects of a job are present) is controlled. He found that job rewards had a large positive effect on overall satisfaction whereas the importance of these rewards has a smaller, but nevertheless significant and independent, negative effect. Finally, Mastekaasa (1984) compared weighted scores and unweighted scores using data on job satisfaction. A multiplicative model
incorporating domain importance was more highly related to overall job satisfaction than scores that did not incorporate importance, particularly when ipsative scores were used.

Although the use of importance weights been explored fairly extensively in research on quality of life, self esteem and job satisfaction, the results of this research have been so mixed that it is not yet possible to arrive at a definitive recommendation whether to weight or not to weight. While some authors have reported finding support for the use of importance scores, evidence from other research suggests that the effectiveness of weighting is doubtful. Marsh (1986) and Trauer and Mackinnon (2001) have argued that the rule of parsimony requires that if no significant benefit can be obtained from using importance ratings, the simplest effective method (usually an unweighted, additive model) should be used instead. Yet even these authors, who are generally unsupportive of weighted scores, are reluctant to abandon the concept entirely, and have recommended further research. Trauer and Mackinnon (2001), despite devoting an entire article to describing their concerns with importance weighting, concluded that the problem lies with the use of multiplicative composites, not importance itself, stating that "we are not implying that all the domains are of equal importance to the individual nor that importance (or value) is not a valid and useful focus of study" (p. 584). Marsh (1986), a critic of both the multiplicative and discrepancy models, said that, although there is little statistical support for their use, importance ratings are too intuitively appealing to ignore and should be studied further. The idea of importance is clearly a difficult one of which to let go.
Complete References


(Eds.), *International perspectives on individual differences: Self perception* (Vol. 2).

Westport, CT: Ablex Publishing.


Hsieh, C.-M. (in press). To weight or not to weight: The role of domain importance in quality of life measurement. *Social Indicators Research, 00*, 1-12.


