VARIATION AND VALIDITY OF BODY DISSATISFACTION VISUAL ANALOGUE
SCALES

by

Lilach Dahoah Halevi

Specialized Honours Bachelor of Science, York University, June 2017

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Arts

in the program of

Psychology

Toronto, Ontario, Canada, 2019

© Lilach Dahoah Halevi, 2019

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

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

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

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

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

Variation and Validity of Body Dissatisfaction Visual Analogue Scales

Master of Arts, 2019

Lilach Dahoah Halevi

Psychology

Ryerson University

Abstract

When measuring state body dissatisfaction, it is common for researchers to use idiosyncratic versions of the Body Dissatisfaction – Visual Analogue Scales (BD-VASs). However, the BD-VAS variants vary in their scale and psychometric properties and impede comparability of results across studies. This thesis aims to review all available BD-VAS variants and their psychometric properties (Study 1), and empirically study the effects of scale modifications on BD-VASs' validity and reliability (Study 2). Study 1 revealed a total of 61 BD-VAS variants. Many of these variants were modified for a single study (77%) and had no supporting reliability and validity evidence (39%). In Study 2, 413 female undergraduate students completed one of six BD-VAS versions varying in extremity and scale polarity. Contrary to predictions, the BD-VAS variants had comparable score distributions, reliability, and validity evidence. While there is unnecessary variability across BD-VASs, such modifications may not greatly impact responses or study results.

Acknowledgements

I would first like to thank my advisor, Dr. Stephen Want. This project would not have been possible without your guidance and support these past two years. Thank you for teaching me to think more critically and inspiring me to be a better researcher. I would also like to thank the members of my examining committee, Dr. Alyssa Counsell and Dr. Eric Hehman, for their contributions in this project. Both of you have played pivotal roles during my research training and I am very grateful to have had you on my committee. Thank you for all your insightful feedback and encouragement throughout my academic career.

Finally, I would like to express my gratitude to my parents and friends for their unconditional and endless support. Thank you for always believing in me – this accomplishment would not have been possible without you!

Dedication

This thesis is dedicated to my late grandfather, Yehoshua Solomon. You have taught me the value of hard work and education and without you I wouldn't be where I am today. Although you will not be there at my graduation, your sacrifice, love, and support will never be forgotten.

Table of Contents

Abstract.....	iii
Acknowledgements.....	iv
Dedication.....	v
Table of Contents.....	vi
List of Tables	ix
List of Figures	x
List of Appendices	xi
Introduction.....	1
Body Dissatisfaction: Theory and Measurement	2
What are Visual Analogue Scales?	3
Variation in VASs for Body Dissatisfaction	5
Variability in the content and quantity of BD-VAS items.....	6
Variability in the extremity and polarity of BD-VAS anchors	7
Variability in the analysis of BD-VASs.....	8
The Unknown Validity of VASs for Body Dissatisfaction.....	9
The Present Studies	10
Study 1	11
Method	11
Literature search and inclusion criteria.....	11
Coding articles	12
Results	13
Variation in the BD-VASs across studies.....	13
Demographic information.....	14
Scale properties	15

Number of items	15
Item content	15
Anchors	16
Scale polarity	16
Item analysis	17
Reliability and validity evidence	17
Most commonly used and validated BD-VAS.....	18
Discussion	19
Multiple definitions of body dissatisfaction	20
Numerous modifications to the scale properties of the BD-VAS.....	21
Limited reliability and validity	23
Study 2	27
Method	28
Participants.....	28
Materials and procedure.....	29
Body Dissatisfaction Visual Analogue Scale (Time 1)	29
Filler task	29
Body Dissatisfaction Visual Analogue Scale (Time 2)	30
Body Areas Satisfaction Scale (BASS)	30
Rosenberg's Self-Esteem Scale (RSES)	30
Sociocultural Attitudes Towards Appearance Questionnaire (SATAQ-4).....	30
Physical Appearance Comparison Scale-Revised (PACS-R).....	31
Body Mass Index (BMI)	31
Demographics questionnaire.....	31
Statistical Analyses	32

Comparing score distributions	32
Reliability.....	33
Convergent validity.....	34
Results	34
Comparing response distributions.....	34
Mixed ANOVAs	35
Kolmogorov-Smirnov tests	36
Reliability.....	37
Internal consistency	37
Test-retest reliability	37
Convergent validity.....	38
Discussion	39
Similar score distributions across BD-VAS variants.....	39
Consistent psychometric properties across BD-VAS variants.....	41
Conclusion and Future Directions	44
Appendices.....	60
Appendix A: Summary of All the Available Body Dissatisfaction Visual Analogue Scale Variants and Their Demographic, Scale, and Psychometric Properties.....	60
Appendix B: The Six Body Dissatisfaction Visual Analogue Scales Assigned to Participants	73
References.....	76

List of Tables

Table 1.	Summary and Frequency of the Items Used in the Body Dissatisfaction Visual Analogue Scales	46
Table 2.	Summary of the Variability in Scale Properties of the Body Dissatisfaction Visual Analogue Scale	48
Table 3.	Demographic information including sample size, mean age, body mass index (BMI), body dissatisfaction (BASS), and ethnicity per condition.	50
Table 4.	Distribution properties and reliability of BD-VAS scores.	52
Table 5.	Intercorrelations between all the measures used to provide convergent validity evidence for the BD-VAS.	54
Table 6.	Intercorrelations between the six versions of the BD-VAS and BASS, RSES, SATAQ-Thin, SATAQ subscales, PACS, and BMI.	55

List of Figures

Figure 1.	An example of a body dissatisfaction VAS	56
Figure 2.	A frequency distribution of the number of BD-VASs variants that have between 1 and 7 items and the number of studies in which they have been used.	57
Figure 3.	The score distributions of the weight, appearance, and the composite body dissatisfaction BD-VAS scores per BD-VAS variant. The plotted vertical lines represent the median values of the distributions.	58
Figure 4.	A Bland-Altman plot that represents the test-retest agreement of BD-VAS scores at time 1 and time 2. The mean difference between the BD-VAS scores is plotted in red and the upper and lower 95% limits of agreement are plotted in blue. The dashed lines represent 95% confidence intervals of those metrics.	59

List of Appendices

Appendix A:	Summary of All the Available Body Dissatisfaction Visual Analogue Scale Variants and Their Demographic, Scale, and Psychometric Properties	59
Appendix B:	The Six Body Dissatisfaction Visual Analogue Scales Assigned to Participants	72

Introduction

Body dissatisfaction is defined as a negative, subjective evaluation of one's own body (Cash, 1994). The experience of body dissatisfaction is common enough among North American women (Swami et al., 2010) as to have been called a form of "normative discontent" (Rodin, Silberstein, & Streigel-Moore, 1984) and is thought to be shaped by sociocultural factors (Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999). Indeed, trait levels of body dissatisfaction, as assessed by multi-item, well-validated scales such as Body Areas Satisfaction Scale from the Multidimensional Body-Self Relations Questionnaire (BASS; Brown, Cash, & Milulka, 1990; Cash, 2018) and the Eating Disorder Inventory – Body Dissatisfaction scale (EDI-BD; Garner, Olmstead, & Polivy, 1983), are correlated with peer and parental preoccupation with weight (Keery, van den Berg, & Thompson, 2004), as well as with exposure to media depicting idealized bodies (Grabe, Ward, & Hyde, 2008).

To attempt to demonstrate causal influences of these sociocultural factors, particularly so-called "thin ideal" media (Want, 2009), researchers have also examined women's responses on state measures of body dissatisfaction after short-term experimental exposure to them. Much of this research, following the influential example of Heinberg and Thompson (1995), uses Visual Analogue Scales (VASs) to measure body dissatisfaction as a state. VASs are visual representations of continua, for example from being very dissatisfied with one's body to being very satisfied. VASs can be reliable, valid, and responsive measures and have been demonstrated to be so for measuring pain and personality (Simms, Zelanzy, Williams, & Bernstein, 2019; Williamson & Hoggart, 2005; Kuhlmann, Dantlgraber, & Reips, 2017). However, little attention has been paid to their psychometric properties for measuring body dissatisfaction. Across studies, Body Dissatisfaction – Visual Analogue Scales (BD-VASs) have greatly varied in terms of their content, number of items, scale anchors, scale polarity, and analysis, with each set of researchers seemingly using their own idiosyncratic versions. This diversity of BD-VASs makes it hard to compare across studies to, for example, examine the relative influence of different sociocultural factors on body dissatisfaction. Moreover, researchers using BD-VASs have often relied on a citation to Heinberg and Thompson (1995) for evidence of their reliability and validity, despite the particular form of the BD-VASs the researchers used bearing little relation to Heinberg and Thompson's original.

Therefore, the goal of the present thesis was to: (1) review and summarize all the available versions of the body dissatisfaction VAS (BD-VASs) that have been used in the literature and what is currently known about their psychometric properties, and; (2) empirically study the implications of potential scale modifications for the validity and reliability of the BD-VAS. The hope is that, in doing so, the thesis will be able to provide some guidance as to how to create BD-VASs to reliably and validly measure state body dissatisfaction.

Body Dissatisfaction: Theory and Measurement

Body dissatisfaction, particularly related to one's appearance and weight, is prevalent among women (Frederick, Sandhu, Morse, & Swami, 2016; Fallon, Harris, & Johnson, 2014). When asked to pick from nine line drawings of women's bodies, ranging from very slender to very heavy, the largest discrepancy between the one chosen to represent their current body and the one representing their ideal body (i.e., the most body dissatisfaction) was primarily found in North and South American women (Swami et al., 2010). Furthermore, as measured with Likert-type scales (1 – very dissatisfied to 5 – very satisfied) women's dissatisfaction with various body parts appears to be relatively stable throughout the lifespan (Tiggemann, 2004). High levels of body dissatisfaction can have deleterious effects, as they have been shown to prospectively predict low self-esteem (Paxton, Neumark-Sztainer, Hannan, & Eisenberg, 2006), are associated with lower sexual satisfaction (Pujols, Meston, & Seal, 2010) and are believed to be a risk factor for depression and eating pathology (Ohring, Graber, & Brooks-Gunn, 2002; Stice & Shaw, 2002).

Women's satisfaction with their bodies is thought to be shaped by several sociocultural influences. According to the Tripartite Influence model (Thompson, Heinberg, Altabe, & Tantleff-Dunn, 1999; van den Berg, Thompson, Obremski-Brandon, & Covert, 2002), women are under strong societal pressures – from peers, family, and the media – to obtain a slim and feminine figure (i.e., thin ideal). Accordingly, researchers find that exposure to thin-ideal media (Grabe, Ward, & Hyde, 2008), parent and peer teasing, preoccupation with, and talk about, weight and shape (Keery, van den Berg, & Thompson, 2004), as well as self-reported pressure from all three sources (Schaefer, Harriger, Heinberg, Soderberg, & Thompson, 2016), are all positively correlated with body dissatisfaction. A mediator of the effect of exposure to pressure from these three sources is internalization of the thin ideal. Internalizing – that is, coming to

accept societal standards of appearance as personal standards – but failing to achieve them, negatively impacts women’s body satisfaction. Researchers have thus found that high levels of internalization of the thin ideal are also positively correlated with trait levels of body dissatisfaction (Schaefer et al., 2016).

The research correlating high levels of body dissatisfaction with what are thought to be their causes and consequences has measured body dissatisfaction as a *trait*. There are numerous, validated measures of trait body satisfaction, for instance the BASS (Brown, Cash, & Milulka, 1990; Cash, 2018) or EDI-BD (Garner, Olmstead, & Polivy, 1983) that have been used to examine these correlations. However, to establish the causal effect of any of the proposed forms of influence on women’s body dissatisfaction, researchers have turned to studying the effects of short-term experimental exposure to these influences (Want, 2009). To do so, researchers are required to use measures that can detect short-term changes; that is, to use measures of *state* body satisfaction. However, to study body satisfaction *well*, it is crucial that researchers precisely measure body satisfaction with state measures that have supporting reliability and validity evidence.

There have been a few attempts to validate scales that measure state body dissatisfaction. The Body Image States Scale (BISS; Cash, Fleming, Alindogan, Steadman, & Whitehead, 2002) is one example of such scale. The BISS is a 6-item scale that evaluates how respondents’ feel about their bodies at a given point in time. But when studying the effects of the aforementioned social influences on body satisfaction, particularly media exposure, it is increasingly common for researchers to use Body Dissatisfaction – Visual Analogue Scales (BD-VASs).

What are Visual Analogue Scales?

VASs consist of straight lines (typically horizontal) with an anchor descriptor at each end, defining the range of the measured construct. Participants are asked to draw a shorter line (if administered on paper) or click to place a marker (if computer administered) bisecting the VAS at the point between the two anchors which represents how they feel. An example of a VAS for body dissatisfaction is shown in Figure 1.

VASs are commonly used across various fields of research, investigating pain (Carlsson, 1983; Williamson & Hoggart, 2005), quality of life (de Boer, van Lanschot, Stalmeier, van

Sandick, Hulscher, de Haes, & Sprangers, 2004), anxiety (van Duinen, Rickelt, & Griez, 2008; Facco, Zanette, Favero, Bacci, Sivoletta, Cavallin, & Manani, 2011), and mood (Monk, 1989; Benedetto & Sheehan, 2014; Steiner, Streiner, & Pham, 2005; Ahearn, 1997). The earliest instance of a VAS for measuring body (dis)satisfaction was in a study by Mock (1993). The original scale was composed of a single item that assessed the state body satisfaction of female cancer patients. The use of the BD-VAS was popularized by Heinberg & Thompson (1995) in a study of the effect of thin-ideal TV commercials on young women's body dissatisfaction. Heinberg and Thompson's BD-VAS was administered to a female, undergraduate sample and measured dissatisfaction with overall appearance and weight. In their study, the BD-VAS demonstrated good convergent validity, with overall appearance ($r = .76$) and weight ($r = .66$) dissatisfaction correlating highly with other measures of body dissatisfaction (i.e., EDI-BD).

In the nearly two-and-a-half decades following Heinberg and Thompson (1995), BD-VASs have become a highly prevalent means of evaluating people's state body dissatisfaction. Much of their appeal to researchers stems from their brevity (often just a single VAS is used) and sensitivity to even small changes in responses (Mayer, 1978; DeVellis, 2003). For those reasons, VASs are especially handy in pre- to post-test research designs. The researchers can administer a quick and short measure that can detect changes in the desired construct after an experimental manipulation and the respondents are unlikely to be able to simply remember their previous responses (which is a concern with Likert-type scales; DeVellis, 2003). Additionally, since many VAS measures have just a single item, the dependent variable can be hidden among other VASs, thereby reducing the amount of emphasis that is placed on the desired construct and consequently reducing experimental demand (Orne, 1962; Want, 2014).

VASs can be reliable and valid measures. Studies have shown that VASs measuring pain or personality are better or comparable to Likert scales in terms of reliability, validity, and responsiveness (i.e., the ability to detect small changes) (Simms et al., 2019; Williamson & Hoggart, 2005; Kuhlmann et al., 2017). For example, Kuhlmann et al. (2017) found that excitement seeking, conscientiousness, and narcissism scales presented in either a Likert scale or VAS format produced nearly identical internal consistency and correlations between the scales. Simms and colleagues (2019) further found that increasing the number of available responses seems to increase the psychometric precision of the measure, although only up to a certain point

(i.e., 6-point Likert scale). These findings show that VASs can demonstrate better psychometric properties than the 5-point Likert scales commonly used in psychological research. At the very least, VASs have been shown to perform just as well as Likert scales with 6 discrete categories or more.

VASs can additionally eliminate the bias in responses associated with scale coarseness. When measuring continuous variables using a scale with discrete categories (e.g., 5-point Likert scale), responses are systematically biased upwards or downwards to fit into the one of the available categories. By pooling responses into categories, we lose valuable information that could help identify small differences in responses (Russell & Bobko, 1992; Russell, Pinto, & Bobko, 1991). Especially when VASs are used in pre- to post-test experimental designs, where evaluations of momentary attitudes or affect are not expected to change drastically within that given timeframe, VASs could help discern any real differences that would have been undetectable with Likert scales. So, VASs can be psychometrically sound measures.

However, VASs specifically for measuring body dissatisfaction are potentially problematic because, while there have been continual efforts to validate longer measures of trait (Brown et al., 1990; Cash, 2018; Garner, Olmstead, & Polivy, 1983) and state (Cash et al., 2002) body dissatisfaction and encourage proper measurement practices in the field of body image (Thompson, 2004), there has been little attention given to the psychometric performance of body dissatisfaction VASs. Researchers using BD-VASs often cite Heinberg and Thompson (1995) and neither cite nor provide any other evidence of reliability and validity, as though Heinberg and Thompson's study established the reliability and validity of all forms of BD-VASs across all samples. This is despite the fact that the construction and usage of BD-VASs has varied greatly from study to study, often deviating from Heinberg and Thompson's use of them in ways that almost certainly affect their psychometric properties.

Variation in VASs for Body Dissatisfaction

Heinberg and Thompson (1995) used two VAS items to measure their participants' weight and overall appearance dissatisfaction, ranging from "none" to "very much" and averaged responses to the two items for analysis. Across studies measuring body dissatisfaction through VASs since Heinberg and Thompson's, there is a large observed variability. For instance, one study might ask participants a single question "How do you feel about your body right now?"

with the anchors on the ends of the VAS labelled “extremely dissatisfied” and “extremely satisfied” (as in Figure 1). Another study might ask participants three questions – “Right now, how satisfied do you feel with your (1) weight/size, (2) body shape and (3) overall appearance?” with the anchors on the three VASs labelled “not at all” and “very much”. Researchers in this second study might then analyse each BD-VAS item separately or analyse an average score. In both studies, the researchers will claim to be measuring the same construct, namely body dissatisfaction. However, these researchers may be making a mistake related to the jingle fallacy – the erroneous assumption that two measures with the same name capture the same construct (Thorndike, 1904); although the measures used in the two studies are believed to assess body dissatisfaction, they may actually be evaluating two distinct constructs. The numerous conceptualizations of body dissatisfaction (represented by the content and quantity of the scale items) and the variation in anchor descriptors (which vary in extremity but also polarity) and analysis (individual vs. averaged VASs items) make it difficult to compare results across studies.

Variability in the content and quantity of BD-VAS items. Variability in the items of the BD-VASs (e.g., “your body”, “your weight/size”, “your appearance”) will clearly reduce comparability between studies. As just one example, some studies using BD-VASs include an item assessing dissatisfaction specifically with facial appearance (e.g., Want & Saiphoo, 2017) while others focus solely on the body. The number of items in each instantiation of the BD-VASs is another source of variability. Longer measures are recommended in practice, as they can better capture a greater amount of unique information associated with the construct researchers are trying to measure (Soto & John, 2019). For instance, researchers can more accurately measure body dissatisfaction using multiple items (in comparison to using a single item) by evaluating satisfaction with numerous body parts (i.e., chest, stomach, thighs, etc.) or assessing different dimensions of body satisfaction (i.e., appearance, weight, body). Scales with a greater number of items also tend to have greater internal consistency than shorter scales; this in turn, will allow researchers to assess body satisfaction more reliably and estimate effect sizes more accurately (Weidman, Steckler, & Tracy, 2017). However, given that one of the major advantages of VASs is their brevity, researchers often use BD-VASs with very few items, propagating the use of BD-VAS with limited or non-existent (i.e., 1-item BD-VASs) reliability.

Variability in the extremity and polarity of BD-VAS anchors. The variability of BD-VAS anchor descriptors across studies is also problematic. The way in which researchers choose to word their anchors is crucial, as different words can convey different meanings, and therefore yield different distributions and mean scores. For instance, anchors indicating extremity (e.g., extremely dis/satisfied) will result in end aversion; people will tend to place their answer away from the extreme ends and towards the middle (Zusman, 1986; Torrance, Feeny, & Furlong, 2001). Furthermore, ill-defined anchors – for example, ones that include jargon, slang, or ambiguous terms – can be construed differently across individuals, and thus won’t accurately measure the desired construct (Marsden & Wright, 2010; Bartoshuk, Duffy, Fast, Green, Prutkin, & Snyder, 2002). For example, studies that evaluated VASs measuring pain intensity revealed that “worst imaginable pain” is an ambiguous anchor (Torrance, Feeny, & Furlong, 2001) and is the least endorsed by patients in comparison to other anchors, such as “worst pain”, “worst pain bearable”, and “worst pain experienced” (Yokobe, Kitahara, Matsushima, & Uezono, 2014). In fact, a study by Yokobe and colleagues (2014) concluded that “worst pain experienced” is the best anchor for the pain intensity VAS, as it conveyed a consistent meaning across participants and allowed them to best express the level of pain they were experiencing.

Scale polarity can pose additional problems for the comparison between studies using BD-VASs. Scales can be either unipolar or bipolar; their range can encompass either the space between the complete absence and presence of a singular construct or the range between two opposing constructs. For example, items in a BD-VAS could solely assess body dissatisfaction with the anchor descriptors “not at all satisfied” to “very satisfied” (i.e., unipolar; e.g., Durkin & Paxton, 2002) or address both body dissatisfaction and satisfaction with the anchor descriptors “very dissatisfied” and “very satisfied” (i.e., bipolar; e.g., Cho & Lee, 2013). Although some argue that unipolar scales are more comprehensible than bipolar scales (particularly for mood; Monk, 1989), others believe that unipolar scales will not be able to properly capture the range of a bipolar construct (DeCastellarnau, 2018). A bipolar scale may also make the construct clearer by providing a clear, neutral midpoint (Schaeffer & Presser, 2003). For instance, one thing that is unclear is what “no (body) dissatisfaction” means to respondents. Respondents could interpret that in one of two ways: as feeling neutral (neither particularly satisfied nor dissatisfied with their bodies) or as feeling completely satisfied. Unipolar body dissatisfaction VASs may add a source of ambiguity and error that is avoided in bipolar scales. Bipolar scales, however, assume

that the two polar ends of the scale are opposites. Although it is unclear whether body satisfaction is a bipolar construct, evidence from emotion and life satisfaction literatures provides some insight.

For momentary affect measures, affect can be interpreted on a bipolar plane ranging from negative to positive emotions (e.g., the circumplex model; Russell, 1980), or as two separate negative and positive unipolar planes (e.g., the bivariate evaluative model; Cacioppo & Berntson, 1994). Under the circumplex model, affect states – including satisfaction and dissatisfaction – can be understood on a two-dimensional space formed by two axes: valence (pleasantness-unpleasantness) and arousal (activation-deactivation). Feelings of satisfaction (0 degrees) are thought to fall directly opposite of feelings of dissatisfaction (180 degrees), leading to the idea that they are polar opposites as opposed to two separate constructs. Indeed, across studies satisfaction consistently falls directly opposite from dissatisfaction (for a review, see Remington, Fabrigar, and Visser, 2000). Research on life satisfaction similarly shows that unipolar measures of life satisfaction and dissatisfaction are strongly and negatively related, lending support to the idea that satisfaction and dissatisfaction are opposing affective states (Davern & Cummins, 2006; Mazaheri & Theuns, 2009).

Variability in the analysis of BD-VASs. Lastly, the analysis of BD-VASs items – whether they are analysed individually or averaged to produce a global score of body dissatisfaction – can impact the way we interpret the results of different studies. On the one hand, if the items in a body dissatisfaction VAS are analysed individually, they might not capture all the dimensions of body dissatisfaction and thus will not accurately measure this construct. On the other hand, averaging across different items assumes that the different items represent different dimensions contributing to the same construct, which may not be the case. For example, Peters, Rhodes, and Simmons (2007) found that evaluations of faces and bodies independently contribute to the rating of a person's attractiveness. This suggests that the appearance of one's body and face can affect perceptions of oneself differently, and potentially impact their satisfaction in different ways. Thus, it may not be appropriate to average VAS items measuring body and face satisfaction. Having no consistent practice across studies for analysing VAS items individually or on average also increases the analytic flexibility available to researchers and the

potential for *p*-hacking, resulting in false positive results (Simmons, Nelson, & Simonsohn, 2011).

In summary, the BD-VASs may vary across different scale properties, including the scale items (content and quantity), anchors (extremity and polarity), and item analysis. These variations can potentially impact respondents' scores and the scales' psychometric properties. Although there is no clear evidence that variation in these scale qualities will definitively impact the interpretation of the construct and scores, they still introduce potential error in our scales that we should try to avoid.

The Unknown Validity of VASs for Body Dissatisfaction

This observed variability in BD-VASs in itself is a problem, since it produces unnecessary sources of variability between studies in this field. However, what makes the problem even worse is the lack of validity evidence provided to support the use of these scale variants. Lack of validity evidence seems to be pervasive across many different fields, including health, education, clinical, and social/personality psychology; researchers have frequently failed to provide sufficient psychometric information regarding the scales they use in their studies (Barry, Chaney, Piazza-Gardner, & Chavarria, 2014; Flake, Pek, & Hehman, 2017; Qualls & Moss, 1996). Barry and colleagues (2014) revealed that research in the field of health and education rarely provides validity evidence for their scales. Most strikingly, the use of newly created and adapted scales is rarely accompanied by validity evidence, with only 37% of the studies reporting validity evidence for the former and 32% of studies for the latter (Barry et al. 2014). Flake, Pek, and Hehman (2017) noted a similar pattern in the field of social and personality psychology; the majority of researchers did not provide structural validity evidence for their scales. A significant portion of researchers relied instead on a reference to a previous validation or a reliability coefficient as their only source of validity. Moreover, 19% of scales that were newly developed and had no previous validation were unaccompanied by psychometric information. The underreporting of reliability and validity evidence hinders the interpretability and comparability of study results, as we do not know how these newly developed or modified scales perform in comparison to previously validated measures.

Even in the case where researchers provide reliability and validity evidence for the used scales, there is an overemphasis on certain types of evidence. Most of the validity evidence

provided for scales comes from the *external* phase of research; this type of evidence involves the assessment of the scale in relation to other theoretically related measures, constructs, or groups (e.g., convergent validity, known groups validity) (Loevinger, 1957). However, validity evidence can be obtained at the substantive and structural phases of research as well (Loevinger, 1957). The *substantive* phase entails the conceptualization of the construct in question and the dimensions it may involve (e.g., item development through expert review or interviews), while the *structural* phase involves the psychometric properties of scales used to measure that construct (e.g., internal consistency and factor analysis). Reported structural validity evidence has primarily focused on the internal consistency of the scale, and more specifically, Cronbach alpha values (Flake, Pek, & Hehman, 2017). Overreliance on these types of validity evidence can mask the poor performance of the scales on other forms of validity. A recent study by Hussey and Hughes (2018) brought this issue to light by exposing the hidden invalidity of commonly used psychological scales; while the scales Hussey and Hughes looked at showed good performance on validity metrics that are commonly reported in the literature (e.g., internal consistency and test-retest reliability), they failed ones that are not often examined, such as factor structure and measurement invariance.

Informal observation suggests that researchers using body dissatisfaction VASs generally create them on the fly to correspond with their specific research questions, and typically cite only a single original study as their validity evidence (with most citing Heinberg and Thompson, 1995). And like other scales used in psychology, the different versions of the BD-VASs may not have reliability or validity evidence to support their use. Thus, it is important to investigate the trends in reporting reliability and validity evidence for the BD-VAS to ensure that these scales in fact evaluate state body dissatisfaction with precision.

The Present Studies

Responding to the variability and unknown validity of BD-VASs, this thesis consists of two parts. In Study 1, I catalogue the extent of variability in BD-VASs. Furthermore, I summarize the existing validity and reliability evidence of all these different versions of this scale and determine which scales and items have been shown to have the most supporting validity and reliability evidence. In Study 2, I empirically study the effect of scale modifications to the validity and reliability of the BD-VAS and make recommendations for scales to be used by future studies.

Study 1

In Study 1, I provide a large-scale summary of the variations seen in BD-VASs across studies. As detailed below, each BD-VAS was coded for its structural properties, such as the number of items, the anchors, the polarity, and so forth. In addition, for each study the reliability and validity evidence given for the BD-VAS were recorded.

Method

Literature search and inclusion criteria. I reviewed articles published up to September 2017 that cited papers that previously used a BD-VAS in order to find the BD-VASs present in the literature. For example, articles measuring body dissatisfaction via a VAS commonly cite Heinberg and Thompson (1995), as it is one of the earlier instances of a BD-VAS. Articles that cited other early uses of VASs to measure body dissatisfaction such as Mock (1993), Ogden and Mundray (1996), Champion and Furnham (1999), and Durkin and Paxton (2002) were also be examined. Additional articles were found on online databases using the terms “visual analog* scale*” AND “body image” or using similar variations (i.e., “body dissatisf*”).

A total of 98 studies were obtained, 87 of which were eligible to be included. In order to be considered a variation of the BD-VAS, the VASs found in these studies had to meet the following criteria. First, articles were excluded if they used VASs to assess components of body image other than satisfaction ($n = 3$). Such constructs included body discomfort (Servian-Franco, Moreno-Dominguez, & Reyes del Paso, 2015), anxiety towards or monitoring of appearance (Monro & Huon, 2005), and how fat, attractive, or sexy participants felt (Ogden & Mundray, 1996). However, exceptions to this rule were made for BD-VASs that included an item that is unrelated to body (dis)satisfaction if it was included along with items related to body dissatisfaction and they were later combined to produce an average score of body dissatisfaction. For example, body dissatisfaction was sometimes represented by a composite score of BD-VAS items measuring participants’ satisfaction with their body type and shape, as well as how attractive and fat they felt (e.g., Fardouly et al., 2015). Second, scales that measured satisfaction with a niche area on the body were excluded ($n = 1$). For instance, the VAS used in Sharp and Tiggemann (2016) was excluded, as it measured women’s satisfaction with the appearance, size, and shape of their genitals. The scope of the scale was too specific for it to be comparable to other BD-VASs in the sample. Finally, studies that adapted the scale format of a BD-VAS to

have discrete categories (e.g., a 10-point scale; Kim & Damhorst, 2010) or include additional labels along the continuous line (e.g., a mid-scale label; Mask, Blanchard, & Baker, 2014) were also excluded ($n = 7$). For these reasons, a total of 11 studies were excluded from the following review.

Coding articles. I coded all 87 articles and an undergraduate research assistant was employed as a second coder. Each study was coded for the following information¹:

1. Sample size (N)
2. Mean age of participants (M_{age})
3. Percentage of Caucasian participants (%C)
4. Percentage of female participants (%F)
5. Country where the study took place
6. All the items presented in the scale
7. The way in which the items were analysed (i.e., items' scores were averaged or analysed individually)
8. The polarity of the scale (i.e., unipolar or bipolar)
9. The anchors of the scale
10. Reliability evidence (cited and calculated)
11. Validity evidence (cited and calculated)

When searching for this information, both coders primarily examined the methods and results sections of each study since those are the most common places for people to report that descriptive and psychometric information. As a result, any information that was reported in other sections of the article (e.g., introduction and discussion sections) was not included in the review.

The coded scale properties – the content and number of items, anchors, scale polarity and method of item analysis – were used to derive all available variants of the BD-VAS. For a scale to be considered a variant of the BD-VAS, it must have unique scale properties that differentiates

¹ In the coding process, the coders also coded for (1) the citation associated with the scale, (2) whether the scale matched the scale in the citation (i.e., whether or not the scale was modified from the version in that citation), and (3) any claims that the scale was modified. However, due to inconsistencies in how the BD-VAS in Heinberg & Thompson (1995) – the most cited scale in our review – was reported in the literature (e.g., Thompson, 1996), I excluded these variables from the descriptive analyses and discussion.

it from the other identified scales. Even a single difference in any of the number or wording of the items, scale anchors, scale polarity, and item analysis, was enough to consider two BD-VASs as separate variants. For example, two BD-VASs that contained the same number of identically-worded items, with the same scale polarity, would be considered different variants if only the scale anchors differed between them.

When coding reliability evidence, we focused on internal consistency (the extent to which the items in a measure correlate, and thus measure the same construct) and test-retest reliability (the consistency of participants' scores on the scale at two separate time points). Interrater reliability evidence was excluded from the review. While it was traditionally important to calculate interrater reliability of paper-administered VASs, the digitization of surveys has made it so that human coders are no longer needed to determine a person's VAS score. Additionally, interrater reliability evidence does not evaluate the scale's ability to yield consistent scores, but instead reflects the raters' ability to accurately assess participants' scores in a given research study.

Similarly, the coders focused on reported evidence of construct validity. As previously discussed, evidence of construct validity can be extracted from three different phases of research: the substantive, structural, and external phases (Loevinger, 1957). Evidence from all three phases was included in our review. Such evidence includes, but is not limited to, exploratory and confirmatory factor analyses, convergent and discriminant validity, and known groups validity.

To ensure the two coders coded articles in a similar and consistent manner, the second coder received in-depth training on how to identify and report information pertaining to all the variables of interest (especially in regard to reliability and validity evidence). A training set of three sample articles was developed to assess the second coder's mastery of the coding scheme. Once the second coder completed the training set, the two coders coded the rest of the articles independently. To ensure the veracity of the coded information, any disagreements between the coders were discussed and resolved.

Results

Variation in the BD-VASs across studies. The review revealed a great deal of variability in the use of BD-VASs. Of the 87 studies that used a BD-VAS, a total of 61 different

scale variants were identified. Strikingly, 47 of the 61 scale variants (roughly 77%) were used in only one study. The remaining 14 BD-VAS variants were not used by many studies: 9 scale variants were used by only two studies, 3 scales were used across three studies, 1 scale by five studies, and 1 by eight studies. All these scale variants and their corresponding studies are outlined in Appendix A. In the following results section, I further discuss the variability in: (1) study demographics, (2) scale properties, and (3) reliability and validity evidence of each scale variant. Additionally, I identify the most used and validated variants of the BD-VAS.

Demographic information. Across studies, there were differences in sample characteristics, including sample size, mean age, country, race and ethnicity, and gender. The average sample size was around 152 participants ($M = 152.51$, $SD = 113.32$, median = 116). Sample sizes ranged from as low as 22 participants, to as large as 595 participants. Most studies had relatively low sample sizes; approximately 70 percent of the studies had a sample size lower than the average sample size ($n = 61$). The average mean age of a sample was 21.80 years old ($SD = 7.67$). While there were some studies that administered the BD-VAS to young children (e.g., Anschutz et al., 2012) or older adults (e.g., Kraus, 1999), the majority reported data from undergraduate participant pools. Nearly all studies were conducted in Western countries, with the most prominent country being Australia ($n = 31$), followed by the United States ($n = 30$), United Kingdom ($n = 9$), Canada ($n = 7$), the Netherlands ($n = 5$), Finland ($n = 1$), Germany ($n = 1$), and South Korea ($n = 1$). Furthermore, most studies had predominantly Caucasian and all-female samples; 55.17% of the samples were at least 50% Caucasian ($n = 48$) and 79.31% of samples were composed of only female participants ($n = 69$).

Thirty-two studies did not report specific ethnic or racial information about their studies; some failed to report any information regarding the racial or ethnic groups the participants belonged to, while others reported having a predominantly Caucasian sample without any associated percentages. Because of this missing information, it is likely that the number of studies containing mostly Caucasian participants is higher than it appears from our current results. This means that study results and psychometric properties of the BD-VAS are mainly based on the responses from relatively small samples of White, female university students from Western, Educated, Industrialized, Rich, and Democratic (WEIRD; Henrich, Heine, & Norenzayan, 2010) societies.

Scale properties. As previously mentioned, there were 61 different variants of the BD-VAS, with each variant having distinct scale properties. Specifically, they differed in the number and content of the scale items, anchors, scale polarity, and item analysis.

Number of items. As expected, most researchers used a BD-VAS with only one or two items ($M = 2.13$, $SD = 1.26$). More specifically, 32.18% of the studies used BD-VAS variants containing only one item ($n = 28$) and 45.98% of the studies used variants containing two items ($n = 40$). Researchers rarely employed a scale with more than four items. At the level of the scale variants, similar trends were observed. One-item ($n = 21$) and two-items BD-VASs ($n = 24$) were more prevalent than longer BD-VASs, making up around 73% of all the scale variants. Figure 2 displays the number of BD-VAS variants that have between one and seven items and displays the number of studies in which each variant has been used. Within each column of Figure 2, each individual box represents a different variant of the BD-VASs with the size of that box representing the number of studies in which that variant has been used, arranged bottom-to-top from used in most to least number of studies. The total height of each column therefore represents the number of studies that have used BD-VASs with one to seven items.

Item content. All of the items across all of the variants of the BD-VAS measured one of the following six constructs: (1) appearance satisfaction, (2) body satisfaction, (3) weight satisfaction, (4) feelings of attractiveness, (5) face satisfaction, and (6) muscle tone satisfaction. All but one of these constructs measure *(dis)satisfaction* with an area on the body; feelings of attractiveness, while thought to be related to body satisfaction, do not directly measure satisfaction with one's body. The different operationalizations of these constructs and their prevalence are outlined in Table 1.

Body, appearance, and weight satisfaction were the most frequently measured constructs across the BD-VAS variants. It was common for researchers to measure global feelings about one's body with a single VAS item. Roughly half (46.43%) of the 21 one-item BD-VASs measured people's satisfaction with their *body* ($n = 13$). In comparison, only two scales (i.e., Mulgrew, Volcevski-Kostas, & Rendall, 2014; Agliata & Tantleff-Dunn, 2004) measured body satisfaction through multiple items assessing satisfaction with specific body parts, such as stomach, arms, shoulders, and chest. Appearance and weight satisfaction were also commonly measured, often in conjunction with each other (i.e., 7 BD-VAS variants used across 19 studies).

In fact, the two most commonly used BD-VASs consisted of two items measuring appearance and weight dissatisfaction. In addition, there were BD-VAS variants measuring appearance ($n = 9$) and weight satisfaction ($n = 1$) alone, or along with other items measuring body satisfaction, face satisfaction, and/or attractiveness ($n = 7$).

Anchors. Table 2 further outlines the variations in the anchors, scale polarity, and item analysis of the BD-VASs across studies and scale variants. Like the scale items, there were numerous ways in which researchers chose to phrase the anchors of their BD-VAS. A total of 27 unique pairs of anchors were found. Nearly half of these anchors (48.15%) were used in only one study ($n = 13$). The most common anchors were “None – Very much” ($n_{studies} = 19$; $n_{variants} = 7$), followed by “None – Extreme” ($n_{studies} = 9$; $n_{variants} = 6$), and “Not at all – Very much” ($n_{studies} = 9$; $n_{variants} = 5$). The first pair of anchors was derived from the BD-VAS used in the Heinberg & Thompson (1995) study; this is an indication that researchers rely on anchors that have been previously used in the literature. Further, the “None – Very much” anchors were the most frequently used anchors (appearing in a totally of 19 studies), even though they only appeared in 7 BD-VAS variants. This trend shows that researchers tend to use BD-VASs with this pair of anchors more than once. Notably, “extreme” anchors were used with BD-VASs. Around 30 percent of the 87 studies used the word “Extreme” or “Extremely” in at least one of their BD-VAS anchors ($n = 26$, 29.89%). Some additional instances of “extreme” anchors include “Extremely dissatisfied – Extremely satisfied” and “No dissatisfaction – Extreme dissatisfaction”.

Scale polarity. Regarding scale polarity, most studies measured either dissatisfaction ($n = 37$) or satisfaction ($n = 19$) with one’s body using unipolar scales. Conversely, it was much less common for researchers to measure body satisfaction on a bipolar scale ($n = 14$); that is, researchers did not seem to view body satisfaction and dissatisfaction as two polar ends on the same spectrum. There were some instances of BD-VASs having multiple scale polarities across different items ($n = 17$). This was a result of either: (1) having items in the BD-VAS that do not measure satisfaction (i.e., how attractive do you feel), and thus using anchors that do not reflect satisfaction (i.e., “Very unattractive” to “Very attractive”); or (2) including a reverse-coded item.

When evaluating scale polarity at the scale variants level, similar patterns emerge. The majority of the BD-VAS variants measured dissatisfaction ($n = 23$). In comparison, unipolar BD-

VAS variants measuring satisfaction ($n = 13$), bipolar BD-VAS variants ($n = 13$), and BD-VAS variants having multiple scale polarities ($n = 12$) were not as common.

Item analysis. To evaluate the variability in item analysis, I focused on BD-VASs with 2 or more items (a total of 40 scale variants), as one-item scales do not have the same analytic flexibility in regard to item analysis as do multi-item scales. Most studies averaged the items of the BD-VAS to create an overall body dissatisfaction score ($n = 44$). It was much less common for researchers to analyze each item individually ($n = 14$). There was also one instance where the researchers averaged some items, but not others; satisfaction with specific body parts, including one's *arms*, *chest*, *stomach*, and *shoulders*, were averaged into an overall body dissatisfaction score, while *overall appearance* and *muscle tone* were analyzed separately (Mulgrew, Volcevski-Kostas, & Rendell, 2014).

Reliability and validity evidence. Reliability and validity evidence provided for each available version of the BD-VASs – both in terms of citations and sample-specific calculations – were coded to determine the psychometric strength of each scale. Trends in the citation and calculation of reliability and validity evidence across studies are discussed in the following sections.

To give support for the BD-VASs' reliability, twenty-nine percent of studies ($n = 25$) provided at least one citation as reliability evidence for their scale and forty-seven percent of studies ($n = 41$) calculated the reliability of their scale. These results show that researchers were more likely to calculate sample-specific reliability evidence, as opposed to citing previous reliability evidence for the BD-VAS. Those that calculated reliability evidence primarily examined the internal consistency of the scale through Cronbach's alpha ($n = 27$, 64.29%) or Pearson's correlation coefficient ($n = 15$, 35.71%). Two studies calculated both Cronbach's alpha and a Pearson's correlation coefficient to examine the interitem correlations or internal consistency of the scale. Test-retest reliability was much less often reported ($n = 2$). Researchers reported either Cronbach's alpha ($n = 1$, 50.00%) or a Pearson's correlation coefficient ($n = 1$, 50.00%) as evidence for test-retest reliability. Finally, two studies reported both internal consistency and test-retest reliability.

The opposite trend was observed for validity evidence; researchers were more inclined to provide citations to establish the validity of the BD-VAS ($n = 50$) rather than to calculate it themselves ($n = 21$). The majority of researchers calculated convergent validity, that is, the degree to which the BD-VAS scores were related to similar constructs (e.g., body dissatisfaction, appearance comparison, self-esteem). All of these relationships were examined using Pearson's correlation coefficient ($n = 20$). One study assessed their scale's structural validity by running a principal component analysis to examine whether their VAS items measuring body dissatisfaction and negative affect belonged to two separate factors (i.e., Mask & Blanchard, 2011).

A significant portion of the examined studies provided citations to support the reliability and validity of their version of the BD-VASs. Twenty-nine percent of studies ($n = 25$) provided at least one citation as reliability evidence for their scale; fifty-seven percent of studies ($n = 50$) provided at least one citation as validity evidence for their scale. This trend seems to suggest that researchers are using BD-VASs that have previously-established reliability and validity. In most of these cases, however, researchers cited the Heinberg and Thompson (1995) study as evidence for their scales' reliability ($n = 15$, 60.00%) and validity evidence ($n = 38$, 76.00%), regardless of whether their scale matched the one used in Heinberg and Thompson's study.

Alarming, twenty-four of the sixty-one variations of the BD-VAS (39.34%) had no supporting reliability and validity evidence. Most of these unsupported scales consisted of one ($n = 17$; 70.83%) or two items ($n = 6$; 25.00%). Some studies that used one of these scale variants did provide citations for its reliability ($n = 10$) and validity ($n = 18$). However, those citations were not associated with the version of the scale that these researchers used, meaning they do not support its reliability or validity.

Most commonly used and validated BD-VAS. The most frequently used and validated variants of the BD-VAS in the literature were those used in Heinberg and Thompson's (1995) study (a two-item variant, measuring dissatisfaction with weight and overall appearance, from none to very much, and combining the responses) and a close variant (i.e., measuring dissatisfaction with *appearance* instead of *overall appearance*). The close variant was used by a total of 8 studies, making it the most frequently used variant. In all those studies, this version of the BD-VAS was administered to Australian, female undergraduate students who are

predominately Caucasian (range from 82.5-95.8% Caucasian participants). Across six of the eight studies, the scale demonstrated acceptable or good internal consistency (i.e., Cronbach alpha values), $n_\alpha = 9^2$, $M_\alpha = 0.83$, range: 0.79-0.88. Scores from this scale were also shown to correlate with thin ideal internalization ($r = .43$, Slater et al., 2012; $r = .46$, Tiggemann & McGill, 2004), appearance anxiety ($r = .74$, Harper & Tiggemann, 2008), and making appearance comparisons ($r_{state} = .38$ and $r_{trait} = .44$; Tiggemann & McGill, 2004).

The original BD-VAS used in Heinberg and Thompson's (1995) scale was used by four other studies ($n = 5$). This scale was also used by solely Australian researchers, apart from its initial use by Heinberg and Thompson (1995) in the United States. Similarly, researchers using this scale also relied on predominately Caucasian and female undergraduate samples, with the exception of two studies that had used younger, teenage samples that included both boys and girls (Hargreaves & Tiggemann, 2003a; Hargreaves & Tiggemann, 2003b). The BD-VAS derived from the Heinberg and Thompson (1995) study also had acceptable internal consistency; Bury and colleagues (2016b) found a mean Cronbach's alpha of .80, range: .76 - .84 (pre- and post-exposure), and Heinberg and Thompson (1995)³ and Hargreaves and Tiggemann (2003b) both found a correlation of .81 between the two BD-VAS items. The BD-VAS scores correlated with measures of body appreciation ($r = -.50$), body surveillance ($r = .49$), and engagement in social comparisons ($r = .34$; Andrew, Tiggemann, & Clark, 2015), as well as another measure of body dissatisfaction ($r_{WEIGHT} = .66$ and $r_{APPEARANCE} = .76$; Heinberg & Thompson, 1995).

Discussion

Researchers often rely on BD-VASs to measure state body dissatisfaction. However, my review revealed that there is no consensus among researchers about the scale properties of the BD-VAS. Even though the BD-VAS developed in Heinberg and Thompson's (1995) study was widely cited, it was only occasionally used by other researchers in its original form. I found 61 distinct variants of the BD-VAS, 47 of which were used in only a single study. Moreover, researchers generally failed to provide the necessary reliability and validity evidence to support

² The number of computed Cronbach alpha values is greater than the number of studies that reported them because three studies had calculated the internal consistency of the scale pre- and post-experimental exposure (i.e., Bury, Tiggemann, & Slater, 2016a; Slater et al., 2012; Tiggemann et al., 2017).

³ For the Heinberg and Thompson (1995) study, the correlation coefficient between the two BD-VAS items was calculated by square rooting the reported shared variance value ($R^2 = .65$).

the use of their particular BD-VAS variant. The modification and generation of new BD-VASs in the field impedes the comparability of BD-VAS scores and results across studies, as well as the development of a BD-VAS with acceptable levels of reliability and validity. In this section, I highlight key findings from the review and provide recommendations for the use and development of BD-VASs.

Multiple definitions of body dissatisfaction. In the absence of a systematically developed BD-VAS, researchers often opt to create or modify existing BD-VASs that proliferate different conceptualizations of body dissatisfaction. All BD-VASs proclaim to measure body satisfaction, but they may tap into different experiences of body dissatisfaction, thereby leading researchers to falsely assume that the various versions of the BD-VAS measure the same construct (i.e., the jingle fallacy; Thorndike, 1904). Although there is considerable overlap among the scales – the BD-VASs measured one or a combination of appearance satisfaction, body satisfaction, weight satisfaction, feelings of attractiveness, face satisfaction, and muscle tone satisfaction – the inclusion (or exclusion) of certain dimensions of body dissatisfaction can influence what the researchers are measuring. For example, research has shown that face and body attractiveness independently predict evaluations of overall attractiveness (Peters, Rhodes, & Simmons, 2007) and thus could possibly have independent effects on how satisfied people are with their overall appearance. Therefore, a scale that measures satisfaction with one's *face* (e.g., Want & Saiphoo, 2017) may measure a different construct in comparison to scales that solely measure *body* satisfaction (e.g., Anschutz et al., 2012). These two scales may also then result in different relationships between their measure of body dissatisfaction and other related variables (e.g., internalization of the thin ideal, self-esteem) and negatively impact our ability to compare results across studies.

Likewise, researchers operationalized each measured construct in different ways that added additional variability between BD-VASs. Numerous operationalizations were especially present in BD-VASs measuring satisfaction with appearance and body. For instance, while some BD-VASs measured people's satisfaction with their *appearance*, others measured their feelings about their *overall appearance* or *physical appearance*. The same was observed for BD-VASs measuring satisfaction with one's body; some scales measured global *body* satisfaction, and others measured satisfaction with *body shape*, *body shape and size*, or multiple areas on the body

(e.g., *arms, chest, shoulders, and stomach*; Mulgrew, Volcevski-Kostas, & Rendell, 2014).

Variability in construct operationalizations can affect participants' interpretations of the construct at hand and influence the strength of effect sizes. A review by Blond (2008) found that studies measuring overall *body* or *muscle* dissatisfaction in men produced smaller effects of media images than studies measuring dissatisfaction with specific body parts. Therefore, it is possible that differences in the phrasing of BD-VAS items can additionally impact study results. However, it is still unclear whether these differences in operationalizations measure significantly different dimensions of body dissatisfaction. This is an empirical question that can be answered in future research.

Numerous modifications to the scale properties of the BD-VAS. Beyond the divergence in measured items between the BD-VASs, I observed variations in BD-VASs' scale properties, including anchors, polarity, and item analysis, that were often left unjustified. First, I found that scale anchors were commonly modified (i.e., 27 different pairs of anchors), often for only a single study. In one study, for example, researchers changed the phrasing of the scale's anchors from "None – Very much" (Heinberg & Thompson, 1995) to "None – A lot". The scale anchors not only varied in phrasing, but also in extremity. Around 30% of BD-VASs assessed "extreme" satisfaction and/or dissatisfaction with participants' bodies. If indeed respondents are susceptible to the aversion bias when responding to items with "extreme" anchors (Torrance, Feeny, & Furlong, 2001), then their responses (and consequently the effects found in the study) will be truncated. Responses of BD-VASs ranging from "extremely" or "very" dissatisfied may not be comparable.

Second, BD-VASs frequently measure either dissatisfaction or satisfaction with people's body (i.e., unipolar scale), but there are a few instances of BD-VASs measuring both dimensions on the scale (i.e., bipolar scale). Preference for unipolar BD-VASs could indicate that researchers view satisfaction and dissatisfaction as two orthogonal constructs, as opposed to two ends of the same spectrum. In other words, researchers may think it is possible to feel both satisfaction and dissatisfaction towards one's body at a given time. Another plausible explanation is that researchers are interpreting the absence of satisfaction as the presence of dissatisfaction, and vice versa. Depending on whether body satisfaction and dissatisfaction are unipolar or bipolar in nature, they may be more reliably captured through BD-VASs with certain polarities.

Finally, BD-VAS items were most frequently averaged to create a single body dissatisfaction score. However, some researchers chose to analyze each item separately. One reason for doing so was that the items were poorly correlated or shared only a small amount of variance (e.g., Brikeland et al., 2005; Hargreaves & Tiggemann, 2002), leading researchers to believe they measured separate constructs and should therefore be analysed separately. The choice to combine BD-VAS items (or leave the items as they are) could then influence the relationship the BD-VAS has with other variables.

Changes to the scale properties of the BD-VAS may seem inconsequential. However, it is currently unknown how these changes may affect responses. Differences in the scales' anchors, polarity, and item analysis may yield different response distributions, thereby increasing the variability across studies. Even when assuming this source of variability does *not* lead to changes in participant responding, this variability is still unnecessary. The choices researchers make about the phrasing of their BD-VAS seem largely subjective and typically unjustified. By making these changes, researchers risk the possibility that these semantic differences do in fact impede comparability across studies.

Of course, there are times when modifying a scale is appropriate, for instance when the current scales have poor psychometric properties, are not generalizable across samples, or do not adequately represent the construct the researchers want to measure (Reise, Waller, & Comrey, 2000). When it comes to BD-VASs, there were several instances where researchers modified their scale to fit the needs of their target sample (Mulgrew & Volcevski-Kostas, 2012; Hargreaves & Tiggemann, 2009; Mulgrew, Volcevski-Kostas, & Rendell, 2014; Agliata & Tantleff-Dunn, 2004) or experimental stimuli (Brown & Tiggemann, 2016). For example, research has shown that men and women experience body dissatisfaction differently; women strive to achieve a thin figure, while men more preoccupied with obtaining a muscular body (Tylka, 2011). As a result, when studying the impact of media exposure on men's body satisfaction, several researchers modified the BD-VAS to measure muscle tone (Mulgrew & Volcevski-Kostas, 2012; Hargreaves & Tiggemann, 2009; Mulgrew, Volcevski-Kostas, & Rendell, 2014) or specific body parts like arms, shoulders, and abdomen that influence men's satisfaction with their bodies (Mulgrew, Volcevski-Kostas, & Rendell, 2014; Agliata & Tantleff-Dunn, 2004). However, changes to the specific scale properties of the BD-VAS (e.g., changing

the anchors from “None – Very much” to “None – Extreme”) were rarely justified and did not appear to be derived from any of the concerns mentioned above. I recommend reducing the amount of unnecessary variability between BD-VASs by only allowing scale modifications to the BD-VAS on theoretical or psychometric grounds. In study two, I explore whether the seemingly innocuous changes to the BD-VASs’ scale properties (anchors, polarity, and item analysis) produce additional variability in participant responses.

Limited reliability and validity. Evidence for the reliability and validity of BD-VASs is limited due to poor reporting practices. I observed low rates of reporting previously-found and sample-specific reliability and validity evidence for the different variations of the BD-VASs. Specifically, 39% of all available versions of the BD-VAS did not have any supporting evidence. This trend is especially concerning because most of the BD-VAS variants have been modified for a single use. This finding is consistent with the reporting of reliability and validity evidence in other fields of research, especially when it comes to newly developed or modified scales (Barry et al., 2014; Flake, Pek, & Hehman, 2017). In addition, I found that it was far more common to cite previous research for a scale’s validity than to provide new validity evidence in a given sample. The overreliance on citations for support of a scale’s validity can be problematic, as it can discourage the calculation of reliability and validity evidence. Citations can also be used inappropriately; the validity evidence provided in previous research may be associated with a different version of the BD-VAS, and thus does not apply to the scale used in the research citing it. Similarly, reliability and validity evidence are contextualized to a specific sample and may not be generalizable beyond that sample.

To ensure a scale produces consistent and accurate scores for the general population, researchers should aim to provide strong evidence of reliability and validity for the scale across large and heterogeneous samples. However, this was not observed in the studies obtained in my review. Many of the samples used in the studies consisted of undergraduate students who were mostly Caucasian and female. This may be because it is much easier to obtain undergraduate students to participate in studies (i.e., because they often do not require monetary compensation and are readily available for testing). The prevalence of mostly Caucasian samples may be due to experimental choices or population constraints. A Caucasian sample may be collected to control for the effects of ethnicity. For example, research has shown that ethnic identity plays a role in

how people perceive the thin ideal and are influenced by it (Rakhkovskaya & Warren, 2016). Thus, to minimize the effect of ethnicity, researchers may choose to recruit only Caucasian participants. The proportion of Caucasian participants can also reflect the countries in which those studies were conducted. A large proportion of studies were conducted by Australian researchers; Australia is a relatively homogenous country with mostly Caucasian citizens (Australia Bureau of Statistics, 2017), thus the Australian samples may be the reason why the samples are primarily White. Finally, the body image literature has primarily focused on women's experience with body image. Especially when it comes to the influence media has on people's body image, women are the primary subjects of research because of the ubiquity and rigidity of societal standards of beauty for women that are presented in the media (Buote, Wilson, Strahan, Gazzola, & Papps, 2011). However, men can experience body dissatisfaction too, and research around their experience is a growing field (Blond, 2008). The use of female-only samples leads to the cumulation of reliability and validity evidence for the use of the BD-VAS for women, but not men.

While the purpose of these studies in my review was not to validate the BD-VAS, the samples used inadvertently impacted the strength of their associated reliability and validity evidence. If a version of the BD-VAS was only ever used in studies with Caucasian, female participants, it lacks the necessary reliability and validity evidence to be used in other populations (e.g., males or ethnic minorities). Therefore, it is inconclusive as to whether the BD-VAS may reliably measure body satisfaction across diverse samples.

The reliability and validity of the BD-VAS is also limited by the type of evidence provided. The reported reliability and validity evidence often involved internal consistency and convergent validity. Overreliance on internal consistency evidence – and more specifically, Cronbach's alpha – is also observed in other fields of psychology (Flake, Pek, & Hehman, 2017). Given the emphasis on internal consistency and convergent validity evidence, researchers can be misled to believe the BD-VAS demonstrates better reliability and validity than it actually has (Hussey & Hughes, 2018). For example, not much is known about whether different groups of people respond to the BD-VAS in a similar manner (i.e., measurement equivalence).

One reason for the for the scale's limited reliability and validity evidence is the prevalent use of short BD-VASs. Researchers often used a BD-VAS with only one or two items; twenty-

eight studies used a one-item BD-VAS and forty studies used a two-item BD-VAS (combined, these account for around 78% of studies). This trend highlights the demand for short measures of body dissatisfaction. However, short measures limit the scope of what is measured, as well as the scale's psychometric properties (Soto & John, 2019). By using shorter measures of body satisfaction, researchers may be sacrificing the extent to which they are capturing the construct of body dissatisfaction in full. One-item measures of body dissatisfaction appear to have face validity, but the scales could be lacking content validity and thus not represent the full breath of body dissatisfaction. The reliability of one-item scales, for instance, cannot be evaluated in the same way one would with longer scales; internal consistency requires at least two items in the scale. The only method of calculating reliability of one-item scales is through test-retest reliability, which researchers do not commonly do.

I thus recommend the use of longer BD-VASs, or at the very least a BD-VAS with two items. A two-item BD-VAS will balance the demands of brevity and using reliable and valid measures. Including an additional item is not more taxing for participants or the length of the study and allows for additional reliability and validity tests of the scale (e.g., internal consistency). This is a temporary solution, however, as we do not know whether these shorter BD-VASs fully capture the breath of body dissatisfaction (i.e., have content validity). Further work is needed to determine the number of items (and potential dimensions) needed to measure body dissatisfaction through VASs.

One limitation of this review is that it is not comprehensive of all the BD-VAS variants available in the literature. The articles included in the review were not collected systematically; I retrieved articles that have previously used a BD-VAS by running a basic literature search and looking at studies that have cited influential work, such as the Heinberg and Thompson (1995) paper. Thus, it is plausible that I failed to include several studies that have used a BD-VAS, particularly in other fields of research (e.g., medicine). Even with the limited pool of studies, the review still revealed a large discrepancy in the use of BD-VASs across studies. Specifically, it revealed numerous different versions of the BD-VAS resulting from seemingly arbitrary changes to the BD-VASs' scale properties, including the scales' anchors and polarity. However, it is still unclear how changes to the BD-VASs' scale properties may affect the scales' responding patterns and validity.

There is a gap in knowledge of how the BD-VAS' psychometric properties may differ between BD-VAS variants varying in scale anchors and polarity. Additionally, it is unknown how the BD-VAS performs on other reliability and validity metrics apart from internal consistency. Study 2 attempts to address this by further examining the test-retest reliability and stability of a subset of BD-VASs, varying in their scale properties, as well as the scales' relationships with other known variables in body dissatisfaction's nomological network (i.e., thin ideal internalization, pressures from peers, family, and media to be thin, BMI, self-esteem, and making appearance comparisons).

Study 2

In the second part of this thesis, I aim to provide some evidence-based recommendations for the use of a short VAS measure for body dissatisfaction. Specifically, I experimentally investigate the extent to which arbitrary research choices regarding scale development might impact the way in which participants respond to the items and the degree to which the scale is psychometrically sound (i.e., its internal consistency, test-retest reliability, and convergent validity). Based on the results from Study 1, I primarily focus on the BD-VAS containing items measuring weight and appearance, as it was used in more studies than any other variant and had the greatest amount of supporting reliability and validity evidence in the literature. In this study, I constructed six variants of this BD-VAS by varying the scale's anchors and polarity; these changes represent the common ways in which researchers altered the BD-VAS (e.g., changing the anchors from "Very dissatisfied" to "Extremely dissatisfied"). Participants obtained from an undergraduate pool were randomly assigned to complete one of six variants of the BD-VAS twice separated by a 5-minute filler task unrelated to body image. This intermediary filler task allowed for the calculation of the BD-VASs' test-retest reliability. Then, participants completed additional measures of trait body satisfaction, self-esteem, thin internalization, sociocultural pressures to be thin, tendency to make appearance comparisons, and body index mass to examine the relationships these measures have with scores on the BD-VASs.

Given the properties of BD-VASs – they are state measures and have no discrete categories – 5 minutes is a sufficient amount of time between the two BD-VAS administrations to calculate test-retest reliability. State body dissatisfaction is much less stable than trait body dissatisfaction and is bound to change from moment to moment. A 5-minute time interval is a short enough time to test the temporal consistency of participants' BD-VAS scores while also ensuring that genuine differences in state body dissatisfaction most likely do not occur. For Likert-type scales a longer wait time between the first and second set of measures is typically adopted to prevent participants from remembering how they previously responded and having that knowledge influence how they respond on the subsequent measure. This is not an issue for VASs, since the absence of discrete categories makes it harder for participants to recall exactly where they placed their response. Finally, a 5-minute time interval additionally reflects the amount of time participants are typically given to view media images in previous experimental

studies (e.g., Want & Saiphoo, 2017). This parallel will provide insight to the stability of BD-VAS scores *without* the presence of experimental stimuli (e.g., media images).

The item distributions of the weight, appearance, and the composite body dissatisfaction scores were plotted to examine any differences in participant responding based on the variation in scale properties. Additionally, the distribution shape and dispersion of the overall body dissatisfaction scores across BD-VAS variants were statistically compared. I hypothesized that BD-VAS variants with different extremities – ranging until “Very” or “Extreme” body (dis)satisfaction – would yield score distributions varying in skew and spread. Specifically, I predicted that a BD-VAS ranging from “Not at all” dissatisfied to “Extremely” dissatisfied would result in a positively skewed distribution, as participants are less likely to report an extreme response, while the item distribution of a BD-VAS with the anchors “Not at all” to “Very” dissatisfied may have a greater spread of scores and more closely resemble a normal distribution. Similarly, a bipolar BD-VAS ranging from “Extremely” dissatisfied to “Extremely” satisfied was predicted to have a greater peak (i.e., leptokurtic distribution) than a bipolar BD-VAS ranging from “Very” dissatisfied to “Very” satisfied. The internal consistency for these two-item scale variants were evaluated using the Spearman-Brown statistic (Eisinga, Grotenhuis, & Pelzer, 2013). Test-retest reliability was assessed for both individual items (i.e., weight and appearance VASs) and their averaged score by correlating BD-VAS scores at time 1 and time 2. The stability of BD-VAS scores was assessed using a Bland-Altman plot and the repeatability coefficient (Bland & Altman, 2003). Lastly, evidence of the BD-VASs’ construct validity was provided by correlating the scores of the BD-VAS with the scores of the longer, previously validated scales of body dissatisfaction and its covariates.

Method

Participants. A total of 523 online responses were collected via Qualtrics. After removing incomplete surveys ($n = 53$), duplicate responses ($n = 6$), surveys with 20% or more missing items within each scale (a cut-off for missing values established in the preregistration) ($n = 19$), inattentive responding ($n = 31$), and those who did not self-identify as female ($n = 1$), the final sample included 413 female undergraduate students. The age of the participants ranged from 18 to 50 years old ($M_{\text{age}} = 19.88$, $SD = 3.90$). Participants’ age, $F(5, 404) = 1.44$, $p = 0.21$, body mass index, $F(5, 304) = 1.18$, $p = 0.32$, and trait body dissatisfaction, $F(5, 396) = 0.55$, $p =$

0.74, did not significantly differ across conditions. Demographic information by condition is displayed in Table 3.

Materials and procedure. Participants completed a set of measures in the following order.

Body Dissatisfaction Visual Analogue Scale (Time 1). Participants were randomly assigned to complete one of the following six versions of the BD-VAS:

1. Not at all dissatisfied – Very dissatisfied (**Dissatisfied (Very)**)
2. Not at all dissatisfied – Extremely dissatisfied (**Dissatisfied (Extreme)**)
3. Not at all satisfied – Very satisfied (**Satisfied (Very)**)
4. Not at all satisfied – Extremely satisfied (**Satisfied (Extreme)**)
5. Very dissatisfied – Very satisfied (**Bipolar (Very)**)
6. Extremely dissatisfied – Extremely satisfied (**Bipolar (Extreme)**)

Each BD-VAS consisted of two items measuring participants' state (dis)satisfaction with their weight and appearance; in other words, the scale evaluated how participants felt about their weight and appearance in *that particular moment*. Although the version of the scale varied between participants, the properties of the scale remained consistent *within* participants, both across items (i.e., weight and appearance) and time points (i.e., prior to or following the filler task). The psychometric properties of the BD-VASs will be discussed in the results section. For the full versions of the BD-VASs, refer to Appendix B.

Filler task. Participants were then asked to engage in a 5-minute filler task. The filler task consisted of a snake game embedded within the online survey. The purpose of the task was to keep the participants occupied for a set amount of time between the first and second administration of the BD-VAS. This allowed for the calculation of the scales' test-retest reliability. The snake game was chosen as it was a neutral task that would likely not change participants' body dissatisfaction. In addition, this task was thought to be engaging enough to dissuade participants from getting distracted and engaging in other activities that may impact their body satisfaction, like browsing social media.

Body Dissatisfaction Visual Analogue Scale (Time 2). After the filler task, participants completed a second set of the same version of the BD-VAS.

Body Areas Satisfaction Scale (BASS). Then, participants filled out the BASS to assess their *trait* body satisfaction. The BASS is a 9-item subscale of the Multidimensional Body-Self Relations Questionnaire (BASS; Brown, Cash, & Milulka, 1990; Cash, 2018) that measures satisfaction with specific areas of the body – for example, the face, thighs, stomach, chest, and so on. Responses are provided on a 5-point Likert scale ranging from “Very dissatisfied” to “Very satisfied”, with greater scores on the scale indicating greater body satisfaction. The BASS has previously shown acceptable internal consistency ($\alpha = .73$) and 1-month test-retest reliability ($r = .74$) in women (Cash, 2018). Additionally, the scale was able to discriminate the experiences of body dissatisfaction between those suffering from an eating disorder (i.e., anorexia nervosa or bulimia nervosa) from healthy controls (Hrabosky et al., 2009). In this sample, the scale demonstrated good internal consistency, $\alpha = 0.85$, 95% CI [0.83, 0.87].

Rosenberg’s Self-Esteem Scale (RSES). Participants’ self-esteem was assessed using the RSES, the most widely used measure of global self-esteem (Rosenberg, 1965). Participants were asked to report their agreement with a total of 10 statements reflecting self-esteem, for instance, “On the whole, I am satisfied with myself” and “At times, I think I am no good at all” (reverse-coded item). Agreement is evaluated on a 4-point Likert scale ranging from “Strongly Disagree” to “Strongly Agree”. Sinclair et al. (2010) showed that the factor structure and internal consistency of the scale was generally consistent across cultures and geographical locations. The scale also demonstrated good convergent validity with measures of self-competence ($r = .92$) and self-liking ($r = .95$). In this sample, RSES showed good internal consistency, $\alpha = 0.88$, 95% CI [0.87, 0.90].

Sociocultural Attitudes Towards Appearance Questionnaire (SATAQ-4). Next, the SATAQ-4 was completed to assess the societal and interpersonal influences on participants’ appearance ideals (Schaefer et al., 2015). The SATAQ-4 contains five subscales that measure (1) internalization of the thin ideal, (2) internalization of the muscular ideal⁴, and pressure from (3)

⁴ The muscular ideal subscale of the SATAQ-4 was not included in the study’s analyses, as it was not predicted to be significantly related to state appearance or weight dissatisfaction in women in comparison to the other SATAQ-4 subscales (e.g., thin ideal internalization).

family, (4) peers, and (5) the media to look thinner, improve appearance, lower body fat, and get in better shape. Participants reported their agreement with statements concerning the above dimensions, such as “I want my body to look very thin”. In the work that initially validated this scale, the SATAQ-4 subscales had internal consistencies that ranged from .87-.95 and showed consistent psychometric properties (i.e., factor structure and internal consistency) across races (i.e., Caucasian vs. non-Caucasian participants) and countries (US vs. Non-US) (Schaefer et al., 2015). In the current sample, the Thin Ideal subscale, $\alpha = 0.78$, 95% CI [0.75, 0.81], Muscular Ideal subscale, $\alpha = 0.89$, 95% CI [0.87, 0.90], Family Pressure subscale, $\alpha = 0.89$, 95% CI[0.88, 0.91], Peer Pressure subscale, $\alpha = 0.86$, 95% CI[0.84, 0.89], and Media Pressure subscale, $\alpha = 0.90$, 95% CI[0.88, 0.91], demonstrated comparable internal consistencies.

Physical Appearance Comparison Scale-Revised (PACS-R). The PACS-R assessed participants’ general tendency to compare their physical appearance to that of others (Schaefer & Thompson, 2014). This includes the comparison of their weight, body size and shape, body fat, and overall appearance. Participants reported how often they make such comparisons (0 = *Never*; 4 = *Always*) across different circumstances. For example, participants provided the frequency of the following occurring: “When I’m out in public, I compare my physical appearance to the appearance of others”. Schaefer and Thompson (2014) reported great internal consistency ($\alpha = 0.97$) for this revised scale. The scale was also positively associated with measures of thin ideal internalization and pressures to be thin from family, peers, and media, and negatively associated with body satisfaction and self-esteem (Schaefer & Thompson, 2014). In the present sample, the scale similarly demonstrated great internal consistency, $\alpha = 0.96$, 95% CI [0.95, 0.96].

Body Mass Index (BMI). BMI was evaluated using self-reported weight (in pounds or kilograms) and height (in feet or centimeters). Participants’ height and weight were obtained through open-ended questions; participants were asked to report both the value and metric of their weight and height. Any height or weight value that did not have an associated metric was excluded from the final analyses.

Demographics questionnaire. Finally, we asked participants about their gender, age, and ethnicity. Participants’ gender was evaluated to ensure our sample consisted of only those that self-identified as female. We evaluated participants’ ethnicity through an open-ended question; this allowed participants to answer in a way that best reflects the ethnicity they identify with. As

a result, we received numerous ethnic identities across participants. To parsimoniously represent participants' ethnicities, we grouped them based on the following ethnic/racial groups: (1) White or European-Decent, (2) East Asian (e.g., China, Japan, Korea), (3) Black or African American, (4) Middle Eastern (e.g., Iran, Turkey), (5) South Asian (e.g., Pakistan, India), (6) Southeast Asian (e.g., Vietnam, Philippines), (6) Hispanic or Latina (e.g., Spain, Latin America), (7) West Indian (e.g., Caribbean), (8) First Nations, and (9) Multi-Ethnic.

Statistical Analyses. In the following section, I describe statistical analyses comparing the score distributions of the BD-VAS variants and calculating their reliability and convergent validity. All of the following analyses were performed in the statistical software R version 3.4.1.

Comparing score distributions. The score distributions of the weight, appearance, and composite body dissatisfaction BD-VAS score were plotted to examine any differences in participant responding based on differences in scale properties. To compare BD-VAS responses across the six scale variations, I initially proposed a 6 (scale variations) by 3 (Appearance, Weight, and averaged Body Dissatisfaction scores) mixed ANOVA (as per the study's pre-registration; <https://osf.io/5zd6n/>). Further, I ran a 2 ("Extreme" vs. "Very" anchors) by 3 (Dissatisfaction, Satisfaction, or Bipolar scales) by 3 (Appearance, Weight, and averaged Body Dissatisfaction scores) mixed ANOVA to examine the effects scale extremity and polarity have on BD-VAS scores. However, these analyses mainly examine the differences in *means* between score distributions, and do not tell us much about the differences in the *shape* or *dispersion* of the distributions (i.e., skewness, kurtosis, variance), which may result from altering the scale properties of the BD-VAS (e.g., anchor extremity). Thus, to compare the shape of the BD-VASs' response distributions, I additionally ran multiple one-sample and two-samples Kolmogorov-Smirnov tests. The Kolmogorov-Smirnov (KS) test is a non-parametric test that compares a response distribution against a normal distribution (one-sample KS test) or another response distribution (two-samples KS test; Goodman, 1954). Thus, one-sample KS tests were run to analyse whether BD-VAS score distributions were statistically different from a normal distribution and two-samples KS tests were run to evaluate whether scores of BD-VASs differing in scale properties produced significantly different distribution shapes. Unlike its parametric test counterparts (e.g., independent samples t-test), a Kolmogorov-Smirnov test does not require the distribution(s) being compared to be normally distributed. It instead calculates the distance

between the empirical distribution functions (regardless of whether the distribution(s) are normal or not normal) of the two compared distributions, with a larger value indicating a larger discrepancy between the distributions. To account for the multiple comparisons and reduce the rate of false positives, the p-values of the Kolmogorov-Smirnov tests were adjusted using the false discovery rate (FDR) method (Benjamini & Hochberg, 1995). Unlike other corrections, such as the Bonferroni correction, FDR is much less conservative, and thus has more power to detect an effect.

Reliability. The reliability of the BD-VAS was determined through both internal consistency and test-retest reliability. First, internal consistency was evaluated using Spearman-Brown coefficient. Research by Eisinga and colleagues (2013) showed that both Cronbach's alpha and Pearson's correlation coefficient tend to underestimate the reliability of two-item scales. The Spearman-Brown coefficient, on the other hand, tends to overestimate reliability, but is less biased in comparison to Cronbach's alpha. Thus, their recommendation is to use the Spearman-Brown correlation adjustment to get at the most accurate estimate of two-item scales' reliability. Second, test-retest reliability was calculated by correlating BD-VAS scores at time 1 and time 2. These correlations will reveal how strongly time 1 BD-VAS scores are associated with time 2 scores.

Third, the stability of BD-VAS scores across those two time points was assessed through a Bland-Altman plot and the repeatability coefficient (Bland & Altman, 2003; Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013). In comparison to the previous reliability analyses, the Bland-Altman plot and repeatability coefficient examines *absolute reliability* – that is, it assesses the agreement, as opposed to the consistency or association, between the two repeated-measure BD-VAS scores (Vaz et al., 2013). So, while a correlation between BD-VAS scores at time 1 and time 2 may reveal that they are strongly related, it cannot conclude whether the scores remained exactly the same from time 1 to time 2. A Bland-Altman plot graphs the average of the scores at the two time points against the difference between those two scores. Additionally, it plots the mean difference between scores and the 95% limits of agreement. Since participants' scores theoretically should not change from the first to second administration of the BD-VAS, these metrics were used to uncover variability that resulted from random and systematic measurement error. For example, if the average difference between BD-VAS scores deviates from zero (i.e.,

no change in scores between time 1 and time 2), it may highlight instances of measurement bias associated with that scale. The 95% limits of agreement contain 95% of the differences in BD-VAS scores across time. In other words, 95% of participants will have a difference in BD-VAS scores that falls within the upper and lower limits of agreement. The absolute distance between the average score difference and the upper or lower limit of agreement represents the repeatability coefficient. The repeatability coefficient is sometimes referred as the Smallest Real Difference (SRD), as it defines the limits where scores vary due to random or systematic error, and thus the minimum difference needed to indicate a meaningful change in the measured construct (Beckerman, Roebroek, Lankhorst, Becher, Bezemer, & Verbeek, 2001). The Bland-Altman plot and its associated statistics were previously used to assess the reliability of VASs measuring acute pain (Bijur, Silver, & Gallagher, 2001).

Convergent validity. Lastly, evidence of the BD-VASs' convergent validity was provided by correlating the scores of the BD-VASs with the scores of the longer, previously validated body satisfaction scale (i.e., BASS), as well as other related variables, including RSES, PACS, the SATAQ-4 subscales (i.e., SATAQ-Thin, SATAQ-Family, SATAQ-Peers, and SATAQ-Media), and BMI. Given that the distributions of the SATAQ-4 subscales, PACS, and BMI were almost never normal, relationships between BD-VAS scores and these measures was assessed using Spearman's rank-order correlation coefficient.

Results

Comparing response distributions. Figure 3 illustrates the score distributions of the six BD-VAS versions at time 1 and Table 4 displays the distributions' mean, median, skew, and kurtosis. By examining the frequency distributions of BD-VAS scores, it can be seen that the different variations of the BD-VAS produced slightly different response patterns. While the means and standard deviations of the BD-VASs' score distributions were seemingly similar, there were some differences in the shape of the distributions. For instance, the polarity of the BD-VAS – that is, whether the scale was measuring satisfaction, dissatisfaction, or both – impacted the skew of the response distributions; BD-VASs measuring body satisfaction on a bipolar scale produced response distributions that were relatively symmetrical ($Skew_{Very} = -0.08$; $Skew_{Extremely} = -0.14$), while BD-VASs measuring sole body satisfaction or dissatisfaction produced moderately negatively ($Skew_{Very} = -0.42$; $Skew_{Extremely} = -0.38$) or positively skewed

distributions ($\text{Skew}_{\text{Very}} = 0.33$; $\text{Skew}_{\text{Extremely}} = 0.12$), respectively. However, based on this information alone, it is unclear whether the score distributions are significantly different from one another. A 6 (scale variations) by 3 (Appearance, Weight, and averaged Body Dissatisfaction scores) mixed ANOVA and two-sample Kolmogorov-Smirnov tests were run to examine whether score distributions from scales with different anchor extremity and scale polarity were significantly different.

Mixed ANOVAs. A 6 x 2 mixed ANOVA was run to assess whether the mean scores of the individual BD-VAS items and overall body dissatisfaction score differed between the six scale versions. Only participants who had responded to both the Appearance and Weight BD-VAS items were included in this analysis ($n = 395$). To run this analysis, three assumptions must be met: (1) normally distributed residuals, (2) homogeneity of variances, and (3) sphericity. Residuals of within- (i.e., item scores) and between-subjects variables (i.e., BD-VAS versions) were found to be normally distributed. A Levene's test showed differences in variance across the different BD-VAS versions, suggesting heterogeneity of variances, $F(5,1179) = 2.987, p = 0.011$. However, given the test's sensitivity to detect even small differences in group variances, a variance ratio (also known as the Hartley's F_{max} ; Pearson & Hartley, 1954) was calculated to provide additional evidence for the homogeneity of variances assumption. This ratio represents the relationship between the largest and smallest variances of scores across the six BD-VAS versions. Ratios must be lower than 2 to indicate a non-significant difference between the variances. The variance ratio was 1.49, providing evidence of homogeneity of variances. Based on these results, I concluded that the assumption of homogeneity of variances for this analysis was met. Finally, the assumption of sphericity was also assessed using a variance ratio of the largest and smallest variances of the *difference scores* between appearance, weight, and body dissatisfaction scores. This ratio yielded a value of 4, indicating that the assumption of sphericity was not met (value greater than 2). However, due to having unbalanced data, corrections for sphericity (e.g., Greenhouse-Geisser correction) were not estimable, and thus the results should be interpreted with caution. The mixed ANOVA revealed no significant main effect of BD-VAS version, $F(5,389) = 0.861, p = 0.507, \eta^2 = 0.00960$, or main effect of item scores, $F(2,778) = 0.233, p = 0.792, \eta^2 = 0.0000745$.⁵ This means that the different BD-VAS versions varying in

⁵ I did not hypothesize that there would be an interaction between BD-VAS version and item type.

extremity and scale polarity did not yield statistically significant differences in mean scores across the different items and composite score.

A second 2 x 3 x 3 mixed ANOVA was run to assess the nuanced effects of BD-VAS scale extremity and polarity on score means of appearance, weight, and overall body dissatisfaction. Within- and between-subjects residuals were normally distributed, thus the normality assumption was met. There was mixed evidence for the assumption of homogeneity of variances; while the Levene's tests comparing variances between different levels of polarity, $F(2,1182) = 4.24, p = 0.015$, and extremity, $F(1,1183) = 4.77, p = 0.029$, seem to suggest the assumption was violated, the variance ratio for polarity and extremity were 1.32 and 1.13 respectively, indicating homogeneity of variances. Finally, as with the previous analysis, the assumption of sphericity was violated. This mixed ANOVA also revealed no significant main effect of item scores, $F(2,788) = 0.23, p = 0.795, \eta^2 = 0.000074$. Additionally, neither the main effect of scale polarity, $F(2,391) = 1.58, p = 0.21, \eta^2 = 0.0070$, or extremity, $F(1,391) = 0.23, p = 0.63, \eta^2 = 0.00050$, were statistically significant. In other words, the independent manipulation of the extremity or polarity of the BD-VAS did not lead to statistically significant differences in mean scores across appearance, weight, and overall body dissatisfaction scores.

Kolmogorov-Smirnov tests. One-sample KS tests were run to assess whether the score distribution of each BD-VAS variant was significantly different from a normal distribution. Results revealed no statistically significant differences between the score distributions of the BD-VAS variants and a normal distribution. Specifically, the Dissatisfaction (Very), $D = 0.125, p = 0.93$, Dissatisfaction (Extreme), $D = 0.14, p = 0.93$, Satisfaction (Very), $D = 0.13, p = 0.93$, Satisfaction (Extreme), $D = 0.17, p = 0.93$, Bipolar (Very), $D = 0.09, p = 0.98$, and Bipolar (Extreme) BD-VAS, $D = 0.12, p = 0.93$, were not statistically significantly different from a normal distribution.

Two-sample KS tests were run between BD-VASs of different extremities to assess differences in the shape of their score distributions. The results show that the extremity of the BD-VAS anchors did not significantly affect the response distributions. The two-sample KS tests revealed there was no significant difference between score distributions of BD-VASs using "Very" or "Extreme" anchors in scales measuring satisfaction, $D = 0.124, p = 0.824$, dissatisfaction, $D = 0.118, p = 0.824$, or both (i.e., bipolar scales), $D = 0.123, p = 0.824$.

Similarly, the polarity of the BD-VAS – whether it measured satisfaction, dissatisfaction, or both on a single continuum – did not significantly impact the way scores were distributed. Score distributions of different polarities were grouped and compared within each anchor type. That is, score distributions of BD-VASs using the anchors “Very” and “Extreme” were analysed separately. Within BD-VASs using the anchor “Very”, there were no statistically significant differences between the distributions of the different scale polarities: satisfaction versus dissatisfaction scales, $D = 0.167$, $p = 0.824$; satisfaction versus bipolar scales, $D = 0.149$, $p = 0.824$; and dissatisfaction versus bipolar scales, $D = 0.130$, $p = 0.824$. The same was observed for BD-VASs using the “Extreme” anchor: satisfaction versus dissatisfaction scales, $D = 0.151$, $p = 0.824$; satisfaction versus bipolar scales, $D = 0.088$, $p = 0.939$; and dissatisfaction versus bipolar scales, $D = 0.163$, $p = 0.824$.

Reliability. In this next section, I discuss my findings for the reliability of the BD-VAS in the form of internal consistency and test-retest reliability. The internal consistency and test-retest reliability found across the different versions of the BD-VAS are outlined in Table 4.

Internal consistency. Internal consistency was calculated using the Spearman-Brown formula, as per Eisinga and colleagues’ (2013) recommendation for two-item scales. The BD-VASs’ internal consistency at time 1 was acceptable, albeit rather low; the Spearman-Brown coefficients across BD-VAS variants ranged from 0.74 to 0.84. At time 2, the scales’ internal consistency improved, with Spearman-Brown coefficients ranging from 0.82 to 0.89. The internal consistencies of the BD-VAS variants were comparable at time 1 and time 2. Overall, weight and appearance satisfaction measured via BD-VASs seem to produce similar scores, thus providing evidence that they measure a similar construct.

Test-retest reliability. Test-retest reliability was first assessed using Pearson correlations between scores at time 1 and time 2. The correlations between time 1 and time 2 BD-VAS scores were high (range: .84-.97), especially for BD-VASs measuring satisfaction; the test-retest reliability for Satisfaction (Very) scales was between .92 and .97, while Satisfaction (Extremely) scales had a test-retest reliability between .93 and .97. A high correlation between BD-VAS scores at time 1 and time 2 is a necessary, but not sufficient piece of evidence that there is good agreement between the two sets of scores. Thus, a Bland-Altman plot was plotted to assess the test-retest agreement of the BD-VASs (see Figure 4).

Overall, the Bland-Altman plot revealed that, irrespective of the BD-VAS variants, the average difference between time 1 and time 2 BD-VAS scores was -0.23 and that 95% of difference scores fell between -16.99 (i.e., time 2 scores decreased by about 17 points) and 16.54 (i.e., time 2 scores increased by about 16.5 points). In other words, the Bland-Altman plot showed that BD-VAS scores can naturally vary by ± 16.77 points. Table 4 further breaks down the mean score differences, 95% limits of agreement, and the repeatability coefficients per BD-VAS variant. None of the BD-VAS variants exhibited concerning measurement bias; the mean difference of BD-VAS scores for all the variants was close to zero. Although the mean difference for the *Dissatisfied (Very)* BD-VAS was larger than the rest (i.e., -2.54), given the range of scores of a VAS (from 0 to 100), this value does not strongly deviate from zero. Once again, the *Dissatisfied (Very)* BD-VAS, as well as the *Bipolar (Very)* BD-VAS, showed greater 95% limits of agreement and repeatability coefficients, meaning that there was a greater discrepancy between their scores at time 1 and time 2 than the other BD-VAS variants. However, this may have been driven by outliers with larger difference scores (Figure 4).

Convergent validity. The intercorrelations between the six versions of the BD-VAS and BASS, RSES, SATAQ-Thin, SATAQ Pressures subscales, PACS, and BMI are presented in Table 6. State body dissatisfaction across the different BD-VAS versions was highly correlated with trait body dissatisfaction (i.e., BASS). Similarly, all of the BD-VASs had a statistically significant relationship with self-esteem (i.e., RSES). However, the *Dissatisfaction (Very)* ($\rho = -0.59$), *Dissatisfaction (Extremely)* ($\rho = -0.53$), and *Bipolar (Very)* BD-VASs ($\rho = 0.50$) had slightly stronger relationships with RSES than the *Satisfaction (Very)* ($\rho = 0.38$), *Satisfaction (Extremely)* ($\rho = 0.36$), and *Bipolar (Extremely)* BD-VASs ($\rho = 0.31$).

The relationships between BD-VAS scores and SATAQ subscales, PACS, and BMI are more variable across the BD-VAS variants. For instance, the *Satisfied (Extremely)* BD-VAS had relatively lower, not statistically significant relationships with those variables; none of the relationships between its scores, thin internalization, pressures from family, peers, and media to be thin, appearance comparisons, and BMI were significant and the Spearman rank-order correlation coefficients ranged from -.16 to -.25. On the other hand, the BD-VASs that revealed the strongest evidence of convergent validity are the *Dissatisfied (Very)* and *Bipolar (Extremely)* BD-VASs, strongly correlating with all the variables or all but SATAQ-Peers respectively. For

the *Dissatisfied (Very)* BD-VAS, the Spearman rank-order correlation coefficients were between .27 and .47; for the *Bipolar (Extremely)* BD-VAS, the Spearman rank-order correlation coefficients were between -.20 and -.49.

Discussion

In study 2, I sought out to determine whether common changes made to the BD-VAS by researchers in the field – for instance, changing the scale’s extremity and polarity – might affect participant responding and the scale’s psychometric properties. Specifically, I compared the score distributions of BD-VASs against a normal distribution, as well as the distributions of BD-VASs using “Very” or “Extreme” anchors, different scale directions (i.e., measuring satisfaction, dissatisfaction, or both on a bipolar scale), and different item analysis techniques (i.e., averaged versus individual item scores). The differences in scale properties across the different BD-VAS variants did not significantly impact the score distribution nor the scales’ psychometric properties as initially predicted.

Similar score distributions across BD-VAS variants. Contrary to my predictions, the extremity of the BD-VAS anchors did not significantly impact score distributions. I expected that scores of BD-VASs using the anchor “Extreme” would be less dispersed and more peaked (i.e., resemble a leptokurtic distribution) than BD-VASs using the anchor “Very”. While there is some evidence for this prediction – kurtosis values generally tended to be larger and standard deviation scores were smaller in BD-VASs using “Extreme”, as opposed to “Very”, as their anchor – there were no significant differences between score distributions of BD-VAS varying in anchor extremity. These results do not support the end aversion described when using “Extreme” anchors (Zusman, 1986; Torrance, Feeny, & Furlong, 2001). In other words, scores near the end of the spectrum were not significantly more likely to occur in BD-VASs using “Very” anchors compared to BD-VASs using “Extreme” anchors.

A possible explanation for the similarity between score distributions of BD-VASs using “Very” or “Extreme” anchors is that the end-aversion bias may not be as pronounced in 100-points VASs as with Likert scales. End-aversion bias in Likert scales (e.g., scales with 5 or 7 response options) can change to the dispersion and shape of response distributions depending on how the end-labels are defined. Assuming people tend to avoid endorsing extreme attitudes and behaviours, employing a Likert scale with an extreme end-label(s) can reduce the amount of

viable responses by 1 or 2 points in unipolar and bipolar scales respectively. This in turn can create significant differences between Likert scales using extreme and non-extreme labels; Likert scales with extreme labels may have reduced response variance and heightened response distribution peaks in comparison to scales with non-extreme labels. VASs, on the other hand, have a great number of possible response options (typically scored from 0 to 100), thus attempts to avoid choosing extreme responses may not result in a drastic shift in the shape and dispersion of response distributions. In other words, end-aversion bias may not significantly impact participants' scores and study results when using VASs.

The same patterns were observed for the effect of scale polarity. The hypothesis that bipolar BD-VASs would produce scores that were normally distributed and that unipolar BD-VASs measuring satisfaction or dissatisfaction with one's body would result in negatively and positively skewed distributions respectively was not fully supported. Once again, when evaluating the skewness of the distributions, they seemed to follow their predicted patterns; satisfaction scales were slightly negatively skewed, dissatisfaction scales were slightly positively skewed, and bipolar scales were relatively symmetrical. However, statistical comparisons of BD-VASs with different polarities revealed that the scale distributions did not significantly vary between satisfaction, dissatisfaction, and bipolar BD-VASs. Further comparisons also showed that the score distributions of each BD-VAS variant resembled a normal distribution. While some researchers have previously advocated for unipolar (Monk, 1989) or bipolar scales (Schaeffer & Presser, 2003) for the measurement of states, in the context of body satisfaction, it is less clear which scale format should be used.

The similarity between score distributions across BD-VAS variants calls into question how participants are interpreting the VAS continua. One explanation for the similarity of score distributions is that there is an additional source of variability in responses among participants that goes beyond the impact of differences in BD-VAS scale properties. Researchers often tacitly assume that the wording of the self-report scale they administer will be universally understood by all of the participants in the same way; however, this may not be the case. For each distinct BD-VAS, respondents may have different interpretations of the scale anchors, as well as the position and distance between meaningful changes in body dissatisfaction. It is also possible that BD-VASs with different anchors and scale polarities are interpreted in similar ways, even though

their underlying meaning is inherently different. For example, participant may have interpreted “Not at all satisfied” as feeling dissatisfied, which would explain the similarities between bipolar and unipolar BD-VASs measuring satisfaction. Another explanation is that perhaps people do not generally pay attention to the wording of the anchors, and instead perceive the two ends of the VAS as the worst and best possible response options. Future research should investigate where people place particular body satisfaction – for example, being moderately satisfied with their body – on the BD-VAS.

Consistent psychometric properties across BD-VAS variants. The six BD-VAS variants generally demonstrated good psychometric properties that seemed unaffected by the differences in scale properties. For instance, the internal consistency was comparable across the different BD-VAS variants at both time 1 (α range: .74-.84) and time 2 (α range: .82-.89). Although BD-VASs measuring *satisfaction* outperformed the other BD-VAS variants in terms of test-retest reliability (r value of .97 for both scales), the other scales had good test-retest reliability as well (r range: .90-.94). Finally, all the BD-VASs correlated strongly with a longer, validated trait body dissatisfaction scale (i.e., BASS). Overall, the BD-VASs seem to produce consistent and related scores and strongly correlate with trait body dissatisfaction, regardless of their scale properties.

However, I also found that BD-VASs are not as stable as researchers may assume. Although there were high associations between time 1 and time 2 BD-VAS scores, the Bland-Altman plot and repeatability coefficients showed that the agreement between the two scores is worse than expected. Specifically, 95% of BD-VAS scores increased or decreased by up to 16.77 points from time 1 to time 2. One explanation for the instability of scores is that participants were more focused on their body when answering the BD-VAS for the second time. By asking participants twice to report how (dis)satisfied they feel about their body within a short amount of time, I might be heightening their attention towards their body and their feelings about it. The 5-minute break in-between BD-VAS administrations gives participants a chance to further think, and potentially ruminate, about the way they look, thus making them feel more dissatisfied with their bodies. Therefore, priming participants to think about their body and their satisfaction or dissatisfaction with it can lead them to change how they feel about their bodies. Alternatively, the BD-VAS may not be as stable of a measure as researchers would like it to be. Due to

measurement error, BD-VAS scores may naturally fluctuate by +/- 16.77 points. This finding can potentially serve a threshold in future studies for detecting meaningful changes in body dissatisfaction. One way to conservatively determine the impact of certain influences (e.g., media images) on body satisfaction would be to conclude that an individual's body satisfaction did not change unless their score increased or decreased by 17 or more points; that is, we cannot be sure of their change in body satisfaction unless it exceeds 95% of the difference scores of people who experience nothing between measurements.

When it comes to the BD-VASs' relationships with other related variables, such as self-esteem and thin internalization, there is greater variability across BD-VAS variants. For example, the BD-VAS variant measuring *satisfaction* with one's body using an "Extreme" anchor was not significantly associated with most covariates that have previously shown to correlate with body dissatisfaction. These variables included internalization of thin ideal, pressures from family, peers, and media to be thin, appearance comparison, and BMI. On the other hand, the *Dissatisfied (Very)* and *Bipolar (Very)* BD-VAS variants demonstrated stronger and statistically significant relationships between state body dissatisfaction scores and its covariates. However, it is unclear whether the strength of the relationships is reflective of the BD-VASs' scale properties or individual differences between samples. While it is possible that measuring state body dissatisfaction using a bipolar BD-VAS or a unipolar BD-VAS measuring dissatisfaction using the "Very" anchor better represented state body dissatisfaction scores and thus were more strongly related to body dissatisfaction covariates, it is also possible that the differences in correlation strength may be due to individual differences between participant samples. Since BD-VAS scores were compared *between-* and not *within-participants*, it is difficult to determine whether these differences are due to scale- or individual-level factors.

Alternatively, BD-VAS variants may have associations of varying strengths and statistical significance due to small sample sizes. A study by Schönbrodt & Perugini (2013) examined at what sample size correlations stabilize across different effect sizes, level of confidence (i.e., the degree to which we allow false positives to occur), and corridor of stability width (i.e., the range of yielded correlation coefficients). Assuming the expected correlations between the BD-VASs and its correlates will yield a moderate effect size (between 0.30 and 0.70) based on previous research ($r_{\text{RSES}} = .36-.57$, Baldwin & Courneya, 1997; Lowery et al.,

2005; Alleva et al., 2014; $r_{\text{SATAQ-Thin}} = .43-.46$; Slater et al., 2012; Tiggemann & McGill, 2004; $r_{\text{PACS}} = .44$, Tiggemann & McGill, 2004), to have a 95% level of confidence and a corridor of stability width of 0.10 each condition requires a minimum of 129 participants. In Study 2, there were between 59 and 80 participants who responded to each of the six BD-VAS variants. Therefore, the BD-VAS variants in Study 2 may not have been strongly correlated with self-esteem, sociocultural pressures to be thin, appearance comparison, and body mass index because the sample sizes per condition were too small to accurately estimate the associations.

Another limitation of this study is that I only tested female participants, and thus not much is known about how the BD-VAS variants perform with male respondents. The rationale for recruiting only female participants stemmed from both theoretical and practical considerations. Given the prevalence of body dissatisfaction among women (Frederick, Sandhu, Morse, & Swami, 2016; Fallon, Harris, & Johnson, 2014), they are often the subjects in body image research. Because of that, I wanted to provide practical scale recommendations for the most studied population in the field. From a practical standpoint, the addition of male participants would have complicated the experimental design. Men and women's experiences of body dissatisfaction are different; while women strive to be thin and tend to experience dissatisfaction with their weight and appearance, men are more likely to be driven to become leaner and more muscular (McCabe & Ricciardelli, 2004; Ridgeway & Tylka, 2005). Further, the BD-VASs that have been previously used with male participants focused on areas that affect men's body satisfaction, such as body shape and size (Humphreys & Paxton, 2004; Hargreaves & Tiggemann, 2009), muscle tone (Hargreaves & Tiggemann, 2009; Mulgrew & Volcevski-Kostas, 2012; Mulgrew, Volcevski-Kostas, & Rendall, 2014), and more specific body parts like their chest, arms, and abdomen (Mulgrew, Volcevski-Kostas, & Rendall, 2014; Agliata & Tantleff-Dunn, 2004). Therefore, administering the BD-VASs used in this study to male participants may not be appropriate or sufficiently reflective of men's experiences of body dissatisfaction. As a result, however, the psychometric value of BD-VASs measuring men's body dissatisfaction is currently unknown.

Conclusion and Future Directions

The purpose of this thesis was to review the different variations of the BD-VAS in the literature and explore whether common changes to the BD-VAS affect the scale's reliability and validity. Study 1 uncovered numerous variations of the BD-VASs across studies, with many of the BD-VAS modifications centering around its scale properties (e.g., anchors, polarity, item analysis). However, Study 2 showed that these changes may not have as much of an impact on participants' responses as anticipated. The BD-VAS variants' score distributions were relatively consistent and showed similar psychometric properties (e.g., internal consistency, test-retest reliability). This means that despite the variability in BD-VASs in the literature, scores on the BD-VAS and study results could be comparable across studies. The comparability of studies is imperative for the reproducibility of study findings and the further development of theory or applications of those findings.

However, even though there may not have been any apparent differences between the BD-VAS scores and their relationships with other variables, it does not mean that researchers should continue to contribute to the variability in BD-VASs. By changing the properties of the BD-VAS without proper justification or validation, we cannot know how the scale will perform relative to its counterparts.

Future research should further examine the smallest noticeable changes in body dissatisfaction using BD-VAS. Based on the findings for the BD-VASs' stability, we now know that for people who were not exposed to any stimuli thought likely to affect their body satisfaction, scores on the BD-VAS could fluctuate by around 16.77 points over a five-minute period. But we do not know by how many points a BD-VAS score would need to change to represent a small, moderate, or large change in body dissatisfaction. Additionally, researchers may also want to provide additional evidence for the validity of the BD-VAS. Particularly, running factor analyses on a pool of BD-VAS items could guarantee that the developed or modified BD-VAS contains the appropriate dimensions of body satisfaction. Also, testing the measurement equivalence of BD-VASs will ensure that BD-VAS values and means are consistent between groups that should not theoretically differ in body satisfaction, for example, people across different age groups.

Table 1

Summary and Frequency of the Items Used in the Body Dissatisfaction Visual Analogue Scale

Items	Number of Studies	Number of Scale Variants
Appearance		
Overall appearance	25	16
Appearance	16	8
Physical appearance	10	7
Overall physical appearance	2	2
Appearance compared to others	1	1
Appearance compared to self	1	1
Looks in comparison to usual self	1	1
Body		
Body	22	17
Body shape	12	11
Body shape and size ^a	7	5
Chest	2	2
Abdomen	1	1
Arms	1	1
Biceps	1	1
Body type	1	1
Body size/shape	1	1
Calf muscles	1	1
Shoulders	1	1
Stomach	1	1
Like body shape	1	1
Want to change body shape	1	1
Weight		
Weight	39	23
Fat	5	4

Body weight	1	1
Attractiveness		
Attractive	6	4
Physical attractiveness	3	3
Attractiveness compared to others	1	1
Feelings of attractiveness	1	1
Multiple		
Weight/Size	6	5
Weight and shape	2	2
Face		
Facial appearance	4	2
Face	1	1
Facial features	1	1
Hair	1	1
Muscle tone		
Muscle tone	2	2
Muscle tone and definition	1	1
Physical fitness	1	1

^aIncludes both “Body shape and size” and “Body size and shape” variations.

Table 2

Summary of the Variability in Scale Properties of the Body Dissatisfaction Visual Analogue Scale

Scale Properties	Number of Studies	Number of Scale Variants
Anchors		
None – Very much	19	7
None – Extreme	9	6
Not at all – Very much	9	5
<i>Missing</i>	6	4
Extremely dissatisfied – Extremely satisfied	5	4
No dissatisfaction – Extreme dissatisfaction	5	3
Not at all satisfied – Very satisfied	4	3
Very dissatisfied – Very satisfied	4	2
Complete dissatisfaction – Complete satisfaction	3	2
Dissatisfaction – Satisfaction	2	2
No <i>construct</i> dissatisfaction – Extreme <i>construct</i> dissatisfaction	2	2
No <i>construct</i> dissatisfaction – Worst <i>construct</i> dissatisfaction	2	2
Not at all – Very much so	2	2
Not at all dissatisfied – Very dissatisfied	2	1
Dissatisfied – Satisfied	1	1
Extremely unsatisfied – Extremely satisfied	1	1
<i>Missing</i> – Very satisfied with my construct	1	1
No – Extreme	1	1
No <i>construct</i> dissatisfaction – Worst possible <i>construct</i> dissatisfaction	1	1
No dissatisfaction – Extreme satisfaction	1	1
None – A lot	1	1
None or no feelings – Extreme feelings	1	1

Not at all – Completely	1	1
Not at all – Very	1	1
Not at all dissatisfied – Extremely dissatisfied	1	1
Not at all satisfied with my <i>construct</i> – Very satisfied with my <i>construct</i>	1	1
Unsatisfied – Satisfied	1	1
Scale Polarity		
Dissatisfaction	37	23
Satisfaction	19	13
Multiple	17	13
Bipolar	14	12
Item Analysis		
Average	44	27
None ^a	28	21
Individual	14	12
Multiple	1	1

^aOne-scale items cannot have further item analysis

Table 3

Demographic information including sample size, mean age, body mass index (BMI), body dissatisfaction (BASS), and ethnicity per condition.

	Dissatisfied (Very)	Dissatisfied (Extremely)	Satisfied (Very)	Satisfied (Extremely)	Bipolar (Very)	Bipolar (Extremely)
<i>N</i>	75	68	59	70	61	80
Age (<i>SD</i>)	19.04 (1.31)	20.16 (5.25)	20.17 (3.33)	20.49 (5.35)	20.18 (3.25)	19.46 (3.44)
BMI (<i>SD</i>)	22.93 (3.61)	23.43 (5.05)	23.96 (5.68)	22.29 (4.69)	22.46 (3.68)	22.29 (3.74)
BASS (<i>SD</i>)	3.07 (0.67)	3.05 (0.75)	3.08 (0.67)	3.21 (0.68)	3.10 (0.70)	3.17 (0.79)
Ethnic background						
White/European- Decent	26	21	16	25	24	32
East Asian	7	8	9	9	10	9
Black/African American	6	5	5	4	4	2
Middle Eastern	0	3	0	3	2	6
South Asian	14	13	11	15	4	15
Southeast Asian	6	3	6	7	5	4
Hispanic/Latina	3	1	3	0	6	2
West Indian	0	2	1	3	3	2
First Nations	0	0	1	0	0	0
Multi-Ethnic	3	7	6	4	3	5

Unidentified	0	5	1	0	0	3
--------------	---	---	---	---	---	---

Table 4

Distribution properties and reliability of BD-VAS scores.

	Satisfied (Very)	Satisfied (Extremely)	Dissatisfied (Very)	Dissatisfied (Extremely)	Bipolar (Very)	Bipolar (Extremely)
<i>N</i>	59	70	75	68	61	80
Time 1 distribution properties ^a						
<i>M</i> (<i>SD</i>)	54.93 (23.58)	54.13 (21.19)	46.28 (25.75)	48.37 (25.16)	51.03 (24.55)	57.77 (21.41)
Median	61	54.5	47	48	52	58
Skew	-0.42	-0.38	0.33	0.12	-0.08	-0.14
Kurtosis	-0.87	-0.23	-0.65	-0.67	-0.87	-0.56
Internal	0.78	0.74	0.83	0.78	0.84	0.78
Consistency						
Time 2 distribution properties ^a						
<i>M</i> (<i>SD</i>)	55.92 (24.38)	55.35 (21.92)	46.35 (28.12)	50.77 (24.31)	51.52 (23.71)	53.93 (22.44)
Median	59.75	56	48	48.5	52	52.5
Skew	-0.34	-0.25	0.27	0.03	-0.03	0.03
Kurtosis	-1.09	-0.44	-0.94	-0.51	-0.75	-1.10
Internal	0.85	0.83	0.89	0.85	0.82	0.85
Consistency						
Test-Retest						

Appearance	0.92	0.93	0.86	0.90	0.88	0.84
Weight	0.97	0.94	0.93	0.93	0.85	0.89
Body	0.97	0.97	0.91	0.94	0.90	0.94
Dissatisfaction (Average)						
Stability ^a						
$M_{\text{difference}}$	1.04	1.19	-2.54	0.62	0.24	-1.33
Lower 95% limit of agreement	-11.62	-9.67	-24.77	-15.70	-19.97	-16.19
Upper 95% limit of agreement	13.69	12.04	19.70	16.94	20.46	13.53
Repeatability coefficient	+/-12.65	+/-10.85	+/-22.23	+/-16.32	+/-20.21	+/-14.86

^aStatistics are based on composite body dissatisfaction scores.

Table 5

Intercorrelations between all the measures used to provide convergent validity evidence for the BD-VAS.

	1	2	3	4	5	6	7	8
1. BASS	—							
2. RSES	0.66***	—						
3. SATAQ-Thin	-0.35***	-0.31***	—					
4. SATAQ-Family	-0.39***	-0.20***	0.33***	—				
5. SATAQ-Peers	-0.37***	-0.28***	0.34***	0.34***	—			
6. SATAQ-Media	-0.44***	-0.33***	0.51***	0.27***	0.37***	—		
7. PACS	-0.52***	-0.42***	0.47***	0.36***	0.43***	0.56***	—	
8. BMI	-0.23***	0.05	0.09	0.39***	0.14**	0.23***	0.18***	—

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 6

Intercorrelations between the six versions of the BD-VAS and BASS, RSES, SATAQ-Thin, SATAQ subscales, PACS, and BMI.

	Satisfied (Very)	Satisfied (Extremely)	Dissatisfied (Very)	Dissatisfied (Extremely)	Bipolar (Very)	Bipolar (Extremely)
BASS	0.81***	0.67***	-0.69***	-0.68***	0.79***	0.79***
RSES	0.38**	0.36**	-0.59***	-0.53***	0.50***	0.31**
SATAQ-Thin	-0.42***	-0.16	0.32**	0.49***	-0.21	-0.32**
SATAQ-Family	-0.19	-0.25	0.47***	0.51***	-0.32*	-0.29*
SATAQ-Peers	-0.24	-0.20	0.39***	0.07	-0.27	-0.20
SATAQ-Media	-0.49***	-0.25	0.27*	0.32*	-0.39**	-0.34**
PACS	-0.58***	-0.23	0.45***	0.45***	-0.46***	-0.49***
BMI	-0.39*	-0.23	0.33*	0.24	-0.41**	-0.31*

* $p < .05$, ** $p < .01$, *** $p < .001$

How do you feel about your body right now?

Figure 1. An example of a body dissatisfaction VAS.

Figure 2. A frequency distribution of the number of BD-VASs variants that have between 1 and 7 items and the number of studies in which they have been used.

Figure 3. The score distributions of the weight, appearance, and the composite body dissatisfaction BD-VAS scores per BD-VAS variant. The plotted vertical lines represent the median values of the distributions.

Figure 4. A Bland-Altman plot that represents the test-retest agreement of BD-VAS scores at time 1 and time 2. The mean difference between the BD-VAS scores is plotted in red and the upper and lower 95% limits of agreement are plotted in blue. The dashed lines represent 95% confidence intervals of those metrics.

Appendices

Appendix A: Summary of All the Available Body Dissatisfaction Visual Analogue Scale Variants and Their Demographic, Scale, and Psychometric Properties

Items	Item Analysis	Direction	Anchors	References	Samples Used	Reliability and Validity Evidence
One-item scales						
Body	NA	Sat	Not at all – Very much	Anschutz et al. (2012)	N = 51; M_{age} = 7.41; %C = N/A; %F = 100; Country: The Netherlands	-
			Not at all satisfied – Very satisfied	Durkin & Paxton (2002)	N = 241; M_{age} = 14.25; %C = N/A; %F = 100; Country: Australia	<i>r</i> = .80 (EDI-BDS; Eating disorder) <i>r</i> = -.62 (FRS; body dissatisfaction) <i>r</i> = -.59 (PASAS; appearance anxiety)
				Durkin, Paxton, & Sorbello (2007)	N = 417; M_{age} = 15.49; %C = N/A; %F = 100; Country: Australia	-
		Dis	Not at all dissatisfied – Extremely dissatisfied	Homan (2012)	N = 92; M_{age} = 19.20; %C = 95; %F = 100; Country: USA	-
			No <i>construct</i> dissatisfaction – Worst possible <i>construct</i> dissatisfaction	DeBraganza & Hausenblas (2010)	N = 61; M_{age} = 19.73; %C = 50.82; %F = 100; Country: USA	-

			No <i>construct</i> dissatisfaction – Worst <i>construct</i> dissatisfaction	Fallon & Hausenblas (2005)	<i>N</i> = 63; <i>M</i> _{age} = 19.75; % <i>C</i> = 100; % <i>F</i> = 100; Country: USA	-
				Hausenblas et al. (2003)	<i>N</i> = 36; <i>M</i> _{age} = 20.80; % <i>C</i> = 88.89; % <i>F</i> = 0; Country: USA	-
	Bipolar		Complete dissatisfaction – Complete satisfaction	Baldwin & Courneya (1997)	<i>N</i> = 64; <i>M</i> _{age} = 51; % <i>C</i> = N/A; % <i>F</i> = 100; Country: Canada	<i>r</i> = .57 (RSE; Self-esteem)
				Kraus (1999)	<i>N</i> = 61; <i>M</i> _{age} = 52.98; % <i>C</i> = 94; % <i>F</i> = 100; Country: USA	<i>r</i> = .51 (BIS; body image)
				Mock (1993)	<i>N</i> = 257; <i>M</i> _{age} = 52; % <i>C</i> = N/A; % <i>F</i> = 100; Country: USA	<i>r</i> = .51 (BIS; body image)
			Dissatisfaction – Satisfaction	Benton & Karazsia (2015)	<i>N</i> = 366; <i>M</i> _{age} = 25.50; % <i>C</i> = 71.9; % <i>F</i> = 100; Country: USA	-
			Very dissatisfied – Very satisfied	Cho & Lee (2013)	<i>N</i> = 88; <i>M</i> _{age} = 21.79; % <i>C</i> = 0; % <i>F</i> = 48.86; Country: South Korea	-
			Extremely dissatisfied – Extremely satisfied	Johnson et al. (2013) – Study 1	<i>N</i> = 432; <i>M</i> _{age} = 27.35; % <i>C</i> = N/A; % <i>F</i> = 50.23; Country: USA	-
Body shape	NA	Sat	Not at all – Very much	O’connor et al. (1996)	<i>N</i> = 22; <i>M</i> _{age} = 36.6; % <i>C</i> = N/A; % <i>F</i> = 100; Country: USA	-
		Dis	Not at all – Very	Champion & Furnham (1999)	<i>N</i> = 203; <i>M</i> _{age} = 13.91; % <i>C</i> = 70; % <i>F</i> = 100; Country: UK	-

Body shape and size	NA	Sat	Not at all – Very much	Prichard & Tiggemann (2012)	N = 184; M_{age} = 18.90; %C = 100; %F = 100; Country: Australia	-
				Thøgersen-Ntoumani et al. (2011)	N = 93; M_{age} = 19.40; %C = 88.17; %F = 100; Country: UK	-
Body weight	NA	Dis	No dissatisfaction – Extreme dissatisfaction	Swami & Smith (2012)	N = 120; M_{age} = 20.63; %C = 100; %F = 100; Country: UK	-
Overall appearance	NA	Sat	None – Extreme	Herbozo & Thompson (2010)	N = 146; M_{age} = 20.28; %C = 56.8; %F = 100; Country: USA	-
				van der Berg & Thompson (2007) – Pilot	N = 98; M_{age} = 21.10; %C = 56; %F = 100; Country: USA	$r = .78$ (BISS; body dissatisfaction)
				van der Berg & Thompson (2007) – Main study	N = 212; M_{age} = 20.47; %C = 58; %F = 100; Country: USA	-
		Dis	Unsatisfied – Satisfied	Mulgrew et al. (2014)	N = 125; M_{age} = 29.51; %C = 93.6; %F = 0; Country: Australia	-
			None – Extreme	Williams et al. (2014)	N = 56; M_{age} = 19.71; %C = 92.9; %F = 100; Country: USA	-
			Dissatisfied – Satisfied	Want, Vickers, & Amos (2009)	N = 76; M_{age} = 20; %C = N/A; %F = 100; Country: Canada	-

Overall physical appearance	NA	Dis	No dissatisfaction – Extreme dissatisfaction	Swami (2011)	N = 82; M_{age} = 24.89; %C = 89; %F = 47.56; Country : UK	<i>r</i> = .43 (PASTAS; physical appearance anxiety)
		Bipolar	No dissatisfaction – Extreme satisfaction	Swami (2016) – Study 2	N = 37; M_{age} = 20.19; %C = 89.2; %F = 100; Country : UK	-
Physical appearance	NA	Bipolar	Extremely dissatisfied – Extremely satisfied	Kim & Park (2016)	N = 116; M_{age} = 23; %C = 38.3; %F = 100; Country : USA	-
Weight/Size	NA	Dis	No <i>construct</i> dissatisfaction – Extreme <i>construct</i> dissatisfaction	Dalley, Buunk & Umit (2009)	N = 172; M_{age} = 21.21; %C = N/A; %F = 100; Country : Unclear	-
Two-item scales						
Appearance Weight	Ind	Sat	Not at all – Very much	Atkinson & Wade (2012) – Study 2	N = 79; M_{age} = 23.59; %C = N/A; %F = 100; Country : Australia	-
				Wade, George, & Atkinson (2009)	N = 100; M_{age} = 24.38; %C = 100; %F = 100; Country : Australia	-
		Dis		Lindeman & Stark (1999)	N = 147; M_{age} = 25.85; %C = N/A; %F = 100; Country : Finland	-
	Avg	Dis	None – A lot	Tiggemann, Slater, & Smyth (2014)	N = 224; M_{age} = 20.54; %C = 90; %F = 100; Country : Australia	α = .81 (internal consistency)

None – Very Much	Bury, Tiggemann & Slater (2016a)	N = 378; M_{age} = 20.10; %C = 82.5; %F = 100; Country : Australia	α = .79 (internal consistency-pre) α = .88 (internal consistency-post)
	Harper & Tiggemann (2008)	N = 90; M_{age} = 20.48; %C = N/A; %F = 100; Country : Australia	α = .80 (internal consistency) r = .74 (PASTAS; appearance anxiety)
	Slater et al. (2012)	N = 102; M_{age} = 20.31; %C = 94; %F = 100; Country : Australia	α = .81 (internal consistency-pre) α = .80 (internal consistency-post) r = .43 (SATAQ-3; internalization of thin ideal)
	Tiggemann et al. (2017)	N = 320; M_{age} = 20.25; %C = N/A; %F = 100; Country : Australia	α = .82 (internal consistency-pre) α = .87 (internal consistency-post)
	Tiggemann & McGill (2004)	N = 126; M_{age} = 20.6; %C = N/A; %F = 100; Country : Australia	r = .38 (state appearance comparisons) r = .44 (PACS; trait appearance comparisons) r = .46 (SATAQ; internalization of thin ideal)
	Tiggemann & Zaccardo (2015)	N = 130; M_{age} = 19.91; %C = N/A; %F = 100; Country : Australia	α = .86 (internal consistency)
	Tiggemann et al. (2013) – Study 1	N = 120; M_{age} = 20.90; %C = 95.80; %F = 100; Country : Australia	α = .84 (internal consistency)
	Tiggemann et al. (2013) – Study 2	N = 114; M_{age} = 20.00;	-

					%C = 94.70; %F = 100; Country: Australia	
Body Muscle tone	Ind	Sat	Not at all – Very much	Mulgrew & Volcevski-Kostas (2012)	N = 90; M _{age} = 26.55; %C = 78; %F = 0; Country: Australia	-
Body Weight	Ind	Sat	Not at all satisfied – Very satisfied	Fett et al. (2009)	N = 77; M _{age} = 21.92; %C = N/A; %F = 100; Country: UK	-
				Geschwind et al. (2008)	N = 60; M _{age} = 18.75; %C = N/A; %F = 100; Country: Unclear	r = .80 (internal consistency- pre)
			Not at all satisfied with my <i>construct</i> – Very satisfied with my <i>construct</i>	Lattimore et al. (2010)	N = 123; M _{age} = 29.16; %C = 93.56; %F = 100; Country: UK	r _{BODY} = -.57 (BSQ-16; weight and shape dissatisfaction - breakfast) r _{BODY} = -.63 (BSQ-16; weight and shape dissatisfaction – evening meal) r _{WEIGHT} = -.60 (BSQ-16; weight and shape dissatisfaction - breakfast) r _{WEIGHT} = -.65 (BSQ-16; weight and shape dissatisfaction – evening meal)
	Avg	Bipolar	Extremely dissatisfied – Extremely satisfied	Smeets et al. (2010)	N = 79; M _{age} = 19.4; %C = N/A; %F = 100; Country: The Netherlands	-
		Sat	Missing – Very satisfied with my <i>construct</i>	Lattimore & Hutchinson (2010)	N = 46; M _{age} = 36.7; %C = 100; %F = 100; Country: UK	r = .85 - .96 (inter-item correlations on each day)

						$r = .85 - .94$ (inter-item correlations between each day's <i>average</i> rating) $r = -.40$ (BSQ-16; weight and shape dissatisfaction)
						Loaded on the same factor (factor loading $> .73$; accounted for 54.43% of total variance) $r = .71$ (EDI-BD; trait body dissatisfaction)
Body shape Weight	Dis		None or no feelings – Extreme feelings	Mask & Blanchard (2011)	$N = 138$; $M_{age} = 19.27$; $\%C = 75.36$; $\%F = 100$; Country: Canada	
	Ind	Bipolar	Extremely unsatisfied – Extremely satisfied	Haedt-Matt et al. (2012)	$N = 45$; $M_{age} = 20.03$; $\%C = 90.5$; $\%F = 100$; Country: USA	-
	Avg	Dis	Not at all dissatisfied – Very dissatisfied	Fitzsimmons-Craft (2017)	$N = 232$; $M_{age} = 18.70$; $\%C = 68.90$; $\%F = 100$; Country: USA	$\alpha = .89$ (internal consistency)
				Fitzsimmons-Craft et al. (2016)	$N = 235$; $M_{age} = 18.70$; $\%C = 68.90$; $\%F = 100$; Country: USA	$\alpha = .80$ (test-retest reliability) $\alpha = .92$ (internal consistency)
			<i>Missing</i>	Roehrig, Thompson, & Cafri (2008)	$N = 139$; $M_{age} = 20.63$; $\%C = 49.6$; $\%F = 100$; Country: USA	$r = .91$ (internal consistency)
		Bipolar	<i>Missing</i>	Harney & Bardone-Cone (2014)	$N = 146$; $M_{age} = 18.83$; $\%C = 65.8$; $\%F = 100$; Country: USA	$r = .63$ (internal consistency) $r = -.50$ (OBCS; body objectification) $r = -.75$ (EDE-Q; eating disorder)

						$r = -.37$ (EAT-26; eating attitudes)
Body shape Weight/Size	Avg	Dis	None - Extreme	Krawczyk & Thompson (2015)	$N = 437$; $M_{age} = 19.98$; $\%C = 48.3$; $\%F = 70.94$; Country: USA	$r = .88$ (internal consistency)
Body shape and size Muscle tone and definition	Ind	Sat	<i>Missing</i>	Hargreaves & Tiggemann (2009)	$N = 104$; $M_{age} = 21.4$; $\%C > 95$; $\%F = 0$; Country: Australia	$r = .22 - .71$ (internal consistency)
Like body shape Want to change body shape	Ind	Multiple	Not at all – Completely	Humphreys & Paxton (2004)	$N = 96$; $M_{age} = 15.6$; $\%C = N/A$; $\%F = 0$; Country: Australia	$r = -.39$ (inter-item correlation)
Overall appearance Weight	Ind	Dis	None – Very much	Hargreaves & Tiggemann (2002)	$N = 401$; $M_{age} = 15.8$; $\%C = N/A$; $\%F = 48.63$; Country: Australia	$r = .76$ (internal consistency-pre)
			No dissatisfaction – Extreme dissatisfaction	Etu & Gray (2010)	$N = 99$; $M_{age} = 19.88$; $\%C = 70.90$; $\%F = 100$; Country: USA	-
	Avg	Dis	None – Very much	Andrew, Tiggemann, & Clark (2015)	$N = 68$; $M_{age} = 20.03$; $\%C = 77.9$; $\%F = 100$; Country: Australia	$r = -.50$ (BAS; body appreciation) $r = .49$ (OBCS-BSS; body surveillance) $r = .34$ (state social comparison)
				Bury, Tiggemann & Slater (2016b)	$N = 120$; $M_{age} = 20.51$; $\%C = 79.2$; $\%F = 100$; Country: Australia	$\alpha = .76$ (internal consistency-pre) $\alpha = .84$ (internal consistency-post)
				Hargreaves & Tiggemann (2003a)	$N = 80$; $M_{age} = 17.2$; $\%C = N/A$; $\%F = 52.5$; Country: Australia	-

				Hargreaves & Tiggemann (2003b)	<i>N</i> = 357; <i>M</i> _{age} = 13.97; % <i>C</i> = N/A; % <i>F</i> = 44.82; Country: Australia	<i>r</i> = .81 (internal consistency)
				Heinberg & Thompson (1995)	<i>N</i> = 138; <i>M</i> _{age} = 22.88; % <i>C</i> = 100; % <i>F</i> = 100; Country: USA	<i>R</i> ² = .65 (shared variance) <i>r</i> _{WEIGHT} = .66 (EDI-BD; eating disorder) <i>r</i> _{APPEARANCE} = .76 (EDI-BD; eating disorder)
Overall appearance Weight/Size	Avg	Dis	No dissatisfaction – Extreme dissatisfaction	Befort & Rickard (2003)	<i>N</i> = 114; <i>M</i> _{age} = 19.50; % <i>C</i> = 90.4; % <i>F</i> = 51.75; Country: USA	-
				Lowery et al. (2005)	<i>N</i> = 433; <i>M</i> _{age} = 18.42; % <i>C</i> = 75.52; % <i>F</i> = 61.66; Country: USA	<i>r</i> = .78 (internal consistency) <i>r</i> = -.36 (RSE; self-esteem)
			Missing	Cattarin et al. (2000)	<i>N</i> = 180; <i>M</i> _{age} = 22.97; % <i>C</i> = 77.8; % <i>F</i> = 100; Country: USA	<i>r</i> = .83 (internal consistency-time 1) <i>r</i> = .92 (internal consistency-time 2)
Overall appearance Physical appearance	Ind	Multiple	None – Extreme	Birkeland et al. (2005)	<i>N</i> = 138; <i>M</i> _{age} = 21.88; % <i>C</i> = 67; % <i>F</i> = 100; Country: USA	<i>r</i> _{PHYSICAL} = .80 (test-retest reliability) <i>r</i> _{OVERALL} = .84 (test-retest reliability)
Physical appearance ^b	Avg	Multiple	None – Extreme	Alleva, Martijn, & Jansen (2016)	<i>N</i> = 70; <i>M</i> _{age} = 22.30; % <i>C</i> = N/A; % <i>F</i> = 100; Country: The Netherlands	<i>α</i> = .89 (internal consistency)
				Alleva, Veldhuis, & Martijn (2016)	<i>N</i> = 70; <i>M</i> _{age} = 20.61; % <i>C</i> = N/A; % <i>F</i> = 100; Country: The Netherlands	<i>α</i> = .90 (internal consistency-pre) <i>α</i> = .95 (internal consistency-post)
Three-item scales						

Appearance Body Weight	Avg	Bipolar	Extremely dissatisfied – Extremely satisfied	Svaldi & Naumann (2014)	N = 30; M _{age} = 46.34; %C = N/A; %F = 100; Country : Germany	α = .92 (internal consistency-baseline) α = .80 (internal consistency-time 1) α = .91 (internal consistency-time 2)
Appearance Facial features Weight	Avg	Dis	None – Very Much	Brown & Tiggemann (2016)	N = 138; M _{age} = 20.10; %C = 77.4; %F = 100; Country : Australia	α = .94 (internal consistency)
Body shape Physical appearance Weight/size	Avg	Dis	None – Extreme	Ata, Thompson, & Small (2013)	N = 342; M _{age} = 19.58; %C = 52.3; %F = 100; Country : USA	α = .93 (internal consistency-pre) α = .96 (internal consistency-post)
Body shape and size Fat^a Physically attractive^a	Avg	Sat	Not at all – Very much	Fardouly et al. (2015)	N = 112; M _{age} = 20.46; %C = 75; %F = 100; Country : UK	α = .74 (internal consistency)
		Dis	None – Very much	Tiggemann & Polivy (2010)	N = 114; M _{age} = 20.07; %C > 90; %F = 100; Country : Australia	α = .76 (internal consistency) r = .83 (pre- and post- manipulation, irrespective of condition)
				Tiggemann & Slater (2004)	N = 84; M _{age} = 20.23; %C = N/A; %F = 100; Country : Australia	r = .50 (appearance comparison)
Fat^a Overall appearance Weight and shape	Avg	Dis	<i>Missing</i>	Hargreaves & Tiggemann (2004)	N = 595; M _{age} = 14.3; %C = N/A; %F = 52.1; Country : Australia	r_s > .52 (inter-item correlations) α = .85 (internal consistency)
Facial appearance Overall appearance Weight and shape	Avg	Bipolar	Extremely dissatisfied – Extremely satisfied	Quigg & Want (2011)	N = 87; M _{age} = 19.31; %C = 100; %F = 100; Country : Canada	α = .87 (internal consistency) $r_{\text{FACIAL/WEIGHT}}$ = .59

						$r_{\text{FACIAL/APPEARANCE}} = .75$ $r_{\text{WEIGHT/APPEARANCE}} = .74$
Overall appearance Physical appearance Physical fitness	Avg	Multiple	None – Very much	Wasilenko, Kulik, & Wanic (2007)	$N = 45$; $M_{\text{age}} = \text{N/A}$; $\%C = \text{N/A}$; $\%F = 100$; Country: USA	$\alpha = .81$ (internal consistency) $r = .74$ (BPSS; body parts satisfaction)
Four-item scales						
Appearance Body shape Feelings of attractiveness Weight	Avg	Sat	Not at all – Very much so	McLean, Paxton, Wertheim (2016)	$N = 246$; $M_{\text{age}} = 13.11$; $\%C = \text{N/A}$; $\%F = 100$; Country: Australia	$\alpha = .94$ (internal consistency-pre) $\alpha = .96$ (internal consistency-post) $r = .77$ (EDI; eating disorder)
Attractive ^a Body shape Body type Fat ^a	Avg	Sat	None – Very Much	Bourn et al. (2015)	$N = 99$; $M_{\text{age}} = 19.35$; $\%C = 100$; $\%F = 100$; Country: Australia	-
Attractive ^a Facial appearance Overall appearance Weight	Avg	Bipolar	Very dissatisfied – Very satisfied	Want & Saiphoo (2017)	$N = 151$; $M_{\text{age}} = 19.05$; $\%C = 45.70$; $\%F = 100$; Country: Canada	$\alpha = .85$ (internal consistency-pre)
				Want et al. (2015) – Study 1	$N = 116$; $M_{\text{age}} = 19.39$; $\%C = 46.55$; $\%F = 100$; Country: Canada	$\alpha = .87$ (internal consistency-pre)
				Want et al. (2015) – Study 2	$N = 177$; $M_{\text{age}} = 18.90$; $\%C = 35.59$; $\%F = 100$; Country: Canada	$\alpha = .85$ (internal consistency-pre) $\alpha = .90$ (internal consistency-post)
Five-item scales						
Appearance Attractive ^a Body Body shape ^a	Avg	Multiple	Not at all – Very much so	Richardson & Paxton (2010)	$N = 194$; $M_{\text{age}} = 12.33$; $\%C = \text{N/A}$; $\%F = 100$; Country: Australia	$\alpha = .91$ (internal consistency)

Weight						
Six-item scales						
Appearance-Others^a Appearance-Self^a Attractive^a Body shape and size Physical appearance Weight	Avg	Bipolar	Dissatisfaction – Satisfaction	Alleva et al. (2014) – Study 1	N = 118; M_{age} = 20.84; %C = N/A; %F = 51.82; Country: The Netherlands	r = .52 (RSE; Self-esteem) r = -.71 (EDE-Q; Eating disorder)
Attractiveness-compared to others^a Body size/shape Looks-compared to usual Overall appearance Physical attractiveness^a Weight	Avg	Multiple	<i>Missing</i>	Lavin & Cash (2001)	N = 66; Median_{age} = 20; %C = 58; %F = 100; Country: USA	α = .86 (internal consistency)
Overall appearance Muscle tone Arms Chest Shoulders Stomach	Multiple ^c	Sat	Not at all – Very much	Mulgrew, Volcevski-Kostas, & Rendell (2014)	N = 180; M_{age} = 14.43; %C = 79; %F = 0; Country: Australia	rs = .45 - .77 (inter-item correlations)
Seven-item scales						
Abdomen Biceps Calf muscles Chest Face Hair Physical appearance	Avg	Dis	No – Extreme	Agliata & Tantleff-Dunn (2004)	N = 158; M_{age} = 21.30; %C = 66; %F = 0; Country: USA	α = .89 (internal consistency)

^aThe following scale item did not measure *satisfaction*, and thus was not measured using the reported anchors. Please refer to the original paper to find the anchors used for these items

^bIn this particular scale, the scale item (i.e., physical appearance) appeared twice, with the first item measuring *dissatisfaction* with their physical appearance and the other measuring *satisfaction*.

^cIn this particular scale, the “Overall appearance” and “Muscle tone” items were analyzed individually and the “Arms”, “Chest”, “Shoulders”, and “Stomach” items were averaged out to create an aggregate upper body dissatisfaction score.

Appendix B: The Six Body Dissatisfaction Visual Analogue Scales Assigned to Participants

Please note that there are six variations of these two questions. Each participant would see only a single pair of questions. Either:

1. Dissatisfaction (Very):

Right now, how dissatisfied do you feel with your overall appearance?

Right now, how dissatisfied do you feel with your weight?

Or

2. Dissatisfaction (Extreme):

Right now, how dissatisfied do you feel with your overall appearance?

Right now, how dissatisfied do you feel with your weight?

Or

3. Satisfaction (Very):

Right now, how satisfied do you feel with your overall appearance?

Right now, how satisfied do you feel with your weight?

Or

4. Satisfaction (Extreme):

Right now, how satisfied do you feel with your overall appearance?

Right now, how satisfied do you feel with your weight?

Or

5. Bipolar (Very):

Right now, how dissatisfied do you feel with your overall appearance?

Right now, how dissatisfied do you feel with your weight?

Or

6. Bipolar (Extreme):

Right now, how dissatisfied do you feel with your overall appearance?

Right now, how dissatisfied do you feel with your weight?

References

- *Agliata, D., & Tantleff-Dunn, S. (2004). The impact of media exposure on males' body image. *Journal of Social and Clinical Psychology, 23*(1), 7-22.
- Aguinis, H., Pierce, C. A., & Culpepper, S. A. (2009). Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods, 12*(4), 623-652.
- Ahearn, E. P. (1997). The use of visual analog scales in mood disorders: a critical review. *Journal of Psychiatric Research, 31*(5), 569-579.
- *Alleva, J. M., Martijn, C., & Jansen, A. (2016). Covariation bias in women with a negative body evaluation: How is it expressed and can it be diminished?. *Journal of Behavior Therapy and Experimental Psychiatry, 50*, 33-39.
- *Alleva, J. M., Martijn, C., Jansen, A., & Nederkoorn, C. (2014). Body language: Affecting body satisfaction by describing the body in functionality terms. *Psychology of Women Quarterly, 38*(2), 181-196.
- *Alleva, J. M., Veldhuis, J., & Martijn, C. (2016). A pilot study investigating whether focusing on body functionality can protect women from the potential negative effects of viewing thin-ideal media images. *Body Image, 17*, 10-13.
- *Andrew, R., Tiggemann, M., & Clark, L. (2015). The protective role of body appreciation against media-induced body dissatisfaction. *Body Image, 15*, 98-104.
- *Anschutz, D. J., Engels, R. C., & Van Strien, T. (2012). Increased body satisfaction after exposure to thin ideal children's television in young girls showing thin ideal internalisation. *Psychology & Health, 27*(5), 603-617.
- *Ata, R. N., Thompson, J. K., & Small, B. J. (2013). Effects of exposure to thin-ideal media images on body dissatisfaction: Testing the inclusion of a disclaimer versus warning label. *Body Image, 10*(4), 472-480.

- *Atkinson, M. J., & Wade, T. D. (2012). Impact of metacognitive acceptance on body dissatisfaction and negative affect: engagement and efficacy. *Journal of Consulting and Clinical Psychology, 80*(3), 416.
- Australia Bureau of Statistics. (2017). *Census of Population and Housing: Reflecting Australia – Stories from the Census, 2016*. Retrieved from <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/2071.0~2016~Main%20Features~Cultural%20Diversity%20Article~60>
- *Baldwin, M. K., & Courneya, K. S. (1997). Exercise and self-esteem in breast cancer survivors: An application of the exercise and self-esteem model. *Journal of Sport & Exercise Psychology, 19*(4), 347.
- Barry, A. E., Chaney, B., Piazza-Gardner, A. K., & Chavarria, E. A. (2014). Validity and reliability reporting practices in the field of health education and behavior: A review of seven journals. *Health Education & Behavior, 41*(1), 12-18.
- Bartoshuk, L. M., Duffy, V. B., Fast, K., Green, B. G., Prutkin, J., & Snyder, D. J. (2003). Labeled scales (eg, category, Likert, VAS) and invalid across-group comparisons: what we have learned from genetic variation in taste. *Food Quality and Preference, 14*(2), 125-138.
- Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. M. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research, 10*(7), 571-578.
- *Befort, C. A., & Rickard, K. M. (2003). Gender similarities in response to figure-size feedback in a selected nonclinical population. *Sex Roles, 49*(1-2), 71-80.
- Benedetto, M. D., & Sheehan, M. (2014). Evaluation of the Cardiac Depression Visual Analogue Scale in a medical and non-medical sample. *Psychology, health & medicine, 19*(1), 93-102.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological), 57*(1), 289-300.

- *Benton, C., & Karazsia, B. T. (2015). The effect of thin and muscular images on women's body satisfaction. *Body Image*, 13, 22-27.
- Beren, S. E., Hayden, H. A., Wilfley, D. E., & Grilo, C. M. (1996). The influence of sexual orientation on body dissatisfaction in adult men and women. *International Journal of Eating Disorders*, 20, 135-141.
- Bijur, P. E., Silver, W., & Gallagher, E. J. (2001). Reliability of the visual analog scale for measurement of acute pain. *Academic Emergency Medicine*, 8(12), 1153-1157.
- *Birkeland, R., Thompson, J. K., Herbozo, S., Roehrig, M., Cafri, G., & Van den Berg, P. (2005). Media exposure, mood, and body image dissatisfaction: An experimental test of person versus product priming. *Body Image*, 2(1), 53-61.
- Bland, J. M., & Altman, D. G. (2003). Applying the right statistics: analyses of measurement studies. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 22(1), 85-93.
- Blond, A. (2008). Impacts of exposure to images of ideal bodies on male body dissatisfaction: A review. *Body Image*, 5(3), 244-250.
- *Bourn, R., Prichard, I., Hutchinson, A. D., & Wilson, C. (2015). Watching reality weight loss TV. The effects on body satisfaction, mood, and snack food consumption. *Appetite*, 91, 351-356.
- Brown, T. A., Cash, T. F., & Milulka, P. J. (1990). Attitudinal body image assessment: Factor analysis of the body self-relations questionnaire. *Journal of Personality Assessment*, 55, 135-144.
- *Brown, Z., & Tiggemann, M. (2016). Attractive celebrity and peer images on Instagram: Effect on women's mood and body image. *Body Image*, 19, 37-43.
- Buote, V. M., Wilson, A. E., Strahan, E. J., Gazzola, S. B., & Papps, F. (2011). Setting the bar: Divergent sociocultural norms for women's and men's ideal appearance in real-world contexts. *Body Image*, 8(4), 322-334.

- *Bury, B., Tiggemann, M., & Slater, A. (2016a). The effect of digital alteration disclaimer labels on social comparison and body image: Instructions and individual differences. *Body Image, 17*, 136-142.
- *Bury, B., Tiggemann, M., & Slater, A. (2016b). Disclaimer labels on fashion magazine advertisements: Impact on visual attention and relationship with body dissatisfaction. *Body Image, 16*, 1-9.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*(3), 401.
- Carlsson, A. M. (1983). Assessment of chronic pain. I. aspects of the reliability and validity of the visual analogue scale. *Pain, 16*(1), 87-101.
- Cash, T. F. (1994). Body-image attitudes: Evaluation, investment, and affect. *Perceptual and Motor Skills, 78*, 1168-1170.
- Cash, T. F., Fleming, E. C., Alindogan, J., Steadman, L., & Whitehead, A. (2002). Beyond body image as a trait: The development and validation of the Body Image States Scale. *Eating Disorders, 10*(2), 103-113.
- Cash, T.F. (2018). *The Multidimensional Body-Self Relations Questionnaire users' manual (fourth revision)*. Available from the author at www.body-images.com.
- *Cattarin, J. A., Thompson, J. K., Thomas, C., & Williams, R. (2000). Body image, mood, and televised images of attractiveness: The role of social comparison. *Journal of Social and Clinical Psychology, 19*(2), 220-239.
- *Champion, H., & Furnham, A. (1999). The effect of the media on body satisfaction in adolescent girls. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association, 7*(3), 213-228.
- *Cho, A., & Lee, J. H. (2013). Body dissatisfaction levels and gender differences in attentional biases toward idealized bodies. *Body Image, 10*(1), 95-102.

- *Dalley, S. E., Buunk, A. P., & Umit, T. (2009). Female body dissatisfaction after exposure to overweight and thin media images: The role of body mass index and neuroticism. *Personality and Individual Differences*, 47(1), 47-51.
- Davern, M., & Cummins, R. A. (2006). Is life dissatisfaction the opposite of life satisfaction?. *Australian Journal of Psychology*, 58(1), 1-7.
- De Boer, A. G. E. M., Van Lanschot, J. J. B., Stalmeier, P. F. M., Van Sandick, J. W., Hulscher, J. B., De Haes, J. C. J. M., & Sprangers, M. A. G. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life?. *Quality of Life Research*, 13(2), 311-320.
- *DeBraganza, N., & Hausenblas, H. A. (2010). Media exposure of the ideal physique on women's body dissatisfaction and mood: The moderating effects of ethnicity. *Journal of Black Studies*, 40(4), 700-716.
- DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: a literature review. *Quality & Quantity*, 52(4), 1523-1559.
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.) Sage Publications, Inc.
- *Durkin, S. J., & Paxton, S. J. (2002). Predictors of vulnerability to reduced body image satisfaction and psychological wellbeing in response to exposure to idealized female media images in adolescent girls. *Journal of psychosomatic research*, 53(5), 995-1005.
- *Durkin, S. J., Paxton, S. J., & Sorbello, M. (2007). An integrative model of the impact of exposure to idealized female images on adolescent girls' body satisfaction. *Journal of Applied Social Psychology*, 37(5), 1092-1117.
- Eisinga, R. N., Grotenhuis, H. F. t., & Pelzer, B. J. (2013). The reliability of a two-item scale: Pearson, cronbach or spearman-brown? *International Journal of Public Health*, 58(4), 637-642. doi:10.1007/s00038-012-0416-3
- *Etu, S. F., & Gray, J. J. (2010). A preliminary investigation of the relationship between induced rumination and state body image dissatisfaction and anxiety. *Body Image*, 7(1), 82-85.

- Facco, E., Zanette, G., Favero, L., Bacci, C., Sivoletta, S., Cavallin, F., & Manani, G. (2011). Toward the validation of visual analogue scale for anxiety. *Anesthesia progress*, 58(1), 8-13.
- *Fallon, E. A., & Hausenblas, H. A. (2005). Media images of the “ideal” female body: Can acute exercise moderate their psychological impact?. *Body Image*, 2(1), 62-73.
- Fallon, E. A., Harris, B. S., & Johnson, P. (2014). Prevalence of body dissatisfaction among a United States adult sample. *Eating Behaviors*, 15(1), 151-158.
- *Fardouly, J., Diedrichs, P. C., Vartanian, L. R., & Halliwell, E. (2015). Social comparisons on social media: The impact of Facebook on young women's body image concerns and mood. *Body Image*, 13, 38-45.
- *Fett, A. K., Lattimore, P., Roefs, A., Geschwind, N., & Jansen, A. (2009). Food cue exposure and body image satisfaction: The moderating role of BMI and dietary restraint. *Body Image*, 6(1), 14-18.
- *Fitzsimmons-Craft, E. E. (2017). Eating disorder-related social comparison in college women's everyday lives. *International Journal of Eating Disorders*, 50(8), 893-905.
- *Fitzsimmons-Craft, E. E., Bardone-Cone, A. M., Crosby, R. D., Engel, S. G., Wonderlich, S. A., & Bulik, C. M. (2016). Mediators of the relationship between thin-ideal internalization and body dissatisfaction in the natural environment. *Body Image*, 18, 113-122.
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370-378. 10.1177/1948550617693063
- Frederick, D. A., Sandhu, G., Morse, P. J., & Swami, V. (2016). Correlates of appearance and weight satisfaction in a US national sample: Personality, attachment style, television viewing, self-esteem, and life satisfaction. *Body Image*, 17, 191-203.
- Garner, D. M., Olmstead, M. P., & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders*, 2, 15-34.

- *Geschwind, N., Roefs, A., Lattimore, P., Fett, A. K., & Jansen, A. (2008). Dietary restraint moderates the effects of food exposure on women's body and weight satisfaction. *Appetite*, 51(3), 735-738.
- Goodman, L. A. (1954). Kolmogorov-Smirnov tests for psychological research. *Psychological Bulletin*, 51(2), 160.
- Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. *Psychological Bulletin*, 134(3), 460-476. 10.1037/0033-2909.134.3.460
- Groesz, L. M., Levine, M. P., & Murnen, S. K. (2002). The effect of experimental presentation of thin media images on body satisfaction: A meta-analytic review. *International Journal of Eating Disorders*, 31(1), 1-16. 10.1002/eat.10005
- *Haedt-Matt, A. A., Zalta, A. K., Forbush, K. T., & Keel, P. K. (2012). Experimental evidence that changes in mood cause changes in body dissatisfaction among undergraduate women. *Body Image*, 9(2), 216-220.
- *Hargreaves, D., & Tiggemann, M. (2002). The effect of television commercials on mood and body dissatisfaction: The role of appearance-schema activation. *Journal of Social and Clinical Psychology*, 21(3), 287-308.
- *Hargreaves, D., & Tiggemann, M. (2003). Longer-term implications of responsiveness to 'thin-ideal' television: support for a cumulative hypothesis of body image disturbance?. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association*, 11(6), 465-477.
- *Hargreaves, D., & Tiggemann, M. (2003). The effect of "thin ideal" television commercials on body dissatisfaction and schema activation during early adolescence. *Journal of Youth and Adolescence*, 32(5), 367-373.
- *Hargreaves, D. A., & Tiggemann, M. (2004). Idealized media images and adolescent body image: "Comparing" boys and girls. *Body Image*, 1(4), 351-361.

- *Hargreaves, D. A., & Tiggemann, M. (2009). Muscular ideal media images and men's body image: Social comparison processing and individual vulnerability. *Psychology of Men & Masculinity*, 10(2), 109-119.
- *Harney, M. B., & Bardone-Cone, A. M. (2014). The influence of body dissatisfaction on set shifting ability. *Cognitive Therapy and Research*, 38(4), 439-448.
- *Harper, B., & Tiggemann, M. (2008). The effect of thin ideal media images on women's self-objectification, mood, and body image. *Sex Roles*, 58(9-10), 649-657.
- *Hausenblas, H. A., Janelle, C. M., Gardner, R. E., & Hagan, A. L. (2003). Affective responses of high and low body satisfied men to viewing physique slides. *Eating Disorders*, 11(2), 101-113.
- *Herbozo, S., & Thompson, J. K. (2010). The effects of ambiguous appearance-related feedback on body image, mood states, and intentions to use body change strategies in college women: An experimental study. *Body Image*, 7(4), 327-334.
- *Heinberg, L. J., & Thompson, J. K. (1995). Body image and televised images of thinness and attractiveness: A controlled laboratory investigation. *Journal of Social and Clinical Psychology*, 14(4), 325-338. doi:10.1521/jscp.1995.14.4.325
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and Brain Sciences*, 33(2-3), 61-83.
- *Homan, K. J. (2012). Attachment to God mitigates negative effect of media exposure on women's body image. *Psychology of Religion and Spirituality*, 4(4), 324.
- Hrabosky, J. I., Cash, T. F., Veale, D., Neziroglu, F., Soll, E. A., Garner, D. M., ... & Phillips, K. A. (2009). Multidimensional body image comparisons among patients with eating disorders, body dysmorphic disorder, and clinical controls: A multisite study. *Body Image*, 6(3), 155-163.
- *Humphreys, P., & Paxton, S. J. (2004). Impact of exposure to idealised male images on adolescent boys' body image. *Body Image*, 1(3), 253-266.

- Hussey, I., & Hughes, S. (2018, November 19). Hidden invalidity among fifteen commonly used measures in social and personality psychology. <https://doi.org/10.31234/osf.io/7rbfp>
- *Johnson, P., Fallon, E. A., Harris, B. S., & Burton, B. (2013). Body satisfaction is associated with Transtheoretical Model constructs for physical activity behavior change. *Body Image, 10*(2), 163-174.
- Keery, H., Van den Berg, P., & Thompson, J. K. (2004). An evaluation of the Tripartite Influence Model of body dissatisfaction and eating disturbance with adolescent girls. *Body Image, 1*(3), 237-251.
- Kim, H., & Damhorst, M. L. (2010). The relationship of body-related self-discrepancy to body dissatisfaction, apparel involvement, concerns with fit and size of garments, and purchase intentions in online apparel shopping. *Clothing and Textiles Research Journal, 28*(4), 239-254. doi:10.1177/0887302X10379266
- *Kim, M., & Park, W. (2016). Who is at risk on Facebook? The effects of Facebook News Feed photographs on female college students' appearance satisfaction. *The Social Science Journal, 53*(4), 427-434.
- Kjeldsen, H. B., Klausen, T. W., & Rosenberg, J. (2016). Preferred presentation of the visual analog scale for measurement of postoperative pain. *Pain Practice, 16*(8), 980-984. 10.1111/papr.12344
- *Kraus, P. L. (1999). Body image, decision making, and breast cancer treatment. *Cancer Nursing, 22*(6), 421-427.
- *Krawczyk, R., & Thompson, J. K. (2015). The effects of advertisements that sexually objectify women on state body dissatisfaction and judgments of women: The moderating roles of gender and internalization. *Body Image, 15*, 109-119.
- Kuhlmann, T., Dantlgraber, M., & Reips, U. D. (2017). Investigating measurement equivalence of visual analogue scales and Likert-type scales in Internet-based personality questionnaires. *Behavior research methods, 49*(6), 2173-2181.

- *Lattimore, P., & Hutchinson, R. (2010). Perceived calorie intake and state body-image satisfaction in women attempting weight loss: A preliminary investigation. *Body Image*, 7(1), 15-21.
- *Lattimore, P., Walton, J., Bartlett, S., Hackett, A., & Stevenson, L. (2010). Regular consumption of a cereal breakfast. Effects on mood and body image satisfaction in adult non-obese women. *Appetite*, 55(3), 512-521.
- *Lavin, M. A., & Cash, T. F. (2001). Effects of exposure to information about appearance stereotyping and discrimination on women's body images. *International Journal of Eating Disorders*, 29(1), 51-58.
- *Lindeman, M., & Stark, K. (1999). Pleasure, pursuit of health or negotiation of identity? Personality correlates of food choice motives among young and middle-aged women. *Appetite*, 33(1), 141-161.
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3(3), 635-694. doi:10.2466/pr0.1957.3.3.635
- *Lowery, S. E., Kurpius, S. E. R., Befort, C., Blanks, E. H., Sollenberger, S., Nicpon, M. F., & Huser, L. (2005). Body image, self-esteem, and health-related behaviors among male and female first year college students. *Journal of College Student Development*, 46(6), 612-623.
- Marsden, P. V., & Wright, J. D. (2010). *Handbook of survey research* (2nd ed.) Emerald.
- *Mask, L., & Blanchard, C. M. (2011). The effects of “thin ideal” media on women's body image concerns and eating-related intentions: The beneficial role of an autonomous regulation of eating behaviors. *Body Image*, 8(4), 357-365.
- Mask, L., Blanchard, C. M., & Baker, A. (2014). Do portrayals of women in action convey another ideal that women with little self-determination feel obligated to live up to? Viewing effects on body image evaluations and eating behaviors. *Appetite*, 83, 277-286.
- Mayer, J.M. (1978). *Assessment of depression*. In M.P.McReynolds (Ed.) *Advances in psychological assessment* (Vol. 4, pp. 358-425). San Francisco: Jossey-Bass.

- Mazaheri, M., & Theuns, P. (2009). Effects of varying response formats on self-ratings of life-satisfaction. *Social Indicators Research*, 90(3), 381.
- McCabe, M. P., & Ricciardelli, L. A. (2004). Body image dissatisfaction among males across the lifespan: A review of past literature. *Journal of Psychosomatic Research*, 56(6), 675-685.
- *McLean, S. A., Paxton, S. J., & Wertheim, E. H. (2016). Does media literacy mitigate risk for reduced body satisfaction following exposure to thin-ideal media?. *Journal of Youth and Adolescence*, 45(8), 1678-1695.
- Mitchell, C. (2001) Assessment and management of chronic pain in elderly people. *British Journal of Nursing*, 10, 296–304.
- *Mock, V. (1993). Body image in women treated for breast cancer. *Nursing Research*, 42(3), 153-157.
- Monk, T. H. (1989). A visual analogue scale technique to measure global vigor and affect. *Psychiatry research*, 27(1), 89-99.
- Monro, F., & Huon, G. (2005). Media-portrayed idealized images, body shame, and appearance anxiety. *International Journal of Eating Disorders*, 38(1), 85-90.
- *Mulgrew, K. E., Johnson, L. M., Lane, B. R., & Katsikitis, M. (2014). The effect of aesthetic versus process images on men's body satisfaction. *Psychology of Men & Masculinity*, 15(4), 452.
- *Mulgrew, K. E., & Volcevski-Kostas, D. (2012). Short term exposure to attractive and muscular singers in music video clips negatively affects men's body image and mood. *Body Image*, 9(4), 543-546.
- *Mulgrew, K. E., Volcevski-Kostas, D., & Rendell, P. G. (2014). The effect of music video clips on adolescent boys' body image, mood, and schema activation. *Journal of Youth and Adolescence*, 43(1), 92-103.
- Myers, T. A., & Crowther, J. H. (2009). Social comparison as a predictor of body dissatisfaction: A meta-analytic review. *Journal of Abnormal Psychology*, 118(4), 683-698.
- 10.1037/a0016763

- *O'Connor, P. J., Lewis, R. D., Kirchner, E. M., & Cook, D. B. (1996). Eating disorder symptoms in former female college gymnasts: relations with body composition. *The American journal of clinical nutrition*, 64(6), 840-843.
- Ogden, J. E., & Munday, K. (1996). The effect of the media on body satisfaction: The role of gender and size. *European Eating Disorders Review*, 4(3), 171-182.
- Ohring, R., Graber, J. A., & Brooks-Gunn, J. (2002). Girls' recurrent and concurrent body dissatisfaction: Correlates and consequences over 8 years. *International Journal of Eating Disorders*, 31(4), 404-415.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17(11), 776.
- Paxton, S. J., Neumark-Sztainer, D., Hannan, P. J., & Eisenberg, M. E. (2006). Body dissatisfaction prospectively predicts depressive mood and low self-esteem in adolescent girls and boys. *Journal of Clinical Child and Adolescent Psychology*, 35(4), 539-549.
- Peters, M., Rhodes, G., & Simmons, L. W. (2007). Contributions of the face and body to overall attractiveness. *Animal Behaviour*, 73(6), 937-942.
- Price, C. I., Curless, R. H. & Rodgers, H. (1999) Can stroke patients use visual analogue scales? *Stroke*, 30, 1357-1361.
- *Prichard, I., & Tiggemann, M. (2012). The effect of simultaneous exercise and exposure to thin-ideal music videos on women's state self-objectification, mood and body satisfaction. *Sex Roles*, 67(3-4), 201-210.
- Pujols, Y., Meston, C. M., & Seal, B. N. (2010). The association between sexual satisfaction and body image in women. *The Journal of Sexual Medicine*, 7(2), 905-916.
- Qualls, A. L., & Moss, A. D. (1996). The Degree of Congruence between Test Standards and Test Documentation within Journal Publications. *Educational and Psychological Measurement*, 56(2), 209-214.

- *Quigg, S. L., & Want, S. C. (2011). Highlighting media modifications: Can a television commercial mitigate the effects of music videos on female appearance satisfaction?. *Body Image*, 8(2), 135-142.
- Rakhkovskaya, L. M., & Warren, C. S. (2016). Sociocultural and identity predictors of body dissatisfaction in ethnically diverse college women. *Body Image*, 16, 32-40.
- Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment*, 12(3), 287.
- Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of Personality and Social Psychology*, 79(2), 286.
- *Richardson, S. M., & Paxton, S. J. (2010). An evaluation of a body image intervention based on risk factors for body dissatisfaction: A controlled study with adolescent girls. *International Journal of Eating Disorders*, 43(2), 112-122.
- Ridgeway, R. T., & Tylka, T. L. (2005). College Men's Perceptions of Ideal Body Composition and Shape. *Psychology of Men & Masculinity*, 6(3), 209.
- Rodgers, R. F., McLean, S. A., & Paxton, S. J. (2015). Longitudinal relationships among internalization of the media ideal, peer social comparison, and body dissatisfaction: Implications for the tripartite influence model. *Developmental Psychology*, 51(5), 706–713.
- Rodin, J., Silberstein, L., & Striegel-Moore, R. (1984). Women and weight: A normative discontent. *Nebraska Symposium on Motivation*, 32, 267-307.
- *Roehrig, M., Thompson, J. K., & Cafri, G. (2008). Effects of dieting-related messages on psychological and weight control variables. *International Journal of Eating Disorders*, 41(2), 164-173.
- Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.
- Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. *Journal of Applied Psychology*, 77(3), 336.

- Russell, C. J., Pinto, J. K., & Bobko, P. (1991). Appropriate moderated regression and inappropriate research strategy: A demonstration of information loss due to scale coarseness. *Applied Psychological Measurement*, 15(3), 257-266.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161.
- Schaefer, L. M., Burke, N. L., Thompson, J. K., Dedrick, R. F., Heinberg, L. J., Calogero, R. M., ... & Anderson, D. A. (2015). Development and validation of the Sociocultural Attitudes Towards Appearance Questionnaire-4 (SATAQ-4). *Psychological Assessment*, 27(1), 54.
- Schaefer, L. M., Harriger, J. A., Heinberg, L. J., Soderberg, T., & Kevin Thompson, J. (2017). Development and validation of the sociocultural attitudes towards appearance questionnaire-4-revised (SATAQ-4R). *International Journal of Eating Disorders*, 50(2), 104-117.
- Schaefer, L. M., & Thompson, J. K. (2014). The development and validation of the physical appearance comparison scale-revised (PACS-R). *Eating behaviors*, 15(2), 209-217.
- Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology*, 29(1), 65-88.
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. *Journal of Research in Personality*, 47(5), 609-612.
- Scott, J. & Huskisson, E. C. (1979) Vertical or horizontal visual analogue scales. *Annals of Rheumatic Disease*, 38, 560.
- Servián-Franco, F., Moreno-Domínguez, S., & del Paso, G. A. R. (2015). Body dissatisfaction and mirror exposure: Evidence for a dissociation between self-report and physiological responses in highly body-dissatisfied women. *PloS one*, 10(4), e0122737.
- Sharp, G., & Tiggemann, M. (2016). Educating women about normal female genital appearance variation. *Body Image*, 16, 70-78.

- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment, 31*(4), 557.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366.
- Sinclair, S. J., Blais, M. A., Gansler, D. A., Sandberg, E., Bistis, K., & LoCicero, A. (2010). Psychometric properties of the Rosenberg Self-Esteem Scale: Overall and across demographic groups living within the United States. *Evaluation & The Health Professions, 33*(1), 56-80.
- *Slater, A., Tiggemann, M., Firth, B., & Hawkins, K. (2012). Reality check: An experimental investigation of the addition of warning labels to fashion magazine images on women's mood and body dissatisfaction. *Journal of Social and Clinical Psychology, 31*(2), 105-122.
- *Smeets, E., Jansen, A., Vossen, E., Ruf, L., & Roefs, A. (2010). Feeling body dissatisfied after viewing thin-ideal pictures is mediated by self-activation. *Body Image, 7*(4), 335-340.
- Soto, C. J., & John, O. P. (2019). Optimizing the length, width, and balance of a personality scale: How do internal characteristics affect external validity?. *Psychological Assessment, 31*(4), 444.
- Steiner, M., Streiner, D. L., & Pham, B. (2005). Validation of a revised visual analog scale for premenstrual mood symptoms: results from prospective and retrospective trials. *The Canadian Journal of Psychiatry, 50*(6), 327-332.
- Stice, E., & Shaw, H. E. (2002). Role of body dissatisfaction in the onset and maintenance of eating pathology: A synthesis of research findings. *Journal of Psychosomatic Research, 53*(5), 985-993.
- *Svaldi, J., & Naumann, E. (2014). Effects of rumination and acceptance on body dissatisfaction in binge eating disorder. *European Eating Disorders Review, 22*(5), 338-345.

- *Swami, V. (2011). Marked for life? A prospective study of tattoos on appearance anxiety and dissatisfaction, perceptions of uniqueness, and self-esteem. *Body Image*, 8(3), 237-244.
- *Swami, V. (2016). Illustrating the body: Cross-sectional and prospective investigations of the impact of life drawing sessions on body image. *Psychiatry Research*, 235, 128-132.
- Swami, V., Frederick, D. A., Aavik, T., Alcalay, L., Allik, J., Anderson, D., ... & Danel, D. (2010). The attractive female body weight and female body dissatisfaction in 26 countries across 10 world regions: Results of the International Body Project I. *Personality and social psychology bulletin*, 36(3), 309-325.
- *Swami, V., & Smith, J. M. (2012). How not to feel good naked? The effects of television programs that use “real women” on female viewers' body image and mood. *Journal of Social and Clinical Psychology*, 31(2), 151-168.
- *Thøgersen-Ntoumani, C., Ntoumanis, N., Cumming, J., Bartholomew, K. J., & Pearce, G. (2011). Can self-esteem protect against the deleterious consequences of self-objectification for mood and body satisfaction in physically active female university students?. *Journal of Sport and Exercise Psychology*, 33(2), 289-307.
- Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York, NY: Columbia University Press. <http://dx.doi.org/10.1037/13283-000>
- Thompson, J. K. (1996). Assessing body image disturbance: measures, methodology and implementation. *Body image, eating disorders and obesity: An integrative guide for assessment and treatment*; Washington, DC: American Psychological Association, 49.
- Thompson, J. K. (2004). The (mis) measurement of body image: ten strategies to improve assessment for applied and research purposes. *Body Image*, 1(1), 7-14.
- Thompson, J.K., Heinberg, L.J., Altabe, M., Tantleff-Dunn, S. (1999). *Exacting beauty: theory, assessment, and treatment of body image disturbance*. Washington (DC): American Psychological Association.
- Tiggemann, M. (2004). Body image across the adult life span: Stability and change. *Body Image*, 1(1), 29-41.

- *Tiggemann, M., Brown, Z., Zaccardo, M., & Thomas, N. (2017). "Warning: This image has been digitally altered": The effect of disclaimer labels added to fashion magazine shoots on women's body dissatisfaction. *Body Image*, 21, 107-113.
- *Tiggemann, M., & McGill, B. (2004). The role of social comparison in the effect of magazine advertisements on women's mood and body dissatisfaction. *Journal of Social and Clinical Psychology*, 23(1), 23-44.
- *Tiggemann, M., & Polivy, J. (2010). Upward and downward: Social comparison processing of thin idealized media images. *Psychology of Women Quarterly*, 34(3), 356-364.
- *Tiggemann, M., & Slater, A. (2004). Thin ideals in music television: A source of social comparison and body dissatisfaction. *International Journal of Eating Disorders*, 35(1), 48-58.
- *Tiggemann, M., Slater, A., Bury, B., Hawkins, K., & Firth, B. (2013). Disclaimer labels on fashion magazine advertisements: Effects on social comparison and body dissatisfaction. *Body Image*, 10(1), 45-53.
- *Tiggemann, M., Slater, A., & Smyth, V. (2014). 'Retouch free': The effect of labelling media images as not digitally altered on women's body dissatisfaction. *Body Image*, 11(1), 85-88.
- *Tiggemann, M., & Zaccardo, M. (2015). "Exercise to be fit, not skinny": The effect of fitspiration imagery on women's body image. *Body Image*, 15, 61-67.
- Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual analog scales: do they have a role in the measurement of preferences for health states? *Medical Decision Making*, 21(4), 329-334.
- Tylka, T. L. (2011). Refinement of the tripartite influence model for men: Dual body image pathways to body change behaviors. *Body Image*, 8(3), 199-207.
- *van den Berg, P., Paxton, S. J., Keery, H., Wall, M., Guo, J., & Neumark-Sztainer, D. (2007). Body dissatisfaction and body comparison with media images in males and females. *Body Image*, 4(3), 257-268.

- van den Berg, P., Thompson, J. K., Obremski-Brandon, K., & Coover, M. (2002). The tripartite influence model of body image and eating disturbance: A covariance structure modeling investigation testing the mediational role of appearance comparison. *Journal of Psychosomatic Research*, 53(5), 1007-1020.
- van Duinen, M., Rickelt, J., & Griez, E. (2008). Validation of the electronic visual analogue scale of anxiety. *Progress in neuro-psychopharmacology and biological psychiatry*, 32(4), 1045-1047.
- Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The case for using the repeatability coefficient when calculating test–retest reliability. *PloS one*, 8(9), e73990.
- Vickers, A. J. (1999). Comparison of an ordinal and a continuous outcome measure of muscle soreness. *International Journal of Technology Assessment in Health Care*, 15(4), 709-716.
- *Wade, T., George, W. M., & Atkinson, M. (2009). A randomized controlled trial of brief interventions for body dissatisfaction. *Journal of Consulting and Clinical Psychology*, 77(5), 845.
- Want, S. C. (2009). Meta-analytic moderators of experimental exposure to media portrayals of women on female appearance satisfaction: Social comparisons as automatic processes. *Body Image*, 6(4), 257-269.
- Want, S. C. (2014). Three questions regarding the ecological validity of experimental research on the impact of viewing thin-ideal media images. *Basic and Applied Social Psychology*, 36(1), 27-34.
- *Want, S. C., Botres, A., Vahedi, Z., & Middleton, J. A. (2015). On the cognitive (in) efficiency of social comparisons with media images. *Sex Roles*, 73(11-12), 519-532.
- *Want, S. C., & Saiphoo, A. (2017). Social comparisons with media images are cognitively inefficient even for women who say they feel pressure from the media. *Body image*, 20, 1-6.

- *Want, S. C., Vickers, K., & Amos, J. (2009). The influence of television programs on appearance satisfaction: Making and mitigating social comparisons to “Friends”. *Sex Roles*, 60(9-10), 642-655.
- *Wasilenko, K. A., Kulik, J. A., & Wanic, R. A. (2007). Effects of social comparisons with peers on women's body satisfaction and exercise behavior. *International Journal of Eating Disorders*, 40(8), 740-745.
- Wee, H., Fong, K., Tse, C., Machin, D., Cheung, Y., Luo, N., & Thumboo, J. (2008). Optimizing the design of visual analogue scales for assessing quality of life: A semi-qualitative study among Chinese-speaking singaporeans. *Journal of Evaluation in Clinical Practice*, 14(1), 121-125.
- Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, 17(2), 267.
- Weinberger, N.A., Kersting, A., Riedel-Heller, S.G., & Luck-Sikorski, C. (2017). Body dissatisfaction in individuals with obesity compared to normal-weight individuals: A systematic review and meta-analysis. *Obesity Facts*, 9(6), 424–441.
- *Williams, G. A., Hudson, D. L., Whisenhunt, B. L., & Crowther, J. H. (2014). An examination of body tracing among women with high body dissatisfaction. *Body Image*, 11(4), 346-349.
- Williamson, A., & Hoggart, B. (2005). Pain: a review of three commonly used pain rating scales. *Journal of clinical nursing*, 14(7), 798-804.
- Yokobe, J., Kitahara, M., Matsushima, M., & Uezono, S. (2014). Preference for different anchor descriptors on visual analogue scales among Japanese patients with chronic pain. *PloS one*, 9(6), e99891.
- Zusman, M. (1986). The absolute visual analogue scale (AVAS) as a measure of pain intensity. *Journal of Physiotherapy*, 32(4), 244-246.