RELATION OF DIETARY RESTRAINT TO BULIMIC SYMPTOMATOLOGY: THE EFFECTS OF THE CRITERION CONFOUNDING OF THE RESTRAINT SCALE

ERIC STICE,[1],* SHARON OZER[2] and MICHELLE KEES[3]

[1]University of California, San Diego, CA 92093, U.S.A., [2]California School of Professional Psychology, San Diego, CA, U.S.A. and [3]State University of New York, Stony Brook, NY 11790, U.S.A.

Summary—Studies indicate that various measures of dietary restraint show inconsistent relations to bulimic symptomatology. The present study tested the assertion that this is because the scales differ in the extent to which they reflect the success or failure of dietary efforts. This study also tested the competing hypothesis that criterion confounding of the Restraint Scale produced the inconsistent findings. Data from 117 undergraduates indicated that both the Restraint Scale and the Dutch Restrained Eating Scale were correlated with bulimic pathology, but that the magnitude to the relations were higher for the former. These results supported the dietary success–failure hypothesis, however, this difference disappeared when the two disinhibited eating items from the Restraint Scale were deleted. Thus, the inconsistent findings in prior research appear to be due to the criterion confounding of the Restraint Scale. Implications for the measurement of restraint are discussed. Copyright © 1997 Elsevier Science Ltd

INTRODUCTION

Investigators have proposed that dietary restraint plays a causal role in the etiology of bulimia nervosa (Heatherton & Polivy, 1992; Polivy & Herman, 1985; Stice, 1994; Striegel-Moore, 1993). Dietary restraint refers to intentional efforts to achieve or maintain a desired weight through reduced caloric intake (Laessle, Tuschl, Kotthaus & Pirke, 1989). Theoretically, restraint leads to binge eating because of both physiological (e.g. altered hormonal reflexes) and psychological mechanisms (e.g. cognitively produced disinhibition; Polivy & Herman, 1985). Binge eating, in turn, is thought to precipitate the full binge–purge cycle of bulimia in at-risk individuals.

There is a large body of research supporting the proposed relation between dietary restraint and bulimia nervosa. First, longitudinal studies have found that restraint predicts future bulimic symptomatology (Greenberg & Harvey, 1986; Marchi & Cohen, 1990) and the onset of bulimia (Kendler *et al.*, 1991; Patton, 1988). Second, both naturalistic and laboratory experiments indicate that dietary restraint increases the likelihood of disinhibited eating (Franklin, Schiele, Brozek & Keys, 1948; Polivy, Herman & MacFarlane, 1994a; Polivy, Zeitlin, Herman & Beal, 1994b).† Further, the majority of bulimics report that the onset of binge eating occurred during a period of voluntary dieting (Hall & Hay, 1991; Lacey, Coker & Birtchnell, 1986), and many bulimics report that they are actively dieting and skipping meals (Davis, Freeman & Garner, 1988; Kendler *et al.*, 1991). Finally, bulimics score higher on restraint scales than controls (Jansen, van den Hout & Griez, 1990; Ruderman & Besbeas, 1992).

Although an extensive body of research has developed around the concept of dietary restraint (Brownell & Rodin, 1994; Lowe, 1993), investigators have raised major questions about the measurement of this construct (Heatherton, Herman, Polivy, King & McGree, 1988; Laessle *et al.*, 1989; Ruderman, 1986). One central problem is that various measures of restraint are differentially predictive of disinhibited eating (Lowe, 1993). There are three widely used measures of restraint: the Restraint Scale (RS; Herman & Polivy, 1980), the Dutch Restrained Eating Scale

---

*Author for correspondence at: Department of Psychiatry, Stanford University, Stanford, CA 94305-5542, U.S.A.
†Although counter-regulatory eating in a laboratory setting and binge eating in the natural environment are qualitatively different, both behaviors can be broadly construed as disinhibited eating.

(DRES; Van Strien, Frijters, Van Staveren, Defares & Deurenberg, 1986b), and the Cognitive Restraint Scale from the Three-Factor Eating Questionnaire (TFEQ-R; Stunkard & Messick, 1985). Although the RS has consistently identified dieters who show disinhibited eating in response to preloads or negative affect inductions (Herman & Mack, 1975; Polivy *et al.*, 1994a), the DRES and the TFEQ-R do not predict such disinhibited eating (Jansen, Oosterlaan, Merckelbach & van den Hout, 1988; Lowe & Maycock, 1988; Wardle & Beales, 1987). Similarly, although studies have found relations between the RS and bulimic symptoms (Hawkins & Clement, 1980; Ruderman, 1985; Ruderman & Besbeas, 1992), some research has failed to find such a linkage for the TFEQ-R (Berland, Thompson & Linton, 1986). Finally, although the DRES and TFEQ-R predict reduced caloric intake in the natural environment (Laessle *et al.*, 1989; Van Strien *et al.*, 1986b; Wardle, 1987; Wardle & Beales, 1987), the RS does not (Kirkley, Burge & Ammerman, 1988; Laessle *et al.*, 1989). Collectively, these studies indicate that the RS is more strongly related to disinhibited eating than the DRES and TFEQ-R, and that the DRES and TFEQ-R are more predictive of decreased food consumption than the RS.

The dietary success–failure hypothesis

Heatherton *et al.* (1988) posit that the inconsistent findings for the various measures of restraint result from the fact that the DRES and TFEQ-R assess "successful food restriction" (dietary success), whereas the RS measures "relatively unsuccessful dieting" (dietary failure). Heatherton's dietary success–failure hypothesis is consistent with the pattern of findings, in that the RS does not predict decreased daily caloric intake (e.g. Kirkley *et al.*, 1988), whereas the DRES and TFEQ-R do (e.g. Laessle *et al.*, 1989). Similarly consistent with this account is the finding that the RS predicts disinhibited eating in the laboratory and natural environment (e.g. Polivy *et al.*, 1994a), and the DRES and TFEQ-R do not (e.g. Jansen *et al.*, 1988). This assertion also converges with Laessle and colleagues' (1989) factor analyses of restraint scales and related constructs. They found that the RS loaded with variables such as the bulimia scale of the Eating Disorder Inventory and the disinhibition scale of the TFEQ, whereas the DRES and TFEQ-R loaded with reduced daily caloric intake. Although these results are consonant with the dietary success–failure hypothesis, the magnitude of the relations between the various restraint scales and bulimic symptoms were never directly compared. Accordingly, the first aim of this study was to provide a test of Heatherton's dietary success–failure hypothesis.

Heatherton's success–failure hypothesis was tested by regressing binge eating criteria on the scales that purportedly measure dietary success and failure. Assuming that binge eating and bulimic symptomatology reflect dietary failure, the RS (a putative measure of dietary failure) should be most predictive of these criteria, and the DRES (a putative measure of dietary success)* should be least predictive. Thus, we examined the predictive utility of these two measures of dietary restraint varying along the success–failure continuum. We elected to use binge eating in the natural environment as the criterion, rather than laboratory analogs, in an effort to enhance the external validity of the findings. Although laboratory experiments can offer strong internal validity, the disinhibited eating observed in standard laboratory paradigms is less extreme than that observed in naturalistic binges (Polivy & Herman, 1985).

The criterion confound hypothesis

An alternative explanation for the fact that the RS is more predictive of disinhibited eating than the DRES or TFEQ-R is criterion confounding. As Wardle and Beales (1987) note, the RS contains items that clearly refer to disinhibited eating: "Do you eat sensibly in front of others and splurge alone?" and "Do you have feelings of guilt after overeating?" (see Appendix for RS

---

*As Heatherton and colleagues (1988) proposed that both the DRES and the TFEQ-R assess dietary success, we decided that it was only necessary to include one of these scales. We elected to use the DRES instead of the TFEQ-R because there was slightly more reliability and validity data for this scale (e.g. Van Strien, Frijters, Bergers & Defares, 1986a; Van Strien *et al.*, 1986b; Wardle, 1987).