The (Mis)measurement of Restraint: An Analysis of Conceptual and Psychometric Issues

Todd F. Heatherton, C. Peter Herman, Janet Polivy, Gillian A. King, and Sheila T. McGree
University of Toronto, Toronto, Ontario, Canada

In this article we examine alleged conceptual and psychometric deficiencies of the Restraint Scale, an instrument intended to identify chronic dieters. These deficiencies include the confounding of restraint with disinhibition, the inapplicability of the scale to obese samples, problems with the factor structure of the scale, and difficulties in completing the scale. We argue that these alleged deficiencies are in most cases chimerical and that the Restraint Scale remains the most useful tool for examining behavioral and other dieter/nondieter differences. Proposed alternatives to the Restraint Scale are examined and found to be inadequate as replacements, although they may be useful for certain purposes. Closer attention to the intended purpose of such instruments may serve to dispel controversy and confusion.

Research on the dynamics of eating has gradually expanded from the examination of behavioral differences as a function of body weight (e.g., Schachter & Rodin's, 1974, survey of obese/normal differences) to include the investigation of parallel differences as a function of attempted weight suppression (relative to initial weight or relative to presumptive biologically defended levels). The Restraint Scale was initially proposed (Herman & Mack, 1975) as a simple and relatively straightforward self-report device for identifying chronic dieters. At the time it was assumed that because chronic dieters were likely to be maintaining a body weight below "set point," identification of such dieters would therefore permit tests—in normal-weight people as well as the obese—of hypotheses derived from Nisbett's (1972) seminal article on the effects of long-term hunger.

Almost since its inception, the Restraint Scale has been subjected to criticism, both psychometric and conceptual. In this article, we review and discuss some of these criticisms of the Restraint Scale, and then consider the alternative scales that have been proposed recently as improvements. The major problems that have been identified are (a) the Restraint Scale's confounding of dietary restriction with disinhibited eating, (b) its apparent inadequacy when applied to the overweight, and (c) its factor structure. These major issues are not entirely separable from one another, but we shall attempt to distill them into their essentials and address them sequentially, along with a number of lesser problems.

Restraint and Disinhibition

Early on, it became clear that the dieters identified by the Restraint Scale were as notable for their lapses of restraint as for their restraint per se (Herman & Mack, 1975; Herman & Polivy, 1975). Indeed, almost all of the research examining the eating behavior of people scoring high on the Restraint Scale has contrasted experimental situations in which restraint has remained intact with situations in which restraint is broken, with consequent overeating. As a result, our view (Herman & Polivy, 1980; Polivy & Herman, 1983) of the restraint construct—or more precisely, of the dieters identified by the scale—has changed so as to acknowledge that most dieters do not succeed in maintaining uninterrupted restriction of intake. The average dieter is more likely to exhibit periods of restraint punctuated by episodes of disinhibited overeating and, in all likelihood, does not achieve significant weight loss relative to physiologically defended levels. (This acknowledgment appeared very soon after the original Restraint Scale experiments; see Hilbacher & Herman, 1977.)

Discussions of the Restraint Scale's purpose—by ourselves and others—have perhaps been insufficiently explicit about the sort of person identified by the instrument. The dieter who succeeds in achieving significant weight loss—presumably by avoiding the splurges of overeating that we have been investigating in the laboratory—may well exist. This is the person who ought to show the long-term deprivation effects of interest to Nisbett (1972). And although this person may obtain a high restraint score, persons who diet less well—whose caloric restrictions are canceled by bouts of caloric excess—may score as high or higher. Indeed, Polivy (1978) found that bulimic anorexia nervosa patients (bingers) score higher on the Restraint Scale than do restricting anorexics (starvers).

The fact that the most successful dieters do not necessarily obtain the highest scores has been viewed by some (Brief, Stunkard, & Hirsch, cited in Stunkard, 1981; Van Strien, 1986) as evidence for the invalidity of the Restraint Scale. These claims focus on two related features of the extant scale: First, as we have already mentioned, it does not give preeminence to "truly" restrained eaters (i.e., eaters whose behavior is characterized exclusively by restraint or restriction); and second, the
dieters who obtain the highest scores may have a propensity to
ward disinhibition, which makes the disinhibition that they dis-
play in the lab (and probably outside of it as well) less surprising.

As for the first claim, it is true that the Restraint Scale does
not select exclusively restrictive dieters; and it follows that the
dieters whom it does select may well not be biologically under-
weight. Brief et al. (cited in Stunkard, 1981) found that re-
strained individuals did not have depleted fat cell size; they
therefore concluded that restrained subjects are not likely to be
below set-point. If the current intention of the scale were to
testify to dieters who are below their biological set points, then
we would agree that the scale's validity is unproved or worse.
However, the claim that individuals with high restraint scores
are underweight or significantly food deprived is simply not
made and has not been made for about a decade. Indeed, this
claim has not been advanced since it became obvious that re-
strained eaters in the high-normal range—or beyond, as in the
case of bulimics (Polivy, 1978)—tend to eat too much to achieve
substantial weight loss. So although a "restrained eater" is no
doubt restrained—and may well have such restraint as a central
trait—such restraint is likely to be accompanied by the afore-
mentioned lapses. (Indeed, we have argued elsewhere—Polivy
& Herman, 1985—that restraint is a major contributor to such
lapses.) It seems that restraint is thus a misnomer, because the
construct we measure involves disinhibition as well. Our reason
for preferring to retain the name, rather than code it to those
(e.g., Stunkard & Messick, 1985; Van Strien, 1986; Van Strien,
Frijters, Bergers, & Defares, 1985) who would measure only
pure restraint without disinhibition, is that most restrained eat-
ers do exhibit occasional disinhibition. The restrained eater
who is exclusively restrained (i.e., the individual who scores
high on Van Strien et al.'s, 1985, and Stunkard and Messick's,
1985, restraint subscales) is not representative of restrained eat-
ers in general, whereas the restrained eater who occasionally
splurges is. The prevalent inability of dieters to achieve and
maintain weight loss is well documented (Stunkard & Penick,
1979; Wilson & Brownell, 1980). These reviews provide ample
testimony to dieters' disinhibitory propensities, with repeated
infractions of the diet rendering weight loss difficult and event-
ual relapse seemingly inevitable. Thus, to argue that the Rest-
raiment Scale confounds true restraint with disinhibition betrays
a mistaken view of the scale's purpose, which is to identify di-
ers. Most dieters (to their regret) display both restraint and dis-
inhibition, and the disinhibition they display is not an arbitrary
attribute, but a direct consequence of their restraint.

The coupling of restraint with disinhibition undoubtedly
makes the experimental demonstration of disinhibition in
restrained eaters less mysterious and perhaps altogether unsur-
prising. Certainly, a decade or more after this research began,
the connection between restraint and disinhibition is no longer
remarkable. Almost since the beginning (e.g., Polivy, 1976),
however, this research has gone beyond the mere demonstration
of the restraint–disinhibition connection to focus on the more
complex questions of precisely when, why, and how disinhibi-
tion occurs in dieters and not simply whether it occurs. The
answers to the questions concerning the details of disinhibited
eating are not all obvious, and much work remains to be
done.

### Table 1

**The Restraint Scale**

<table>
<thead>
<tr>
<th>Question</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How often are you dieting?</td>
<td>CD</td>
</tr>
<tr>
<td>2. What is the maximum amount of weight (in pounds) you have ever lost in one month?</td>
<td>WF</td>
</tr>
<tr>
<td>3. What is your maximum weight gain within a week?</td>
<td>WF</td>
</tr>
<tr>
<td>4. In a typical week, how much does your weight fluctuate?</td>
<td>WF</td>
</tr>
<tr>
<td>5. Would a weight fluctuation of 5 lbs affect the way you live your life?</td>
<td>CD</td>
</tr>
<tr>
<td>6. Do you eat sensibly in front of others and splurge alone?</td>
<td>CD</td>
</tr>
<tr>
<td>7. Do you give too much time and thought to food?</td>
<td>CD</td>
</tr>
<tr>
<td>8. Do you have feelings of guilt after overeating?</td>
<td>CD</td>
</tr>
<tr>
<td>9. How conscious are you of what you're eating?</td>
<td>CD</td>
</tr>
<tr>
<td>10. How many pounds over your desired weight were you at your maximum weight?</td>
<td>WF</td>
</tr>
</tbody>
</table>

**Note.** CD = Concern for Dieting. WF = Weight Fluctuations.

### Restrained and Obesity

A second set of criticisms of the Restraint Scale has centered
on its alleged inapplicability to the obese. These criticisms fall
roughly into three classes: first, that the Restraint Scale does
not accurately measure dieting tendencies in obese respon-
dents; second, that the factor structure of the scale differs for
normal and obese respondents; and third, that restrained obese
subjects do not behave in a manner parallel to that of restrained
normal-weight subjects, thereby casting doubt on the assertion
that differences in restraint are superior to differences in degree
of overweight as explanations for behavior.

**Do the Obese Obtain Spuriously High Restraint Scores?**

It has been recognized from the beginning (Herman & Polivy, 1975) that the Restraint Scale contains items assessing both
weight fluctuation (WF) and subjective concern for dieting
(CD). (These two factors correspond, in most analyses, to Items
2, 3, 4, and 10, and Items 1, 5, 6, 7, 8, and 9, respectively; see
Table 1.) On the basis of an observation by Bray (1976) that the
obese exhibit greater "spontaneous" weight fluctuations, Drew-
nowski, Riskey, and Desor (1982) argued that obese individuals
would tend to obtain high scores on the Restraint Scale even if
they did not engage in chronic dieting. Ruderman (1983,
1985b, 1986) likewise has claimed that high restraint scores
among the obese do not necessarily mean that they are serious
dieters. High restraint scores may result from their greater ex-
tent of weight fluctuation, which is an attribute of overweight
people unconnected with dieting (or any other behavior).

The consistently significant correlation between percentage
overweight and restraint (Lowe, 1984: $r = .38$; Ruderman,
1983: $r = .37$; 1985b: $r = .38$; Wardle, 1980: $r = .39$) suggests
that the obese do indeed obtain higher restraint scores than do
normal-weight individuals. However, it is unclear whether the
obese's higher scores are due to their concern for dieting, greater
weight fluctuations, or both. Whereas Blanchard and Frost
(1983) reported that the WF factor is more highly correlated
with overweight ($r = .48$) than is the CD factor ($r = .29$), Lowe
(1984) reported that the correlation between overweight and CD \( r = .43 \) is significantly higher than the correlation between overweight and WF \( r = .14 \). Furthermore, Lowe (1984) found that the correlation between overweight and WF is eliminated when the CD factor is used as a partial correlate. Lowe also conducted a direct comparison between normal-weight and obese subjects and found that they differed only in CD, not in WF. Ruderman (1985b) reported finding partial correlations opposite to those of Lowe (1984) such that WF remains significantly correlated with overweight when CD is partialed out, whereas CD does not remain significantly correlated with overweight when WF is partialed out. On balance, the evidence does not clearly support the claim that obese subjects' higher restraint scores are simply a matter of greater weight fluctuation.

Although it may be that obese people exhibit greater spontaneous weight fluctuations for reasons of physiology, as Bray (1976) suggested—and it is certainly the case that if everyone fluctuates, say, 5%, then the heavier the individual, the greater the absolute weight fluctuation—it may nevertheless also be the case that these exaggerated weight fluctuations occur because heavier individuals are more likely than are lighter individuals to engage in active dieting. Such dieting, along with the sporadic excessive eating to which dieting conduces (Herman & Polivy, 1980; Polivy & Herman, 1985), will contribute directly to weight fluctuations in a way that is directly reflective of re-straint, and not an artifact of physiology or the law of initial constraint scores are simply a matter of greater weight fluctuation.

Obesity and the Factor Structure of the Restraint Scale

Ruderman (1983), Johnson, Lake, and Mahan (1983), and Lowe (1984) found that the Restraint Scale displays a different factor structure when the sample consists largely or completely of overweight subjects. When the subject population consists mainly of obese individuals, factor analyses tend to extract more than the two basic factors (CD and WF) of the Restraint Scale (Ruderman, 1986). However, if one considers that perhaps as many as 90% of obese individuals in some samples are restrained (Herman & Polivy, 1980), then a factor analysis on obese subjects may well amount to a factor analysis on only a restricted (high) range of restraint scores. The decreased variability of responses among highly restrained subjects might produce lower interitem correlations than one might ordinarily expect and consequently produce extraneous factors (Gorsuch, 1983).

Moreover, although factor analyses based on restricted samples will usually weaken the item intercorrelations and thereby produce additional factors, even unrestrained (non truncated) samples may suffer if the scores are not normally distributed. When the method of sampling introduces substantial skew into the distribution of scores, the correlations among items may be lowered (Brewer & Hills, 1969), thereby increasing the probability that additional factors will be extracted.

Support for the contention that the factor structure for samples of obese subjects may be distorted by truncation or skew may be found in the reduced reliability coefficients for obese (Johnson et al., 1983; Ruderman, 1983) and bulimic (Johnson, Corrigan, Crusco, & Schlundt, 1986) samples. For example, Ruderman (1983) found that coefficient alpha was .86 in the normal-weight sample but only .51 in the obese sample. The mean item/total-scale correlations on which the alpha coefficients were based were .56 for normal-weight subjects and .22 for obese subjects. These lower correlations for the obese sample (attributable to the truncated range of item scores) may have weakened the factor structure of the Restraint Scale. Likewise, Johnson et al. (1986) found a relatively low coefficient alpha (.57) for their 26 bulimic subjects, who obtained very high restraint scores, again suggesting reduced variability as a function of a truncated range of responses. When Johnson et al. (1986) added these 26 bulimic subjects to the already restricted sample explored by Johnson et al. (1983), a principal-components analysis of the Restraint Scale resulted in four factors. In effect, Johnson et al. (1986) used an even more skewed sample than had Johnson et al. (1983) and thereby produced yet another extraneous factor.

To demonstrate the factor analytic consequences of studying a sample with a restricted range of restraint scores, we conducted a series of factor analyses on our recent data sets obtained in our lab (Heatherton, Polivy, & Herman, 1987; Herman, Polivy, Lank, & Heatherton, 1987). A principal-components analysis for the entire sample, using the Kaiser criterion for retention of factors and varimax rotation, revealed the two factors (WF and CD) typically found during factor analysis. We then divided the sample using a median split (15) and conducted principal-components analyses on the two subgroups. These analyses resulted in a five-factor solution for restrained subjects (high scorers) and a four-factor solution for unrestrained subjects (low scorers). It seems clear that a factor analysis conducted on a truncated range or on a sample with a skewed distribution of scores may result in the extraction of spurious factors (Brewer & Hills, 1969; Gorsuch, 1983). And it follows that the greater the proportion of obese (or eating-disordered) subjects in the sample—and therefore presumably the more skewed or truncated the distribution of scores—the more factors may be extracted during factor analysis.

Whether these additional factors represent spurious statistical artifacts or real differences in the structure of the Restraint Scale for obese samples seems to be basically a matter of interpretation. In any event, two conclusions seem to be in order: First, the factor structure of the Restraint Scale (and presumably most other scales) will change if the sample characteristics change in a systematic way; and second, this sort of sample-based variation in factor structure is not a distinctive "prob-
lem” with the Restraint Scale but rather a general issue in the interpretation of factor analyses derived from unusual samples.

Obesity and Response to a Preload

Ruderman and Wilson (1979) found evidence that obese restrained subjects behaved differently from normal-weight restrained subjects in response to a preload challenge. Normal-weight restrained subjects showed the “standard” (Herman & Polivy, 1980) counterregulation effect, with those receiving a preload consuming more ice cream ad lib in a subsequent taste test than did those who did not receive a milk shake preload. Obese restrained subjects were found to display direct compensation, albeit weakly, after a milk shake preload (Ruderman & Christensen, 1983). Ruderman (1986) has argued that either the Restraint Scale does not accurately identify dieters among the obese (a complaint that we have already addressed) or that restraint theory’s predictions concerning response to a preload are wrong when applied to the obese (or both).

We believe that obese people who score high on the Restraint Scale are in fact dieters (as we discussed earlier), and furthermore, we adhere to our original contention that independent of weight status, restrained individuals are more likely than are unrestrained individuals to counterregulate. Counterregulation is said to occur when subjects eat more after a larger preload than after a smaller one (or none at all). Presumably, the larger preload sabotages the dieter’s current diet intentions and makes further dieting seem (temporarily) to be not worth the effort. Recently, we (Herman & Polivy, 1984) have attempted to construct a spatial, or boundary, model of this situation. In this model, the dieter is characterized by a diet boundary (located between hunger and satiety), which represents the upper limit on episodic intake prescribed by the diet. If the preload exceeds this limit, disinhibition will occur; and if the effect of a preload that just exceeds this limit is compared with the effect of a preload that falls short of the limit, it is likely that more eating will occur after the former (larger) preload than after the latter (smaller) preload, resulting in the counterregulation effect. However, the effect of a given preload on a given individual will depend crucially on the (perceived) size of the preload compared with the permissible limit as dictated or represented by the diet boundary. If the larger preload exceeds this limit and the smaller preload does not, then counterregulation is likely to occur, as we have seen. But dieters do not exhibit counterregulation as a consistent, traitlike characteristic, regardless of the pertinent circumstances. For instance, if the larger preload, like the smaller preload, falls short of the diet boundary, we would expect less eating after the large preload (because following a large preload there is less “room” before the diet boundary is reached). By the same token, if the smaller preload and the larger preload both exceed the diet boundary, we should again expect less eating after the larger preload than after the smaller preload (because after the larger preload there is less room before the next operative boundary, satiety, is reached). These complex effects (i.e., direct regulation or counterregulation depending on whether the two preloads being compared are on the same or opposite sides of the diet boundary) were recently demonstrated in a study (in normal-weight restrained eaters) conducted by Herman, Polivy, and Esses (in press).

The effects of various preloads, then, depend on their relation to the diet boundary; and whether obese and normal-weight dieters show the same (counter)regulatory effects following particular preloads is not simply a matter of the preload sizes in isolation; rather, it depends on the diet boundaries in obese and normal-weight dieters. As yet, there is no direct evidence on this point, but it seems entirely plausible that obese and normal-weight dieters may differ substantially in how much (rich) food they will allow themselves on any given occasion.

It might be the case that the diet boundary for obese dieters is more stringent than that for normal-weight dieters; after all, the obese have more to be concerned about and require a stricter diet if they are to achieve their goal (significant weight loss), whereas normal-weight dieters are more likely to be concerned about maintenance or relatively mild further weight reduction. Accordingly, if the diet boundary is not at all permissive, then it seems quite likely that even a small preload may disinhibit obese dieters, thereby destroying the counterregulation effect if this small preload is compared with a large preload that also disinhibits obese dieters but brings them closer to satiety.

It is also conceivable that some obese people may have a less stringent diet boundary than do normal-weight dieters. In either case, however, it is clear that the same preloads that produce counterregulation in normal-weight dieters will not necessarily produce counterregulation in obese dieters. This does not mean that the Restraint Scale is unable to detect obese dieters; nor does it mean that the phenomenon of counterregulation does not apply to obese dieters. It simply means that counterregulation will appear in obese dieters following different, more appropriate preload comparisons. The most explicit test of this argument would involve (a) attempting to ascertain directly from obese and normal-weight dieters how much rich food it would take to exceed their personal quotas and (b) subjecting both obese and normal-weight restrained eaters to a graded series of preloads (e.g., none, small, large, and very large). For both obese and normal-weight dieters, we would predict that counterregulation will be evident when comparisons are made across their respective diet boundaries, whereas normal compensation will be evident in comparisons involving preloads on the same side of the boundary. Obese and normal-weight dieters may well differ, however, in the critical breakpoint; that is, for normal-weight dieters, the diet boundary may occur between the small and large preload, whereas for obese dieters, it may be located below the small preload. Ideally, such a study would demonstrate the presence of counterregulation in obese dieters as well as account for prior failures to observe counterregulation in obese dieters exposed to preloads not selected so as to span the relevant diet boundary.

The Restraint Scale is useful for examining dieter/non dieter differences in behavior: Such differences may be detected in normal-weight and obese samples, although, as we have seen, the precise nature of these differences may depend on factors that vary from normal-weight to obese samples, such as the location of the diet boundary. Recognition of such subtleties does not demand abandoning the Restraint Scale or confining it to the
identification of normal-weight dieters. Rather, it requires a more active attempt to understand the basis for dieter/nondieter differences and to expect variation in such differences as a function of more fundamental processes.

Our research began with an attempt to explain obese/normal differences in terms of dieter/nondieter differences. But the very first study in our series (Herman & Mack, 1975) found that normal-weight dieters counterregulate, whereas prior studies on the obese (e.g., Schachter, Goldman, & Gordon, 1968) had led us to expect neither regulation nor counterregulation, but equivalent amounts eaten regardless of preload size. The unexpected discovery of counterregulation led in turn to many studies on dieter/nondieter differences in (dis)inhibition situations and a corresponding neglect of the effort to explain the behavior of the obese, except insofar as being obese affects one's likelihood of becoming a dieter (and vice versa).

The Bifactorial Structure of the Restraint Scale

Our earliest reports (Herman & Polivy, 1975) noted that the Restraint Scale is composed of two sorts of items, corresponding to what have since become known as the CD and WF factors. In numerous factor analyses, a rather stable two-factor solution has emerged; in most cases, six questions have been associated with CD and four with WF (Blanchard & Frost, 1983; Drewnowski et al., 1982; Heatherton, 1986a; Lowe, 1984; Polivy, Herman, & Howard, in press; Ruderman, 1983). (See Table 2 for a comparison of factor structures.) As we mentioned earlier, Johnson et al. (1983, 1986), Ruderman (1983), and Lowe (1984) found a more complex factor structure when the sample scores were skewed by obese or bulimic subjects, with Questions 6 and 7 appearing most variable.

A frequent complaint regarding the Restraint Scale is that because there are two factors, it is impossible to determine whether it is CD or WF that is responsible for the behavior displayed by restrained eaters. In short, the Restraint Scale is alleged to ignore the classical test construction principle that a single scale ought to measure a single construct (Briggs & Cheek, 1986).

The principle that a single scale ought to measure a single construct does not mean, however, that a scale ought not to contain two or more correlated factors. Rather, it simply requires that the component factors be statistically and conceptually related; that is, they should measure different aspects of the same construct.

The CD and WF factors of the Restraint Scale are not orthogonal to each other. The average reported correlation between the two factors is .48 ($SD = .09$) (Blanchard & Frost, 1983: $r = .53$; Drewnowski et al., 1982: $r = .17$; Heatherton, 1986a: $r = .66$, 1986b: $r = .62$; Herman & Polivy, 1975: $r = .48$; Klajner, Herman, Polivy, & Chhabra, 1981: $r = .66$; Lowe, 1984: $r = .28$). In some of these instances the correlation between the two factors may well be underestimated because, as was pointed out by Herman and Polivy (1982), the mechanics of factor analysis dictate that the correlation between two factors be minimized during rotation. Thus, when the factor structure derived from one sample (e.g., Drewnowski et al., 1982) is applied to different samples (Heatherton, 1986b; Herman & Polivy, 1982), the inter factor correlations increase substantially. It is clear from the relatively strong correlation between CD and WF that the two are sufficiently related to satisfy the demand that a single scale ought to measure a single construct, yet dissimilar enough so that they do not supply redundant information.

It has become conventional for those who regard the Restraint Scale as being composed of two independent factors to use these factors individually as predictors of behavior, other questionnaire responses, and so on. However, although such exercises have often demonstrated the predictive superiority of one factor over the other, there is no consensus as to which factor is paramount. Lowe (1984) and Ruderman (1983, 1985a, 1985):

Note: CD = Concern for Dieting, WF = Weight Fluctuations. nr = not reported.
1985b, 1986) have suggested that the CD factor is the most important factor, whereas Blanchard and Frost (1983) have argued the merits of the WF factor. According to Briggs and Cheek (1986), “it makes sense to continue subdividing a large global factor into smaller, more precise subfactors as long as the distinctions are conceptually meaningful and empirically useful” (p. 111). It remains possible that the separate correlates of CD and WF will eventually be elucidated, but in the absence of clear evidence as to which factor is likely to be more useful on any given measurement occasion, it seems prudent to use the whole scale, which has never been shown to be inferior to either subscale alone.

Considered as a single entity, the Restraint Scale displays acceptable levels of reliability, and—as long as one avoids respondent samples displaying marked skew or an attenuated range of scores (see earlier section, Obesity and the Factor Structure of the Restraint Scale)—the mean interitem and item-total correlations appear to be within acceptable ranges according to the criteria of Briggs and Cheek (1986; Polivy et al., in press). These robust whole-scale reliabilities have been reported by others (Blanchard & Frost, 1983; Johnson et al., 1983; Ruderman, 1983) and support our contention that the Restraint Scale assesses a unitary construct with adequate internal reliability.

Is the Restraint Scale Difficult to Complete?

Wardle (1986) has recently reported that a significant number of people are unable to complete the Restraint Scale because of the inherent difficulty of questions related to current weight and weight fluctuations, especially for unrestrained subjects. Wardle mentioned that up to two thirds of men and 40% of women failed to complete all items on the scale. Wardle (1980) had earlier reported that 22 of 90 British subjects (24%) had failed to complete all of the items on the Restraint Scale. It would appear that these difficulties may be cultural in origin, as North American samples have not been reported to have had difficulties with any of the items. Blanchard and Frost (1983) found completion rates of from 95% to 99% in two large samples of American college students, whereas Heatherton (1986a) found that 205 out of 216 Canadian respondents (95%) filled out all of the items from the scale. As weight is typically measured in stones in Great Britain, whereas the Restraint Scale seeks information in pounds, some respondents may have had difficulty with the unfamiliar units of measurement; this difficulty can be solved easily by simple translation of the scale, as has been demonstrated by Booth (personal communication, April 13, 1987). Another consideration may be that respondents in the British samples were generally less concerned about their weight and thus attended to weight fluctuations to a lesser extent. The overall restraint scores in Wardle’s samples tend to be slightly lower than those of American samples, suggesting a lower overall concern for weight status and dieting among the British.

Alternative Measures of Restraint

Three-Factor Eating Questionnaire

One might consider it mildly ironic, given that one of the primary criticisms of the Restraint Scale is its bifactorial structure, that the most frequently mentioned replacement for the Restraint Scale is Stunkard and Messick’s (1985) Three-Factor Eating Questionnaire (TFEQ). Whereas the Restraint Scale has two correlated factors, the TFEQ possesses three factors (Cognitive Restraint, Disinhibition, and Hunger), one of which (Hunger) appears to be unrelated to the others, displaying if anything a slight negative correlation with Cognitive Restraint for both dieters and “free eaters.” Regarded as a single scale, the TFEQ certainly violates the cardinal psychometric rule concerning the assessment of a unitary construct (Briggs & Cheek, 1986; Guilford, 1954; McNemar, 1946).

Furthermore, although some authors suggest that factorially derived scales (such as the TFEQ) are superior to empirically derived scales (such as the Restraint Scale), the predictive validity of the TFEQ remains to be demonstrated. Lançon and Goodstein (1982) have reminded us that “usefulness, in any predictive sense, is not an intrinsic property of factorially derived scales, but must be demonstrated empirically” (p. 93). In a recent study in our laboratory (Heatherton, 1986b), typical counterregulatory findings were obtained using the Restraint Scale; this pattern failed to emerge, however, when the same subjects were assigned to condition by means of overall TFEQ scores. This inability of the TFEQ to discriminate those who would become disinhibited from those who would not was evident when each of its separate subfactors was used as a predictor.

The TFEQ’s predictive divergence from the Restraint Scale is also evident in studies on the relation between restraint and emotionally disinhibited eating, which has been well chronicled and consistently replicated. Laboratory studies have demonstrated that dysphoric moods consistently disrupt the diets of restrained subjects, resulting in excessive eating. At the very least, the behavior of restrained subjects has been found to be significantly different from that of unrestrained subjects following mood or anxiety manipulations (Frost, Gooolkasan, Ely, & Blanchard, 1982; Herman & Polivy, 1975; Herman et al., 1987; Ruderman, 1985a). Baucom and Aiken (1981) observed disinhibition of eating by dysphoric mood in dieters who were identified simply by asking them about their current dieting status (rather than by administering the Restraint Scale). Cooper and Bowskill (1986) have recently found that bulimics and current dieters report being in a dysphoric state before bingeing; however, those rated high in TFEQ Cognitive Restraint were no more likely to be in a dysphoric than nondysphoric state before (over)eating. Evidently, the Stunkard and Messick Cognitive Restraint subscale is not a simple substitute for the Herman and Polivy Restraint Scale.

3Herman, Polivy, and Heatherton (1987) have noted that physical fear differs from more general dysphoric moods in its effect on restrained and unrestrained individuals. Although physical fear normally decreases food intake in nondieters (and normal-weight subjects), it does not significantly increase the eating of restrained (or obese) individuals. Dysphoric mood manipulations normally increase the eating of restrained (or obese) subjects significantly but do not significantly suppress the eating of normal-weight or unrestrained subjects. This observation has yet to be specifically tested but is highly consistent in previous literature.
It is our contention that much can be learned from the apparent inability of the TFEQ to replicate previous Restraint Scale findings. A surface analysis would suggest that the Cognitive Restraint factor of the TFEQ measures a tendency to be aware of or focused upon restricting one's food intake. Thus the scale would identify those individuals who are successfully reducing food intake. Stunkard (1981) noted that the need for a new measure of restraint was a result of the failure of the Restraint Scale to correlate negatively with fat cell size. Stunkard reasoned that if individuals were successfully limiting intake, their fat cells ought to be depleted. The predicted negative relation between fat cell size and the TFEQ was revealed by Brief et al. (cited in Stunkard, 1981). Thus it appears that the TFEQ measures successful food restriction, whereas the Restraint Scale measures relatively unsuccessful dieting. (This distinction reverts to the restraint vs. restraint-plus-disinhibition controversy addressed at the beginning of this article.)

Despite the empirical covariation of restraint and disinhibition, Stunkard and Messick (1985) have proposed that it is the Disinhibition factor of the TFEQ—and not the TFEQ Cognitive Restraint factor—that identifies individuals who will binge. Even if one neglects the empirical restraint-disinhibition connection, however, one must nevertheless contend with the logical impossibility of disinhibited eating in the absence of (prior) restraint. Disinhibition can be defined only as the lifting or suspension of some sort of inhibition (i.e., restraint). Without initial inhibition or restraint, it is difficult to imagine disinhibition. We have argued (Herman & Polivy, 1984; Polivy & Herman, 1985) that disinhibited, or binge, eating results from the temporary collapse of a person's diet boundary (which acts as an episodic limit on intended consumption). As we noted earlier, Heatherton (1986b) recently failed to find classic counter-regulation when subjects were classified on the basis of a median split of the Disinhibition factor of the Stunkard and Messick (1985) scale, whereas a highly significant interaction between restraint and preload size emerged when subjects were classified on the basis of a median split of the Restraint Scale. Stunkard and Messick's (1985) contention that the disinhibition exhibited by those who score high on the Restraint Scale is equivalent to high scores on their “tendency-toward-disinhibition” factor must be demonstrated explicitly, with the same patterns of results occurring when subjects are classified using either scale. Note in this context that Heatherton (1986b) found the association between TFEQ Cognitive Restraint and Restraint Scale total scores to be stronger ($r = .68$) than that between TFEQ Disinhibition and Restraint Scale total scores ($r = .48$), although both were significant (contrary to what was obtained by Stunkard & Messick, 1985). Thus the Cognitive Restraint and Disinhibition subfactors of the TFEQ are clearly related to the Restraint Scale, although hardly identical to it. However, to the extent that the TFEQ (or some subset of its subfactors) is related to the Restraint Scale, it is not yet clear whether it can adequately replace the Restraint Scale in the sense of successfully predicting eating behavior as well or better. And to the extent that the TFEQ (or its component subfactors) diverges from the Restraint Scale, it is likewise not yet clear what behaviors or other variables of interest will be differentially predictable by the TFEQ.

One area in which the Restraint Scale and the TFEQ achieve roughly similar results concerns weight gain during depression. Polivy and Herman (1976) and Zielinsky (1978) noted that clinically depressed dieters tend to report gaining weight when depressed, whereas unrestrained dieters report losing weight. Weissenburger, Rush, Giles, and Stunkard (1986), using the TFEQ, also were able to discriminate depressive weight gainers from depressive weight losers, although they found that “in the discriminant function analysis, [the] disinhibition [factor] emerged as the only powerful discriminator of weight gain from weight loss and, in the split-sample replication, as the only stable predictor” (p. 280). Weissenburger et al. concluded that the Polivy and Herman (1976) and Zielinsky (1978) results were probably attributable to the disinhibition component of the Restraint Scale rather than to restraint per se. However, disinhibition is not psychometrically separable from restraint in the Restraint Scale; and even in the Weissenburger et al. data, weight gain is significantly correlated with the TFEQ restraint factor, although not quite as strongly as with TFEQ disinhibition. To the extent that the TFEQ does achieve predictive success, then, it seems to do so with the same confound of restraint and disinhibition that characterizes the Restraint Scale. Moreover, these studies of restraint and depressive weight gain are entirely a matter of correlations between self-report measures; the jury is still out on behavioral prediction with the TFEQ and, more crucially, on whether the differential application of three separate TFEQ components—or some combination—can achieve predictive success that is reliably superior to that achieved by unitary restraint.

A final difficulty with the TFEQ is its inclusion of a subscale of perceived hunger. On reflection, it is not as clear as it might be what one should predict regarding the association between restraint status and perceived hunger. Numerous studies have demonstrated that restrained individuals do not differ from unrestrained individuals on simple analogue hunger rating scales (cf. Kirschenbaum & Tomarken, 1982). Although Nisbett (1972) provided the impetus for restraint research with the theory that obese individuals were chronically hungry, Nisbett used the term hunger interchangeably with the term deprivation to indicate a physiological state that is below some set point for body weight. It seems worth distinguishing between deprivation (as a state of energy deficit, perhaps chronic) and hunger (as a sensation presumably arising from such a deficit). Schachter (1968) and Bruch (1961) have both argued that obese subjects differ crucially from nonobese people in that the obese are relatively insensitive to internal state; we might thus regard the obese—or any dieters, following Nisbett—as deprived, but not necessarily phenomenologically hungry. Indeed, it may well require more acute deprivation for a restrained individual to experience hunger than for an unrestrained individual to experience hunger.

3The Restraint Scale contains items representing both chronic dietary restriction as well as excessive eating. Some (e.g., Van Strien, 1986) have suggested that this aspect of the Restraint Scale creates a selection bias for those individuals prone to binge eating or disinhibition. Note, however, that binge eating or disinhibition does not appear as a separate component in factor analyses; this suggests that disinhibition is as integral a part of chronic dieting as is weight fluctuation.
ence hunger, because the restrained eater is in some sense habituated to inordinate periods of deprivation. Heatherton (1986b) found that TFEQ perceived hunger was slightly negatively correlated with the restraint score ($r = -0.08$). The assumption that perceived hunger ought to be related to disinhibition or even to cognitive restraint demands that hunger as an experience be coordinated to food deprivation; for many people, particularly dieters and the obese, this is simply not the case.

The TFEQ, as its name implies, is not a unitary scale. High total scores on the scale do not clearly identify a particular coherent syndrome or type of individual. Certainly, in terms of Stunkard and Messick's own analyses and arguments, the TFEQ is not a substitute for the Restraint Scale in the sense that it measures the same thing, only better. Insofar as the TFEQ attempts to separate cognitive restraint from disinhibition and treat them as independent predictors, it is forcing an artificial distinction that seems unlikely to improve the accurate prediction of eating behavior. And the inclusion of the Hunger factor seems likely to add only noise to total TFEQ scores, while not adding much (if any) predictive power.

**Dutch Eating Behavior Questionnaire (DEBQ)**

The DEBQ was developed by Van Strien (1986) and her colleagues to provide a homogeneous scale to test the three primary theories of (over)eating and obesity: psychosomatic, externality, and restraint. To do this, a factorially derived scale was developed that sought information about the extent of emotional, external, and restrained eating. Wardle (1986) has recently advocated the use of the DEBQ because it is relatively uncontaminated by weight. (Wardle regarded scores on the Restraint Scale as artificially inflated for the overweight, but see our earlier discussion of the connection between [over]weight and restraint.) Furthermore, for Wardle's British population, where the Restraint Scale may be less appropriate (see our discussion regarding the difficulty of completing the Restraint Scale), the DEBQ may be easier to fill out because it purports to measure restraint without items asking questions about body size or weight fluctuations.

Van Strien, Frijters, van Staveren, Defares, and Deurenberg (1986) have presented data in support of the predictive validity of the DEBQ Restraint scale: Those scoring higher in DEBQ restraint also report eating less food. However, validation of self-reported restraint should ideally be in terms of actual eating behavior rather than in terms of another self-report measure of eating.

In any case, one is struck by the similarities between the DEBQ and the TFEQ. Both measure three ostensibly independent aspects of eating, rather than a unitary construct; indeed, it might be said that whereas the DEBQ and the TFEQ both measure various styles of eating (restrained, disinhibitory, external, and so on), the Restraint Scale simply identifies dieters, without distinguishing between the various eating styles that they may display and which are regarded as interrelated. Thus both the DEBQ and the TFEQ use as their measure of restraint a measure of the degree of food restriction, rather than a measure of the full range of behaviors (such as disinhibition and guilt over excessive eating) exhibited by most dieters and considered indiscriminably as components of the Restraint Scale. This discrepancy with respect to measurement strategy may explain the divergence of findings obtained with the different scales; for example, both the DEBQ and the TFEQ appear to be unrelated to self-reports of emotional eating (Cooper & Bowskill, 1986; Wardle, 1986), whereas the Restraint Scale shows a strong association (e.g., Frost et al., 1982; Herman & Polivy, 1975; Ruderman, 1986).

**A Synthesis**

We believe that the differences among the different measures of restraint are as much conceptual as psychometric. The Restraint Scale, we suggest, measures the extent to which people (a) display (over)concern with their weight and (b) chronically diet to control it. Polivy and Herman (1983) have stressed just how difficult the task of dieting may be and have recently suggested that the accumulation of dieting failures may lead to more pathological eating disorders (Polivy & Herman, 1985). Dieters, then, are not typically "successful." The TFEQ and the DEBQ, by attempting to isolate successful caloric restriction, do not appear to measure the same behavioral tendencies as does the Restraint Scale. They are designed to measure successful dieting, whereas the Restraint Scale is designed to identify dieters.

Our initial labeling (as restraint) of the tendency of some people to become concerned with their body weight and to undertake dieting was perhaps unfortunate; certainly, it seems to have led to some confusion. Restraint, rather than referring to a single behavioral tendency, is a multifaceted syndrome involving both a propensity to restrict food intake as well as a tendency to splurge. Just as many anorexics are as notable for their bulimic episodes as for their presumably more characteristic caloric restriction, so descriptions of "normal" dieters must not neglect the disinhibitory tendencies to which their restrictive practices give rise (Polivy & Herman, 1987).

While acknowledging our bias, we believe that restraint, as we have defined it here, has proved to be an empirically useful and conceptually important construct. Ruderman (1986), who has reviewed many of the studies using the Restraint Scale, has concluded that the theory associated with this scale provides the best explanation for the patterns of restriction and overindulgence often seen in the laboratory, as well as clinically. (These patterns, it should be mentioned, were not merely explained in terms of restraint; before the restraint construct was introduced, the very existence of these patterns was barely acknowledged, at least by experimentalists.) Ruderman, in her review, disputed the applicability of the Restraint Scale to the obese; but we have addressed this concern at sufficient length earlier in this article.

The ultimate question of concern is that of which scale is appropriate for measuring restraint. The answer, we believe, depends critically upon the questions that researchers are attempting to answer. If the research focus involves the effects of simply restricting food intake, then the appropriate subscales of the TFEQ or DEBQ probably provide adequate measure-
THE (MIS)MEASUREMENT OF RESTRAINT

4 If the concern of the research, however, is to broaden our knowledge of behaviors (such as counterregulation) that characterize dieters irrespective of their extent of weight loss, then the Restraint Scale must be regarded as the instrument of choice.

Unless the current conceptual confusion is resolved, there is no doubt that the field of dieting research will continue to suffer. It is the responsibility of researchers to make clear what types of dieting issues they are addressing and to select their measures accordingly.

4 We cannot wholeheartedly recommend the use of either the Dutch Eating Behavior Questionnaire or the Three-Factor Eating Questionnaire until full behavioral validation has been provided.

References


Received June 17, 1987
Revision received August 18, 1987
Accepted August 20, 1987

---

**Delworth Appointed Editor of Professional Psychology: Research and Practice, 1989–1994**

The Publications and Communications Board of the American Psychological Association announces the appointment of Ursula M. Delworth, University of Iowa, as editor of *Professional Psychology: Research and Practice* for a 6-year term beginning in 1989. As of January 1, 1988, manuscripts should be directed to

Ursula Delworth
University of Iowa
College of Education
360 LC
Iowa City, Iowa 52242

Manuscript submission patterns for *Professional Psychology: Research and Practice* make the precise date of completion of the 1988 volume uncertain. The current editor, Norman Abeles, will receive and consider manuscripts until December 31, 1987. Should the 1988 volume be completed before that date, manuscripts will be redirected to Delworth for consideration in the 1989 volume.