
Meta-Analysis of Experimental Manipulations: Some Factors Affecting the Velten Mood Induction Procedure

Randy J. Larsen
University of Michigan

Laura M. Sinnott
Grinnell College

Manipulation validity may be influenced by many factors, including specific aspects of experimental procedure. Meta-analysis is an ideal tool for assessing factors that influence manipulations because it is capable of equating, combining, and gauging the impact of different experimental procedures on a given manipulation. This use of meta-analysis is demonstrated by meta-analyzing experiments employing the Velten mood manipulation. This manipulation has been criticized for creating experimental demand. Procedures relevant to experimental demand include whether an honest or deceptive cover story is used and whether a self-report or non-self-report manipulation check measure is used. Effect sizes are smaller when a deceptive (vs. honest) cover story is used and when a non-self-report (vs. self-report) manipulation check measure is used. Discussion concerns how the Velten manipulation may be compromised and how meta-analysis may be employed to evaluate factors influencing manipulation validity.

Most researchers agree that concerns about validity apply to psychological measurement. The same concerns apply equally to the issue of psychological manipulations. In both cases—measurement and manipulation validity—it is assumed that some empirical operation of the researcher is connected to some theoretical entity. We assume, for example, that measurements on the Wechsler Adult Intelligence Scale (empirical operation) covary with intelligence (theoretical entity). Similarly, we might assume that having subjects read various positive or negative emotional statements (empirical operation known as the Velten mood induction procedure; Velten, 1968)¹ covaries with subjects' moods (theoretical entity). With measurement validity the assumption is that some theoretical entity causes covariation in the empirical-level operation (Messick, 1980). With manipulation va-

lidity the assumption is that the empirical-level operation (i.e., the manipulation) causes covariation in some theoretical entity. In both cases, the issue of validity refers to the adequacy of the link between the empirical operation of the researcher and the theoretical entity under investigation.

MANIPULATION VALIDITY

Expanding on Cronbach and Meehl's (1955) discussion of measurement validity, the scientific meaning of a manipulation is elaborated and made explicit by building a nomological network of theoretical and empirical relations around it. The meaning of a manipulation is then conveyed by the network of assertions in which it appears. These assertions posit relations between theoretical and observable terms and represent hypotheses to be tested. The minimal nomological network consists of two observable terms and a theoretical term—for

Authors' Note: This work was supported, in part, by Research Scientist Development Award #KO1-MH00704 and Grant #RO1-MH42057 from the National Institute of Mental Health to the first author and a David Ross Award from the Purdue Research Foundation to the second author. We would like to thank Alice Eagly for helpful advice during the conception of this project and Blair Johnson for access to specialized software that made the project more manageable. We would also like to thank Margaret Kasimatis for assistance in the coding of studies and Laura Spera for programming a database for the articles used in this meta-analysis. Finally, we are indebted to Norman Miller, Brian Mullen, and seven anonymous reviewers who provided extensive and extremely helpful comments and advice on earlier versions of this article. Correspondence concerning this article should be addressed to Randy J. Larsen, Department of Psychology, 580 Union Drive, University of Michigan, Ann Arbor, MI 48109-1346.

PSPB, Vol. 17 No. 3, June 1991 323-334
© 1991 by the Society for Personality and Social Psychology, Inc.

example, $O_1 \rightarrow T \leftarrow O_2$. Here the hypothesis to be tested is whether or not a relation exists between observables O_1 and O_2 , assuming that they are both tied in some manner to the theoretical construct (T). O_1 might be a mood manipulation technique, such as the Velten procedure. O_2 might be a measure of current mood. O_1 and O_2 may be empirically related because both are tied to the theoretical construct of mood, or they may be unrelated because one (or both) is not tied to mood.

What influences the probability of obtaining an effect from an experimental manipulation? Meehl (1978) argues that three categories of concerns influence the probability of finding an effect: (a) theoretical concerns, (b) measurement concerns, and (c) concerns over the auxiliary conditions of the experiment. An effect may not be found because the theory that posits the effect is incorrect. An effect may not be found because the measures used are inadequate. Or an effect may not be found because the conditions of the experimental procedure do not sufficiently represent an adequate application of the manipulation. In considering the Velten manipulation, we will focus not so much on the first two concerns above² as on the third concern: Are there specific procedural conditions that compromise the validity of the Velten mood induction procedure?

What conditions might threaten the validity of the Velten mood manipulation? Several emotion researchers have been concerned with how demand characteristics³ created by experimental conditions might lessen the probability that the Velten mood induction procedure actually produces changes in mood per se (e.g., Berkowitz & Troccoli, 1986; Buchwald, Strack, & Coyne, 1981; Clark, 1983; Coleman, 1975; Isen, 1984; Isen & Gorgoglione, 1983; Kenealy, 1986; Polivy & Doyle, 1980; Thayer, 1989; Wispe, 1980). The Velten procedure is rather transparent. As Kenealy (1986) points out, it seems hard to believe that, after being asked to read such statements as "I feel happy" and "It's great to be alive" and then being asked to rate how "happy" he or she is "at this moment," the subject would not be able to infer that the experimenter expects the manipulation to alter the subject's mood. The intent of the manipulation might be clear enough to subjects so that they apprehend the expectation that these statements should influence their mood. If the experimenter further made this expectation clear by explicitly informing subjects that the *intent* of the statements was to influence their mood, then such an honest cover story might compound the demand characteristics in the experiment. Alternatively, if the experimenter used a deceptive cover story to disguise the intent of the Velten statements, then that might lessen the demand engendered by this manipulation.

Another aspect of the experiment that may influence the validity of the Velten technique is whether the effects are obtained only with self-report manipulation check measures. It has been argued that self-report mood measures are more susceptible to faking (Clark, 1983) and thus more amenable to influence by demand characteristics than non-self-report measures. If the Velten manipulation produces mood effects on both self-report and non-self-report manipulation check measures, then the scientific judgment that this manipulation actually influences mood would receive some support.

In this article we will use meta-analysis to focus on how these two features of experiments (type of cover story used and type of manipulation check measure) may relate to the impact of the Velten mood induction procedure. Our purpose is to examine and compare studies that have employed the Velten technique under these conditions. We argue that these conditions are differentially related to experimental demand characteristics. If the manipulation effect sizes in the Velten literature are found to vary as a function of these aspects of experimental procedure, then we are in a position to comment on the Velten manipulation vis-à-vis factors related to demand characteristics.

We do not mean to imply that the impact of the Velten manipulation is equivalent to its validity. Rather, impact is a *necessary* but not sufficient condition for demonstrating the validity of a manipulation. If a manipulation yields a significant effect on a manipulation check measure, this does not *necessarily* demonstrate the validity of that manipulation.⁴ The psychological meaning of the manipulation may still be unclear. If one is attempting to induce a depressed mood, for example, then it must be shown that it was depressed mood and not some other psychological process (e.g., demand characteristics, anxious mood, lowered positive affect) that produced the effect on the manipulation check measure. In other words, discriminant validity must be evaluated in order to thoroughly establish the construct validity of a manipulation. Nevertheless, because impact is a necessary condition for validity, we will focus our meta-analysis on the effect sizes of manipulation check measures in studies that employ the Velten manipulation under various conditions relevant to experimental demand.

META-ANALYSIS OF MANIPULATION STRENGTH

Meta-analysis is an ideal tool for assessing factors that may affect experimental manipulations⁵ for at least three reasons. First, it uses a common metric capable of equating and combining the results of many studies that employ a given manipulation and include some form of

manipulation check measure. Second, it can assess the effects of variations in experimental procedure on the strength of the manipulation. Finally, meta-analysis describes and summarizes how a given manipulation has been used within a research area. This descriptive information may be useful to researchers working with the theoretical construct that the manipulation supposedly influences. For example, it would be useful for researchers interested in computing power analyses when preparing experimental investigations (Cohen, 1977). Additionally, such information would provide an overview of procedural variations in how a given manipulation is employed.

In using meta-analysis to evaluate the impact of the Velten mood manipulation, we begin with the assumption that this manipulation potentially *does* have an effect. That is why it is being used by researchers. What needs to be determined, however, is whether or not variation in *how* it is applied influences the size of its effect.

THE VELTEN TECHNIQUE FOR MOOD MANIPULATION

Because mood is thought to influence a wide range of behaviors, a large number of studies in diverse areas of research employ the Velten manipulation in a variety of ways. In our search of the general mood induction literature we recovered a total of 341 publications, most from mainstream psychology journals but others from journals outside of psychology. Although studies employing laboratory mood manipulations are quite numerous and scattered, conditions that influence the strength of these techniques are not well understood. Leventhal and Tomarken (1986, p. 596) state that "the induction of mood states is perhaps the major methodological problem" that emotion researchers face.

We should make it clear that we will *not* be evaluating the impact of the Velten manipulation on some theoretical construct, such as the impact of Velten-induced mood on memory (e.g., Blaney, 1986). This would be a traditional application of meta-analysis in theory testing by means of integration of mean effect sizes. Other meta-analyses demonstrate such a theory-testing approach, such as Carlson, Charlin, and Miller's (1988) meta-analysis of the positive mood and helping literature. Our objective is to demonstrate an application of meta-analysis to evaluating the conditions that influence manipulation strength; therefore, we will include studies of the impact of the Velten manipulation on *any* theoretical construct. The only requirement is that the study include an independent manipulation check on the Velten mood induction procedure. We will consequently be examining the

effect size of the Velten mood manipulation on the manipulation check measure, not the effect size of the manipulation on the theoretical construct under investigation.

METHOD

Sample

The search for articles containing laboratory studies employing the Velten mood induction procedure consisted of three stages. First, a computer search of *Psychological Abstracts* was conducted for the years 1950 through 1989 using the key words *mood*, *emotion*, or *affect* in conjunction with *induction* or *manipulation*. The second phase was a hand search of journals known to publish work on mood-related phenomena. This search was conducted for the years 1985 through 1989 of *Journal of Personality and Social Psychology*, *Personality and Social Psychology Bulletin*, *Journal of Personality*, *Psychophysiology*, *Motivation and Emotion*, *Archives of General Psychiatry*, *Journal of Research in Personality*, and *Journal of Experimental Social Psychology*. Finally, an ancestry search of major articles was performed. Applying this sampling method yielded 341 published articles.

Articles considered relevant for the present meta-analysis were those published in the English language, with data gathered on human, nonclinical samples. Studies that did not report the use of a control or comparison group, did not include a manipulation check measure,⁶ or did not report adequate statistics for our computation of effect sizes were excluded from our analyses as well.

Study Characteristics

The general purpose of our meta-analysis is to account for the variance in the strength of the Velten mood induction procedure across primary research studies. For our purpose, this entails clarifying the relationship between the Velten technique and certain features of the experiments. To do this, it is necessary to devise a coding scheme that captures those aspects of the experiment believed to influence the impact of the Velten manipulation. These coding categories then serve as blocking factors to partition variance in the Velten manipulation effect sizes. We coded the studies along the following four dimensions: (a) whether the Velten technique or some other mood induction procedure was used, (b) within the Velten category, whether the standard or some modified form of the Velten technique was used, (c) the type of cover story used, and (d) the type of manipulation check measure used.

Type of induction procedure employed. Researchers have used a number of techniques to experimentally manip-

ulate moods. In addition to the Velten technique, other techniques commonly used include facial (or postural) posing, false feedback, films and slides, free gifts, imagery, music, suggestion, and hypnosis. Within studies employing the general Velten-type procedure a category entitled *modified Velten* was created to encompass those studies that adapted the Velten procedure in some manner, such as using only a subset of the original mood-inducing statements or administering the technique in a group setting where the subjects read the mood induction statements silently to themselves. Excellent reviews of the procedures and details of various types of mood induction techniques are provided by Goodwin and Williams (1982). A more recent discussion of mood induction procedures in general is provided by Singer and Salovey (1988).

Use of a cover story. As discussed above, research employing the Velten technique has been criticized for its susceptibility to demand characteristics. We therefore coded whether experimenters employed deception in their manipulations of mood or were honest in informing subjects that the intent of the Velten statements was to manipulate mood. This enabled us to determine whether differences in effect sizes would vary as a function of whether the experimenter made his or her expectations explicitly known to the subject through an honest cover story (i.e., telling the subject that the purpose of the manipulation is to induce an emotion). An honest cover story in a mood induction experiment can make explicit the expectations of the experimenter, thus contributing to demand characteristics.

Type of manipulation check. Self-report questionnaires, such as mood adjective rating scales, are widely used as manipulation check measures. However, a number of studies have employed non-self-report manipulation check measures, such as subjective probability estimates of positive and negative events, psychomotor tasks, physiological measures, or ratings of facial or expressed emotion. These coding categories were derived from narrative reviews of manipulation check measures used in mood induction research (Clark, 1983; Singer & Salovey, 1988).⁷ Different types of manipulation check measures vary with respect to the response system they assess, their susceptibility to influence by experimental demand characteristics, and their fakeability.⁸ It has been suggested that self-report measures are most susceptible to demand characteristics (Clark, 1983). We therefore coded for the type of manipulation check measure to see whether effect sizes associated with the Velten technique varied as a function of whether a self-report or non-self-report manipulation check measure was used.

Calculation of Effect Sizes

As indicated, the index of mood manipulation strength is the effect size. An effect size is interpreted as the standardized mean difference between two comparison groups. Our analyses use the *d* statistic (Cohen, 1977; Hedges & Olkin, 1985), which is the effect size *g* (Glass, McGaw, & Smith, 1981) corrected for the sample size on which each effect is based. Effect sizes were computed using means and pooled standard deviations. When these were not reported, *t*s and *F*s were converted to effect sizes.⁹

Most studies report mood manipulation results for more than one comparison. For example, it is not uncommon for a study to have an analysis of variance design with induction conditions defined by three types of Velten statements: a series of positive Velten statements, a series of negative Velten statements, and a series of neutral Velten statements. In such cases two effect sizes can be computed, one comparing the positive induction group with the neutral induction control group and another comparing the negative induction group with the neutral induction control group.¹⁰ Authors also routinely use more than one type of manipulation check measure, allowing a set of effect sizes to be computed for each comparison. This redundancy in data points introduces a small amount of statistical bias into the analyses (Glass et al., 1981). However, we felt that the information sacrificed in order to avoid redundancy would even further bias or limit our results. We therefore decided to include all reported effects in this meta-analysis.

Groups used for comparisons in effect size computations varied across studies. Most comparisons examined mood manipulation groups versus neutral induction control groups. However, other studies compared preinduction with postinduction measures of affect. When effect sizes could not be computed in either of these ways and the data were available, comparisons were made between positive and negative mood manipulations groups.¹¹ In all cases, a positive-signed effect size indicates the presence of the intended mood manipulation effect.

Categorical meta-analytic models allow the statistical comparison of effects across categories of experiment characteristics using homogeneity statistics similar to analyses of variance (Hedges, 1982; Rosenthal & Rubin, 1982). To explain the notation used in this article, H_t is the total variation in effect sizes. H_b represents between-category variation. Overall category variation within groups is represented by H_w , whereas H_{wi} represents the variation within each group or characteristic of a category. In comparing the impact of different manipulation procedures, an overall H_t was computed. This was bro-

ken into components, H_b and H_w . Using a χ^2 distribution, a value of H_b larger than the critical value would indicate a significant difference among the various procedures. When the between-groups effect was significant, post hoc tests were computed to specify the nature of the significant effects.

RESULTS

Effect Size Analyses of Velten Mood Induction Procedure

Of the 841 mood manipulation effect sizes computed over *all* of the mood induction studies in our sample, the Velten and modified Velten techniques are the most commonly used, accounting for 45% of the effect sizes obtained. Out of the 62 articles reporting a Velten-type technique¹² we obtained a total of 380 effect sizes that were based on 16,599 observations. Across the total of 380 individual effect sizes, the average was .7596, $p < .05$. This indicates that, on average, Velten-type mood induction groups scored 0.76 *SD* higher on manipulation check measures than the comparison/control groups.

We then partitioned the effect sizes into those obtained with the standard Velten technique and those obtained when it was modified and compared them. This contrast revealed that the standard Velten procedure results in a slightly larger manipulation effect ($d = .8819$) than the modified forms of the Velten procedure ($d = .6745$), $\chi^2(1) = 3.98$, $p < .05$. This is particularly interesting because the standard and the modified Velten methods are conceptually similar but procedurally different, suggesting that changing the procedural conditions of the experiment appears to alter the efficacy of the Velten technique.

Effect Size Analyses Within Velten Mood Induction Techniques

Studies using some modified form of the Velten procedure result in mood manipulation effects that are somewhat weaker than those obtained in studies using the standard Velten technique. Selecting only those effect sizes obtained with studies using the standard Velten and modified Velten procedures and further partitioning the variance in effect sizes according to other experiment characteristics may provide some insight into the cause of this difference. Remember that these two mood induction techniques (the Velten and modified Velten) are conceptually similar. Hence, the possibility exists that characteristics of experimental procedure may influence the impact of the Velten manipulation. The characteristics relevant to our purposes were (a) the use of an explicit cover story prior to the experimental manipulation of mood and (b) the type of manipulation

TABLE 1: The Velten Technique: Frequencies and Effect Sizes by Cover Story Use

<i>Technique</i>	<i>Cover Story Use</i>	<i>d</i>	<i>n</i>	<i>Percentage of Total</i>
Velten	Deceptive	.46*	8	4.6
	None reported	.89*	106	60.9
	Honest	.89*	60	34.5
				100.0
Modified Velten	Deceptive	.55*	70	34.0
	None reported	.70*	119	57.8
	Honest	1.04*	17	8.2
				100.0

* $p < .05$.

check measure used.

Use of a cover story. We coded each study on cover story use by noting when investigators reported attempts to deceive subjects about the purpose of the experimental manipulation and when the experimenters were explicitly honest with the subjects about the intent of the manipulation (i.e., to alter mood). A third category was used (*no cover story reported*) when authors did not mention whether a cover story was presented to the subjects. Across both applications of the Velten procedure (standard and modified), almost two thirds of the mood effects fell into this third category. Interestingly, only a small number of investigators reported using deception when the standard Velten procedure was employed, whereas researchers who modified the Velten procedure were more likely to report the use of deception. This pattern was reversed with respect to reported use of honest cover stories (see Table 1 for frequency counts). One possible explanation for this difference in experimental procedures is that investigators who modified the Velten technique for their own use also made more efforts at deception in order to avoid the experimental demand characteristics for which the standard Velten procedure has been criticized (Buchwald et al., 1981; Polivy & Doyle, 1980).

An examination of both applications of the Velten technique reveals an interesting, albeit unsurprising, trend in the strength of mood manipulation effect sizes. These are presented in Table 1. In both cases, effects are weakest when the subjects were deceived and strongest when the subjects were explicitly told that the purpose of the manipulation was to induce the intended emotion. The difference between effect sizes for these two procedural variations on cover story use could be interpreted as the result of demand characteristics. In other

words, about one half of the mood manipulation effect associated with the use of honest presentation to subjects of the experimenter's expectations might be the result of subjects' cooperation with explicit experimenter demands and not due to actual mood changes per se. Effect sizes for those cases when investigators failed to report cover story use fell between the two extremes of deceptive and honest cover story use. Again, this is not surprising and most likely indicates that this average effect represents a combination of deception and honesty with experimental participants. It is worth noting that authors should report this auxiliary procedural detail, as it indeed may be important in interpreting experimental outcome or in evaluating the link between the Velten manipulation and the theoretical construct of mood.

The between-category effect of cover story use was significant for both the standard and modified versions of the Velten technique, $H_b(2) = 7.56$ and 36.70 , respectively, both $ps < .02$. Post hoc contrasts showed that effects associated with deceptive versus honest cover stories differed for both the standard Velten and modified Velten methods, $\chi^2(1) = 7.31$, $p < .01$, and $\chi^2(1) = 35.02$, $p < .001$, respectively. Contrasts performed to examine differences in mood effects between the two Velten applications for the three categories of cover story use are presented in Table 2. Only when no cover story was reported did the effect sizes differ significantly; the Velten technique resulted in stronger mood induction effects in this category of studies. Given the difference between effects associated with honest and deceptive cover stories using the standard Velten procedure, this latter result might be due to the obvious nature of the standard Velten mood induction technique. Another possible explanation is that the standard administration of the Velten technique, as described by Velten (1968), provides a form of suggestion to the subject. In the standard Velten manipulation subjects are told that as they read the statements, they should "try to feel the mood" suggested by those statements. Although this is not tantamount to an explicitly honest cover story, these instructions nevertheless may be equivalent to the experimenter's implicitly revealing his or her expectations for the manipulation to the subject. This lessens the difference between the honest-cover-story and the no-cover-story-reported conditions for the standard Velten administrations. Another possible explanation for the larger effect of the standard Velten than the modified Velten in the experiments where no cover story was reported is that many of the modifications consist in having subjects read fewer of the emotion statements (e.g., 10 statements instead of the standard 60 in some studies). The fact that subjects read more emotion statements in the unmodified version of the Velten manipu-

TABLE 2: The Velten Technique Versus the Modified Velten Technique: Post Hoc Comparisons for Cover Story Use

Cover Story Use	χ^2	df
Deceptive	.28	1
None reported	21.04*	1
Honest	2.58	1

* $p < .05$.

TABLE 3: The Velten Technique: Frequencies and Effect Sizes by Type of Manipulation Check Measure Used

Technique	Type of Manipulation Check	d	n	Percentage of Total
Velten	Self-report	1.16*	96	55.2
	Subjective probability task	.23	7	4.0
	Ratings of expressed emotion	.46*	9	5.2
	Psychomotor tasks	.60*	58	33.3
	Physiological measures	.00	4	2.3
				100.0
Modified Velten	Self-report	.70*	181	87.9
	Subjective probability task	1.04*	4	1.9
	Ratings of expressed emotion	.42	2	1.0
	Psychomotor tasks	.41*	15	7.3
	Physiological measures	.52	4	1.9
				100.0

* $p < .05$.

lation may result in a stronger induction due to a larger dose of the manipulation.

Type of manipulation check measure. Mood induction effect sizes vary in how they are assessed by primary researchers. Table 3 displays the frequencies of each of five types of manipulation check measures when the Velten techniques (either standard or modified) are employed. In each case, self-reported mood manipulation check measures are the most frequently used, followed by psychomotor tasks.

Effect size analyses indicate that self-report measures of Velten-manipulated mood yield stronger effects than all other measures, with the exception of the average of four effect sizes from the modified Velten category (subjective probability task cell, see Table 3). Although this cell yields a large effect, it is based on only four effect sizes and so might not be truly representative of mood effect sizes obtained through this particular combination of manipulation and manipulation check measure. Mood manipulations measured by psychomotor tasks exhibit moderate and significant effects for both categories of Velten techniques, as do manipulation check measures that employ ratings of expressed emotion,

although this is significant only for the standard Velten procedure. Also noteworthy is the finding that the four effect sizes associated with the standard Velten technique used in combination with physiological manipulation check measures yielded an average effect size of zero. But again, this is based on only a small number of effect sizes in this cell.

Again, the between-class statistic was significant for both the standard Velten and the modified Velten techniques, $H_b(4) = 170.58$ and 27.09 , respectively, both $ps < .001$. Because the Velten procedure has been criticized for its potential to alert subjects to the experimenter's expectations, we thought that self-report measures would be the type of manipulation check measure most likely to yield strong induction effects. Therefore, in performing post hoc contrasts, we were particularly interested in the comparison between self-report and all other manipulation check measures combined. The data somewhat support this notion; self-report manipulation check measures do yield stronger mood effects than other measures, though only when the standard Velten procedure is employed, $\chi^2(4) = 140.49$, $p < .001$. For the modified Velten procedure this contrast was not significant, $\chi^2(4) = 1.89$.

Self-report manipulation checks by cover story use. When we partitioned the variance in mood manipulation effect sizes on the basis of cover story use and type of manipulation check measure, we found strong induction effects when they were assessed by self-report measures in combination with honest cover stories. Because the combination of these two experimental procedures (honest cover story plus a self-report manipulation check) seems to especially allow subjects to respond to experimental demand characteristics, we thought it would be interesting to further partition mood induction effect size variation by cover story use when only self-report manipulation check measures were used. The number of mood effects and the mean effect size for each category are presented in Table 4. The most striking aspect to notice is the way effect sizes in the no-cover-story-reported and honest categories increase when examining only the effect sizes assessed by self-report measures. This difference in effect sizes associated with the use of self-report measures can be most readily seen by comparing the effect sizes in Table 1 with the effect sizes in Table 4.

The between-groups effect is significant for both the standard Velten and the modified Velten procedures, $H_b(2) = 25.48$ and 63.36 , respectively, both $ps < .001$. Post hoc contrasts comparing each type of cover story use with the other types combined, within each Velten category, were all significant at $p < .05$. Because Velten versus modified Velten did not interact with the effects of type

TABLE 4: The Velten Technique With Self-Report Manipulation Check Measures: Frequencies and Effect Sizes by Cover Story Use

Technique	Cover Story Use	Self-Report		Percentage of Total
		d	n	
Velten	Deceptive	.46*	8	8.3
	None reported	1.25*	52	54.2
	Honest	1.10*	36	37.5
				100.0
Modified Velten	Deceptive	.54*	67	37.0
	None reported	.73*	102	56.4
	Honest	1.28*	12	6.6
				100.0

* $p < .05$.

of cover story used, we collapsed across the Velten categories and present mean effect sizes in Figure 1. This figure presents the effect sizes obtained with only self-report measures and those obtained with only the non-self-report measures. Post hoc contrasts compared deceptive with honest cover story effect sizes for self-report and non-self-report measures separately. As Figure 1 clearly illustrates, the use of deceptive versus honest cover stories makes very little difference when manipulation impact is assessed with a non-self-report measure (χ^2 is nonsignificant). However, when a self-report measure is used to assess manipulation impact, there is a strong effect for honest versus deceptive cover story, $\chi^2(1) = 55.5$, $p < .001$, such that honest cover stories are associated with significantly larger effect sizes. These findings should serve to alert investigators using the Velten-type mood inductions to the problems associated with obtaining and reporting an induction effect when using only self-report manipulation check measures. In order to assert clearly that the mood manipulation did indeed lead to changes in subjects' moods per se, consideration should be given to aspects of experimental procedure, such as use of an honest or deceptive cover story.

DISCUSSION

This article demonstrates the application of meta-analytic techniques in the evaluation of factors that influence the Velten mood manipulation. We argue that issues of validity are as relevant to psychological manipulations as they are to psychological measurement. In both cases, some empirical operation of the experimenter is assumed to reflect or alter some theoretical construct. The adequacy of the link between the empirical level and the theoretical level is what is meant by the notion of validity. In the case of manipulation validity, we

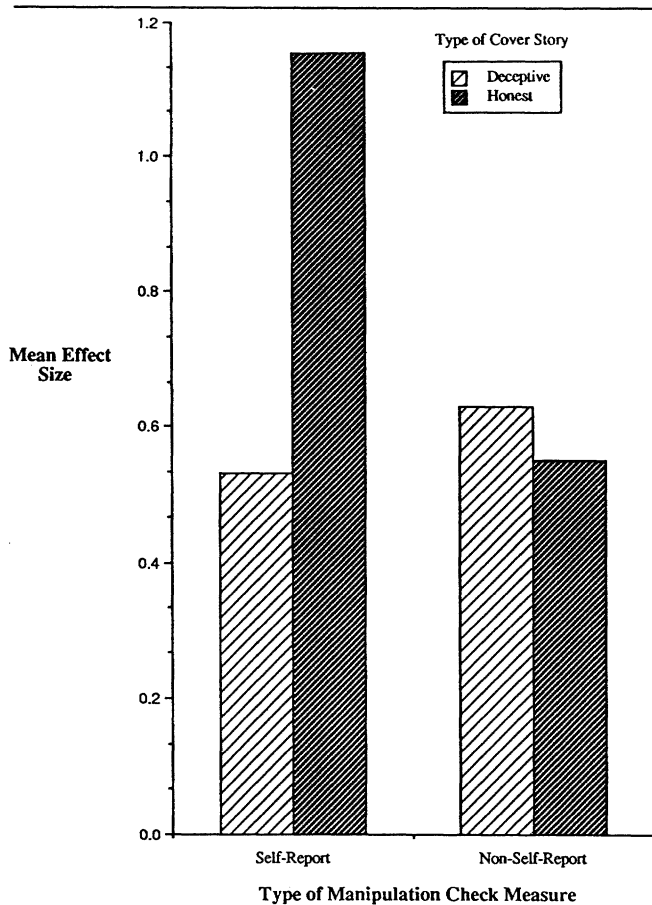


Figure 1 Combined Velten and modified Velten effect sizes by type of cover story used and type of manipulation check measure used.

start with the assumption that, in general, the manipulation does have some effect. The important question is whether variation in *how* the manipulation is applied influences the magnitude of its effect and, if so, what interpretation should be made of such variation.

In our demonstration of meta-analytic evaluation of the Velten mood induction procedure we chose a literature that is particularly large and diverse. Our purpose was to examine whether specific variations in experimental procedure moderate the impact of the Velten mood induction procedure. We focused on experimental procedures related to the validity of the Velten mood induction procedure for two reasons. First, it is the most widely used manipulation in mood induction research, accounting for almost half of the effect sizes in our sample. It thus contains the large number of effect sizes necessary to assess moderators of its strength. Second, the Velten technique has been the topic of ongoing debate concerning the effects of demand characteristics in mood induction research (e.g., Berkowitz & Troccoli, 1986; Buchwald et al., 1981).

We chose two experimental characteristics that we believed were most likely to be related to the elicitation of experimental demand. One procedural characteristic concerned whether or not the experimenter explicitly told the subjects an honest or a deceptive cover story. We felt that making the expectations of the experimenter explicitly known to the subjects through an honest cover story would serve to accentuate experimental demand characteristics. A deceptive cover story, in contrast, should serve to lessen the subjects' knowledge of the expectations of the experimenter. The second aspect of procedure that we thought related to demand characteristics was whether a self-report or non-self-report manipulation check measure was used. Self-report mood measures are ostensibly more fakeable (Clark, 1983) than non-self-report measures.

In a narrative review of the Velten literature, Kenealy (1986) concluded that self-report mood measures show consistently significant effects, whereas evidence for the behavioral effects of this manipulation is much more equivocal. Our quantitative integration is consistent with Kenealy's conclusion. Another narrative reviewer (Clark, 1983), however, concluded that the Velten technique *does* affect behavioral variables, especially those associated with depression. Despite his generally optimistic narrative review, Clark (1983) nevertheless suggests that "it is still possible that *some* subjects fake their responses, particularly on measures of self-reported [affect]" (p. 39). Our meta-analysis supports Clark's more cautiously phrased conclusion; over studies, the average effect of Velten-type manipulations is significantly smaller for non-self-report than self-report measures.

In using the Velten techniques, the largest effect sizes are obtained when using an honest cover story in conjunction with a self-report manipulation check measure. On the surface, this specific combination of experimental characteristics appears particularly susceptible to experimental demand effects, implying that the manipulation impact obtained in this manner may not be *entirely* due to mood changes per se. However, the use of a deceptive cover story in conjunction with a non-self-report manipulation check measure results in smaller, albeit still significant, manipulation effect sizes. This combination appears *most resistant* to demand, implying that there is a significant mood induction effect for the Velten techniques even when the procedures minimize demand characteristics. Our results thus contradict the conclusion of Buchwald et al. (1981) that "the use of the Velten induction procedure to manipulate mood experimentally . . . is unwarranted" (p. 478).

Our interpretation of the effects of type of manipulation check measure is that self-report measures are more susceptible to compliance with experimental demand. Another review of the Velten procedure, based on a narra-

tive examination of the literature (Berkowitz & Troccoli, 1986), concluded that it is not appreciably affected by demand characteristics, even though subjects may nevertheless exhibit a social approval motive. Berkowitz and Troccoli (1986) argue that demand compliance is different from social approval motivation, defined as a tendency of subjects to refrain from "any socially disapproved reactions to the stimuli presented to them" (p. 338). However, if it is socially disapprovable to *not* react in the manner expected of you, *and* the expectation that the Velten statements will alter your mood is transparent, then this social approval motive appears tantamount to compliance with demand. Provided that experimental procedures foster the communication of the experimenter's expectations (through an honest cover story or a suggestion to "feel the mood" of the Velten statements) *and* enable the subject to comply easily (by using a self-report manipulation check measure), then the increased effects obtained under these conditions seem most parsimoniously explained as resulting from demand.

Another interpretation is possible, however. It could be argued that self-report measures show stronger effects because they tap emotional responses more directly (i.e., are more valid measures of mood) than the non-self-report measures used in this literature (measures of emotional expression). The question of differential validity of self-report and non-self-report measures of mood, like all questions of validity, has to be answered with a judgment based on evidence and reasoning. Evidence suggests that, for example, when asked to *fake* depressed mood, the faking group scores as more depressed than a Velten depression induction group only when self-report measures are used, not when behavioral measures are used (Alloy, Abramson, & Viscusi, 1981). Instructions to fake a mood state thus appear to influence self-report measures more than non-self-report measures.

Reasoning also suggests that larger effect sizes for self-report measures are due to influence by demand characteristics. For example, in her narrative review of the Velten literature, Kenealy (1986) reasoned that "if demand characteristics are largely responsible for the effects found in studies that have used Velten's technique, then consistent self-reports of mood induction would be expected" (p. 328, emphasis added). Although we do not believe that demand characteristics "largely" account for the Velten effects, we believe that a demand explanation is the most parsimonious interpretation of why larger effects are found with honest cover stories and self-report measures than with deceptive cover stories and non-self-report measures.

Because the average effect size associated with experimental conditions that ostensibly *minimize* demand is

itself significantly different from zero, it might be best to use the term *additional effect* associated with demand characteristics. Like other reviewers who have addressed the issue of demand characteristics in the Velten technique (e.g., Buchwald et al., 1981; Goodwin & Williams, 1982; Kenealy, 1986), we conclude that demand conditions probably serve to somewhat inflate an already significant mood induction effect. We might conceptualize demand as creating a degree of additional effect on top of a true mood induction effect, a conclusion endorsed in primary research (Polivy & Doyle, 1980).

Beyond these conclusions regarding the Velten mood induction procedure, we wish to return to a discussion of applying meta-analysis to a literature employing a given manipulation. One limitation of meta-analysis is that it necessarily combines effect sizes from primary studies that differ in ways unrelated to the coding categories under investigation. For example, within our category of self-report manipulation check measures, we combined effects from studies that used different types of self-report measures, ranging from standard adjective checklists and rating scales (e.g., the Profile of Mood States [POMS], the Mood Affect Adjective Checklist [MAACL]) to study-specific mood scales. It is likely that these different self-report manipulation check measures differ in their sensitivity to mood changes, a difference averaged over in our application of meta-analysis. Such a limitation, however, applies to all meta-analyses where the effect sizes are based on nonidentical measures, which includes the majority of meta-analyses. In interpreting the results from any meta-analysis one must keep in mind what is being lost along with what is being gained through integration.

On the more positive side, meta-analysis brings to light important methodological issues involving the manipulation. In studies that involve mood manipulations (or any manipulation, for that matter) there are two conceptually separate aspects to the analyses. One aspect is to establish whether the manipulation was successful. The second aspect is to examine the dependent variable of interest to determine how it is influenced by the manipulated variable (e.g., mood). We agree with Kenealy's (1986) statement that an "important aspect of mood induction research . . . is *ascertaining whether the mood manipulations have been successful* and that subjects' mood states are different in each of the experimental groups before proceeding to study the behavior under investigation" (pp. 319-322, emphasis added). Some researchers (e.g., Clark, 1983; Teasdale & Taylor, 1981) establish a criterion of minimal mood change that the subjects must meet before their data are included in the analyses of the dependent variables of interest. Having subjects meet some minimal mood induction effect before in-

cluding their data from the experiment proper can be a strong form of ensuring manipulation validity but may threaten the internal validity of the experiment.¹³

A meta-analysis of a manipulation literature may be instructive in how primary researchers might best proceed in experimental research. For example, an awareness of how the cover story influences the size of the effect, especially in conjunction with a self-report manipulation check measure, should guide researchers in evaluating results in light of the particular procedural variations used in the experiment. Experimenters worried about demand characteristics in the Velten procedure might try various techniques to lessen or at least detect demand. For example, subjects might be instructed to base their self-report mood ratings on "how they actually feel rather than on how they think the experimenter would like them to feel" (Clark, 1983, p. 46). Alternatively, researchers might use a debriefing questionnaire, administered after the experiment is "officially" over and remuneration has been given, that emphasizes honesty and inquires whether or not the subjects believed that the induction did in fact alter their mood. Another technique for reducing demand is to lead subjects to believe that the experiment is over and then gather observations on the effects of the manipulation. These latter two options should be used only after consideration of the ethical issues involved.

Meta-analysis can also suggest useful lines of inquiry for primary research concerning a given manipulation. For example, it would be useful to have studies that directly compare the Velten technique with other mood manipulations (e.g., imagery, films) within the same sample, where all other procedural conditions are held constant. Similarly, studies that compare different manipulation check measures under identical experimental conditions would be useful in identifying the most reliable, valid, and sensitive measures of manipulated mood. Additionally, primary researchers might want to compare manipulation check measures that are more closely related to the induction with those that are conceptually distinct from it. For example, the Velten technique has subjects verbalize statements regarding mood, and self-report manipulation check measures then have subjects endorse verbal statements regarding mood. Comparing effects with this manipulation check against those obtained with a manipulation check measure that is conceptually distinct from the Velten manipulation (i.e., nonverbal) may uncover something like a method effect.

Meta-analysis may also bring to light important theoretical questions that are implicit in a literature on a given manipulation. For example, primary researchers using the Velten manipulation almost never discuss a

theoretical rationale for how and why this technique actually works to produce changes in emotional state. What are the specific psychological processes through which the Velten procedure is affectogenic? Theories of emotion might suggest likely processes for explaining how the Velten manipulation actually works (e.g., spreading activation of mood in memory structures). Specific Velten studies might vary in ways that are differentially relevant to theoretical processes hypothesized to be responsible for mood change (e.g., how many Velten statements are read might influence the degree of spreading activation). A meta-analysis that initially focused on factors that influence the impact of manipulation compiles information that may prove useful for examining other important theoretical questions about the construct being manipulated.

NOTES

1. The standard Velten mood manipulation is to have subjects first read silently, then aloud, at their own pace, 60 self-referent statements that contain either elated or depressed content. As described by Velten (1968), subjects are instructed to "try to feel the mood suggested by the statements" (p. 474). An example of an elated statement is "This is great, I really do feel good right now, I am elated about things." An example of a depressive statement is "I have too many bad things in my life." Velten (1968) also provided 60 affectively neutral, non-self-referent statements to use for control group subjects. An example of a neutral statement is "This book or any part thereof must not be reproduced in any form."

2. Certainly the effect size of a manipulation is influenced both by the strength of the manipulation and by the validity and sensitivity of the manipulation check measure. Ideally, in order to evaluate the former, one would want a collection of primary studies that use perfectly valid and equally sensitive manipulation check measures. In lieu of this ideal situation, one would at least desire a subset of primary studies that use the same manipulation check measure. Because our application of meta-analysis concerns experimental procedures that influence the validity of the Velten manipulation, we will necessarily aggregate over different manipulation check measures within our coding categories of interest. This is not different from other meta-analyses, where many types of dependent variables, all supposedly measuring the same construct, are aggregated to yield an average effect size. Nevertheless, if one were interested in differential validity or sensitivity of specific forms of dependent variables or specific manipulation check measures, then these distinctions could form coding categories within a separate meta-analysis. We did, in fact, attempt to analyze our collection of effect sizes by looking for subsets of effects within our coding categories that were based on identical manipulation check measures. The largest number of effect sizes with a manipulation check measure in common was 10 effect sizes based on the Mood Affect Adjective Checklist in the standard Velten category, which grew even smaller once we broke these down by cover story use. Consequently, we judged that the number of effect sizes based on common manipulation check measures was too small to yield a reliable comparison of specific manipulation check measures. We therefore aggregated over different types of specific manipulation check measures in order to more reliably analyze the effects of the procedural conditions of interest.

3. Demand characteristics are defined by Orne (1962) as the sum total of cues that convey the experimental hypothesis to the subject and thus potentially influence the subject's behavior. In applying the concept to manipulations, we use the notion of demand in a slightly

different way, to refer to the cues that convey the intent or expected action of the manipulation to the subject. Manipulation demand increases the possibility that subject performance on a manipulation check measure is due, not entirely to true effects of the manipulation, but to subject compliance with experimenter expectations. Any aspect of the experiment that contributes to the possibility that subjects may simply comply with experimenter expectations is thus called a demand characteristic.

4. It is possible for a mood manipulation to have an insignificant effect on a manipulation check measure and yet a significant effect on the dependent variable of interest (e.g., Schwarz & Clore, 1983). Such findings leave the researcher with difficult questions: Is the dependent measure more reliable or more sensitive to mood changes than the manipulation check measure? Is the effect on the dependent variable due to something other than mood? Is the timing of the manipulation check measure important in assessing mood change due to laboratory mood inductions?

5. We do not mean to imply that our application of meta-analysis is qualitatively different from previous applications of these techniques. We are using meta-analysis to examine central tendency, variability, and the prediction/explanation of effect sizes, just as in most typical applications of meta-analytic techniques. The uniqueness of our study lies, not in the technology of our application, but in its focus on the factors that influence an experimental manipulation.

6. Many field studies that use some naturalistic form of mood manipulation (e.g., finding a dime in a phone booth) do not include a manipulation check measure and therefore are not represented in this meta-analysis.

7. The distinction between what is a manipulation check measure and what is a dependent variable is occasionally blurred. In some studies, for example, psychomotor tasks were used as manipulation checks, whereas in other studies identical tasks formed the sole dependent variable of the experiment. We decided on the stated categorization of manipulation check measures a priori before coding began, on the basis of narrative reviews of the literature and our own cursory reading of portions of this literature. Thus, we did not simply take the authors' classification of their variables (as manipulation check or as dependent variable) at face value. Instead, if a variable fell into one of our a priori categories of manipulation check variables, we computed an effect size, regardless of whether the authors called the variable a manipulation check measure.

8. A related but separate issue concerns the validity of manipulation check measures themselves. Although self-report measures are most likely vulnerable to demand characteristics, non-self-report measures may have a more indirect relation to mood than self-report measures. That is, different types of manipulation check measures are likely to be differentially sensitive to mood changes. This could be the topic of a separate meta-analysis.

9. When primary researchers reported that an effect was nonsignificant, it was assumed to be zero and included in the analyses. This is a conservative procedure, as it most likely serves to lower the overall effects. In our sample of all Velten and modified Velten studies, 11% of the effect sizes were reported as nonsignificant and assumed to be zero in our analyses. All analyses reported in this article were done both with and without the assumed-zero effect sizes. None of our results changed in any important ways as a function of deleting the assumed-zero effect sizes. That is, significance levels of our results did not change as a function of deleting the assumed-zero effects; categorical effects that were significant or insignificant when the assumed-zero effect sizes were included were still significant or insignificant (respectively) when the assumed-zero effect sizes were deleted from analyses. The reason is most likely that the assumed-zero effect sizes were spread fairly evenly across our coding categories. We report the results obtained by including the assumed-zero effect sizes, as this is the more conservative approach to this issue and it allows a larger number of effect sizes to be included in our comparisons.

10. When we collapsed over coding categories, there was no significant main effect for whether the mood induction was positive or negative. Within coding categories, the effect of type of induction

(positive or negative) remained insignificant for all categories except the standard Velten assessed with self-report manipulation check measures. Here there was a small main effect for type of induction, suggesting a small disordinal interaction between type of induction and this particular coding category. However, because this interaction effect is much smaller than the main effect for either the Velten versus modified Velten comparison or the self-report versus non-self-report comparison, averaging over type of induction does not cover up any large differential effects due to type of mood induced.

11. Effect sizes based on comparisons between affect and control groups are conceptually different from those based on comparisons between positive induction and negative induction groups. This distinction may play an important role in determining whether experiment effects are due to increases in positive affect, decreases in negative affect, increases in negative affect, or decreases in positive affect. Indeed, this issue is worthy of a meta-analysis in its own right. However, in examining factors that influence the Velten technique, the problem of having to separately analyze comparisons of positive and negative groups is not such a large problem, as most of the researchers using this technique include a control group, perhaps because Velten (1968) was wise enough to provide 60 neutral statements along with the 60 elated statements and 60 depressed statements. Going over the Velten studies carefully, we were able to calculate effect sizes using control group comparison on 90% of the effect sizes. The remaining 10% of the effect sizes based on positive versus negative induction group comparisons were fairly evenly distributed among the categories on which we coded. In most cases, we were thus able to calculate effect sizes using the more conservative comparisons (positive vs. neutral, negative vs. neutral).

12. A list of the 62 articles included in this meta-analysis can be obtained on request from the first author.

13. The use of data only from those subjects for whom the Velten produces a change in mood may potentially introduce a subject selection bias into the data and threaten the internal validity of the experiment. This selection procedure, which is not uncommon in this literature (e.g., Teasdale & Taylor, 1981), is useful only if it can be assumed that subjects who are responsive to the Velten represent a random sample from the population under study. This assumption is unlikely to be true. For example, Scheier and Carver (1977) found a positive relation between dispositional self-consciousness and responsivity to the Velten depression induction.

REFERENCES

- Alloy, L. B., Abramson, L. Y., & Viscusi, D. (1981). Induced mood and the illusion of control. *Journal of Personality and Social Psychology, 41*, 1129-1140.
- Berkowitz, L., & Troccoli, B. T. (1986). An examination of the assumptions in the demand characteristics thesis: With special reference to the Velten mood induction procedure. *Motivation and Emotion, 10*, 337-349.
- Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99*, 229-246.
- Buchwald, A. M., Strack, S., & Coyne, J. C. (1981). Demand characteristics and the Velten mood induction procedure. *Journal of Consulting and Clinical Psychology, 49*, 478-479.
- Carlson, M., Charlin, B., & Miller, N. (1988). Positive mood and helping behavior: A test of six hypotheses. *Journal of Personality and Social Psychology, 55*, 211-229.
- Clark, D. M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behavioral Research and Therapy, 5*, 27-49.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Orlando, FL: Academic Press.
- Coleman, R. E. (1975). Manipulation of self-esteem as a determinant of mood of elated and depressed women. *Journal of Abnormal Psychology, 84*, 693-700.
- Cronbach, L. J., & Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.
- Goodwin, A. M., & Williams, J. M. (1982). Mood induction research: Its implications for clinical depression. *Behavior Research and Therapy*, 20, 373-382.
- Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, 7, 119-137.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Isen, A. M. (1984). Toward understanding the role of affect in cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 179-236). Hillsdale, NJ: Lawrence Erlbaum.
- Isen, A. M., & Gorgoglione, J. M. (1983). Some specific effects of four affect-induction procedures. *Personality and Social Psychology Bulletin*, 9, 136-143.
- Kenealy, P. M. (1986). The Velten mood induction procedure: A methodological review. *Motivation and Emotion*, 10, 315-335.
- Leventhal, H., & Tomarken, A. J. (1986). Emotion: Today's problems. *Annual Review of Psychology*, 37, 565-610.
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Orne, M. T. (1962). On the social psychology of the psychological experiment with particular reference to demand characteristics. *American Psychologist*, 17, 776-783.
- Polivy, J., & Doyle, D. (1980). Laboratory induction of mood states through the reading of self-referent mood statements: Affective changes or demand characteristics. *Journal of Abnormal Psychology*, 89, 286-290.
- Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, 92, 500-504.
- Scheier, M. F., & Carver, C. S. (1977). Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 35, 625-636.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513-523.
- Singer, J. A., & Salovey, P. (1988). Mood and memory: Evaluating the network theory of affect. *Clinical Psychology Review*, 8, 211-251.
- Teasdale, J. D., & Taylor, R. (1981). Induced mood and accessibility of memories: An effect of mood state or of induction procedure? *British Journal of Clinical Psychology*, 20, 39-48.
- Thayer, R. E. (1989). *The biopsychology of mood and arousal*. New York: Oxford University Press.
- Velten, E. (1968). A laboratory task for induction of mood states. *Behavior Research and Therapy*, 6, 473-482.
- Wispe, L. (1980). The role of moods in helping behavior. *Representative Research in Social Psychology*, 11, 2-15.