DO PROSPECTIVE RATINGS CORRECT RETROSPECTIVE DISTORTIONS
BASED ON NEGATIVE SOCIAL STEREOTYPES
OF PREMENSTRUAL SYNDROME?

by

Mary L. Stone

A thesis submitted in partial fulfillment
of the requirements for the degree
of
MASTER OF SCIENCE
in
Psychology

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1994
DEDICATION

To a Father in Heaven who loves me and who gives me all good things—a mind and a desire with which to learn, constant opportunities for growth and improvement, and loving guides to lead and support me along the way.

M. L. Stone
ACKNOWLEDGMENTS

I would like to thank my thesis chairman, Dr. David M. Stein, for his tireless efforts in my behalf, in teaching me the research process, and for his willingness to take me by the hand in learning much that was new to both of us. I also thank the other members of my thesis committee, Dr. Steven R. Hawks and Dr. Elwin C. Nielsen, for their scholarly review of this paper.

This project was funded in part by the Women and Gender Research Institute on the Utah State University campus.

Mary L. Stone
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>6</td>
</tr>
<tr>
<td>Retrospective Versus Prospective Reports</td>
<td>6</td>
</tr>
<tr>
<td>Speculations about Causes of the Rating</td>
<td>6</td>
</tr>
<tr>
<td>Discrepancy</td>
<td>7</td>
</tr>
<tr>
<td>Beliefs and Social Stereotypes about Extreme</td>
<td>8</td>
</tr>
<tr>
<td>PMS/LLPDD Symptoms</td>
<td>8</td>
</tr>
<tr>
<td>Memory Issues in Retrospective Reports</td>
<td>13</td>
</tr>
<tr>
<td>Summary</td>
<td>18</td>
</tr>
<tr>
<td>III. METHODS</td>
<td>20</td>
</tr>
<tr>
<td>Overview</td>
<td>20</td>
</tr>
<tr>
<td>Subject Definition</td>
<td>22</td>
</tr>
<tr>
<td>Incentives</td>
<td>23</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>23</td>
</tr>
<tr>
<td>Prospective Use of the DAF</td>
<td>26</td>
</tr>
<tr>
<td>Retrospective Use of the DAF</td>
<td>27</td>
</tr>
<tr>
<td>Reliability and Validity of the DAF</td>
<td>29</td>
</tr>
<tr>
<td>Case Presentation of Negative PMS/LLPDD</td>
<td>30</td>
</tr>
<tr>
<td>Stereotypes</td>
<td>30</td>
</tr>
<tr>
<td>Procedures</td>
<td>31</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>34</td>
</tr>
<tr>
<td>IV. RESULTS</td>
<td>37</td>
</tr>
<tr>
<td>Confirmation of DAF Reliability</td>
<td>37</td>
</tr>
<tr>
<td>Sample Characteristics</td>
<td>37</td>
</tr>
<tr>
<td>Screening-Data Analyses</td>
<td>39</td>
</tr>
<tr>
<td>Analyses Involving Pretest Data</td>
<td>44</td>
</tr>
<tr>
<td>Posttest Analyses and Within-Subjects Design</td>
<td>47</td>
</tr>
</tbody>
</table>
## V. DISCUSSION

<table>
<thead>
<tr>
<th>Hypotheses and Findings</th>
<th>52</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limitations of the Present Study and Research Recommendations</td>
<td>56</td>
</tr>
<tr>
<td>Implications</td>
<td>58</td>
</tr>
</tbody>
</table>

### REFERENCES

| REFERENCES | 59 |

### APPENDICES

| APPENDICES | 66 |

- **Appendix A:** IRB Statement and Consent Form | 67 |
- **Appendix B:** Permission Notification and Daily Assessment Form | 72 |
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>46</td>
</tr>
<tr>
<td>5</td>
<td>48</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
</tr>
</tbody>
</table>

1. Criterion/Symptom Correspondence  
2. Rating-Group Scores on Screening Inventories  
3. Completion-Group Scores on the Screening Total  
4. Case-Presentation-Group Scores on Pretest Inventories  
5. Case-by-Completion Scores on Pretest Inventories  
6. Scores Across Time on Pre/Posttest Inventories
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Basic study design</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Pretest case and completion</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>Time effects on retrospective symptoms</td>
<td>51</td>
</tr>
</tbody>
</table>
ABSTRACT

Do Prospective Ratings Correct Retrospective Distortions Based on Negative Social Stereotypes of Premenstrual Syndrome?

by

Mary L. Stone, Master of Science
Utah State University, 1994

Major Professor: Dr. David M. Stein
Department: Psychology

The Daily Assessment Form (DAF) retrospectively assessed symptoms of late luteal phase dysphoric disorder (LLPDD), both at the beginning and at the end of a 2-month time interval. Ninety-four women between the ages of 18 and 45 entered the study. Half of the subjects viewed a negative-case, stereotypic presentation of LLPDD prior to the pretest. Sixty-eight subjects qualified on the pretest to complete the study. Half of the remaining no-case subjects and half of the remaining negative-case subjects completed 8 weeks' prospective ratings via the DAF. Forty-eight subjects remained to complete the posttest at the end of that time interval. Pretest Total and Pretest Criteria scores revealed significant main effects for case-presentation condition ($F[1,63] = 7.08, p = .01$) and ($F[1,63] = 8.34, p = .01$) and completion level ($F[1,63] =$
6.76, \( p = .01 \) and \( F[1,63] = 3.76, \ p \ .06 \). Effect sizes equalled 0.48 and 0.45 for case presentation and 1.92 and 1.98 for completion.
CHAPTER I
INTRODUCTION

The National Institute of Mental Health (NIMH) has stated that a useful definition of premenstrual syndrome (PMS) is a constellation of mood, behavior and/or physical symptoms which have a regular cyclical relationship to the luteal phase of the menstrual cycle, are present in most (if not all) cycles, and remit by the end of menstrual flow with a symptom-free interval of at least one week each cycle. (Workshop, 1983, quoted in Harrison, Rabkin, & Endicott, 1985, p. 789)

More specific NIMH guidelines for the diagnosis of PMS include

"the cyclic occurrence of [at least five] symptoms that are of sufficient severity to interfere with some aspects of life and which appear with consistent and predictable relationship to menses," with a) marked change of about 30% in the intensity of symptoms measured intermenstrually from days 5 to 10 as compared with that premenstrually (with 6 days before menstruation), and b) prospective documentation of these changes for at least two consecutive cycles. (Workshop, 1983, quoted in Gise, Lebovits, Paddison, & Strain, 1990, p. 230)

It is generally acknowledged that some women report experiencing symptoms of PMS, while others do not. However, what exactly are they reporting or failing to report? Most clinicians tend to rely on patients' retrospective reports (see the following paragraph) of symptoms in making a diagnosis. Retrospective reports are likely used for the sake of economy in order to provide symptom relief by offering a treatment as quickly as possible. However, patients' reports of the mere presence of symptoms commonly
associated with PMS may be an insufficient basis for making a diagnosis of the syndrome. As noted by Rubinow, equally important are severity, timing, symptomatic baseline about which symptoms fluctuate, and method of establishing menstrual linkage of symptoms ("Diagnosis of PMS," 1987). Do women have a sufficiently accurate recall of their symptoms so that they can provide truly valid symptom reports?

Evidence pointing to women's apparent difficulty in adequately self-reporting their PMS symptoms is seen in the frequent discrepancies between retrospective versus prospective reports of premenstrual symptomatology (Cameron, Kuttesch, McPhee, & Curtis, 1988; Christensen, Oei, & Callan, 1989; Endicott & Halbreich, 1982; Endicott, Nee, Cohen, & Halbreich, 1986; Freeman, Sondheimer, & Rickels, 1988; Freeman, Sondheimer, Weinbaum, & Rickels, 1985; Gise et al., 1990; Halbreich et al., 1989; Harrison, Endicott, Rabkin, & Nee, 1984; Kersey, Kruikov, Shannon, & Boyd, 1989; McMillan, Ghadirian, & Pihl, 1989; McMillan & Pihl, 1987; O'Boyle, Severino, & Hurt, 1988; Rapkin, Chang, & Reading, 1988; Youdale & Freeman, 1987). Retrospective reports typically involve one or more of the subject's most recent cycles or her past premenstrual symptomatology. Prospective reports focus on current symptoms. They are usually completed daily. Thus, they are less reliant on subjects' long-term memory. Rubinow indicated that 50 to 70% of
historical reports of PMS are not subsequently corroborated by prospective reports over 2 to 3 months ("Diagnosis of PMS," 1987).

More specifically, prospective reporting involves a woman's daily recordings of her temperature (i.e., for determining ovulation), the presence or absence of menstruation, and the presence and severity of physical and/or emotional symptoms (Severino & Moline, 1989). Such information provides a symptomatic baseline about which symptom severity fluctuates, as well as an objective look at how changes in symptomatology do or do not correlate with hormonal changes. Such an objective view would serve either to substantiate or to disqualify women's prior conceptualizations of premenstrual symptomatology. It may also provide a "corrective learning experience" (Bandura, 1977, p. 78) for those women whose prior (i.e., retrospective) notions of PMS were not in line with their more immediate objective view.

A number of studies have documented differences between women's retrospective and prospective reports about PMS, but little parallel research has been conducted on the "official," American Psychiatric Association (APA) equivalent of PMS, which is called late luteal phase dysphoric disorder (LLPDD), as listed in the Diagnostic and Statistical Manual of Mental Disorders (Third Edition - Revised) (DSM-III-R) (APA, 1987). Research studies dealing
with retrospective versus prospective assessment of LLPDD are practically nonexistent. Such studies are needed to support or refute results obtained from previous studies, which typically lack well-conceptualized and/or widely-agreed-upon definitions of PMS.

Severino and Moline (1989) described differences between PMS and LLPDD along three dimensions: (a) symptom presence, (b) timing, and (c) severity. First of all, at least one symptom of LLPDD must be of an emotional or affective nature (i.e., affective lability, anger or irritability, anxiety, and depression) (APA, 1987). Such is not necessarily the case with PMS. Second, symptoms of LLPDD must appear within 1 week prior to and remit within 2 or 3 days following the onset of menses, whereas symptoms of PMS may begin at any time between ovulation and menses. Finally, symptoms of LLPDD are generally more debilitating than PMS, causing functional impairment in at least one major life area (i.e., work, usual social activities, relationships with others) (APA, 1987). A provisional diagnosis of LLPDD is based on a year's history during which symptoms are present in most cycles and confirmed, as is PMS, by 2 months' prospective daily ratings (APA, 1987; Severino & Moline, 1989).

Also lacking in the literature are studies addressing the factors which may affect or influence subjects' recall of both past and current symptomatology, as reflected in
retrospective versus prospective ratings. As will be addressed in the literature review, one hypothesis accounting for the discrepancies found between retrospective and prospective reports of PMS or LLPDD is that women's retrospective recall is more prone than their prospective reports to be influenced by negative expectations or social stereotypes of PMS or LLPDD (Cameron et al., 1988; Endicott & Halbreich, 1982; Youdale & Freeman, 1987). That is, relative to their prospective ratings, women may base retrospective reports primarily on recall of their own "worst" episodes and what they have been taught about negative social stereotypes regarding PMS or LLPDD.

Examples of such stereotypes are provided by Paige (1973). One can infer from her article that intellectual unreliability, emotional instability, and a state of being unfit for strenuous physical exertion are three main categories of stereotypes typically employed in regard to PMS.

Prospective evaluation may provide compelling, immediate information about PMS or LLPDD and correct women's tendency to make extreme and negative retrospective reports. The question of whether training/experience with prospective ratings may correct extreme beliefs and expectations about PMS or LLPDD (i.e., that it usually involves many diverse psychological symptoms at an extreme level of severity) has not been studied to date.
CHAPTER II
REVIEW OF THE LITERATURE

Retrospective Versus Prospective Reports

The review of the literature will address the question of whether retrospective or prospective reports provide the most valid representation of premenstrual-syndrome (PMS) or late-luteal-phase-dysphoric-disorder (LLPDD) symptoms. Hypotheses accounting for discrepant results found between measurement techniques will be presented along with pertinent literature addressing that topic.

It has been suggested by some writers (Cameron et al., 1988; Endicott & Halbreich, 1982; Youdale & Freeman, 1987) that the validity of patients' reports of symptoms of PMS or LLPDD may be enhanced if patients were to monitor them daily rather than provide only retrospective reports. The question of the general validity of retrospective- versus prospective-assessment techniques in psychology and medicine has encompassed many domains. Those have included, for example, chronic pain (Linton & Gotestam, 1983; Linton & Melin, 1982), labor and postdelivery pain (Niven & Gusbers, 1984; Norvell, Gaston-Johansson, & Fridh, 1987; Rofe & Algom, 1985), alcohol consumption (Hilton, 1989), head trauma (Hunter, Philips, & Rachman, 1979), seizure frequency (Glueckauf, Girvin, Braun, & Bochen, 1990), nutritional status (Vobecky, Vobecky, & Froda, 1988), physical activity
(Wilbur, Miller, Dan, & Holm, 1989), and acute coronary pain (Pakula & Milvidaite, 1983).

The recurrent, aversive, physical, and emotional qualities of PMS or LLPDD share more conceptual overlap with the chronic pain state than with any of the other domains cited above. Results from studies of retrospective versus prospective assessments of chronic pain do, in fact, parallel results obtained from retrospective versus prospective assessments of PMS conducted to date. Generally, the data show that, relative to prospective assessment, retrospective reports of chronic pain tend to involve an overestimation of severity of symptoms, whereas retrospective reports of PMS or LLPDD tend to involve greater numbers of symptoms, a possible overestimation of severity of symptoms, and inappropriate timing of symptoms.

Speculations about Causes of the Rating Discrepancy

A number of hypotheses have been offered in the PMS literature to account for discrepancies found between retrospective and prospective reports of the syndrome. Those discrepancies seem to be generally in the direction of retrospective overestimation of number and severity of symptoms. One hypothesis attributes discrepancies to faulty recall based on extreme, negative expectations or stereotypical beliefs involving more numerous and severe symptoms than are actually acknowledged prospectively.
Beliefs and Social Stereotypes about Extreme PMS/LLPDD Symptoms

A number of authors have referred to the notion that menstrual beliefs influence experience of menstrual symptomatology (Caplan, McCurdy-Myers, & Gans, 1992; Walsh, 1987). Brooks-Gunn (1986) indicated that "we do not know to what extent" that is the case (p. 385). However, it has been hypothesized that premenstrual syndrome is [purely] the result of our expectations and attitudes toward menstruation, and that the reports of premenstrual symptoms are simply reflective of stereotypic beliefs concerning menstruation rather than of symptoms actually experienced. (Rolker-Dolinsky, 1987, p. 116)

Ainscough (1990) and Gallant and Hamilton (1992) have supported that view as accounting for discrepancies between retrospective and prospective reports. Ainscough attributed those discrepancies to a "widespread cultural belief that premenstrual negative affect is part of a woman's normal experience" (p. 43). Gallant and Hamilton referred to the retrospective diagnosis of LLPDD as being less compelling...than other diagnoses based on retrospective reports [due to the role of] cultural stereotypes about the debilitating effects of menstruation on women's psychological and behavioral functioning, and the negative emotional characteristics attributed to being "premenstrual." (pp. 726-727)

Both the popular literature of the media and the medical literature of the health-care system have been cited
as vehicles for the promulgation of those stereotypes. Chrisler and Levy (1990) conducted a content analysis of 78 magazine articles in press between 1980 and 1987 in order to evaluate their descriptions of premenstrual syndrome. The authors found

a strong bias in favor of reporting negative menstrual cycle changes. Articles are generally negative in tone and present a confusing array of symptoms and contradictory treatment recommendations. The media coverage of PMS supports the stereotype of the maladjusted woman. (p. 89)

Rittenhouse (1991), as well, placed much of the blame for propagating a negative premenstrual stereotype on the popular literature, "which problematizes both women and their cycles [and takes] for granted that the majority of women" suffer from PMS (pp. 416, 417). Jurgens and Powers (1991), however, implicated the health-care system as being "partly responsible for women's misconceptions and discomforts about their bodies" (p. 39). They indicated that

medical literature reflects and reinforces negative images and promulgates false information regarding natural female processes, which have often been addressed within the idiom of "disease" and "dysfunction." (p. 39)

A number of studies have lent credibility to the notion that menstrual beliefs influence symptom reports of PMS or LLPDD. A landmark study was conducted by Ruble (1977). She administered the Moos Menstrual Distress Questionnaire (MDQ) (Moos, 1968),

consisting of 48 items, 46 of which form eight clusters
of symptoms, [to] 48 women undergraduates at Princeton University, aged 18 to 24, who were not taking oral contraceptives at the time of the study nor had taken them within the previous 3 months. (p. 291)

Ruble indicated that "variability in the length of their cycles did not exceed 2 weeks" (p. 291).

"Women’s perceptions of their cycle phase were separated experimentally from actual cycle phase" (Ruble, 1977, p. 291). T tests were utilized to show that "women who were led to believe that they were premenstrual reported experiencing a significantly higher degree of several physical symptoms, such as water retention" (p < .01), pain (p < .05), change in eating habits (p < .025), and sexual arousal (p < .05) "than did women who were led to believe they were intermenstrual" (Ruble, 1977, p. 291).

Means and standard deviations for the premenstrual group (n = 15) were as follows: M = 2.62, SD = 0.29 for water retention; M = 2.32, SD = 0.17 for pain; M = 2.93, SD = 0.51 for change in eating habits; and M = 3.60, SD = 0.42 for sexual arousal. Means and standard deviations for the intermenstrual group (n = 14) were as follows: M = 1.54, SD = 0.12 for water retention; M = 1.88, SD = 0.17 for pain; M = 1.57, SD = 0.27 for change in eating habits; and M = 2.50, SD = 0.40 for sexual arousal.

Another study, conducted by Parlee (1974), provided support for the existence of social stereotypes regarding PMS. She compared responses of men and women on the Moos Menstrual Distress Questionnaire (Moos, 1968). Subjects
were asked to complete the inventory in line with their expectations of what women purportedly experience during the three phases of the menstrual cycle. Findings indicated that males, like females, anticipated greater symptom presence both premenstrually and menstrually than during the intermenstruum. Males and females both tended to rank order similar categories of symptoms they believed showed the greatest change in severity throughout the cycle (i.e., pain, water retention, and negative affect). However, males tended to report anticipating greater symptom severity overall than did their female counterparts.

Parlee (1974) suggested that expectations regarding the menstrual cycle are "learned through a myriad of social sources" (p. 238) and that "the beliefs seem to be more strongly held (more extreme) by those with little opportunity to acquire falsifying information" (p. 239). She seems to allude to the idea that women's own personal experiences with menstruation possibly temper views imposed by associations, communications, and social contacts. Men's views, on the other hand, remain untempered because of a lack of such experience. Therefore, their views fall more in line with society's extreme expectations. As mentioned previously, Paige (1973) listed emotional instability, intellectual unreliability, and a state of being unfit for strenuous physical exertion as examples of such negative stereotypes.
A similar study was conducted by Clarke and Ruble (1978) in which three groups of adolescents were assessed regarding their beliefs about menstruation.

The subjects were 18 postmenarcheal girls (mean age 12-11), 18 premenarcheal girls (mean age 12-5), and 18 boys (mean age 12-10). All were in the sixth to eighth grades of four junior high schools in a white, middle-to upper-class area. (p. 232)

The Moos Menstrual Distress Questionnaire (1968) was again used as the dependent variable for assessing what subjects thought "'girls in general' experience during the menstrual phase and during the intermenstrual phase" of the cycle (p. 232).

A 3 x 2 repeated measures ANOVA (group x phase) showed that all groups rated girls' experience of symptoms significantly (p < .001) higher in the menstrual phase than in the intermenstrual phase, for all variables except Arousal, which was higher for the intermenstrual phase. (Clarke & Ruble, 1978, p. 232)

The authors likely used an F test to determine level of statistical significance. "The largest phase differences were for Pain, Behavioral Change, and Negative Affect" (p. 232). Such results obtained from an adolescent sample and paralleling those obtained from Parlee's (1974) adult sample indicate that menstrual attitudes develop early in our culture and are fairly uniform among males and females. Such would allude to a cultural or learning basis for such attitudes, beliefs, and expectations.

In summary, substantial evidence in the literature supports the notion that culture-bound, stereotypic beliefs regarding PMS or LLPDD are highly influential in directing
women's views of their own personal experiences of the disorder as well as society's more general views. Though women's personal experiences may diminish the effect of those stereotypes somewhat, it can be plainly seen that symptom reports, whether concurrent or historical, can be altered by the mere mention of the syndrome. Both men and women, boys and girls tend to view symptoms of PMS in a similar light, one which is extreme in its negativity.

Memory Issues in Retrospective Reports

As mentioned previously, faulty memory has been implicated as a causal factor of the noted discrepancy between retrospective and prospective reports. It may, indeed, play a crucial role in furthering the impact of stereotypical beliefs on subjects' ability to recall past symptomatology. A number of studies support the notion of faulty memory and its impact on retrospective reports of PMS or LLPDD.

For example, Cameron et al. (1988) evaluated 10 women both retrospectively and prospectively on 25 symptoms of panic, general anxiety disorder, and LLPDD. Mean age of subjects was 34.3 (SD = 7.1). Prospective ratings were made daily by patients for at least one menstrual cycle. Clinicians' ratings were also made on a weekly basis. Retrospective ratings were made by subjects following completion of both daily (i.e., subjects' ratings) and
weekly (i.e., clinicians’ ratings) prospective ratings. Discrepancies were found between retrospective and prospective measures. Retrospective reports indicated greater symptom fluctuations across the menstrual cycle with higher severity pre-/perimenstrually than either method of prospective reports. The authors attributed discrepancies to faulty recall based on negative expectations as well as upon nonspecific autonomic arousal at the premenstruum.

Cameron et al. (1988) indicated:

It is possible that the symptoms only show minimal variation (as indicated by the prospective ratings), but that...women have learned from sources other than their own subjective experience to expect more or more severe symptoms in the premenstruum; thus, in the retrospective ratings, which are temporarily more removed from the actual experience than are the prospective daily ratings, subjects recall their symptoms as more severe at that point in the menstrual cycle. In other words, there is little actual fluctuation in symptomatology, but only an exaggeration due to an error of memory based on inappropriate expectation. (p. 173)

In another study, Youdale and Freeman (1987) administered both the 95-item retrospective and the shorter prospective Premenstrual Assessment Forms (PAFs) (Halbreich, Endicott, Schacht, & Nee, 1982) to two groups of women, 19 self-defined severely premenstrually symptomatic and 26 self-defined premenstrually asymptomatic. Mean age of the symptomatic group was 32.11 (SD = 4.9). Mean age of the asymptomatic group was 29.84 (SD = 5.68). Depressive syndrome was evaluated retrospectively with one PAF report involving the subjects’ last three premenstrual phases.
Subjects were evaluated prospectively with nine PAFs: one premenstrual administration, one menstrual administration, and one intermenstrual administration for each of 3 months.

Data revealed that retrospective reports were more severe than prospective reports. The retrospective assessment demonstrated that 95% of self-defined symptomatic subjects qualified as having depressive syndrome and that 58% of self-defined asymptomatic subjects qualified. Prospective-assessment data revealed that 74% of self-defined symptomatic subjects qualified as having depressive syndrome and that 27% of self-defined asymptomatic subjects qualified. Other PAF subtypes were also lower prospectively than retrospectively.

The authors offered a number of speculations to account for discrepancies between retrospective and prospective estimates of symptomatology. They indicated that "retrospective perception of premenstrual symptomatology is based on stereotypical beliefs about the menstrual cycle" (Youdale & Freeman, 1987, p. 422). The authors offered no explanation, however, to clarify exactly what they meant by that. Overestimation of symptoms by those women without severe symptomatology and the possibility that retrospective reports may be more indicative of overall depression as opposed to purely premenstrual symptomatology were also offered as possible explanations to account for the aforementioned results.
One other important point mentioned by Youdale and Freeman (1987) was that 53% of symptomatic and 27% of asymptomatic subjects met criteria for clinical PAF categories based on their intermenstrual reports. The authors suggested that women might be using the PAF rating scale as a measure of symptom severity rather than as a measure of change from base rate. Such would necessitate extra care in instructing women as to how to use the PAF.

To date, only one study has systematically examined the effects of daily self-monitoring of PMS on subjects' overall experience and view of their premenstrual symptomatology. Endicott and Halbreich (1982) administered the PAF to 48 women both prior to and following administration of a prospective measure which included "21 items of particular interest in ongoing studies by the authors" (p. 110). Prospective ratings were made on a daily basis by each subject for at least one menstrual cycle.

Discrepancies were found between ratings. The second PAF rating was lower than the first. For instance, the pre-PAF mean on the PAF Summary Scale for Depressed Mood and Ideation was 26; the post-PAF mean was 23. Also, the number of subjects whose pre-PAF ratings met criteria for Major Depressive Syndrome was 41; the number of post-PAF subjects was 27. The prospective ratings were also lower than the first PAF rating. Criteria for determining PMS included average premenstrual daily ratings being two points higher
than average postmenstrual daily ratings as well as subjects' exhibiting no frequency of depressive changes throughout the cycle. Overall, there was only a 59% confirmation of retrospective ratings by prospective ratings. Thirteen of 15 "severe" diagnoses were confirmed as well as 6 of 14 "moderate," and 5 of 12 "mild." Three of seven women who had not complained of PMS were found to exhibit depressive changes of a syndromal type. High levels of confirmation were limited to those women reporting severe PMS.

Endicott and Halbreich (1982) indicated women may initially describe a scenario of their "worst case" (p. 110) of PMS when providing a retrospective report of symptoms over a period of past months. They suggested that subjects typically have no baseline against which to compare symptoms over time and that such reports cannot take into account the severity or timing of fluctuations in symptoms. Therefore, a lack of ongoing, systematic attention to critical premenstrual variables would hinder a woman's ability to recall symptoms. Rating symptoms on a daily basis may have served to eliminate any effects of faulty memory on symptom reports and shed lasting illumination on subsequent retrospective reports.

Possible ways to improve upon the aforementioned study encompass a number of considerations. First of all, having subjects rate premenstrual symptoms prospectively for at
least 2 months is a guideline for diagnosis of LLPDD provided in DSM-III-R (APA, 1987). Subjects in Endicott’s and Halbreich’s study may have provided ratings for as short a period as 1 month. Second, assessing subjects for the more rigorously defined symptoms of LLPDD in place of the more loosely conceptualized symptoms of PMS would provide greater consensus and more precise focus for operationally defining the disorder. Finally, experimental induction of negative expectations or stereotypical beliefs through presentation of a negative-case study of PMS or LLPDD could be employed as a means for testing the validity of the aforementioned hypothesis that expectations and beliefs influence recall and reports of symptoms.

Summary

In summary, discrepancies have been noted between retrospective and prospective assessments of PMS or LLPDD. Inherent in the literature is the general belief that prospective assessments provide a much more valid evaluation of PMS or LLPDD symptoms. One hypothesis attributes retrospective- versus prospective-rating discrepancies to faulty recall based on negative expectations or stereotypical beliefs. Learning how to conduct a prospective documentation of symptoms is thought to possibly correct for retrospective-reporting errors.

A study addressing the validity of that hypothesis could assess the extent to which learning to prospectively
assess PMS or LLPDD symptoms alters retrospective ratings over time. Also, the study could incorporate a procedure for reinforcing negative expectations or stereotypical beliefs about PMS or LLPDD among women with the disorder. The extent to which such a procedure affects retrospective ratings could be examined. Further, the degree to which the prospective-rating process alters those biased reports could be evaluated.

Specifically, one or more negative-case stereotype(s) of LLPDD would be presented to half of a group of subjects attesting to symptoms of LLPDD just prior to their making retrospective reports of their own symptomatology. The other half of the subjects would be given no case example(s) before giving their retrospective ratings of LLPDD. In addition, prospective ratings would then be collected from representative samples of each of those two groups, while other subgroups would simply continue the retrospective-rating process. Thus, the study would examine whether prospective ratings correct reinforced, negative-case stereotypes and retrospective ratings.
Overview

The basic study design included three main factors: (a) case-presentation condition, (b) rating condition, and (c) time (i.e., a repeated-measures' factor). Four interactions of interest were also considered possible: (a) case-presentation condition x rating condition, (b) case-presentation condition x time, (c) rating condition x time, and (d) case-presentation condition x rating condition x time (see Figure 1).

Retrospective assessments of late-luteal-phase-dysphoric-disorder (LLPDD) symptoms were obtained for subjects both at the beginning and at the end of a 2-month time interval. Those assessments evaluated changes in retrospective reports of symptomatology across time. That was the within-subjects' factor. However, half of those subjects were presented with a negative-case, stereotypic definition of LLPDD by means of videotape, while the other half received no case definition. The case presentation preceded subjects' first retrospective ratings. Also, half of the subjects in the negative-case-definition group and half of those in the no-case-definition group were administered 8 weeks' prospective ratings of LLPDD, while the other members of both groups continued retrospective-only ratings. Thus, two between-subjects' factors (i.e.,
<table>
<thead>
<tr>
<th>Group # 1 No-Case-Presentation, Retrospective-Only-Rating (NoC/R) Condition</th>
<th>Group # 2 No-Case-Presentation, Retrospective-Plus-Prospective-Rating (NoC/R+P) Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time 1</strong> Retrospective Rating 1</td>
<td><strong>Time 1</strong> Retrospective Rating 1</td>
</tr>
<tr>
<td>Two-Month Time Lapse</td>
<td>Two Months’ Prospective Ratings</td>
</tr>
<tr>
<td><strong>Time 2</strong> Retrospective Rating 2</td>
<td><strong>Time 2</strong> Retrospective Rating 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group # 3 Negative-Case-Presentation, Retrospective-Only-Rating (NegC/R) Condition</th>
<th>Group # 4 Negative-Case-Presentation, Retrospective-Plus-Prospective-Rating (NegC/R+P) Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time 1</strong> Retrospective Rating 1</td>
<td><strong>Time 1</strong> Retrospective Rating 1</td>
</tr>
<tr>
<td>Two-Month Time Lapse</td>
<td>Two Months’ Prospective Ratings</td>
</tr>
<tr>
<td><strong>Time 2</strong> Retrospective Rating 2</td>
<td><strong>Time 2</strong> Retrospective Rating 2</td>
</tr>
</tbody>
</table>

Figure 1. Basic study design.

Note: 2 (no-case presentation versus negative-case presentation) x 2 (retrospective-only ratings versus retrospective-plus-prospective ratings) x 2 (retrospective-rating period 1 versus retrospective-rating period 2) design.
negative-case presentation versus no-case presentation and retrospective-only ratings versus retrospective-plus-prospective ratings) were employed.

Another main factor (i.e., completion level) emerged during the course of the study. That made possible a number of other interactions: (a) case-presentation condition x completion level, (b) rating condition x completion level, and (c) case-presentation condition x rating condition x completion level.

Subject Definition

Female subjects were recruited from several large undergraduate classes at Utah State University. Subjects between the ages of 18 and 45 were selected based on the fact that women only experience premenstrual syndrome (PMS) or LLPDD (Severino & Moline, 1989). Those subjects who had undergone oophorectomies involving the surgical removal of both ovaries were excluded from participation. Subjects should also have been free of oral-contraceptive use so as to eliminate the confounding effects of artificial hormones on the body’s natural hormonal cycle. Pregnant women were also omitted from the study. Finally, subjects needed to have been menstruating regularly (i.e., they must have had at least one menstrual period during the 2 months prior to their participation in the study). Such specific criteria ensured physiological homogeneity within the sample and a lack of extraneous influences.
Incentives

Informed consent to complete the screening inventory was obtained from each subject before admission to the study. (See Appendix A.) Female students received extra credit in line with the agreements they had with their instructors regarding participation in research. Those subjects whose results on the screening inventory indicated that they possibly experienced LLPDD and whose subsequent pretest scores indicated that they likely attributed their symptoms to the disorder were asked to complete the study. Equal proportions of various incentives (i.e., more extra-credit points and 10 dollars per subject who completed the study—provided by the Women & Gender Research Institute on Utah State University campus) were offered across groups.

Instrumentation

The Daily Assessment Form (DAF) (Rivera-Tovar & Frank, 1990) was chosen for the proposed study because of its emphasis on symptoms of LLPDD as well as for its parsimony and versatility. (See Appendix B.) Also, a review of the literature uncovered no other instrument developed specifically for the assessment of LLPDD. It was, therefore, necessary to make use of the DAF. The only other alternative would have been to develop a similar inventory.

The DAF evaluates the presence and intensity of LLPDD using 33 items of physical (e.g., "low energy, fatigue, feel
unable to move") and emotional (e.g., "irritable, angry, impatient") symptomatology, some of which are associated with specific LLPDD criteria and some of which are not. Some of the items are considered to be "positive" (e.g., "feel excited, bursts of energy"), whereas most of the items are considered to be "negative" (e.g., "irritable, angry, impatient"). None of the "positive" items are associated with any of the LLPDD criteria. Items are rated on a six-point, Likert-type, severity scale. Scores range from 1, indicating the total absence of a given symptom, to 6, indicating the symptom to be extremely apparent or noticeable.

One or more item(s) from the DAF is/are associated with each of the 10 LLPDD criteria listed in DSM-III-R (APA, 1987). For cases involving more than one item per criterion, the mean of those items involved is calculated to produce a criterion score. APA (1987) guidelines for diagnosing LLPDD indicate that a subject must experience symptoms falling within at least five criteria. One of those criteria must be affective. Endicott and Halbreich (1982) also noted that only those retrospective reports made by subjects initially complaining of severe-intensity PMS were later substantiated by subjects' prospective reports. Therefore, a minimum of five criterion scores, one of them being affective, must be of moderate to extreme intensity for subjects to be considered as possible candidates for
study inclusion. It was thought that using a moderate score as the cut-off point for study inclusion would help to avoid exclusion of false negatives, though more false positives may have entered the initial phase of the study. Such were likely screened out later.

As mentioned above, the inventory includes both symptoms which are and symptoms which are not relevant to making a diagnosis of LLPDD. The irrelevant symptoms helped to disguise the true focus of the inventory. Thus, subjects likely experiencing symptoms of the disorder could be identified without specifically educating them about all of the essential features of LLPDD. That aspect of the study was important as the investigator wanted to avoid contaminating initial ratings of symptoms or giving subjects cues about what specific symptoms were of interest to the investigator at that point in the study.

A personal phone conversation with Heide Reppert, secretary to DAF coauthor, Ellen Frank, Ph.D. (November 26, 1991), revealed a likely change in wording of future versions of the DAF. Rather than having subjects rate their degree of "change" of symptoms from 1 day to the next, the authors were simply going to require subjects to report their current severity of symptoms for each day. The present study made the anticipated correction so as to eliminate any confusion caused by the previous wording.
Prospective Use of the DAF

Rivera-Tovar and Frank (1990) provided a description of their use of the DAF as it pertains to the prospective assessment of LLPDD. They indicated:

The diagnosis of late luteal phase dysphoric disorder was determined by applying a rigorous percent-change criterion. Scores on the items on the Daily Assessment Form that corresponded to each of the 10 symptoms of the DSM-III-R criteria were summed and averaged over the 7 days before menses (premenstrual week) and the 7 days after the cessation of menses (postmenstrual week). For each symptom, the difference between the premenstrual average and the postmenstrual average was divided by the postmenstrual average and expressed as percent change. A subject met the criteria for a given cycle if the averages for at least five of the 10 symptom areas showed a 30% or greater premenstrual increase in severity and if all postmenstrual averages were less than 3. In an attempt to exclude cases of chronic symptoms that were heightened premenstrually (premenstrual magnification), symptoms with a postmenstrual average score higher than or equal to 3 (signifying "mild" distress) were excluded. A positive diagnosis also required that at least one of the five symptoms be one of the first four symptoms listed in the DSM-III-R criteria (mood lability, irritability, anxiety, or depressed mood) and that the subject meet the criteria during at least two menstrual cycles. (p. 1635)

The prospective use of the DAF allows the researcher to make a differential diagnostic determination as to whether symptoms are truly confined to the late-luteal phase or whether they are more prevalent throughout the cycle, with or without a premenstrual exacerbation of symptoms. Criteria for making such a determination are included in the above instructions by Rivera-Tovar and Frank (1990).

Because the present study was limited to an 8-week, prospective-rating period, the author was able to collect
pre- and postmenstrual ratings for only one cycle for some of the subjects. Therefore, premenstrual days were averaged together, and postmenstrual days were averaged together for each cycle to determine that cycle's percent-change scores. Also, all premenstrual days were averaged together, and all postmenstrual days were averaged together in determining overall percent-change scores. So, it was possible to determine if a subject had met prospective-rating criteria based on each cycle's ratings as well as combined ratings.

Retrospective Use of the DAF

Retrospective-symptom reports by subjects on the DAF differed from the above application in that subjects were required to recall and report the premenstrual average intensity of symptoms of LLPDD from the prior 2 months' cycles. Again, a moderate score was used as the cut-off point for determining if a subject had experienced sufficient intensity of at least five symptoms, one of them being affective, corresponding to DSM-III-R criteria for LLPDD (APA, 1987; Endicott & Halbreich, 1982).

A retrospective symptom-report format was also employed in the initial screening of subjects. The inventory was presented simply as a list of "physical and emotional symptoms" experienced by women "within the past 2 months" (see Appendix A) versus questioning them about their past two premenstrual or late-luteal phases, as in the above application. Thus, the researcher was able to collect
pertinent screening data while avoiding subjects' self-selection based on their preconceived notions of PMS.

Scoring for the retrospective application of the DAF was done in two ways. The first, the **Total** scoring method, incorporated all symptoms on the inventory. Severities of all symptoms considered to be "negative" were totalled at face value. Severity scores for all symptoms considered to be "positive" (i.e., "feel excited, bursts of energy," "feel more efficient, increased orderliness," and "sense of well-being, more enjoyment of things") were reverse scored prior to their being added into the Total score. That was done for the purpose of gauging just how negatively a subject might have perceived and reported her past symptomatology.

In developing the computer-scoring program, the author of the present study took into account the possibility that not all subjects would complete all 33 items on the inventory. Therefore, a mean score, the Total score divided by the total number of items completed, served as the true Total score for retrospective assessments on the DAF.

The second method of retrospective scoring resulted in a **Criteria** score which incorporated only those symptoms specifically related to LLPDD criteria (see Table 1). As mentioned previously, in cases where more than one symptom corresponded to a given criterion, a mean score was obtained for those symptoms so that 10 criteria scores, some of them being means, were summed to total a grand Criteria...
Table 1

Criterion/Symptom Correspondence

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Symptom(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>(10+11+17)/3</td>
</tr>
<tr>
<td>5</td>
<td>(28+31)/2</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>12 or 21</td>
</tr>
<tr>
<td>9</td>
<td>19 or 29</td>
</tr>
<tr>
<td>10</td>
<td>(7+8+15+22+26)/5</td>
</tr>
</tbody>
</table>

Note. Information constituting this table was provided by the authors of the DAF.

Reliability and Validity of the DAF

Reliability and validity data regarding the DAF were provided by Dr. Rivera-Tovar by telephone on May 19, 1992 and again through the mail on May 22, 1992. Four subscales (i.e., negative affect, physical symptoms, agitation, and arousal) emerged as a result of factor analyzing items on the DAF. Internal consistency of those subscales ranged from what was considered to be an adequate score of .67 to a
high score of .96. Test-retest reliability or stability scores of the subscales were statistically significant. They ranged from .44 to .66. Dr. Rivera-Tovar indicated that intercycle variability of symptoms may have lowered the values somewhat. Dr. Rivera-Tovar considered prevalence rates of LLPDD when administering the DAF to a normative (i.e., 4.6%) versus a clinical (i.e., 62%) sample to be indicative of the validity of the instrument.

Case Presentation of Negative PMS/LLPDD Stereotypes

Specifically, the case presentation consisted of an interview situation in which a "subject" (i.e., a confederate of the interviewer) was retrospectively assessed for LLPDD using the DAF. The "subject" responded to all symptoms that could be perceived in a "negative" light (e.g., "irritable, angry, impatient") at a severe-to-extreme intensity level. In addition, she avowed all symptoms that could be perceived in a "positive" light (e.g., "feel excited, bursts of energy") at a negligible level of severity. Thus, the negative definition showcased a rather dysfunctional, severe example of LLPDD, leading subjects to infer that it was typical of most patients. As mentioned in the overview, the video clip was presented to half of the subjects in each of the rating groups just prior to their completing the pretest.
Procedures

As mentioned previously, the screening inventory consisted of the DAF presented simply as a list of "physical and emotional symptoms" experienced by subjects "within the past 2 months." (See Appendix A.) Each of the study groups (see Figure 1) consisted of subjects whose scores on the screening inventory indicated they were experiencing LLPDD. Also, as noted, sufficient numbers of symptoms (i.e., at least five, one of them being affective) and intensity of symptoms (i.e., moderate to extreme) corresponding to DSM-III-R criteria were necessary to qualify subjects as evidencing LLPDD (APA, 1987; Endicott & Halbreich, 1982). Again, higher intensity of premenstrual complaints of subjects is a necessary criterion for study inclusion, Endicott and Halbreich having discovered a lack of confirmation of symptoms by those subjects complaining of lower intensity PMS.

It was difficult to anticipate exactly how many individuals needed to be screened to fill each cell with 20 to 30 subjects. However, the greater demands placed on subjects by the prospective ratings necessitated greater numbers of subjects being assigned to the prospective-rating condition. Approximately 500 women were screened in all. It was planned that at least 12 subjects would be retained within each cell by the end of the study. That number was thought to lend adequate statistical power to the data.
Subjects were informed by telephone that their scores on the screening inventory qualified them for further study participation and that they might be offered an additional incentive (i.e., possibly more extra-credit points and a possible monetary incentive) for continuing the study. The experimenter scheduled to meet with subjects individually for the purpose of briefing them as to what their study participation would involve.

The pretest was administered to all subjects. Subjects who had met screening-inventory criteria had been randomly assigned to one of four treatment groups. (See Figure 1.) All groups were required to complete the DAF as it related to their two most recent premenstrual or late-luteal phases. Half of the subjects were presented with the video clip just prior to their taking the first retrospective DAF. Subjects in the no-case-presentation conditions simply completed their symptom ratings. A 2-month interval then lapsed before subjects were again required to complete the inventory retrospectively. Subjects in the retrospective-plus-prospective-rating conditions, whether having viewed the video tape or not, completed both retrospective administrations of the DAF as well as 8 weeks’ prospective ratings in between.

The dependent variable for determining whether a subject would be retained for further study inclusion was the total number of DAF criterion scores, one of them being
affective, of moderate to extreme intensity (APA, 1987; 
Endicott & Halbreich, 1982). A proportion of the subjects 
who had passed the screening-inventory criteria were 
expected to attribute their symptomatology to something 
other than LLPDD. They would, therefore, have reported 
fewer, lower intensity symptoms on the first retrospective 
assessment of LLPDD and been dropped from study 
participation at that point (i.e., pretest disqualifiers). 
However, data were retained for all subjects who completed 
the pretest so that the effect of the video clip on symptom 
attributions could be ascertained. That was done to reveal 
the validity of the hypothesis that symptom reports of PMS 
or LLPDD are influenced by negative social stereotypes.

It was also anticipated that the dropout rate for the 
retrospective-plus-prospective-rating groups would be more 
severe than for the retrospective-only-rating groups because 
of the time demands placed on the former subjects. 
Therefore, additional subjects were assigned to that more 
demanding rating condition. It was the concern that, due to 
the likelihood of higher drop-out rates, the retrospective-
plus-prospective-rating groups would become a more select 
sample of subjects than the retrospective-only-rating 
groups, the former demonstrating a greater degree of overall 
compliance and motivation than the latter. Pretest 
comparisons were later made to determine initial differences 
between those subjects who failed to complete their
prospective ratings and were, therefore, not administered the posttest (i.e., posttest dropouts) and those subjects who subsequently completed the study.

Data Analysis

Verification of the reliability of the DAF was sought by means of calculating Cronbach's alpha coefficients for determining internal consistency. Correlation coefficients were also calculated for verification of stability or test-retest reliability of the instrument by means of comparing pre- and posttest scores of the control group.

A number of methods were used to analyze demographic data from the sample. Age data were considered by employing separate two-by-two analyses of variance (ANOVAs) to determine significant differences between subjects in the case-presentation (i.e., no-case versus negative-case) and rating groups (i.e., retrospective-only versus retrospective-plus-prospective). Also, a one-way ANOVA was utilized to determine differences in age between those subjects meeting different levels of study completion (i.e., pretest disqualification versus posttest dropout versus full study).

Other demographic characteristics of the sample (i.e., marital status, children, religion, and racial background) were examined. A frequency count was used to reveal the percentage of subjects falling within each category of a given characteristic. Contingency tables revealed
differences between subjects in the case-presentation (i.e., no-case versus negative-case) and rating conditions (i.e., pretest-disqualification, retrospective-only, and retrospective-plus-prospective) with regard to each of those variables.

Analyses of variance were also employed in reference to screening data. Separate two-by-three ANOVAs for case presentation (i.e., no case and negative case) and rating (i.e., pretest disqualification, retrospective only, and retrospective plus prospective) determined initial comparability of groups. One-way ANOVAs were also utilized to determine screening-score differences between subjects meeting different levels of study completion (i.e., pretest disqualification, posttest dropout, and full study) and between subjects having taken the inventory during different academic quarters.

A 2-by-2-by-2 analysis of covariance (ANCOVA) for case presentation, rating, and completion was to be employed in analyzing pretest data. However, due to an empty cell, two 2-by-2 comparisons were made instead. The first of those involved case presentation (i.e., no case and negative case) and rating (i.e., retrospective only and retrospective plus prospective). The second comparison involved case presentation (i.e., no case and negative case) and level of completion (i.e., posttest dropout and full study).
Finally, a multivariate-analysis-of-variance (MANOVA) approach to a repeated-measures test was employed. That was used to determine posttest differences between case presentation (i.e., no case and negative case) and rating (i.e., retrospective only and retrospective plus prospective) as well as to determine differences between pretest and posttest scores according to case-presentation and rating assignments and the passage of time.
CHAPTER IV
RESULTS

Confirmation of DAF Reliability

The present study sought to verify internal consistency of the Daily Assessment Form (DAF) (Rivera-Tovar & Frank, 1990). Therefore, Cronbach’s alpha coefficients were calculated. An alpha coefficient of .92 was found for the pretest when considering all 33 items on the inventory. A coefficient of .90 for the pretest was found when only those items specifically associated with LLPDD criteria were considered. Stability or test-retest reliability was also assessed by comparing the first and second retrospective ratings for the NoC/R group. Those ratings were administered approximately 2 months apart. A correlation coefficient of .59 was discovered for Pretest/Posttest Total scores and .62 for Pretest/Posttest Criteria scores.

Sample Characteristics

The 94 subjects participating in the study had a mean age of 22.05 (SD = 5.84). Two-by-two analyses of variance (ANOVAs) showed that no significant age differences were evident between subjects assigned to different case-presentation (F[1, 64] = 0.00, p > .98) or rating groups (F[1, 64] = 1.53, p > .22). Also, no interaction was present (F[1, 64] = 0.13, p > .72). Neither did a one-way ANOVA reveal a significant difference in age between those
subjects who (a) failed to meet criteria for study inclusion, (b) subsequently dropped out of the study, or (c) completed the full study ($F[2, 91] = 0.71, p > .49$).

As has been mentioned in the Methods section, an attempt was made to solicit the participation of older subjects so that the age range was more representative of adult women generally. That was accomplished by including women involved in academic programs during summer quarter as well as those involved in more advanced undergraduate courses. However, the mean age of participants overall tended to fall near the lower end of the age continuum of adult women. A likely bias was the inclusion of large numbers of freshmen enrolled in the larger classes.

A number of other sample characteristics (i.e., marital status, children, religion, and racial background) were examined. A frequency count revealed the vast majority of subjects to be single (i.e., 75.5%), white (i.e., 93.6%), Latter-Day-Saint (LDS) (i.e., 90.4%) females with no children (i.e., 85.1%). The majority of subjects deviating from that pattern were those who had had opportunities for marriage (i.e., 22.3% married at the time of the study, 2.1% divorced) and child rearing (i.e., 14.9%). Deviations in religion (i.e., 9.6% non-LDS) and racial background (i.e., 3.2% Hispanic, 3.2% Asian) were infrequent.

Pearson chi-square tests revealed no significant differences between case-presentation or rating groups with
respect to any of those demographic variables (i.e., for case-by-marital status \( \chi^2(2) = 1.46, p > .48 \); for rating group-by-marital status \( \chi^2(4) = 1.35, p > .85 \); for case-by-children \( \chi^2(1) = 0.12, p > .72 \); for rating group-by-children \( \chi^2(2) = 0.88, p > .64 \); for case-by-religion, \( \chi^2(1) = 0.62, p > .43 \); for rating group-by-religion \( \chi^2(2) = 4.08, p > .13 \); for case-by-racial background \( \chi^2(2) = 6.09, p > .04 \); and for rating group-by-racial background \( \chi^2(4) = 4.27, p > .37 \)).

**Screening-Data Analyses**

Separate two-by-three ANOVAs involved a case-presentation condition (i.e., no case versus negative case) and a rating condition (i.e., pretest disqualification versus retrospective only versus retrospective plus prospective). The Screening Total score and the Screening Criteria score were dependent variables. Those analyses were conducted to assess initial comparability of groups with regard to screening-inventory scores. It should be noted that the **Screening Total** score incorporated all questions on the screening inventory. The **Screening Criteria** score incorporated only those questions specifically related to the 10 late-luteal-phase-dysphoric-disorder (LLPDD) criteria listed in **DSM-III-R** (APA, 1987).

The means and standard deviations for the pretest-disqualification-, retrospective-only-, and retrospective-plus-prospective-rating groups are presented in Table 2.
Table 2
Rating-Group Scores on Screening Inventories

<table>
<thead>
<tr>
<th>Rating Group</th>
<th>n</th>
<th>Total</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest disqual.</td>
<td>26</td>
<td>3.32</td>
<td>3.73</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.42</td>
<td>0.39</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrospective only</td>
<td>25</td>
<td>3.79</td>
<td>4.13</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.48</td>
<td>0.57</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retrospective plus prospective</td>
<td>43</td>
<td>3.58</td>
<td>3.85</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.46</td>
<td>0.49</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Higher scores indicate higher symptomatology. Boldface scores differ significantly from the others.

The somewhat larger sample size for the retrospective-plus-prospective-rating group reflects a priori planning for a possibly higher rate of subsequent dropouts for that group over the course of the study, as that group was subjected to greater task demands. Significant differences were noted between subjects initially assigned to rating groups on both the Screening Total score ($F[2, 88] = 6.97, p \geq .002$) and the Screening Criteria score ($F[2, 88] = 4.74, p \geq .011$).
Group contrasts on the Screening Total revealed that the pretest-disqualification-rating group \( (M = 3.32, \ SD = 0.42) \) scored significantly lower than the means of 3.79 \( (SD = 0.48) \) and 3.58 \( (SD = 0.46) \) for retrospective-only- and retrospective-plus-prospective-rating groups, respectively. Mean scores for the latter two groups did not differ significantly. Those results were expected, as subjects in the pretest-disqualification group later disqualified themselves from further participation because of their low, subsequent pretest scores. Because of their low degree of symptomatology, subjects in that rating group were not assigned to any research condition.

Findings for the mean Screening Criteria breakdowns, however, were unexpected. Despite attempts to assign qualified subjects to experimental conditions on a strict rotating basis, the mean of 4.13 \( (SD = 0.57) \) for the retrospective-only-rating group was significantly higher than the mean of 3.85 \( (SD = 0.49) \) for the retrospective-plus-prospective-rating group. Such a difference between mean scores for those rating groups suggested the need to statistically control for initial Screening Criteria differences in subsequent analyses.

There was one other expected finding on the screening inventory. Screening Total differences discovered through a one-way ANOVA for those subjects meeting different levels or degrees of completion, as described in the following
paragraph, were significant \( F[2, 91] = 4.90, p < .01 \).

Scheffé posthoc tests revealed that the Screening Total score for the pretest-disqualification group \( (M = 3.32, SD = 0.42) \) was significantly lower than that for the posttest-dropout group \( (M = 3.66, SD = 0.55) \) or that for the full-study group \( (M = 3.66, SD = 0.45) \). (See Table 3.) That trend did not persist with the Screening Criteria scores \( F[2, 91] = 1.81, p < .17 \).

Table 3
Completion-Group Scores on the Screening Total

<table>
<thead>
<tr>
<th>Completion Group</th>
<th>n</th>
<th>Screening Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest disqualification</td>
<td>26</td>
<td>3.32</td>
</tr>
<tr>
<td>M</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Posttest dropout</td>
<td>20</td>
<td>3.66</td>
</tr>
<tr>
<td>M</td>
<td>3.66</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Full study</td>
<td>48</td>
<td>3.66</td>
</tr>
<tr>
<td>M</td>
<td>3.66</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.45</td>
<td></td>
</tr>
</tbody>
</table>

Note. Higher scores indicate higher symptomatology.
Boldface scores differ significantly from the others.

What these scores indicate is that those subjects who disqualify themselves (i.e., pretest-disqualification group)
by failing to meet LLPDD criteria report less numerous and severe symptoms on their screening inventories. Thus, there were some initial differences on the screening inventory between disqualifiers and those subjects whose subsequent pretest scores qualified them for further study participation. Pretest scores for the posttest-dropout and full-study groups qualified those subjects for further inclusion in the study. However, subjects in the posttest-dropout group failed to meet study requirements (i.e., they primarily failed to complete their prospective ratings). They were, therefore, dropped from the study prior to administration of the posttest.

On the other hand, no other significant differences with regard to screening-inventory scores were found. For example, the two-by-three ANOVA revealed no significant mean differences due to case-presentation-group assignment on either the Screening Total score ($F[1, 88] = 0.96, p > .32$) or the Screening Criteria score ($F[1, 88] = 2.39, p > .12$). Neither did a one-way ANOVA reveal a significant difference in mean scores for subjects who had taken the screening inventory during different academic quarters on either the Screening Total score ($F[2, 91] = 1.68, p > .19$) or the Screening Criteria score ($F[2, 91] = 0.44, p > .64$). No interaction effect was present between case presentation and rating on either Screening Total scores ($F[2, 88] = 0.26, p > .77$) or Screening Criteria scores ($F[2, 88] = 0.61, p > .54$).
In summary, with the exception of an apparent lack of comparability of the pretest-disqualification group, there appears to have been a fairly homogenous assignment of subjects into different treatment cells. Also, these data suggested the need to statistically control for Screening Criteria differences in analyses of pretest data.

**Analyses Involving Pretest Data**

Because of screening-score differences among study groups, the present author chose to report adjusted-mean pretest values: Pretest scores were regressed onto the combination of Screening Criteria scores, case-presentation group assignment, rating-group assignment, and the group-assignment indices. The predicted or adjusted-mean pretest score for each subject was then computed from the overall regression model. Subgroup means and standard deviations were calculated based on the resulting predicted Pretest Total and Pretest Criteria scores. Those means were then used in some of the calculations that follow.

Pretest scores were analyzed by means of a 2-by-2-by-2 analysis of covariance (ANCOVA) controlling for Screening Criteria scores. Interactions and main effects were examined for case-presentation group (coded no case and negative case), rating group (coded retrospective only and retrospective plus prospective), and completion level (coded posttest dropout and full study). It was discovered that no three-way interaction could be examined because of an empty
cell (i.e., no-case-presentation group, retrospective-only-rating group, plus posttest-dropout-completion level).

It was, therefore, necessary to analyze the pretest data by means of two separate two-by-two ANCOVAs. The first involved case presentation (i.e., no case versus negative case) and rating (i.e., retrospective only versus retrospective plus prospective) using the Pretest Total score, incorporating all questions on the pretest inventory, and the Pretest Criteria score, incorporating only those questions specifically related to the 10 late-luteal-phase-dysphoric-disorder (LLPDD) criteria.

A main effect was noted for the case-presentation condition on both the Pretest Total score ($F[1, 63] = 5.67$, $p > .020$) and the Pretest Criteria score ($F[1, 63] = 7.19$, $p < .01$). Effect sizes of 0.42 and 0.42 were found for Pretest Total and Pretest Criteria scores, respectively. However, as expected, no significant differences were found for initial rating-group assignment on either the Pretest Total score ($F[1, 63] = 0.00$, $p > .95$) or the Pretest Criteria score ($F[1, 63] = 0.19$, $p > .66$). Neither was there an interaction effect evident between case presentation and rating on either the Pretest Total scores ($F[1,63] = 0.11$, $p > .74$) or the Pretest Criteria scores ($F[1,63] = 0.43$, $p > .51$). Means and standard deviations for case-presentation groups and adjusted for Screening Criteria differences are listed in Table 4.
Table 4  
**Case-Presentation-Group Scores on Pretest Inventories**

<table>
<thead>
<tr>
<th>Case-Presentation Group</th>
<th>n</th>
<th>Pretest Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
</tr>
<tr>
<td>No case</td>
<td>35</td>
<td>3.92</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.33</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative case</td>
<td>33</td>
<td>4.07</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Higher scores indicate higher symptomatology.

The second two-by-two ANCOVA involved case presentation (i.e., no case versus negative case) and completion (i.e., posttest dropout versus full study) using Pretest Total and Pretest Criteria scores. A significant main effect was again noted for the case-presentation condition on the Pretest Total score ($F[1, 63] = 7.08, p < .010$) and the Pretest Criteria score ($F[1, 63] = 8.34, p < .005$). Effect sizes of 0.48 and 0.45 for case presentation were found for Pretest Total and Pretest Criteria scores, respectively.

A second main effect was noted for the completion factor with both the Pretest Total ($F[1, 63] = 6.76, p < .012$) and the Pretest Criteria scores ($F[1, 63] = 3.76, p < .057$). Effect sizes of 1.92 and 1.98 were associated with
the Pretest Total and Pretest Criteria scores, respectively. Means and standard deviations for case-by-completion adjusted for Screening Criteria differences are listed in Table 5. (Also see Figure 2.) It was evident that no interaction effect existed between case presentation and completion for either Pretest Total scores ($F[1,63] = 1.89, p > .17$) or Pretest Criteria scores ($F[1,63] = 0.79, p > .37$).

**Posttest Analyses and Within-Subjects Design**

A multivariate-analysis-of-variance (MANOVA) approach to conducting a repeated-measures analysis compared pretest-and posttest-retrospective ratings associated with case-presentation (coded no-case and negative-case) and rating assignments (coded retrospective-only and retrospective-plus-prospective). A significant main effect for time was found. Specifically, a decrease in symptom severity was noted across time for both Pre/Posttest (i.e., Retrospective) Total scores ($F[1, 44] = 15.52, p > .00$) and Pre/Posttest (i.e., Retrospective) Criteria scores ($F[1,44] = 18.49, p > .00$). Effect sizes of 0.56 and 0.63 were found for Retrospective Total and Retrospective Criteria scores, respectively. Means and standard deviations adjusted for Screening Criteria scores are listed in Table 6. (Also see Figure 3.)
Table 5
Case-by-Completion Scores on Pretest Inventories

<table>
<thead>
<tr>
<th>Case-by-Completion</th>
<th>n</th>
<th>Total</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>No case</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest dropout</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>3.57</td>
<td>3.81</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>0.25</td>
<td>0.33</td>
</tr>
<tr>
<td>Full study</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>4.04</td>
<td>4.44</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>0.21</td>
<td>0.28</td>
</tr>
<tr>
<td>Negative case</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest dropout</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>3.70</td>
<td>3.96</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>0.27</td>
<td>0.36</td>
</tr>
<tr>
<td>Full study</td>
<td>24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>4.18</td>
<td>4.62</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>0.28</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note. Higher scores indicate higher symptomatology.

No other significant contrasts were found either between or within groups in regard to Pre/Posttest Total scores (i.e., case-by-time $[F(1, 44) = 0.10, p > .75]$), rating-by-time $[F(1, 44) = 1.22, p > .27]$, and case-by-rating-by-time $[F(1, 44) = 0.03, p > .87]$). Neither were there significant differences found regarding Pre-/Posttest Criteria scores (i.e., case-by-time $[F(1, 44) = 0.17, p >$
Figure 2. Pretest case and completion.

Note: Means are adjusted for Screening Criteria scores.
Table 6

Scores Across Time on Pre/Posttest Inventories

<table>
<thead>
<tr>
<th>Time Period</th>
<th>n</th>
<th>Total</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>48</td>
<td>3.97</td>
<td>4.37</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.49</td>
<td>0.63</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td>48</td>
<td>3.62</td>
<td>3.89</td>
</tr>
<tr>
<td>M</td>
<td></td>
<td>0.75</td>
<td>0.88</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Higher scores indicate higher symptomatology.

.68], rating-by-time \([F\{1,44\} = 0.69, p > .41]\), and case-by-
rating-by-time \([F\{1,44\} = 0.17, p > .68]\).

One other important point regarding prospective ratings
is that they confirmed retrospective reports for only 6 of
24 subjects (i.e., 25%) who completed that aspect of the
study.
Figure 3. Time effects on retrospective symptoms.

Note: Means are adjusted for Screening Criteria scores.
Hypotheses of the present study fell under three main categories. First of all, it was hypothesized that the negative-case presentation of late-luteal-phase-dysphoric-disorder (LLPDD) symptoms would model negative social stereotypes about the disorder among subjects. Thus, it was suspected that initial retrospective-symptom ratings for the two negative-case-presentation groups (i.e., groups 3 and 4) (see Figure 1) would reflect both more numerous and severe symptoms compared to the retrospective ratings of the no-case-presentation groups (i.e., groups 1 and 2).

Indeed, higher severity ratings were noted for those subjects who saw the case presentation compared to those who did not. It seems possible that the case presentation truly did impress subjects with a more negative, stereotypical view of their symptoms, thus elevating their initial retrospective scores. However, the impact of the case-presentation condition was no longer evident by the time subjects completed posttest inventories.

A behavioral-learning approach to explaining the impact of the negative-case presentation could also be considered. Perhaps subjects were merely patterning their responses to the inventory after those responses that had been modelled by the "subject" in the video clip.
Second, an interaction effect was expected between case-presentation condition and type of ratings. Relative to all other groups, subjects in the negative-case-presentation, retrospective-only-rating condition (NegC/R) were expected to continue to report more severe symptoms over time. Subjects in the negative-case-presentation, retrospective-plus-prospective-rating condition (NegC/R+P), however, were expected to show the greatest within-group shift in symptom ratings over time, relative to all other groups, due to experiencing the prospective-rating process. The shift would reflect fewer and less severe symptoms. That interaction was not found.

As was mentioned previously, to date, only one study has systematically examined the effects of daily self-monitoring of premenstrual syndrome (PMS) on subjects' overall experience and view of their premenstrual symptomatology (Endicott & Halbreich, 1982). One of the intentions of this research was to duplicate the results of that study, with the exception that the more rigorously defined criteria for LLPDD would be substituted for the more loosely conceptualized criteria of PMS. That might impact efforts of the health profession also to develop more studies in line with LLPDD criteria versus those of PMS.

In the present study, rating assignment per se, however, did not seem to impact subjects' final retrospective scores. Endicott and Halbreich had indicated
in their 1982 article that studies utilizing nonclinical populations such as college students were not likely to duplicate their results, which were based more specifically on a treatment-seeking population of women suffering extreme-intensity PMS.

Also, no significant within-group changes over time were predicted for (a) subjects exposed to the no-case-presentation, retrospective-only-rating condition (NoC/R) and (b) subjects exposed to the negative-case-presentation, retrospective-only-rating condition (NegC/R). That was anticipated because those subjects would have experienced no prospective-rating training, which would correct subjects' a priori response biases. Again, such was not the case.

It would appear that the passage of time had a far greater influence on pre-to-posttest-severity ratings than did either case-presentation or rating-group assignment. What would account for such a decrease in severity of ratings for the entire sample across time? Kazdin (1980) referred to such "spontaneous remission" as being those "changes made without receiving formal treatment in a given investigation" (p. 143). He also referred to "multiple influences that may impinge on clients," whether or not they have been assigned to a formal treatment condition (p. 143).

A possible explanation for the decrease in symptom severity over time in the present study could be that pretest inventories were administered at the beginnings of
academic quarters and the posttest inventories at the quarters' ends for all subjects. Perhaps the stress present at the beginnings of academic quarters somehow accentuates subjects' perceptions and recall of their symptomatology. Perhaps that stress level was universally reduced at the ends of academic quarters so that posttests of all women reflected that change in stress.

The above assumption seems to cast considerable doubt on the validity of one-time retrospective reports of premenstrual symptomatology. Do retrospectively administered inventories truly measure what they purport to measure, or do confounding factors such as immediate subjective experience (e.g., stereotypical views, stressors present at the time of administration) serve to distort memory?

Another possible explanation for the apparent reduction in symptomatology among subjects is the multiple exposures to the testing materials through screening and initial retrospective assessments. In completing the first retrospective symptoms ratings, subjects could have been made more aware of their symptoms. They might, therefore, have begun attending to those symptoms during the 2-month time interval between pre- and post-retrospective assessments. Such self-monitoring on the part of subjects may have produced an effect similar to that induced by prospective ratings on symptom experience and report.
One final hypothesis emerged during the course of the study when it became clear that dropout rates for the retrospective-plus-prospective-rating groups were fairly high. It was presumed that there might exist some difference between dropouts and completers that could be detected in the pretest scores. It was the suspicion of the experimenter that pretest scores for full-study completers would be less severe, indicating an initially higher degree of physical and emotional health and, therefore, a heightened ability to complete the prospective inventories.

However, those subjects who subsequently completed the study tended to report higher severity on their ratings than did those subjects who failed to complete the study. It could have been that subjects who truly did experience higher symptomatology might have had a more vested interest in completing the study, thus accounting for higher pretest scores among those subjects meeting full-study completion.

Limitations of the Present Study and Research Recommendations

A main limitation of the present study was the use of a population of college undergraduates attending a Utah university. It is an empirical question as to whether the results from using such a group would generalize to other populations (e.g., older populations, those with a greater percentage of women who had been married and had had children, non-LDS women, and those from various racial
backgrounds, as well as treatment-seeking, clinical populations). In light of the aforementioned limitation of the present study, it is suggested that other populations of women be studied for the purpose of determining generalizability of this study’s results.

Another possible limitation of this study is the fact that only self-defined symptomatic subjects were inducted into the study. That may have greatly limited the amount of change in severity of ratings for those subjects who were exposed to the negative-case presentation of LLPDD symptoms. It is suggested that self-defined asymptomatic subjects be admitted into future studies examining the effects of such a case-presentation so as to ascertain the full impact of induced negative stereotypes on subjects’ ratings of their symptoms.

Another limitation of the study is that pre- and postmenstrual ratings were collected for as short a time as one menstrual cycle for some of the subjects. That was also a limitation of the 1982 study by Endicott and Halbreich, which the present study hoped to correct. However, that proved to be an unrealistic goal considering the 8-week, prospective-rating interval during each academic quarter. Therefore, a possible way to improve upon both studies would be to gather ratings from at least two complete cycles as is suggested in DSM-III-R (APA, 1987).
Implications

Results of the research at hand hold implications for both the medical profession as well as the target population of young adult females. Establishing a method for the accurate assessment of LLPDD would further the work of the clinician in both diagnostic and treatment efforts. Those clients truly suffering from the disorder should, therefore, have a greater opportunity for accurate diagnosis and successful treatment.
REFERENCES


Cosponsored by the Center for Studies of Affective Disorders and the Psychobiological Processes and Behavioural Medicine Section, Clinical Research Branch, National Institute of Mental Health, Rockville, MD.

APPENDICES
Appendix A

IRB Statement and Consent Form
Statement of the PI to the IRB for Proposed Research Involving Human Subjects

Proposal Title  Do Prospective Ratings Correct Distortions Associated with Negative-Case Stereotypes of Premenstrual Syndrome (PMS) or Late Luteal Phase Dysphoric Disorder (LLPDD) Combined with Retrospective Ratings?

Principal Investigator  Dr. David Stein, Ph.D.
Dept.  PSY
UMC  2810
Ext.  3274

Student Researcher  Mary Stone
Dept.  PSY
UMC  2810
Ext.  1460

This study will examine whether prospective ratings seem to correct distortions associated with negative-case stereotypes of premenstrual syndrome (PMS) or late luteal phase dysphoric disorder (LLPDD) combined with retrospective ratings.

A. Human subjects who participate in this research will be asked to do the following: Initial screening of subjects will involve completion of a newly-developed, self-report inventory, the Daily Assessment Form (DAF) (see attachment), presented simply as a list of "physical and emotional symptoms" "experienced by women within the past two months."

Subjects who meet screening-inventory criteria will then be randomly assigned to one of four groups. All groups will be required to complete the DAF as it relates to their two most recent premenstrual or late-luteal phases. A two-month interval will then lapse before subjects are again required to complete the inventory retrospectively. Equal numbers of subjects from the retrospective-only-rating and retrospective-plus-prospective-rating groups will be assigned to either a no-case-presentation group or a negative-case presentation group.

The negative-case-presentation groups will be exposed to a video clip presenting a "negative" (severe/chronic) definition of LLPDD just prior to taking the first retrospective DAF. Specifically, a "subject" (a confederate of the interviewer) will be retrospectively assessed for LLPDD using the DAF. The "subject" will respond to all LLPDD-relevant symptoms at a moderate-to-extreme-severity level. In addition, she will avow all symptoms irrelevant to
LLPDD at a mild level of severity. Thus, the "negative" (chronic/severe) definition will showcase a rather poorly functioning, severe example of LLPDD, leading subjects to infer that it is "typical" of most patients. Subjects in the no-case-presentation condition will simply complete their symptom ratings.

Subjects in the retrospective-plus-prospective-rating conditions will complete both retrospective administrations of the DAF as well as two-months’ prospective ratings.

B. The Potential Benefits to be gained from the proposed research are: Results of the proposed research hold implications for both the medical profession as well as the target population of young adult females. Establishing a method for the accurate assessment of late luteal phase dysphoric disorder (LLPDD) would further the work of the clinician in both diagnostic and treatment efforts. Those clients truly suffering from the disorder should, therefore, have a greater opportunity for accurate diagnosis and successful treatment.

As of yet, only one study has systematically examined the effects of daily self-monitoring of PMS on subjects’ overall experience and view of their premenstrual symptomatology. One of the intentions of the proposed research is to duplicate the results of that study, with the exception that the more-rigorous criteria for LLPDD will be substituted for the more-loosely-defined criteria of premenstrual syndrome (PMS), furthering the efforts of the clinical profession to develop more studies in line with LLPDD criteria versus those of PMS.

Another intention of the proposed research is to establish the validity of the hypothesis that faulty recall, based on negative expectations or stereotypical beliefs, underlies discrepancies found between retrospective and prospective assessments of PMS or LLPDD. Experimental induction of such expectations or beliefs via presentation of a negative-case example will be employed for that purpose.

C. The risk(s) to the rights and welfare of human subjects involved are: No or minimal risk(s) to subjects is/are foreseen. The primary potential risk is loss of confidentiality. All data gathered on subjects will have identifying information removed, and will be placed in a database containing number-coded, group data. The potential compromises to confidentiality are minimal, and are thus, attended to by the researchers.

D. In addition, the following safeguards/measures to mitigate/minimize the identified risks will be taken: NA
E. The informed consent procedures for subjects will be as follows: (Explain procedures to be followed, and attach an example of the informed consent instrument.) A copy of the consent form (see attachment), affixed to each DAF, will be distributed to female subjects recruited from several large undergraduate classes convening on Utah State University campus. Subjects will be asked to review the form and then to sign and date it at will, handing it in to their instructors separately from the inventory. Subjects will be asked not to put any identifying information on the inventory. A special I.D. code will be printed on both the consent form and the inventory so as to identify a given set of responses without jeopardizing confidentiality. Subjects will complete the DAF in the privacy of their homes, and return it to class the following day.

F. The following measures regarding confidentiality of subjects will be taken: As noted previously, a special I.D. code will be assigned to each subject so that only the principal investigator and the research assistants will be able to match up responses on the inventory to a given subject.

Dr. David Stein, Ph.D
Principal Investigator

Mary Stone
Student Researcher
Consent Form

If you are FEMALE, are BETWEEN the ages of 18 AND 45, have NOT undergone SURGERY for the removal of both ovaries, are NOT currently taking any form of ORAL CONTRACEPTIVE, are NOT PREGNANT and HAVE EXPERIENCED MENSTRUATION (HAVE HAD A PERIOD) DURING THE PAST 2 MONTHS, PLEASE CONTINUE. OTHERWISE, PLEASE HAND THE PACKED BACK IN. THANK YOU!

The purpose of this study is to examine how well a newly-developed, self-report inventory, the Daily Assessment Form (DAF), relates to some important physical and emotional symptoms that affect many women. This inventory should aid the efforts of both clinicians and researchers in determining symptom presence.

In this study, you are asked to volunteer approximately 5 minutes of your class time to complete the DAF. IF YOU CHOOSE NOT TO DO SO, PLEASE HAND THE PACKET BACK IN. THANK YOU!

FOR THOSE WHO DO CHOOSE TO COMPLETE THE DAF, PLEASE FILL THIS CONSENT FORM. THEN USE THE ATTACHED COMPUTER SHEET IN COMPLETING BOTH SIDES OF THE INVENTORY. FILL IN THE CIRCLE CORRESPONDING TO THE MAXIMUM DEGREE OF SEVERITY FOR EACH SYMPTOM YOU EXPERIENCED WITHIN THE PAST 2 MONTHS. THE SYMPTOM(S) NEED NOT HAVE PERSISTED THROUGHOUT THE ENTIRE 2-MONTH INTERVAL, BUT MUST HAVE BEEN PRESENT AT SOME POINT DURING THE RATING PERIOD TO RECEIVE A SCORE GREATER THAN 0. PLEASE HAND THE COMPLETED PACKET BACK IN WHEN YOU ARE FINISHED. THANK YOU!

All of the answers that you give will be held in complete confidence. This means that no one but the principal investigator (Dr. David Stein, Ph.D.) and his research assistants will have access to this information. A special I.D. code will be assigned to each participant for the sake of maintaining confidentiality. PLEASE DO NOT PUT ANY IDENTIFYING INFORMATION ON THE COMPUTER SHEET, OR MARK THE INVENTORY IN ANY WAY!

Completing the DAF may make you eligible for further study participation, for which an incentive may be provided. However, you may withdraw your consent to participate at any time, and will suffer no penalty for doing so.

Any questions regarding this research may be directed to Dr. David Stein (USU PSY Dept., 750-3274).

NAME
AGE
PHONE #
SIGNATURE
TODAY'S DATE
Appendix B

Permission Notification and Daily Assessment Form
Mary Stone
69 1/2 N. 200 W. #4
Logan, Utah 84321

Dear Ms. Stone:

Thank you for your interest in our Daily Assessment Form (DAF) which you hope to use for a study of retrospective versus prospective assessment approaches in diagnosing LLPDD. Dr. Frank and I have no objection to your using the instrument in your thesis research provided you furnish us with a copy of your study results.

I wish you success in your research endeavor.

Very truly yours,

Ana Rivera-Tovar, Ph.D.
Daily Assessment Form
(Pearlstein T, Rivera-Tovar A & Frank E, 1986)

Please use the attached computer sheet in completing both sides of this inventory. Fill in the circle corresponding to the maximum degree of severity for each symptom you experienced within the past 2 months.

0-None (feature not present at all)
1-Minimal (only slightly apparent to you, others would probably not be aware of it)
2-Mild (definitely apparent to you and perhaps to others who know you well)
3-Moderate (clearly apparent to you and/or others who know you well)
4-Severe (very apparent to you and/or others who know you well)
5-Extreme (the degree of severity is so extreme that it is very apparent to you OR even people who do not know you well might notice)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1) Irritable, angry, impatient</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2) Low energy, fatigue, feel unable to move</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3) Feel overwhelmed</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4) Mood swings, feel labile, emotions feel out of control</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5) Anxious, jittery, nervous</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6) Feel excited, bursts of energy</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7) Headaches</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8) Breast tenderness or swelling</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9) Forgetful, confused</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10) Depressed, sad, low, blue, tearful</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>11) Decreased self-esteem, insecure</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>12) Decreased appetite, eat less</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>13) React intensely to &quot;+&quot; or &quot;-&quot; daily events</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>14) Feel more efficient, increased orderliness</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>15) Abdominal pain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>16) Distractable, difficulty concentrating</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>17) Pessimistic, gloomy, despondent</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>18</td>
<td>Guilt feelings, brood over events</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>Decreased sleep</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>20</td>
<td>Active, restless, &quot;can’t sit still&quot;</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td>Increased appetite, eat more</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>Abdominal bloating or swelling</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>23</td>
<td>Desire to be alone</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>24</td>
<td>Sensitive to criticism or rejection from others, easily hurt</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>More affectionate feelings, more sexual interest</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>26</td>
<td>Back, joint, or muscle pain</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>27</td>
<td>Poor performance, impaired functioning at home, work or school</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>28</td>
<td>Loss of interest or pleasure in usual activities</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>29</td>
<td>Increased sleep, naps, stay in bed</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>30</td>
<td>Sense of well-being, more enjoyment of things</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>31</td>
<td>Avoid social commitments</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>32</td>
<td>Lowered motor coordination, feel clumsy, minor accidents</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>33</td>
<td>Less sexual interest</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>