INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

DEFENSIVE PROJECTION IN THEMATIC APPERCEPTIVE STORIES

by

JEANNE M. CLANCY

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1997

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

8-26-97
Date

Charles P. Smith
Chair of Examining Committee

3/29/97
Date

Executive Officer

Charles P. Smith

Alan Gross

Herbert Saltzstein

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

DEFENSIVE PROJECTION IN THEMATIC APPERCEPTIVE STORIES

by

Jeanne M. Clancy

Advisor: Professor Charles P. Smith

This research investigates the validity of P. Cramer's (1991b) Defense Mechanism Manual (DMM) for scoring defensive projection in the Thematic Apperception Test (TAT) by attempting to differentiate between TAT stories from subjects who were angry and defensive, and from subjects who were pretending to be angry but had no reason to be defensive. This experiment is a modified replication and extension of the P. Cramer (1991a) study of anger and defense mechanisms.

The subjects were 89 men and 129 women college students, ages 17 to 23 years, from subject pools at 2 public colleges in New York City. In small groups, subjects wrote stories to TAT pictures 1, 5, 6GF, 7BM, 8, and 18GF, under one of 3 conditions, and then filled out a questionnaire regarding their affect at the beginning and end of the experiment.

Control subjects wrote stories with no interventions. Subjects in the anger condition were subjected to criticism of their stories, intended to induce anger. The control and anger conditions were comparable to Cramer's. In a third condition, simulated-

anger, subjects were asked to imagine that their stories had been subjected to the same criticisms given in the anger condition. All stories were scored blind for defensive projection using the DMM, and for aggression using categories from the Hall and Van de Castle (1966) and the Whitman, Pierce, Maas, and Baldridge (1961) systems.

Data were analyzed by 2 (gender) by 3 (condition) MANOVAS. Subjects in the anger and simulated-anger conditions reported slightly more anger on the questionnaire than those in the control condition (multivariate $p = .09$; univariate $p = .004$). For TAT aggression and projection scores, there were no significant differences among conditions. The failure to replicate Cramer's findings of increased aggression and projection in her anger condition precluded the demonstration of predicted differences between the anger and simulated-anger conditions. Some support for gender differences in use of projection was shown (men > women, multivariate $p < .02$, univariate $p = .056$).

Results were inconclusive regarding the validation of P. Cramer's scoring system. Possible reasons include differences in subject characteristics, and group versus individual data collection.

Acknowledgements

This dissertation is dedicated to the memory of a dear friend, Kathleen Connolly-Weinert, and to the two most cherished people in my life, my son, David Moran, and my daughter, Siobhan Moran. To my children I owe many thanks. I thank them and their friends for being my "guinea pigs" as I worked out my research design. I also deeply appreciate all they gave up, having a mother who was so involved in graduate school. I thank them, for always believing that I could do this, and for never letting life get boring.

I owe an enormous debt of gratitude to my mentor, Charles Smith, who has persevered with me through the multiple obstacles that I faced. He has given of his time with amazing generosity, has continued to provide encouragement for me to finish, and has provided for me, a model of excellence in research. I also thank Judy Smith for her consistent graciousness throughout my interruptions of their schedule.

I am grateful to my committee and outside readers, Alan Gross, Herb Saltzstein, Howard Ehrlichman, and Suzanne Ouellette who took time from their busy schedules to assist me in finishing my dissertation, and who were willing to participate in my oral defense outside of regular semester time. I thank Howard Ehrlichman who also provided tremendous assistance in obtaining

subjects.

To my brother, James Clancy, and mother, Rita Clancy I am grateful for their financial support, which enabled me to return to school full time. Without their help, none of this would have been possible.

I thank Rudy Palmieri for his tremendous assistance in the interscorer reliability testing, and for his emotional support to start me on the road to this degree.

I would like to thank the staff at my job, These Our Treasures, for their patience during the completion of this dissertation. I thank all those who helped code material, count words, rate pictures, locate needed articles and who provided me with encouragement to continue.

Lastly, I would like to acknowledge all the undergraduates who were subjects in this research.

Table of Contents

List of Tables

Chapter I

Introduction

Purpose

The present research is intended to contribute to the literature on Thematic Apperception Test (TAT) interpretation by investigating the validity of a scoring system for measuring defensive projection in TAT stories (Cramer, 1991b). Following Sherwood (1981, p. 445), and Cramer (1987, p. 599), defensive projection is defined as the attribution of unacceptable internal psychological states to others as a means of avoiding psychological threat.

According to Cramer (1991b) defensive projection is illustrated by the following story told by an 11 year old boy to TAT card 17BM (man clinging to a rope):

> Once there was a story about a man and he was building a house and then some warriors came along and broke down the house and were trying to kill him. Luckily he escaped and went away on his horse. Trying to get into the government building, he climbed up a rope and got in and killed the emperor (p. 81).

In terms of Cramer's scoring system, defensive projection is indicated by the addition (to a picture of a single figure) of ominous characters or objects (warriors), themes of pursuit, entrapment, and escape (he escaped)

and apprehensiveness of death, injury, or assault (trying to kill him). The present research is designed to validate and improve upon Cramer's measure.

Significance of the Research

A reliable and valid TAT measure of defensive projection would constitute a contribution to personality measurement. Such a measure could be of use both for clinical and research purposes. Attempts have been made to measure defenses by means of self-report inventories. However, the nature of an objective measure is such that the stimuli are relatively clear-cut and the possible responses are limited to a choice between a few alternatives. A subject may have personal conflicts, for which defensive responses may be critical, that may lie outside the boundaries set by the limited stimuli and responses of an objective measure (Jacobson, Beardslee, Hauser, Noam, & Powers, 1986). For example, the Defense Mechanism Inventory or DMI (Gleser & Ihilevich, 1969) asks the subject to choose from five possibilities which would be his or her most and least likely responses to a given situation. In addition to employing a limited number of situations and response alternatives, the self-report measure assumes that a person can and will accurately report his or her characteristic responses. However, in psychoanalytic theory defenses are assumed to operate outside of conscious awareness, so the TAT would

appear to be a more promising method of assessment than a self-report measure, both because subjects who are telling stories are usually not aware that they are revealing things about themselves (Murray, 1938), and because the range of threatening situations and response alternatives is less restricted.

An understanding of the manner in which defensive tendencies are expressed in stories would contribute to the literature on TAT interpretation. Bellak (1950, 1975) suggests that analyzing defenses appearing in TAT stories is a way to increase available information, particularly from stories that are very brief.

For clinical purposes, a clearer understanding of defenses as expressed in the TAT could be helpful in understanding the dynamics of a patient's problems. For example, TAT stories may reveal what is threatening to a patient and how the patient defends against the threat.

For research purposes, a valid measure of projection could provide a way to investigate the development and operation of defenses. For example, Cramer (1987, 1991b), used her system to study the development and use of three defenses over an age range from childhood to early adulthood. Research into other issues such as stress reduction, outcome of different psychotherapy treatment modalities, and the relation of defenses to other personality variables (e.g., field dependence-

independence) would be facilitated by a valid measure of defenses.

Background Research

Derivation of TAT scoring systems. There are basically two approaches to the development of systems of content analysis: theoretically based systems and empirically derived systems. The first of these methods involves specifying a set of a priori scoring criteria that are derived from some theory. For example, many TAT scoring systems used for clinical purposes have been derived from psychoanalytic theory. In contrast, McClelland and his associates (McClelland, Atkinson, Clark, & Lowell, 1953) introduced an alternative procedure for identifying scoring categories indicative of the construct to be measured. Their method, empirical derivation, contrasted the stories of a group in whom a motive was aroused with the stories from a non-aroused group. Whatever differences in imagery were found in the aroused group, as compared with a non-aroused group, constituted an empirically derived set of imagery criteria reflective of the motive.

Unfortunately Cramer (1987, 1991b) does not describe her derivation procedures in detail. However, her system appears to be a combination of theoretical derivation, empirical derivation, and categories from previous systems, specifically Bellak (1975) and Haworth (1963).

With respect to empirical derivation, Cramer (1987) states:

> ... the analysis of empirical findings of themes and response styles which seemed to differentiate between older and younger children,...illustrated some aspect of the theoretical description of one of the three defenses (e.g., Fenichel, 1945; A. Freud, 1946) and/or had been used in one of the few previously published attempts to formulate a scoring system for defenses based on projective test data... (p. 601).

Cramer contrasted older and younger children because of her focus on the development of defense mechanisms. However, she does not report the statistical significance of differences between stories of older and younger children. Nor does she report, on a category by category basis, which categories were identified by this procedure.

Systems for scoring defenses in the TAT. Murray indicates that the TAT may be scored for "projectivity," which he defines as "the disposition to project unconsciously one's own sentiments, emotions and needs into others" (Murray, 1938, p. 148). However, Murray regards projectivity as a form of egocentricity rather than as a defense. Two prior systems for scoring defenses in the TAT, those of Bellak and Haworth, will be

described briefly before Cramer's (1991b) system is presented.

Bellak. Bellak's (1975) approach derives from that of Murray but is based more explicitly on ego psychology and develops more fully the notion that story interpretation may be enhanced by scoring for defenses (see especially Bellak, 1975, chap. VI). Defensive projection is one of many defenses that he considers scorable. He provides definitions, clinical significance, and examples of defenses in stories. However, there are few specific criteria to guide scoring, and the system requires that the scorer have extensive training in the TAT as well as a theoretical understanding of ego psychology and defense mechanisms.

Haworth. Haworth (1963) created a more explicit system for scoring defenses in the Children's Apperception Test (CAT). The system (Schedule of Adaptive Mechanisms in CAT Responses) encompasses 10 categories of defenses and adaptive mechanisms. They are: reaction and undoing, isolation, repression and denial, deception, symbolization, projection and introjection, fear and anxiety, regression, weak controls, and identification. For each category she presents a checklist of items which are scored if present in the story. Haworth, in her initial presentation of the scoring system (1963), and in her book about CAT

interpretation (1966), does not explain why she combines projection and introjection in one category. The items for Haworth's category "Projection and Introjection" are:

1. Attacker is attacked, "eat and be eaten"
2. Innocent one is eaten or attacked
3. Child is active aggressor (bites, hits, throws; do not include verbal or teasing attacks)
4. Characters blame others
5. Others have secrets or make fun of somebody
6. S adds details, objects, characters, or oral themes
7. Magic or magical powers (Haworth, 1963, p. 183)

Scores for each category are summed over stories. The subject's score is compared to a standard critical score for that category. The final score is the number of categories which equal or surpass the critical score, and ranges from 0-10.

Cramer. Cramer's (1979) previous work with the DMI, and the findings of previous researchers (e.g., Vaillant, 1971) led her to propose an age- or stage-related development of defense mechanisms. However, she felt that a TAT measure of defenses would be preferable to the DMI. She developed the Defense Mechanism Manual (DMM) in order to investigate the development of three defense mechanisms: denial, projection, and identification. The seven scoring categories in the DMM, that Cramer (1991b)

developed for defensive projection are:

1. Attribution of aggressive or hostile feelings, emotions, or intentions to a character or other feelings, emotions or intentions that are normatively unusual.
2. Addition of ominous people, ghosts, animals, objects, or qualities.
3. Magical or circumstantial thinking.
4. Concern for protection from external threat.
5. Apprehensiveness of death, injury, or assault.
6. Themes of pursuit, entrapment, and escape.
7. Bizarre or very unusual story or theme (p. 221).

A story is scored for the presence of each of these categories as many times as it occurs (Cramer, 1991b, p. 43).

Cramer (1991b) presents information about the psychometric properties of scores from the DMM. She reports split-half reliabilities for 40 control subjects from a college sample of 80 students who told stories to eight TAT cards. The reliability coefficient for projection for the control group only, corrected for attenuation (Spearman Brown formula), was r = .68; for the combined experimental and control groups, total N = 80, the coefficient was r = .77.

Interrater reliability for the DMM was determined by comparisons among four coders on scores from different

subject groups used in various studies. The interrater reliability coefficients for projection ranged from .71 to .90.

Cramer (1991b) presents results from four studies that she believes support the validity of the DMM. In the first study two Ph.D. clinical psychologists were given 33 stories, of which 11 represented high scores for projection, 11 high scores for denial, and 11 high scores for identification. Raters were asked to indicate which of the three defenses was most prominent for each story. Each judge was able to identify 29 of 33 stories correctly.

In the second study, TAT story segments (N = 450) representing the seven categories for each defense were presented along with neutral segments to two judges. These judges rated each segment as illustrative of denial, projection or identification. Of the 140 segments representing projection, one judge accurately identified 87 percent, the other 82 percent.

In the third validity study (Dollinger and Cramer, 1990) the DMM was used to score TAT stories from a group of boys who had witnessed a fatal lightening strike. Scores were used to attempt to distinguish between boys who were high-defensive and those who were low-defensive. For the low-defensive boys there was good agreement between self-report and parent reports of the child's

fears. For the high-defensive boys self-report of fears did not correspond to their parents' ratings of psychological distress. Cramer suggests that these findings support the self-protective nature of defenses, and the validity of the DMM.

The fourth study that Cramer (1991b) presents was one in which she intercorrelated scores for the three defense measures using two populations of elementary and high school children. The correlations ranged from $\underline{r}$ = .08 (denial with projection) to $\underline{r}$ = .37 (projection with identification). She suggests that these results indicate a small amount of common variance, but that the three systems are primarily measuring different characteristics (discriminant validity).

Another validation study done by Hibbard, Farmer, Wells, Difillipo, Barry, Korman, and Sloan (1994) had a group of college students and a group of psychiatric patients write stories to six TAT cards. The stories were scored for denial, projection, and identification, using the Defense Mechanism Manual, and each defense was scored for more primitive versus more mature indices of the particular defense. Hibbard et al. (1994) report support for criterion, divergent, and construct validity as demonstrated by the finding that the psychiatric population used more primitive defenses as well as more primitive levels of the defenses that were used.

Cramer's (1991a) study, of anger and the use of defense mechanisms in a college sample, will be reviewed in detail because it served as the basis for the present research. The purpose of the study was to investigate the theory that defense mechanisms develop in a stage-related manner, specifically, that in a college sample, the higher level defenses of projection and identification would be used more often than the simpler defense of denial. Her subjects were 40 male and 40 female college students, equally split between a control group and an experimental group. The subjects individually told stories to eight TAT cards. The first four cards were:

6GF A woman sitting on edge of sofa looks over her shoulder at older man.

14 Silhouette of a man (or woman) against a window.

10 Young woman's head against a man's shoulder.

5 Middle aged woman standing on the threshold of a door looking into a room.

The cards were presented to all subjects in the order listed above. The second set was presented in random order, they were:

1 Young boy with a violin

7BM An older man looking at a younger man who is staring into space.

8GF Young woman sits with her chin in her hand looking off into space

18GF Woman with her hands squeezed around the throat of another woman whom she appears to be pushing backwards across a stair banister.

Cards 8GF and 18GF were slightly modified from the standard TAT pictures. In card 8GF, equations were added to the blackboard behind the young woman in order to make this picture more similar to card 1 with respect to the amount of achievement imagery elicited. In card 18GF, in which a woman has her hands around the neck of another woman, the hair of one woman was extended to cover the left hand of the second woman. Cramer (1991a) adapted 18GF in order to reduce possible aggressive connotations of the picture. Cramer (1991a) mentions in a footnote (p.42) that these pictures are modified, but she does not discuss why reduced aggressive content is desirable. Presumably the modification of 18GF is intended to reduce the baseline amount of aggressive imagery produced by the picture so that group differences in aggressive imagery will be detectible. These two modified pictures are reproduced in Appendix A.

Cramer's (1991a) procedure was a replication of a Bellak (1944) study. In both the Bellak and Cramer studies, the experiment was described to subjects as dealing with creative imagination in college students,

and this was followed by standard TAT instructions (Murray, 1943). The first four stories were told under the same (neutral) conditions in both the experimental and control groups. In the experimental group, after the fourth story was completed, and after stories 5, 6, and 7, comments were made to the subjects about the inadequacy of their stories. For example, after the fourth story the experimenter said, "These stories are about the worst I have ever heard. Could you try to get some better ones" (Cramer, 1991b p. 42). At the end of the testing session, subjects filled out a questionnaire to assess their affective experience during the testing session. The stories were transcribed, with all identifying data removed and were scored for projection, denial, and identification using the DMM. To verify the effectiveness of her manipulation, Cramer also scored these stories for aggression, using scoring systems based on Hall and Van de Castle (1966) and Whitman, Pierce, Maas and Baldridge (1961).

Cramer's (1991a) study used two aggression scoring systems, aggression intensity and aggression frequency. The aggression intensity scale was taken from Hall and Van de Castle (1966), who explicitly describe their scale as a nominal scale. This scale should not have been used as an interval scale in statistical analyses, and her results regarding aggression intensity are not

interpretable, and will not be presented. Therefore, the present study will use only the aggression frequency measure. However, the categories of aggression listed by Hall and Van de Castle were used along with the categories of Whitman et al. (1961) in the present study (and in Cramer's) to identify individual occurrences of aggression which were summed for the aggression frequency score. The criteria for scoring each category of the aggression systems are presented in Appendix B.

Cramer (1991a) found that the experimental manipulation was successful in arousing anger, according to the post-experimental questionnaire which indicated a significant increase in anger for the experimental group. Also, in the TAT stories there was more aggressive content in the aroused group as compared with the control group.

As expected, Cramer (1991a) found that defenses were expressed more often in the aroused group than in the control group. Her results also showed that projection and identification were used more often than denial in this college age sample. These findings supported her hypothesis concerning the development of defense mechanisms, namely that denial is more frequent in early childhood, projection is predominant during adolescence, and identification becomes prominent in late adolescence.

Cramer's (1991a) study also showed a significant

interaction with men using projection more often than women in the experimental group. Her study also showed significant correlations between aggressive content in the TAT stories and defense scores, for the experimental group. In the experimental group, the correlation between aggression frequency and projection was $r = .85$ ($p < .001$). Cramer suggests that the correlation between aggression and projection indicates that those who were most angered by the experimental criticism also used defenses more frequently. However, as will be discussed subsequently, the measures for aggression and projection have some categories in common and therefore there is a built in correlation.

Critique of Cramer. An empirically designed scoring system should show significant discrimination among the criterion groups (Messick, 1978). In Cramer's (1987) initial formulation of the categories for the DMM she used 42 subjects in three age groups, with mean ages of 5 years 2 months, 9 years 1 month, and 11 years 9 months. She describes her method of developing the scoring categories for Denial, Projection and Identification as follows:

> These categories were derived from the analysis of empirical findings of themes and response styles which seemed to differentiate between older and younger children...(1987, p. 601).

This description implies that the scoring manual was empirically derived. However, Cramer never states whether the differences she alludes to were statistically significant or whether statistical analysis was employed.

Having identified a set of scoring categories for the three defenses, Cramer (1987) conducted another study with a new sample of children in four age groups with means of 5 years 8 months, 9 years 10 months, 14 years 6 months, and 16 years. She was able to replicate her previous finding of age differences in defense use, that is, that denial was used more often by the youngest age group, identification was used more often by the oldest age group and projection was used more often by the two middle age groups. This replication of age differences does not, however, establish whether the categories validly reflect the particular defense named.

There are several other questions in regard to Cramer's (1991b) system. They are as follows:

1. Cramer does not provide information for individual categories. She doesn't discuss the basis for the derivation of each of the seven categories. This lack of information makes it difficult to evaluate her results.

2. A comparison between the Haworth (1963) and Cramer (1991b) projection scoring systems shows some overlap in categories. However, each author has

categories which are not represented by the other. There are also categories representing defenses other than projection in the Haworth system which match categories of projection in the Cramer system. For example, Haworth's defense regression is scored if the story contains references to ghosts, witches, or a haunted house and symbolization is scored if items are broken. Both the references to ghosts and decrepit or broken things are scored in Cramer's second category for projection. Furthermore, Cramer does not explain why she omits some of Haworth's categories. The lack of more consistent overlap between the Cramer and Haworth systems suggests that some of the categories in one or both may be incorrect. There may also be a need to develop other categories which would more accurately represent defensive projection.

3. Cramer's scoring system relies on expressions of hostility as indicators of projection. However, her scoring system does not provide for projection in response to other kinds of threat (e.g., unacceptable sexual impulses).

4. Cramer does not discuss the conditions under which defenses would be expected to be aroused. Cramer's (1991a) study of anger and defense mechanisms provides no discussion of the possibility that the subjects who are angered are not using a defense but are instead

expressing anger directly in the stories. For example, a story containing an assault would be scored in category five (Apprehensiveness of death, injury or assault). While Cramer (1991b) explains this category as the projection outward of dangerous internal impulses, she does not explain why an assault should be considered projection rather than expression of simple anger or aggression. (It should be pointed out that Cramer scores some of the same phrases for both aggression and projection, thus building in a correlation between the two sets of scores.) The high correlations that Cramer (1991a) found between aggression scores and projection scores for her experimental group, raises the question as to whether the DMM is identifying defensive projection or simple aggression.

Present Research

The present research investigated the validity of the Cramer (1991b) system for scoring defensive projection in TAT stories. It was a modified replication and extension of the Cramer (1991a) study of anger and defense mechanisms in college students. For reasons to be explained, the subjects told stories to only six TAT cards, rather than to eight cards as in the Cramer study. The present research had a control group, corresponding to Cramer's, and an experimental group (anger), similar to Cramer's, which experienced an intervention intended

to produce anger and arouse defenses. The present research added a third condition in which subjects were asked to imagine that they had experienced the same anger manipulation that was experienced by those in the anger condition. The stories of all groups were scored using Cramer's system for defensive projection. The third condition was added to test the ability of the scoring system for projection to distinguish between a group which is angry and defensive and a group which is pretending to be angry but has no reason to be defensive.

All stories were scored for defensive projection, and aggression frequency, following the procedure used by Cramer. The scores for both variables for the simulated-anger group were compared to those of the anger group. Unlike the anger group, the subjects who were told to imagine anger were not expected to show an increase in their use of projection, first, because the directions they were given, (to pretend anger) made the expression of anger socially acceptable for the situation, and second, because the directions to this group provided no real threat to their self-esteem. The simulated-anger group was expected to show an increase in aggression scores but not in projection scores. If the anger group shows higher projection scores than the simulated-anger group, this would constitute evidence for the validity of Cramer's (1991b) scoring system. However, if the

projection scores for both experimental groups are similar and significantly higher than the control group, then issues regarding the confounding of projection and aggression must be considered. If, for any of the seven categories in Cramer's projection scoring system, an increase in the simulated-anger group is shown, then the validity of these individual categories is called into question.

The Cramer (1991a) study revealed a sex difference for projection, specifically that men had higher projection scores than did women. Therefore, the present data were also analyzed for sex differences.

In a study of repression, Weisskopf-Joelson, Asher, Albrecht, and Hoffman (1957) suggested that the stronger the repression of an unacceptable impulse, the more likely it is to be expressed with inappropriate stimuli (i.e., calm stories to aggressive pictures and aggressive stories to pictures judged to be non-aggressive.) The six pictures used in this study were rated for aggression "cue strength" by ten raters (five men and five women with a minimum of a Master's degree in psychology, education, science or engineering). Cue value is defined by Peterson and Schilling (1983) as "the tendency of the test stimuli to evoke...certain cognitive and affective responses in the subject" (p. 265). In the present research, the judges rated the level of aggression that

they perceived in each of the six TAT cards. The cue value (referred to as "label" by Weisskopf-Joelson et al., 1957) was needed in order to investigate the possibility of a relationship between strength of defense (i.e., projection), and the expression of an impulse to inappropriate stimuli.

Hypotheses

Aggression. Due to the experimental manipulations, aggression scores should be higher for the anger and simulated-anger groups than for the control group. Therefore, it is hypothesized that:

> Hypothesis 1. For the post-criticism stories, mean aggression frequency should be higher for the anger and simulated-anger groups than for the control group.

Cramer's (1991a) study showed no significant effect of gender on aggression frequency. Therefore, there is no hypothesis regarding gender and aggression scores.

Projection. If Cramer's (1991b) scoring system is a valid measure of defensive projection then it is hypothesized that:

> Hypothesis 2. For the post-criticism stories, the mean of the defensive projection scores for the anger group should be significantly higher than the mean for the control group.
>
> Hypothesis 3. For the post-criticism stories, the

mean of the defensive projection scores for the anger group should be significantly higher than the mean for the simulated-anger group.

If the mean of the projection scores for the anger group and the control group are not significantly different then it will indicate a failure to replicate. If the mean of the projection scores for the anger group and the simulated-anger group are both significantly higher than the control group, then the validity of the scoring system is called into question.

The previous hypotheses deal with a total projection score for the post-criticism stories for each person. It is also possible to make comparisons among groups with respect to each scoring category for the post-criticism stories (something Cramer does not report). For each of the seven categories of the Cramer scoring system, means will be determined for each of the three groups. If these individual categories are valid components of defensive projection, then it is hypothesized that:

Hypothesis 2a. For the post-criticism stories, the mean of the defensive projection scores for each scoring category for the anger group will be significantly higher than the mean for the corresponding category in the control group.

Hypothesis 3a. For the post-criticism stories, the mean of the defensive projection scores for each

category for the anger group will be significantly higher than the mean for the corresponding category in the simulated-anger group.

If there are individual scoring categories in which the anger and simulated-anger groups have scores which are similar to each other and significantly higher than the control group, then the validity of that particular category as a component of projection would be called into question.

A picture by picture analysis is also employed to determine the "pull" or cue value for projection of the pictures used in this study. Cramer (1991a) does not report any information regarding defensive projection scores by picture. However, it will be of interest to find out whether some pictures elicit projection more effectively than others.

Sex differences in the use of defensive projection have been shown in previous studies using the DMI (e.g., Cramer, 1979; Cramer & Carter, 1978). In Cramer's (1991a) study men in the experimental group had higher projection scores than women. Based on these findings it is hypothesized that:

Hypothesis 4. For the post-criticism stories, for the anger condition, the mean of the defensive projection scores for men will be higher than the mean for women.

Relation between projection and aggression scores. A positive correlation between projection scores and aggression scores would replicate Cramer's (1991a) findings. Therefore, it is hypothesized that:

Hypothesis 5. For the post-criticism stories, for the anger group, there should be a positive correlation between projection scores and aggression scores.

Hypothesis 6. For the post-criticism stories, the positive correlation between projection scores and aggression scores should be greater for the anger group than for the simulated-anger group.

For the simulated-anger group, a finding of no correlation between projection scores and aggression scores would provide support for the Cramer scoring system. If, however, scores for projection and aggression in the simulated-anger group show a significant positive correlation, then this would call into question the validity of the Cramer scoring system, because there is no reason in this group to expect defenses to increase with aggression.

The theory that the stronger the repression of an unacceptable impulse, the more likely it is to be expressed to an inappropriate stimulus (Weisskopf-Joelson et al. 1957) leads to the following hypothesis.

Hypothesis 7. The mean of the aggression

scores of high-defensive subjects (those who score high on projection) will be higher for pictures rated as having low cue for aggression than for pictures rated as having high cue for aggression.

Chapter II

Method

Overview

The present study included a control group and two experimental groups. All subjects told stories to the same six TAT cards. Subjects in the control group told stories under a neutral classroom condition. These stories provided a baseline for comparison with the two experimental groups. The anger group was subjected to an experience designed to induce anger and/or threaten self-esteem. Subjects in the simulated-anger group were asked to imagine that they had just been through the same anger manipulation as the anger group. All stories were scored blind for defensive projection and aggression.

Pretesting. Pretesting was carried out in order (a) to try the pictures and scoring systems used in the Cramer (1991a) study, and (b) to test the most appropriate type of instructions to use with the simulated-anger group. The two instructions which were tested were to "write aggressive stories", or "to pretend they had been through an experience which had angered them." Instructions to write aggressive stories produced stories with action-oriented and frequently murderous aggression. The following story told to card 8GF is an example:

> Mary is out on bail for killing her husband.

Her lawyer and friends assure her that it was justifiable. He had beaten her for years. She thought he would kill her that night. She had no other choice.

She doesn't care anymore. She ponders her fate. She has lost the only man she ever loved. By her own hand.

What can the court do to her she hasn't already done to herself.

The instructions to pretend anger produced stories that were more varied in content. The following is an example of a story written to card 8GF with instructions to pretend anger:

This woman has just been dumped by her boyfriend. She is currently reviewing in her mind the conversation with her boyfriend. Although he broke up with very generic statements - we just don't match, etc- she is blaming herself for the breakup, and trying to dissect the relationship to identify behavior that she could have changed. She is feeling very negative about herself. She will do this for some time, ultimately coming to the realization that there is no blame that she should internalize, perhaps also getting angry, until she puts this behind her and gets on with her life.

The instructions to pretend anger were chosen as the more

appropriate for the present study so that (a) the instructions in the simulated-anger group would be as similar as possible to the instructions in the anger group and (b) the instructions would parallel those used by Cramer (1991a) and (c) a wide range of content would be obtained.

Subjects

As in Cramer's (1991a) study, the subjects were male and female college students. The students were drawn from college subject pools, in which they were required to participate in two hours of research as part of the requirement for their Introductory Psychology course. The experiment fulfilled one hour of their requirement. Because subjects were in the process of taking an Introductory Psychology course, they would not be expected to have had extensive experience with projective testing. The two colleges from which subjects were obtained were four year colleges in the City University of New York (CUNY) system.

Standard procedures of each college were followed for recruiting students. That is, sign up forms (see Appendix C), were posted at the location used for that purpose. The forms listed the time and place of the experiment and the title of the study as: "Creative imagination in college students". The following description was provided on the sign up sheet: "In small

groups, subjects will write brief creative stories to six pictures and fill out a brief questionnaire." There were seven spaces listed under Men and seven for Women. Students selected which group they chose to join. There was no indication on the sign up forms that there were three conditions. The data were collected at various points during the fall and spring semesters 1994-95.

Because of the large number of subjects who were foreign speaking, and the need for a basic proficiency in written English to complete the study, the stories of all subjects who listed a language other than English as their primary language were reviewed for English proficiency. Two stories from each of these 58 subjects were read by two judges to determine whether their English proficiency was sufficient to complete the task required by this experiment. If either of the two sample stories was shorter than 30 words, or if the ideas presented were not clear, the remaining four stories for that subject were screened by the two judges. Grammatical errors were not considered unless they made the story incomprehensible. Using this procedure, six subjects were dropped because of insufficient proficiency in written English. A cut-off of 30 words was chosen as a minimum necessary story length based on the recommendation of Walker and Atkinson (1958) that stories which are shorter than 30 words are unscorable.

A subject needed to have at least four scorable stories, or two-thirds of their data, to be included (Smith, Feld, & Franz, 1992). One subject had three stories which were shorter than 30 words each, and was removed.

All subjects passed the following criteria in order to be included in the final analysis: at least four of their six stories had to be at least 30 words in length, and the stories needed to present an understandable flow of ideas. Because the main requirement for inclusion was whether the subject had a sufficient proficiency with written English, it was not automatically assumed that English speaking students had the required proficiency. The stories of any English speaking subject who had one or more stories shorter than 30 words were also reviewed. However, none of the English speaking students was removed from the analysis for this reason.

In summary, there was a total of 286 students who signed up and participated in the experiment. Of that number, 218 (89 men and 129 women) were included in the final analyses. There were 68 subjects (55 women and 13 men) who were dropped for the following reasons: 27 because they were outside the age range of 17 to 23 years of age, 8 because they provided incomplete data on the post-experimental questionnaire, 6 because their use of written English was considered insufficient, 1 because of

insufficient story length, 1 because of lack of cooperation, and 25 subjects because their groups were not gender mixed.

The subjects included in the final analysis were divided into the six cells as shown in Table 1. Because subjects signed up for their groups based on the posted times, it was not possible randomly to assign individual subjects to conditions. The groups were then assigned to conditions based on the following considerations. First, as testing proceeded, sign-up groups were assigned to conditions to maintain an approximately equal number of subjects in each condition. Second, time of testing session affected assignment of sign-up group. Where possible, subjects assigned to the control and simulated-anger conditions were run before the anger condition in order to reduce the possibility of subjects divulging the

Table 1

Subjects by Cell

	Condition			
Gender	Control	Anger	Simulated Anger	Total
Men	30	31	28	89
Women	49	30	50	129
Total	79	61	78	218

anger manipulation. Groups consisting of only one male subject were not assigned to the anger condition in order not to impose an unpleasant experience unfairly on women more than men.

The initial plan of the study was to use a minimum of 162 subjects equally divided between men and women, that is, 27 subjects in each of six cells. The initial number of 162 was arrived at by means of a power analysis intended to test an hypothesis concerning defensive projection scores by means of a 2 x 3 ANOVA. Assuming an alpha level of .05, and, using an estimate based on previous studies, a medium effect size (*f* = .25, Cohen, 1988, chap. 8) calls for a minimum of 162 subjects to allow for tests of main effects and interaction with power of at least .80 for each test.

Because of the higher ratio of women in the subject pool, it was necessary to run the groups until the male cells reached a minimum of 27 subjects. This resulted in an excess number of female subjects. After data collection, it was decided to use all available data, even though this would mean a different *n* for each cell.

Measures

The present study used six pictures instead of the eight used by Cramer (1991a). This was done in order to reduce the time necessary for a group testing session. All subjects viewed the six TAT cards in this order: 5,

6GF, 1, 8GF (as modified by Cramer), 7BM, and 18GF (as modified by Cramer). Descriptions of these pictures were given earlier (see page 11). Cramer (1991a) randomized card order for the four post-criticism cards. However, because there appears to be no advantage in randomizing the order of the last four pictures, this was not done in the present study. Giving pictures in the same order for each subject was done to make it easier to compare the extent to which each picture contributes to the projection score.

The cards that Cramer (1991a) used were balanced for number of figures in each picture; half the cards had one figure and half had two. Although, Cramer (1991a) reports that she found no clear effects attributable to number of persons in pictures, the six pictures in the present study were also balanced as to the number of figures in each picture. Male and female characters in the pictures are also balanced as closely as possible. At least one male character is present in three of the pictures, and at least one female character is present in four of the pictures.

Prior to data collection, ten judges rated the six pictures for their apparent level of aggression using a scale of zero (no aggression) to seven (high aggression). This was done in order to conduct an analysis of interaction of cue value of the pictures and the level of

defensiveness of subjects. The pictures received the mean ratings for level of aggression listed in Table 2.

Post-experimental questionnaire. As in the Cramer (1991a) study, a brief questionnaire was administered to subjects at the end of the testing session. This assessed how the subjects perceived their affect at the beginning and at the end of the testing session. In the present study the subjects were asked to circle on a scale of 0 (not at all) to 7 (very much) the way they experienced the following affects: angry, anxious, happy, and sad. The questionnaire was slightly different than that used by Cramer (1991a) in which the subjects were asked to choose one of three responses (i.e., yes, no, a little) to describe how they

Table 2

Mean Ratings of Level of Aggression

Aggression	TAT Code	Description of Picture
.2	8GF	Woman with blackboard and equations
.4	1	Boy with violin
2.0	7BM	Older and younger man
3.3	5	Woman looking into room
3.8	6GF	Woman looking over her shoulder at man
5.9	18GF	Woman with hands around throat of other woman

experienced the four affects. The questionnaire was changed in the present study in order to secure more specific information and a greater range of scores. The questionnaire was also used to secure information about the age, sex, year in school and the primary language of subjects. The subjects were also asked what they thought the experiment was about.

TAT scoring. An associate of the experimenter put a code number on all stories, so that each scorer was blind to group condition. All stories for a single picture were scored before going on to the next picture. All stories for each picture were shuffled so that the order of subjects was different for the scoring of each picture. There were two scoring systems used: Cramer's (1991b) system for scoring projection from the Defense Mechanism Manual, and aggression frequency (Hall & Van de Castle, 1966; Whitman et al., 1961).

Cramer's (1991b) scoring system for projection has seven categories, and each is scored as many times as it occurs in a story. However duplicates of a scorable word or phrase are not scored (P. Cramer, personal communication, January 30, 1993).

For the aggression frequency score, each occurrence of aggression in a story is given a value of one. An item was scored for aggression frequency if it matched the criteria of either the Hall and Van de Castle (1966)

system or the aggression scoring system of Whitman et al. (1961). Hall and Van de Castle define aggression as an act intended to harm or annoy another, or the desire to do so. Their aggression scoring system includes categories ranging from a "covert feeling of hostility or anger without any overt expression" to "an aggressive act which results in the death of a character" (Cramer, 1991b, p. 43). The Whitman et al. system is a seven point scale (0-6), encompassing three levels of aggression. Each level of aggression is split between human and non-human objects of the aggression. The highest level, category six, of the Whitman et al. system is "death or death threat by stabbing, shooting, pushing, striking, hit by car, drowning, illness, warfare, animal attack, mutilation, drive to insanity, violence" toward a human. Category five is "equivalent destruction of animate or inanimate objects other than human" (p. 220).

Whitman et al. (1961) do not provide examples of the aggression scoring system. However, they cite the scoring systems of Saul and Sheppard (1956) and Gottschalk, Gleser, Springer, Kaplan, Shanon and Ross (1960) as the basis for their system. A later manual of Gottschalk and Gleser (1969) provided multiple examples for their categories. The scorers for the present study used these examples from Gottschalk and Gleser as a guide in using the aggression frequency system. The aggression

frequency score was the number of occurrences of aggression identified by either the Hall and Van de Castle or the Whitman et al. scoring system.

Cramer (1991b) provides some sample stories to demonstrate her scoring system, but there were no published materials specifically designed for scorer training. However, upon request, Cramer provided a sample of 50 stories with expert scoring that she uses for training others in the use of this system. For the present research, the independent scorer for the projection and aggression scoring systems was trained by the author by using Cramer's published sample stories, as well as the training materials that she provided. The practice stories that were used were those secured during pilot testing and those of subjects who were removed from the final analysis for various reasons.

Interrater reliability was determined by comparing the author's scoring with that of the independent coder, a Ph.D. psychologist, on a subsample of the data consisting of 25% (n = 55) of the sample. The sample was chosen by selecting every fourth subject from the master list of subject codes.

Procedure

The study was conducted in a classroom setting. The session began with all subjects reading and signing the consent form (see Appendix D) to participate in the

study. The TAT stories were written rather than oral. Cramer's (1991a) subjects were tested individually and her subjects produced stories orally which were recorded on tape. The change to written stories in the present study was done to accommodate group rather than individual testing. Group administration was chosen to facilitate data collection. In a previous study, Cramer (1987) collected stories by written as well as oral means. She does not report any differences based on the two methods of data collection. Also, it has been suggested (Juni, 1980) that individual testing by an authority may restrict a subject's freedom to use defenses.

As in the Cramer (1991a) study, the experimenter began with the explanation that the experiment was an investigation of creative imagination in college students. Cramer states that she followed this explanation with the standard Murray (1943) instructions. The present study used the Murray (1943, p. 3) instructions adapted for group use. See Appendix E.

The subjects then had 4 minutes to write each story. After three minutes had elapsed, the experimenter announced "There is one minute left." The subjects used forms which listed, at the top of the page, the following four questions (adapted for group administration) from Murray's instructions.

What has led up to the event shown in the picture?

Describe what is happening at the moment.

What are the characters thinking and feeling?

What is the outcome?

For reasons to be explained, for all groups the pages on which each story was written were collected before the next story was begun.

Control group. After initial instructions, subjects in the control group wrote their stories with no further intervention. They then filled out the post-experimental questionnaire.

Experimental group I. The anger group began in the same manner as stated above. After the first story was written and collected, and while subjects were writing the second story, the experimenter stood in front of the class and, in an obvious fashion, silently read through the first set of stories. (The experimenter needed to convey to the subjects that she had read their stories.) After collecting the second story, the experimenter made negative comments to the group about their ability to write stories. The comments are adapted as closely as possible from those used by Bellak, (1944), and Cramer, (1991a). The negative statements were as follows: "I've been reading through these stories, and honestly, they are about the worst I've ever read. Could you try to write some better ones?" The comment after the third

story was, "These are no better than the first set." At the end of the forth story, the experimenter stated, "There are still no ideas in them, no life. Could you try to put some effort into this?" The final negative comment, at the end of the fifth story was, "This stuff is pretty hopeless for college students. See if you can do something with this last picture."

After all stories were completed, the subjects completed the post-experimental questionnaire. Subjects were then debriefed. After the debriefing, the experimenter requested that subjects not share information regarding the study with other students.

As part of the initial plan for data collection, the groups were to be scheduled to run as close together as was feasible, and the anger groups were scheduled to run last. This was planned in order to reduce the possibility of subjects divulging information about the study to other prospective subjects which may have contaminated results. However, the high absentee rate (25-50%) necessitated running many more groups than was anticipated, and at times when subjects were available. However, at the end of each testing session, the experimenter asked if any subject had been told about the experiment beforehand. No participant admitted knowing anything about the anger manipulation.

Experimental group II. The simulated-anger group

received the same initial instructions as the prior two groups. At the end of their second story they were given the following instructions.

> Before you begin writing the next story, I want you to imagine that I have been reading your stories, and that I said to you "I've been reading through these stories, and honestly, they are about the worst I've ever read. Could you try to write some better ones?" Imagine that this statement has made you understandably very angry. Pretend that this is the way you are feeling as you write your next story.

At the end of the third story, the following instructions were given: "I want you to imagine that after reading your last set of stories, I state that these are no better than the first set." After the fourth story, the group was given the following instructions: "Again I want you to imagine that after reading more of your stories I tell you that there are still no ideas in them, no life, and could you try to put some effort into this?" At the end of the fifth story, the experimenter gave the following instructions: "Again I want you to imagine that after reading more of your stories I tell you that this stuff is pretty hopeless for college students, and see if you can do something with this last picture."

Ethical Issues

All students were debriefed after each session and the experimenter remained to answer questions, in the remaining time. Deception was a necessary component of the present study, because full disclosure of the purpose of the study would have made the experimental manipulations meaningless. The deception involved concealing the full purpose of the study from all subjects until they were debriefed, and, for one experimental group, making derogatory comments about the quality of their stories. The comments, which were intended to produce anger and/or threat to self-esteem, posed the most serious ethical issue for this study. The question was whether subjects would be harmed by the manipulations. The intervention was based on Cramer's (1991a) study. She reports no adverse consequences with her intervention. The state of anger was maintained during the writing of four stories and the post-experimental questionnaire. This took approximately 25 minutes. Subjects were then debriefed. The full text of the debriefing is presented in Appendix F. The debriefing was patterned after the process debriefing technique (Lepper, Ross, & Lau, 1986; Lord, Lepper, & Preston, 1984) in order to reduce the perseverative effects of inaccurate feedback. Although this intervention was temporarily unpleasant, no lasting

harmful effects were expected following the debriefing.

During debriefing, the experimenter explained the purpose of the study and the need for deception. It was stressed that the negative comments made about their stories were pre-planned and had no relation to the quality of their stories. In order to emphasize the pre-planned nature of the criticism, the experimenter shared with subjects the experimenter's packet which included a copy of the scripted criticisms.

The experimenter approached the debriefing in a two-fold manner. First, an explanation of perseverance of inaccurate feedback was given with an example, and second, a technique was suggested to help remove perseverative effects. That technique, suggested by Lord, Lepper, and Preston (1984) is to have subjects "consider the opposite." This technique involved having the subjects think about what they liked best about their stories, and verbalize this for the group. They were also asked to discuss what their response might have been if, instead of criticism, the experimenter had said, "These are the best stories I have ever read." It was expected that a debriefing of this type would be able to deal with subjects' anger. Every attempt was made to insure that any disturbance on the part of subjects was satisfactorily resolved.

All subjects were given a copy of their consent form

which contained the phone number of the experimenter and they were encouraged to use this if they had any questions or comments they wished to make about the research. No student used this option to call the experimenter. The consent form also contained the phone number of the office of the head of the subject pool for each college.

Subjects were also offered the option of receiving a summary of findings from the researcher, and at the end of the debriefing, addressed an envelope provided by the experimenter for this purpose.

Chapter III

Results

Prior to a report of tests of hypotheses, information will be presented concerning the sample, interscorer reliability, and the effectiveness of the experimental manipulations.

Description of Sample

The subjects ranged in age from 17 to 23 years old, with a mean age of 19.1 years. The ages of subjects are presented in Table 3.

According to a 2 (gender) by 3 (condition) ANOVA, comparing ages of subjects in each condition, there were no significant main effects for gender, $F(1, 212) = 1.17$, $p < .28$; or condition, $F(2, 212) = .26$, $p < .77$; and no interaction, $F(2, 212) = .15$, $p < .87$. Men had a mean

Table 3

Ages of Subjects

Age	Men	Women	Total	Percent
17	3	7	10	4.6
18	31	48	79	36.2
19	28	39	67	30.7
20	13	19	32	14.7
21	6	10	16	7.3
22	5	6	11	5.1
23	3	0	3	1.4

age of 19.2, SD = 1.39, and women had a mean age of 19.0, SD = 1.22. The mean ages for each condition were: control = 18.9, SD = 1.20; anger = 19.1, SD = 1.24; and simulated-anger = 19.1, SD = 1.43.

Cramer's (1991a) subjects were mostly juniors and seniors who were paid to participate in her study. Because there appeared to be no reason to believe that either years in college or being paid was relevant to testing Cramer's hypotheses, no attempt was made in this research to replicate these aspects of her study. Moreover, because subjects were being drawn from subject pools it was expected that the majority would be freshman and sophomores. In fact, the majority of subjects, 84%, listed themselves as freshmen or sophomores. The mode for year in school was first year, with 55% of subjects reporting themselves to be freshman. Five students did not list their year in school. Two of these were 18 years old, two were 19 years old and one was 21 years old. A 2 (gender) by 3 (condition) ANOVA revealed no significant main effects or interactions for year in school, gender, $F(1, 212) = .94$, $p < .33$; condition, $F(2, 212) = .39$, $p < .68$; interaction $F(2, 212) = .13$, $p < .89$.

Of the 40 groups tested, five (25 subjects) consisted of all women and were dropped from the analysis for that reason. The remaining groups contained both men

and women, and averaged between seven and eight participants.

The proportion of women to men (1.8 to 1) signing up for the study matched the gender makeup of the subject pools, (reported by the head of the subject pool at one of the colleges as approximately 2:1 women to men). The ratio of women to men used in the final analysis was 1.4 to 1.

The ethnic diversity of New York City was reflected in the obtained sample. There were 58 subjects (20 men, 38 women) who listed a language other than English as their primary language. This represented 27% of the sample, and included 17 different languages: Russian (15 subjects), Spanish (12), Chinese (12), Hindi (3), Hebrew (2), Polish (2), Filipino (2), Portuguese (1), Rumanian (1), Yoruba (1), Swedish (1) Armenian (1), Haitian (1), Bengali (1), Korean (1), Tamil (1), Urdu (1).

Interscorer Reliability

A subset (n = 55) of stories from the total sample was scored by the experimenter and an independent scorer. Total scores were compared by using a Pearson Product Moment correlation. Category scores were compared by using percent of category agreement. Category agreement was calculated in the following manner (Smith, Feld & Franz, 1992):

$$CA = \frac{2\ (\#\ \text{of agreements on presence of category})}{(\#\ \text{present, scorer 1}) + (\#\ \text{present, scorer 2})}$$

Aggression scores. The interscorer reliability for total aggression frequency scores was r = .91 for the Whitman et. al system and r = .95 for the Hall and Van de Castle system (Table 4). The overall category agreements for the Whitman et al. categories ranged from 67% to 86% and for the Hall and Van de Castle categories ranged from 69% to 87% (Table 5). The interscorer reliabilities for each picture separately for aggression scores are also listed in Table 4. Cramer (1991a) reports an interscorer reliability of r = .97 for aggression intensity (i.e., the Hall and Van de Castle system). Her interscorer reliability for her second aggression score, (i.e., aggression frequency) was r = .90.

As previously stated, the present study did not use aggression intensity as a dependent variable as did Cramer (1991a). This was done because the Hall and Van de Castle system is described by its authors as a nominal scale, but was used by Cramer as if it were an interval scale. However, the dependent variable that Cramer

Table 4

Interscorer Reliabilities for Aggression and Projection Scores by Picture

	Picture						
Scoring System	1	2	3	4	5	6	Total
Aggression Hall & Van de Castle	.98	.88	.93	.86	.96	.99	.95
Aggression Whitman et al.	.83	.89	.86	.94	.85	.88	.91
Projection	.89	.86	.83	.88	.86	.90	.89

labels "aggression frequency" was determined, both by the Cramer and the present study, by using the categories of Hall and Van de Castle and the categories of Whitman et al. to identify instances of aggression. Therefore category agreements for both the Hall and Van de Castle and the Whitman et al. systems were computed in the present study and are listed in Table 5.

There was some overlap in content of categories between the Hall and Van de Castle and Whitman et al systems. For example, category 8 of Hall and Van de Castle identifies instance of "death by murder," and category 6 of Whitman et al. identifies instances of "death by murder or illness or accident". In order to eliminate problems, caused by overlapping categories, in computing reliability or category agreement, the Hall and Van de Castle system was chosen as the primary system.

That is, all instances of aggression were first identified by the Hall and Van de Castle categories, and then the Whitman et al. categories were used to identify any remaining instances of aggression present in the story. Thus, for each instance of aggression: (a) it was scored in only one system, (b) if it could be scored under both systems, the Hall and Van de Castle system was chosen. Therefore, the category agreements listed in table 5, represent no overlap of scores per categories.

Cramer (1991a) does not state whether her reported interscorer reliability of $\underline{r}$ = .90 for aggression frequency is for the Whitman et al. system only, or for a combination of the two scoring systems.

Projection scores. The interscorer reliability for total projection scores (sum of six stories) was $\underline{r}$ = .89 (Table 4). The overall category agreements ranged from .83 to .90 (Table 6). Interscorer reliabilities by story are listed in Table 4. Category agreements for each story are listed in Table 6.

Post-Experimental Questionnaire

The effect of the experimental manipulations was assessed by means of responses to the post-experimental questionnaire and by aggression scores obtained from the TAT stories. If the experimental manipulations were

Table 5

Percent of Category Agreement (CA) for Aggression Scoring Systems

Category	Story 1	2	3	4	5	6	Overall CA[a]
Hall & Van de Castle System							
1	86	89	88	93	92	75	87
2	67	73	69	69	86	83	75
3	67	69	81	77	76	83	76
4	*	78	0	*	50	73	69
5	73	0	93	*	100	100	85
6	67	75	100	67	100	80	80
7	73	50	86	100	0	80	75
8	93	100	*	100	87	80	87
Whitman et al. System							
1	64	50	80	89	100	0	73
2	90	89	82	79	86	84	85
3	0	*	100	*	100	*	67
4	100	75	67	67	100	89	86
5	100	0	*	*	*	*	67
6	91	89	67	80	88	85	85

- Notes. * indicates category did not occur in stories about this picture.

CAs of 0% or 100% represent few scores, averaging between 1 and 2 scores for that category for that story.

[a] Final column is category agreement calculated for the sum of all 6 pictures.

Table 6

Category Agreement (CA) for Projection by Story

Category	Story 1	2	3	4	5	6	Overall CA[a]
1	80	80	76	69	77	75	76
2	85	67	86	100	86	85	84
3	75	80	79	67	67	40	73
4	76	74	100	100	75	55	76
5	90	96	86	100	89	87	89
6	80	67	*	*	*	80	78
7	80	100	*	100	*	67	83

Note. * indicates category did not occur in stories about this picture.

CAs of 100% represent few scores, averaging between 2 and 3 scores for that category for that story.

[a] Final column is CA calculated for the sum of each category present in all 6 pictures.

successful, the amount of anger reported on the post-experimental questionnaire should be higher for the anger condition than for the control or the simulated-anger condition. A 2 (gender) x 3 (condition) MANOVA was employed to compare the means of four affect scores (angry, anxious, happy, sad). This analysis was carried out in two ways. First, the dependent variable was the affect scores that the subjects reported feeling at the end of the experiment (end scores), and second, the dependent variable was each subject's end score minus the

affect score that they reported feeling at the start of the experiment (difference scores).

For end scores there was a significant main effect for condition, for the affect "angry," multivariate $F(8, 418) = 2.29$, $p < .03$, univariate $F(2, 212) = 7.35$, $p = .001$. The mean of the scores for the affect angry was higher in experimental group I (anger group, $M = 1.85$, $SD = 2.38$) and experimental group II (simulated-anger, $M = 1.81$, $SD = 2.11$) than in the control group ($M = .70$, $SD = 1.20$). The remaining affects, anxious, happy, and sad, showed no effect for condition. Affect scores are presented in Table 7.

Moreover, for end scores, there was a trend toward a significant main effect for gender for the affect "sad", multivariate $F(4, 209) = 1.95$, $p = .10$, univariate $F(1, 212) = 4.62$, $p < .04$, with women ($M = 2.2$, $SD = 2.53$) expressing more sadness than men ($M = 1.5$, $SD = 2.08$).

Because the assumption of homogeneity of variance was violated for the affect "angry", this analysis was rerun using the square roots of all the affect scores. This did not correct the problem with the variance, so the analysis was redone using the square roots of scores for "anxious", "happy", and "sad" scores, and the square root of the logarithm of "angry" scores. (In all cases in which logarithms were used, one was added to each score before computing the logarithm, in order to

Table 7

Post-Experimental Questionnaire Mean Affect Scores (End Scores)

	Condition		
Gender	Control	Anger	Simulated Anger
Men			
angry	1.0	2.0	1.8
anxious	2.1	2.8	2.2
happy	2.1	2.7	2.0
sad	1.2	1.4	1.8
Women			
angry	.5	1.7	1.8
anxious	2.0	2.7	2.5
happy	2.6	2.1	2.3
sad	2.2	1.9	2.5

N = 218

accommodate zero scores.) Using the square root of the logarithm for "angry" corrected the problem with the variance. This analysis reduced the significant main effect for condition to only a trend, multivariate $F(8, 418) = 1.72$, $p = .09$; univariate $F(2, 212) = 5.58$, $p = .004$. The trend toward a main effect for gender for the affect "sad" was also weakened slightly, multivariate $F(4, 209) = 1.83$, $p = .12$; univariate $F(1, 212) = 4.08$, $p = .045$. This analysis of end scores indicated that while there was an increase in anger in the expected direction

(i.e., the anger group should have become angrier than the control group) the anger manipulation was not as successful as was intended.

A second 2 (gender) by 3 (condition) MANOVA was run for affect scores, using difference scores (affect at end - affect at beginning). For condition, the multivariate F only approached significance $F(8, 418) = 1.8$, $p = .068$. A trend was shown on a univariate level, anger scores showed a significant difference, $F(2,212) = 4.57$, $p = .011$. However, the pattern was the same as that found in the end score analysis with the following marginal means for the affect anger for each condition: control, .37 (SD = 1.30); anger, 1.23 (SD = 2.39); and simulated-anger 1.23 (SD = 1.91).

As in the analysis for end scores for affect, there was no significant effect for gender, multivariate $F(4, 209) = .39$, $p = .82$; and no interaction effect, multivariate $F(8, 418) = .75$, $p = .65$.

Test of Hypotheses Concerning Aggression

No differences in mean pre-criticism TAT aggression scores were expected among the three conditions. A 2 (gender) by 3 (condition) MANOVA showed no main effects for gender or condition and no interaction effect for aggression frequency scores for the pre-criticism phase (first two pictures). Multivariate Fs were: for gender $F(2, 211) = .63$, for condition $F(4, 422) = 1.10$, and for

gender by condition $F(4, 422) = 1.82$. The p values were .53, .36, and .20 respectively.

However, because the assumption of homogeneity of variance was violated for both variables in the above analysis, it was rerun using the logarithms of the scores. Using logarithms the problem with the variances was corrected and the analysis still showed no significant main effect for condition or sex and no interaction effect.

As stated in Hypotheses 1, the aggression frequency scores from post-criticism stories were expected to be higher in the anger and simulated-anger groups than in the control group. This hypothesis was tested by means of a 2 (gender) by 3 (condition) MANOVA comparing means for total scores (average of stories three to six) for aggression as well as projection scores. A main effect for condition was expected.

This analysis showed no main effect for condition, multivariate $F(4, 422) = 1.51$, $p = .20$. A univariate F test was also not significant for aggression frequency. This analysis showed a trend toward an interaction effect, for aggression frequency. In the simulated-anger condition men expressed more aggression than women, and for the anger group women expressed more aggression than men, multivariate $F(4, 422) = 2.10$, $p = .08$, and univariate $F(2, 212) = 3.65$, $p < .03$. Because of a

problem with homogeneity of variance for one of the variables (projection) this analysis was carried out using the logarithms of the scores. There were again no main effects for condition for aggression scores, and the trend toward an interaction effect was essentially the same, multivariate F (4, 422) = 2.04, p = .09, univariate F(2, 212) = 2.89, p < .03.

In addition the same hypothesis was tested comparing post-criticism aggression scores by means of a 2 (gender) x 3 (condition) MANCOVA with the pre-criticism aggression scores used as a covariate. A main effect for condition was expected. However, there was no main effect for condition shown with multivariate F(4, 418) = 1.54, p = .19, univariate F(2, 210) = 1.26, p = .29. Thus, hypothesis 1 was not supported.

There were no hypotheses regarding gender differences for aggression scores. There were no gender differences shown for aggression scores.

Test of Hypotheses Concerning Projection

Pre-criticism scores. No difference in mean pre-criticism projection scores was expected among the three conditions. A 2 (gender) by 3 (condition) MANOVA was carried out with aggression frequency and projection scores for the first two stories as dependent variables to determine whether the three conditions differed. There was no main effect for condition for the pre-

criticism projection scores, multivariate $F(4, 422) = 1.10$, $p = .36$. There was a trend toward an interaction effect for projection, with women in the control condition ($M = 1.8$, $SD = 1.19$) scoring higher than men ($M = 1.3$, $SD = 1.27$), and in the simulated-anger condition men ($M = 1.7$, $SD = 1.37$) scored higher than women ($M = 1.1$, $SD = .87$), multivariate $F(4, 422) = 1.82$, $p = .12$, univariate $F(2, 212) = 3.54$, $p = .03$. When this analysis was run with logarithms of scores, the pattern of effects was the same. The trend toward an interaction effect became stronger, but still not significant, multivariate $F(4, 422) = 2.34$, $p = .054$, univariate $F(2, 212) = 4.39$, $p < .02$.

Post-criticism scores. For the post-criticism scores, Hypotheses 1 through 4, comparing mean total projection scores (average of pictures 3 to 6) by condition and gender, were tested by means of a 2 (gender) x 3 (condition) MANOVA with projection and aggression as dependent variables. Main effects were expected for condition (hypotheses 1-3) and for gender (hypothesis 4). There were no significant effects for condition, multivariate $F(4, 422) = 1.51$, $p = .20$. However, there was an effect for gender which closely approached significance, multivariate $F(2, 211) = 4.06$, $p < .02$, and univariate $F(1, 212) = 3.68$, $p = .056$. Stories of men showed more projection ($M = 1.78$, $SD =$

1.15) than the stories of women (M = 1.51, SD = .94). However, because the assumption of homogeneity of variance was violated for the dependent variable, projection, this analysis was also carried out using the logarithm of the scores. This method corrected the problem with the variances, and showed the same nearly significant effect for gender, multivariate F(2, 211) = 4.08, p < .02, univariate F(1, 212) = 3.62, p = .058.

In addition the MANCOVA previously described under the heading Aggression Scores also provided an analysis of the post-criticism projection scores. This was a 2 (gender) x 3 (condition) MANCOVA with the pre-criticism scores for each dependent variable used as covariates. The logarithms of projection scores were used in the analysis because of previously observed problems with variance. Higher projection scores were expected for the anger group than for the simulated-anger group (hypothesis 3) and for the control group (hypothesis 2). Again, with the first two stories used as covariates, there was no main effect for condition, multivariate F(4, 418) = 1.54, p = .19, univariate F(2, 210) = 1.25, p = .29. The trend toward a main effect for gender shown in the previous analysis was not shown when the pre-criticism stories were added as covariates. There was no interaction effect for projection.

Analyses by scoring category. The scoring system

for defensive projection is made up of seven categories. Hypotheses 2a and 3a, regarding the main effect of condition on each scoring category, was tested by a 2 (gender) x 3 (condition) ANOVA for each of the seven categories. There were no significant main effects for condition in any of the seven analyses. The F statistics for each category are listed in Table 8. Descriptions of the seven categories of projection scale are listed on page 8.

As seen in Table 8, category three of the projection scoring system (Magical or circumstantial thinking) showed a significant main effect for gender, with men (M

Table 8

F Values From ANOVAs for Projection Scoring Categories

Category	Condition df 1,212	Gender df 1,212	Condition X Gender df 2,212
1	1.93	1.52	1.76
2	1.61	.08	.63
3	.07	4.72*	.08
4	1.14	.01	.84
5	.39	4.57*	7.40**
6	1.57	5.65*	1.51
7	.25	3.62	1.28

Note. *p < .05; **p = .001

= .49, SD = 1.01) using this category more often than women (M = .26, SD =.56, see Table 9).

As Table 8 shows, category 5 (Apprehensiveness of death injury or assault) also showed a significant main effect for gender with men (M = 1.84, SD = 1.61) using this category more frequently than women (M = 1.42, SD = 1.32). There was also an interaction effect for this category (see Table 10). For men, the simulated-anger group showed the highest scores of the three conditions. For women, the simulated-anger group showed the lowest scores of the three conditions.

As Table 8 also shows, for category six (Themes of pursuit, entrapment, and escape), an infrequently used

Table 9

Mean Projection Scores for Category 3

	Condition			
Gender	Control	Anger	Simulated Anger	Total
Men				
M	.50	.48	.50	.49
SD	1.23	1.00	.79	1.01
Women				
M	.20	.27	.30	.26
SD	.46	.52	.68	.56
Total				
M	.32	.38	.37	
SD	.84	.80	.72	

Table 10

Mean Projection Scores for Category 5

	Condition			
Gender	Control	Anger	Simulated Anger	Total
Men				
M	1.40	1.61	2.57	1.84
SD	1.33	1.23	2.00	1.61
Women				
M	1.51	1.77	1.12	1.42
SD	1.29	1.33	1.30	1.32
Totals				
M	1.47	1.69	1.64	
SD	1.30	1.27	1.73	

category, there was a significant difference for gender with men (M = .17, SD = .48) using this category more than women (M = .05, SD = .26, see Table 11).

Category seven (Bizarre or very unusual story or theme) was also an infrequently used category. However, as Table 8 shows, there was a trend toward a gender difference F(1, 212) = 3.62, p = .058. The marginal means were .12 (SD =.56) for men and .02 (SD = .15) for women.

Correlations between projection and aggression scores. Pearson product moment correlations were calculated between projection scores and aggression frequency scores for each of the three conditions (listed below in Table 12). This was a test of hypotheses 5, that is, that there should be a significant positive

Table 11

Mean Projection Scores for Category 6

	Condition			
Gender	Control	Anger	Simulated Anger	Total
Men				
M	.07	.16	.29	.17
SD	.25	.45	.66	.48
Women				
M	.06	.00	.08	.05
SD	.24	.00	.34	.26
Totals				
M	.06	.08	.15	
SD	.25	.33	.49	

correlation between projection scores and aggression scores for the anger condition, and hypothesis 6, that this correlation should be greater for the anger condition than for the simulated-anger condition.

Hypothesis 5 is supported by the significant positive correlation between projection and aggression scores for the anger group. Hypothesis 6 was not supported. On the contrary, the correlation between projection and aggression scores for men in the anger group was significantly lower than that of the simulated-anger group (projection with aggression, $z = -17.4$, $p = .0001$). For women, the correlation between projection and aggression for the anger group was not significantly different than the correlation of the simulated-anger group (projection with aggression, $z = -.18$, $p = .35$).

Table 12

Correlations Between Projection and Aggression Scores by Cell

Simulated Gender Anger		Control	Anger
Men			
Projection with Aggression Frequency	.66	.59	.82
Women			
Projection with Aggression Frequency	.49	.73	.75

Note. All correlations are significant at $p \leq .001$ level.

Analysis of cue value and defenses

The relationship between level of projection and appropriateness of aggressive responses to the cue of the pictures (hypothesis 7) was investigated by a 2 (projection) by 2 (cue value) ANOVA, with aggression frequency as the dependent variable. The two groups for projection were determined by a median split of projection scores obtained in the study. The ratings for aggression cue strength were used to separate the six pictures into two groups: high cue for aggression, and low cue for aggression. The aggression scores for those pictures which were rated high in aggressive cue (pictures 1, 2, and 6) were compared to the aggression scores for those rated low in aggressive cue (pictures 3,

4, and 5). An interaction effect would have supported the theory of Weisskopf-Joelson, et al. (1957) which suggests that the stronger the repression of an unacceptable impulse, the more likely it is to be expressed with an inappropriate stimuli. Thus, those who scored high in projection would have the higher aggression scores to the pictures rated least aggressive (low cue value). This was not supported, there was no interaction effect, $F(1, 154) = .14$, $p = .71$. The means for projection and cue are listed in Table 13. The pattern demonstrates why there was no significant interaction effect (hypothesis 7).

Analyses by Picture

There was also a picture-by-picture analysis to determine the stimulus pull for projection and aggression of the pictures used in this study. The means for aggression and projection scores were compared for the individual pictures used in the post-criticism phase of the study. This was done by a 2 (gender) by 3 (condition) MANOVA for each picture. A 2 (gender) by 3 (condition) MANCOVA was also run with aggression and projection scores for the first two pictures (pre-criticism) for each dependent variable used as

Table 13

Means for Aggression Scores by Cue Value and Level of Projection for the Control Condition

	Aggression Frequency	
Cue	Low Projection	High Projection
Low	1.59	2.42
High	2.44	3.39

covariates. The F statistics (obtained in the MANOVAS) for each picture are summarized in Table 14. The means and standard deviations for each variable for the last four pictures are given in Table 15.

Picture three. For aggression frequency for picture three (boy with a violin) the MANOVA showed a significant main effect for condition, multivariate F (4, 422) = 2.81, $p < .03$, univariate $F(2, 212) = 4.73$, $p = .01$ (logarithms of scores were used). The simulated-anger group received the highest aggression frequency scores of the three conditions, with a mean of 3.38, SD = 2.08, the control group's mean was 2.72, SD = 1.80, and the anger group mean was 2.49, SD = 2.20. This relationship became stronger when the scores for the first two pictures were used as a covariate, multivariate $F(4, 418) = 3.71$, $p = .006$; univariate $F(2, 210) = 6.57$, $p = .002$. For aggression frequency, there was no main effect for gender and no interaction effect for picture three.

Table 14

F Values for Aggression and Projection by Picture

	Condition	Gender	Condition x Gender
	Picture 3		
Multivariate F	2.81*	2.12	.28
Univariate F			
Aggression Frequency	4.73*	3.46	.49
Projection[a]	2.20	3.41	.28
	Picture 4		
Multivariate F	1.17	3.07*	1.19
Univariate F			
Aggression Frequency[a]	1.59	4.75*	1.19
Projection[a]	1.38	.32	2.30
	Picture 5		
Multivariate F	.84	3.26*	1.36
Univariate F			
Aggression Frequency	.16	1.23	2.53
Projection	.92	.91	.55
	Picture 6		
Multivariate F	1.26	3.29*	1.88
Univariate F			
Aggression Frequency	1.68	.01	3.02
Projection	1.93	3.88*	2.58

Note. *p < .05.
[a] Indicates that logarithms of scores were used because of problems with variances.

Table 15

Means & SDs for Aggression and Projection by Gender and Condition for Pictures 3 to 6.

	Aggression Frequency			Projection		
Gender	Control	Anger	Simulated Anger	Control	Anger	Simulated Anger
			Picture 3			
Men						
M	2.90	2.68	3.96	1.17	1.36	1.82
SD	2.25	2.33	1.97	1.51	1.58	1.96
Women						
M	2.61	2.30	3.06	.92	1.13	1.14
SD	1.48	2.09	2.09	1.15	1.87	1.13
			Picture 4			
Men						
M	.87	.74	1.50	.60	.58	1.29
SD	1.33	1.41	1.84	.97	.88	1.68
Women						
M	1.16	1.67	1.40	.65	1.10	.76
SD	1.30	2.12	1.41	.81	1.40	1.00
			Picture 5			
Men						
M	2.47	1.90	2.61	1.60	1.68	1.61
SD	1.85	1.62	2.33	1.50	1.70	1.64
Women						
M	2.29	3.10	2.50	1.20	1.80	1.26
SD	1.68	2.54	1.83	1.31	1.97	1.40
			Picture 6			
Men						
M	4.67	4.45	4.68	3.00	3.26	3.43
SD	2.28	2.23	2.26	1.91	1.91	2.06
Women						
M	4.45	5.53	3.90	2.57	3.43	2.14
SD	2.05	2.56	2.06	1.70	2.10	1.73

For projection, there was no main effect for condition or gender, and no interaction effect demonstrated. There were no changes in findings when the scores of the first two stories were used as covariates.

Picture four. Logarithms of scores were used for both dependent variables for the analyses of picture four (woman with blackboard and equations). For aggression frequency, there was a significant main effect for gender, multivariate $F(2, 211) = 3.07$, $p = .048$, univariate $F(1, 212) = 4.75$, $p = .03$ with women expressing more aggression than men. This effect was slightly stronger when the first two stories were used as a covariate, multivariate $F(2, 209) = 3.62$, $p = .028$, univariate $F(1, 210) = 6.33$, $p = .013$. The MANOVA and MANCOVA showed no main effect for condition and no interaction effect for picture four.

For projection, the MANOVA and MANCOVA showed no significant main effects for condition or gender, and no interaction effects for picture four.

Picture five. For aggression frequency for picture five (older and younger man), the MANOVA and MANCOVA showed no main effect for condition or gender and no interaction effects.

For projection, The MANOVA and MANCOVA showed no main effects or interaction effects for picture five.

Picture six. For aggression frequency for picture six (woman with hands around throat of other woman), the

MANOVA and MANCOVA showed no significant main effects or interaction effects.

For projection, the MANOVA showed a significant main effect for gender, multivariate $F(2, 211) = 3.29$, $p = .039$, univariate $F(1, 212) = 3.88$, $p = .05$, with men ($M = 3.22$, $SD = 1.95$) using projection more than women ($M = 2.60$, $SD = 1.86$). This effect was weakened when the first two stories were used as covariates: multivariate $F(2, 209) = 3.20$, $p = .043$, univariate $F(1, 210) = 3.18$, $p = .076$. There was no main effect for condition and no interaction effect for projection for picture six.

In sum, for the dependent variable aggression, picture three showed a significant main effect for condition, (simulated-anger > control > anger) and picture four showed a gender difference (women > men). For projection, only picture six showed any significant findings (men > women), and this finding was weakened by adding covariates to the analysis.

Rated versus obtained cue strength. The ratings of cue for aggression which were previously used can be compared to the obtained aggression scores from the control group (Table 16). Based on the obtained aggression scores, the cue value for aggression of the pictures used in the study can be best described as falling into a low cue (8GF), high cue (18GF) and moderate cue value (1, 5, 6GF, and 7BM). The pictures judged as

Table 16

Comparison of Ratings for Cue for Aggression With Obtained Aggression Scores for the Control Group*

Aggression Ratings	TAT Number	Aggression Frequency	Order**
.2	8GF	1.1	4
.4	1	2.7	3
2.0	7BM	2.4	5
3.3	5	2.2	1
3.8	6GF	2.2	2
5.9	18GF	4.5	6

Note. *N = 79.
**Order of presentation of pictures to subjects.

lowest and highest aggressive cue in the ratings were also the lowest and highest in the obtained scores. TAT card 1 (boy with a violin) was given a low rating for aggression but the obtained scores place it in the middle range of aggressive cue strength.

The cue value of each picture for projection was not rated as was the cue value for aggression. However, the obtained scores, for projection give us information about the projection cue value for each picture. Card 8GF showed the lowest mean for projection scores (M = .63) and card 1 ranked second (M = 1.01). The other four pictures had means of 1.35 (7BM), 1.72 (5), 1.46 (6GF), and 2.73 (18GF). Ehrenreich (1990) in a study of responses

elicited by five TAT cards (1, 2, 3BM, 12M, and 13MF) found that within his range of TAT pictures, card 1 tended to elicit intermediate levels of defenses. This is similar to the results obtained in the present study.

Story Length

The frequency of content analytic categories tends to be positively correlated with the length of the material being analyzed (Smith, Feld, & Franz, 1992). Therefore the number of words per story was determined for all subjects.

Length of pre-criticism stories. Table 17 shows means and standard deviations of the two pre-criticism stories combined and of the four post-criticism stories combined. A 2 (gender) by 3 (condition) ANOVA with word count of the pre-criticism stories as the dependent variable showed no significant main effects for condition, $F(2, 212) = 1.38$, $p = .26$, or for gender, $F(1, 212) = 1.87$, $p = .17$. However, there was an interaction effect, $F(2, 212) = 4.12$, $p < .02$. Women wrote longer stories than men in the control and anger conditions, and shorter stories in the simulated-anger condition. This interaction effect was not expected because all subjects wrote pre-criticism stories under the same neutral classroom conditions. The occurrence of this unexpected interaction indicates the need to control for story length

Table 17

Average Story Length for Pre- and Post-Criticism Stories

	Condition							
	Control		Anger		Simulated Anger		Total	
Gender	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Men								
M	71.1	80.7	72.0	83.8	77.6	83.1	73.5	82.5
SD	22.9	22.5	25.0	24.4	18.3	19.9	22.3	22.2
Women								
M	78.8	86.5	85.1	99.3	70.3	82.5	77.0	87.9
SD	22.2	21.5	20.1	23.4	20.1	18.7	21.6	21.8
Total								
M	75.9	84.3	78.9	91.4	73.0	82.7		
SD	22.6	21.9	23.5	24.9	19.7	19.0		

Note. Story length refers to number of words per story.

of pre-criticism stories in analyses of the post-criticism stories.

In view of this problem, the analysis of aggression and projection scores of post-criticism stories, that provided information to test the first three hypotheses, was performed with the addition of word count for stories 1 and 2 as the covariate. Thus, a 2 (gender) by 3 (condition) MANCOVA was conducted with aggression and projection scores of the post-criticism stories as the dependent variables and the word count for pre-criticism stories as the covariate. This MANCOVA produced only one adjustment to the findings of the original MANOVA (see page 56) namely, the trend toward a main effect for

projection for gender (M > W) became significant, multivariate $F(2, 210) = 4.48$, $p = .012$, univariate $F(1, 211) = 6.84$, $p = .010$.

As in the original MANOVA, the additional MANCOVA showed no significant effects for condition, multivariate $F(4, 420) = 1.78$, $p = .132$, univariate $F(2, 211) = 1.56$, $p = .212$ for aggression, and univariate $F(2, 211) = 1.53$, $p = .217$ for projection. Also, the MANCOVA showed no significant interaction effects, multivariate $F(4, 420) = 1.02$, $p = .398$, univariate $F(2, 211) = 1.70$, $p = .185$ for aggression, and univariate $F(2, 211) = .67$, $p = .511$ for projection. Thus, the addition of pre-criticism story length as a covariate did not change the original findings regarding hypotheses 1, 2 and 3.

Length of post-criticism stories. A 2 (gender) by 3 (condition) ANOVA was carried out with the word count of the post-criticism stories (3-6) as the dependent variable. A significant main effect was found for condition $F(2, 212) = 3.74$, $p < .03$, with the anger group producing the longest stories, $M = 91.4$, $SD = 24.9$. A main effect was also found for gender $F(1, 212) = 4.63$, $p < .04$ with women showing the larger word count. There was a nonsignificant interaction effect, $F(2, 212) = 2.31$, $p = .10$.

For the post-criticism stories, for all conditions combined, the average story length was 85.7 words, as

compared to the pre-criticism average of 75.6 words. Such an increase in average word count might be expected in the later stories due to (a) increased familiarity with the task (b) the fact that pictures 3 to 6 were different, and/or (c) the fact that the experimental manipulations may have affected story length in the anger and simulated-anger conditions.

Because one of the ways subjects may have responded to criticism about their stories may have been to write longer stories, it was decided to investigate whether the increase in story length occurred based on gender or condition. Therefore, a 2 (gender) by 3 (condition) ANOVA was carried out on the average change in story length from pre-criticism to post-criticism stories (thus taking into account the length of the pre-criticism stories). For this analysis, the effect for condition approached significance, $F(2, 212) = 2.87$, $p = .059$ with the anger group showing the largest increase ($M = 13.0$) in words per story. The simulated-anger and control conditions had mean increases of 9.8 words and 8.4 words respectively. There was no main effect for gender, $F(1, 212) = 1.97$, $p = .16$, and no interaction effect, $F(2, 212) = 2.32$, $p(2, 212) = .10$. This analysis provides marginal evidence for an effect of the manipulation in the anger group, as was also suggested by a previous analysis of affect scores from the post-experimental questionnaire.

Correlations were computed in order to determine the relationship between story length and scores for aggression frequency and projection for each cell (Table 18).

For aggression frequency, there were significant positive correlations with story length in all but one cell. For projection, there were significant positive correlations in four of the six cells. The positive correlations shown in Table 18 between story length and TAT scores may appear to suggest that subsequent analyses of post-criticism TAT scores should be corrected for differences in post-criticism story length. This possibility has been rejected by the author, however, because story length, for stories 3 through 6, especially

Table 18

Correlations Between Length of Story (Word Count) and Projection and Aggression Scores by Cell, Stories 3-6

	Control	Anger	Simulated Anger
Men			
Aggression	.43*	.30	.40*
Projection	.42*	.18	.27
Women			
Aggression	.46**	.67**	.59**
Projection	.42**	.59**	.46**

Note. *p < .05. **p < .01.

in the two experimental conditions is viewed as an effect of the experimental manipulations; that is, story length is a dependent variable like projection and aggression scores. In fact, story length may be the mechanism by which the aggression and projection scores increased in the original study (Cramer 1991a). To remove post-criticism story length from projection and aggression scores would amount to removing some part of the effect of the manipulations on these scores.

Chapter IV

Discussion

The present research was an investigation of the validity of the Defense Mechanism Manual (DMM, Cramer, 1991b) for scoring defensive projection in the Thematic Apperception Test. The experiment was a modified replication of the Cramer (1991a) study of anger and defense mechanisms in college students. Based on the psychodynamic view of projection as the externalizing of unacceptable internal states, the DMM uses categories that rely on some level of projected aggression (e.g., attribution of aggression or hostile feeling to a character). The primary research question was whether the DMM was identifying defensive projection or merely direct aggression. Determining a difference between projection and aggression was attempted by comparing two groups: a group that was angered and expected to be defensive and a group that was pretending to be angry and not expected to be defensive.

I will summarize the following: description of the sample, interrater reliability, effectiveness of experimental manipulations, and tests of this study's hypotheses.

Sample Characteristics

This study used a total of 89 men and 129 women ranging in age from 17 to 23 years. Most students were in

their freshmen year (84%) and were required to participate in research as part of an Introductory Psychology course at one of the senior colleges in the CUNY system. Of the sample, 27% listed a language other than English as their primary language.

By comparison, Cramer's (1991a) subjects were mostly juniors and seniors at a private college in Massachusetts, who volunteered and were paid for their participation. Cramer doesn't discuss the culture or language dominance of her subjects. In summary there are differences in subjects between the Cramer study and the present study based on, year in school, motivation to participate, and, possibly, language and cultural diversity, also data were collected from Cramer's subjects individually and from subjects in the present study in small groups.

Interscorer Reliability

The interscorer reliability in the present study for aggression frequency (r = .91) was comparable to Cramer's (r = .90). The interscorer reliability in the present study for projection was r = .89. Cramer (1991b) reports interscorer reliabilities, for three different pairs of raters on multiple subject samples, that range from .71 to .90. The interscorer reliability, for the present study, falls within the upper range of what Cramer's scorers had achieved. However, this took a great deal of effort to achieve. While this degree of agreement increased a sense

of confidence regarding scoring stories reliably, in retrospect, it limits what can be said about the usefulness of the scoring system. In other words, based on the training methods used in the present study, the defense mechanism projection can be scored reliably by two scorers using the DMM who are previously trained in psychology and with a great deal of practice (approximately 450 practice stories were scored). The present study suggests that the time required to master the projection scoring system is a limitation of the Defense Mechanism Manual.

Post-Experimental Questionnaire

The effect of the experimental manipulations was determined in two ways: first, through the change in affect reported on the post-experimental questionnaire and second, through the amount of aggression expressed in stories. The Post-experimental questionnaire will be discussed here. The aggression scores will be discussed subsequently with the tests of hypotheses.

On the post-experimental questionnaire, experimental group I (anger) and group II (simulated-anger) reported more anger at the end of the experiment than the control group. This finding was significant for "end" scores, but was reduced to a trend when the analysis was conducted using logarithms. Difference scores (affect at end minus affect at beginning) showed the same pattern as end

scores, that is experimental groups reported more anger than the control group, and the difference approached significance. These findings suggest that the experimental manipulations had an effect in the intended direction. However, the effect was marginal, that is, the increase in anger was not as strong as expected, and not as clear as that obtained in Cramer's study.

Test of Hypotheses

Aggression scores. The first hypothesis concerns the expected increase in aggression scores for the experimental conditions. Hypothesis 1, that for the post-criticism stories aggression frequency should be higher for the anger and simulated-anger groups than for the control group, was not supported. Analyses revealed no significant differences attributable to condition for aggression frequency scores obtained from the post-criticism stories. This was the case whether or not pre-criticism scores were used as a covariate.

The fact that aggression scores were not significantly higher in the experimental conditions than in the control condition appears to support the finding, in the post-experimental questionnaire, that the experimental manipulations were not sufficiently effective in producing increased levels of anger in the subjects.

Projection scores. Hypothesis 2, that for the post-criticism stories the mean of the defensive projection

scores for the anger group should be significantly higher than the mean for the control group, was not supported.

The failure to support hypotheses 1 and 2 also indicates that this study failed to replicate Cramer's (1991a) findings of significant differences between the anger group and the control group for aggression and projection scores.

Hypothesis 3, that for the post-criticism stories, the mean of the defensive projection scores for the anger group should be significantly higher than the mean for the simulated-anger group was not supported. This was a key hypothesis for this study. Hypotheses 1 and 2 are indicators of whether or not this study replicated Cramer's (1991a) findings. Hypothesis 3 was an attempt to broaden the scope of the Cramer study to determine whether the DMM could distinguish between anger and defensiveness. However, due to the failure to clearly replicate Cramer's findings as demonstrated in the first two hypotheses, there are limited possibilities for interpreting the findings regarding hypothesis 3. One possibility is to consider the analysis for each projection category separately. Although, overall analyses did not support hypotheses 1, 2, or 3, some support may be provided by the analysis of individual categories or pictures.

Analysis of projection by scoring category.

Hypothesis 2a and 3a dealt with the individual categories

of the DMM. For the post-criticism stories, the mean of the defensive projection scores for each scoring category for the anger group should have been significantly higher than the mean for the corresponding category in the control group (hypothesis 2a) and the simulated-anger group (hypothesis 3a). Analyses showed no significant main effects for projection by condition for any of the seven categories. Although, for all categories, the projection scores of the anger group were at least the same, or slightly higher than the control group.

Although, the primary point of the previous analysis of categories was to determine which of the categories could distinguish between conditions, there were main effects for gender for three categories, and an interaction effect. Gender differences were shown for the following categories: category 3 (magical or circumstantial thinking) category 5 (death, injury, or assault) and category 6 (themes of pursuit entrapment and escape, see Tables 12, 13, & 14). In all cases men used the particular category more frequently than did women.

Category 5 showed an interaction effect. Men in the simulated-anger condition showed the greatest use of this category, while women in the simulated-anger condition showed the least use of this category.

Analysis of Pictures

Although, there were no hypotheses regarding

pictures, an analysis of each picture was done to examine whether any of the pictures produced the pattern of aggression and projection scores which was proposed in the first three hypotheses.

Cramer (1991a) reports changing picture 8GF (by adding equations to the blackboard behind the seated girl) in order to match the achievement pull of this picture to that of card 1 (boy with violin). In the initial ratings of aggression by the 10 judges, cards 1 and 8GF received very similar ratings (.4 and .2 respectively). The ratings by the judges support the similarity of these two pictures. However, the obtained scores place the cue value for aggression of card 1 closer to cards 5, 6GF, and 7BM (see Table 16). This may suggest problems in a priori judgements of pictures. Moreover, if these pictures appear non-aggressive to 10 judges but draw a moderate amount of aggression, this may provide some support for the claim that subjects are not completely aware of what they are disclosing while writing TAT stories (Cramer, 1996; Murray, 1938). In other words, a picture that appeared not aggressive to judges, but drew fairly aggressive stories is cuing something that the judges are not aware of.

Picture 3 (boy with violin) is considered to be one that draws stories of parent-child relationships. If this is what comes to mind when judges view the pictures, they

may make an automatic assumption that parent-child relationships are nonaggressive. But, in fact, many of the parent-child stories contain sufficient conflict to be scored for aggression. Differences between judges' ratings of cue value and actual story content is an area for further investigation.

Picture 3 was the first of the four post-criticism pictures. It was also the only picture that showed a main effect for condition for aggression frequency (simulated-anger group had the highest scores). This picture may be more sensitive to outward hints concerning possible responses, such as that of the instruction to pretend to be angry.

Picture 4 (girl on desk, staring into space) produces very little aggressive imagery. However the main effect for gender showed that women wrote more aggressive stories than men. This picture was rated the lowest in aggressive cue strength by the 10 judges, and obtained the lowest aggression scores by the subjects as well. Many of the stories to this picture were romantic or achievement oriented. This picture has limited value if aggressive imagery is sought.

Pictures 5 (faces of two men) showed no main or interaction effects for either dependent variable. The obtained scores and the ratings by judges both place this picture in a middle range for aggressive cue strength.

Picture 6 (woman with hands around throat of other woman) showed a gender difference for projection (men > women). This picture also produced the highest levels of aggression of all the pictures. It appears that this picture triggers a sense of danger that was expressed in clearly aggressive stories, or stories about illness and death, which would also be scored as aggression by the Whitman et al. (1961) criteria.

Test of gender differences. Projection has been considered to be a defense likely to be used by men more often than women (Cramer, 1991a, 1991b, 1996). The theoretical basis in psychoanalytic theory lies in the process of externalization (considered a more masculine orientation) necessary for the defense of projection, as opposed to a defense that is considered more internal (and feminine) such as repression. There have been some research findings to support the greater use of projection by males in the TAT (Cramer, 1991a) as well as scores from an objective measure (DMI, Cramer, 1979; Cramer & Carter, 1978; Gleser & Ihilevich, 1969).

Hypothesis 4, that for the post-criticism stories, for the anger condition, the mean of the defensive projection scores for men will be higher than the mean for women, was not supported. A nearly significant main effect for gender was shown in the 2 (gender) by 3 (condition) MANOVA, with men using projection more than

women. This marginal difference did not hold when the projection scores of the first two stories were used as a covariate. The gender difference that Cramer (1991a) found was demonstrated in her experimental (anger) group only. In her control group, men and women expressed the same amount of projection. In the present study, the nearly significant main effect for gender occurred because men in the control and simulated-anger condition used more projection than did women. In the anger condition, the projection scores for men were slightly lower than those for women. Although, the results do not support hypothesis 4, or replicate Cramer's findings, they are consistent with psychoanalytic theory.

When the projection score was analyzed by category, it was shown that men used projection more frequently than did women for the following three categories: category 3, magical or circumstantial thinking; category 5, apprehensiveness of death injury or assault; and category 6, themes of pursuit, entrapment or escape. An analysis of the last four pictures showed gender differences for only one picture. For picture 6 (woman with hands around throat of other women), men used projection more often than women.

Cramer's (1991a) study found no overall gender differences in aggression scores. The present study also found no overall gender differences in aggression scores.

However, the analysis of individual pictures showed that for picture four (woman with blackboard and equations), women expressed aggression more frequently than did men. It is interesting to note that picture four received the lowest rating for aggression by the judges, and received the lowest scores for aggression of all six pictures. However, there was more aggression present in the stories of women then of men for this picture. Picture four was the only picture with a single female character. Although, picture six (woman with hands around throat of another woman) had two women as characters, the depicted scene is fairly obviously aggressive, and produced high aggressive scores for both genders. It is possible that the single figure in picture four interacted in some way with the gender of the subjects, producing more aggressive stories for women.

Correlations between projection and aggression scores. Hypotheses 5 through 7 deal with the relationship between projection and aggression scores. Hypothesis 5, stated that for post-criticism stories, for the anger group, there should be a positive correlation between projection scores and aggression scores. This hypothesis was supported. Correlations between aggression and projection scores were determined for each cell. For the anger condition, the correlations between aggression frequency and projection were .59 (men) and .73 (women).

Both were significant at the $p = .001$ level. In fact the correlations between aggression and projection scores were significant for all conditions (see Table 12).

The Cramer (1991a) study revealed a significant correlation between projection and aggression scores in the experimental group only (aggression frequency and projection $r = .85$, $p < .001$). For her control group, the correlation between projection and aggression scores was not significant $r = .29$. Cramer proposes that "those subjects who were most angered by the experimental criticism were also the ones who made the greatest use of all three defense mechanisms" (1991, p. 52).

Hypothesis 6, that for the post-criticism stories, the positive correlation between projection scores and aggression scores should be greater for the anger group than for the simulated-anger group, was not supported. For women, there was no difference between the correlations of the projection and aggression scores of the anger and simulated-anger condition. For men, the findings were opposite to the predictions, that is, the correlations between projection and aggression scores were significantly lower in the anger condition.

Hypothesis 7, was that the mean of the aggression scores for high-defensive subjects would be higher for pictures rated as having low cue for aggression than for pictures rated as having high cue for aggression. This

hypothesis was a test of the Weiskopf-Joelson et al. (1957) theory that those who were highly defensive show inappropriate responses to cue strength. For the present study "highly defensive" was operationally defined as those subjects who scored above the median on defensive projection. Hypothesis 7 would have been supported by an interaction effect in the 2 (cue) X 2 (projection) MANOVA. There was no interaction effect and hypothesis 7 was not supported. Aggression scores were higher for those pictures judged to have a higher aggressive cue value, regardless of whether the subject had scored high or low on projection.

Story Length

There are three noteworthy findings regarding length of stories. First, there was a significant interaction effect for story length in the pre-criticism stories (1-2). This interaction effect, found under neutral conditions, suggests that assignment of subjects may not have equated the groups as desired. This interaction effect also suggests the necessity of removing the effect of pre-criticism story length by adding it as a covariate in the analysis of post-criticism scores. When this was done, a clearly significant main effect for gender was revealed, with men using projection more often than women. This finding conforms to the psychodynamic view that men use this defense more often than women.

Second, the analysis of length showed that the anger group wrote significantly longer stories than the other two groups in the post-criticism phase of the experiment, and that the increase in story length from pre- to post-criticism was nearly significant for the anger group. These findings suggest that story length should not be considered an artifact to be controlled, but as a dependent variable. Viewed as a dependent variable, the effects shown for story length support the position that the experimental manipulations had an effect, but it was not as strong as intended.

Third, there were strong positive correlations between post-criticism story length and post-criticism aggression and projection scores. These correlations may appear to suggest the need to control for story length in analyses of post-criticism TAT scores. However, it is possible that the increase in story length was a direct result of the intervention in the present research, and might have been the mechanism by which aggression and projection scores increased in Cramer's (1991a) original study. However, since Cramer does not report any information about story length, it is not possible to determine whether or not this is the case. In the present study it was deemed inappropriate to remove the effects of story length on aggression and projection scores, because story length should not be viewed as an artifact, but

rather as a response to the experimental manipulations, that is, another dependent variable.

The intervention that the anger group experienced was a criticism of their work. Therefore, it is also possible that the subjects heard the intervention not so much as criticism, but as exhortations to write better stories. If this were the case, then longer stories may be an attempt to conform to the requirements of the task as well as an expression of anger.

Effectiveness of Experimental Manipulations

Generating anger in the "anger" condition was an essential prerequisite of the present research. A cursory look at the lack of significant differences between the anger and control condition, suggests that the subjects in the present study were not made angry by the experimental manipulations. However a comparison between Cramer's (1991a) post-experimental questionnaire and the one used in this study suggests another possible interpretation.

The post-experimental questionnaire for the Cramer (1991a) study had four response possibilities regarding affect: yes, a little, no, and don't know. Each subject chose a response for the four affects: angry, anxious, happy, and sad. Cramer found that there were no differences between experimental and control groups as they began the experiment, but found that at the end of the experiment significantly more subjects in the

experimental group reported anger than in the control group, $\chi^2(1) = 9.94$, $p < .01$. For the purpose of analysis, Cramer combined "yes" and "a little" as one score, and "no" and "don't know" as the second score. The post-experimental questionnaire of the present study was slightly different. For each affect, the subjects reported their level of anger by circling a number on a scale which ranged from 0 to 7. To compare the post-experimental scores of the present study with those of Cramer's, the present data were split into two categories. Those who reported no anger (choosing zero) and those who reported some level of anger (choosing 1 through 7). When the data are reorganized in this manner the following comparisons can be made.

Of the 40 subjects in Cramer's (1991a) experimental (anger) group, 14 (35%) indicated some level of anger at the end of the experiment. In the present study 30 (49%) of the 61 subjects in the anger group indicated some level of anger at the end of the experiment. This percentage would appear to indicate that there was a level of self-reported anger in the "anger" condition of the present research that may be comparable to or greater than Cramer's "anger" condition.

However, the comparison becomes more complicated when the pre-experimental affect scores are considered. For Cramer's (1991a) anger group only 5% of subjects (2 of

40), as opposed to 30% (18 of 61) in the present research, indicated that they felt any anger at the beginning of the study. Moreover, in Cramer's control group, only one subject (3%) indicated any level of anger at the beginning or end of the experiment. For the 79 subjects in the control condition of the present research, 16% (13 subjects) indicated some level of anger at the beginning of the experiment and 29% (23 subjects) indicated anger at the end of the experiment.

For the present study, the higher levels of anger reported at the beginning of the study, as well as the higher levels of anger in the control group, may have contributed to an inability to detect differences between conditions.

The finding that the anger group wrote significantly longer stories than the control or simulated-anger groups, also provides support for the view that the manipulations did have a limited effect. However, this effect was not strong enough to create a significant difference in aggression scores between the control condition and the anger condition.

Relation Between Projection and Aggression Scores

Significant positive correlations between aggression and projection scores occurred for all conditions (Table 12). This may have occurred because of (a) overlap in aggression and projection categories; (b) higher anger

produced higher aggression and more projection; or (c) a confound of concepts of aggression and projection.

First, there is overlap in some of the categories between the aggression and projection systems. For example, the first category for projection is the "attribution of aggressive or hostile feelings or intentions..." This would also be scored as aggression. This is also true of the projection category 5, the occurrence of "...death, injury or assault". This overlap would be expected to produce a built-in positive correlation.

Second, in Cramer's (1991a) study, there were significant correlations in her "anger" group between projection and aggression scores. However, the correlations in Cramer's control group were not significant. Cramer proposes that it is the greater anger of the experimental group that increases aggressive content and the use of defenses. If Cramer's position is that subjects who are angry will produce greater aggression in stories and use projection more frequently, and the subjects in all conditions of the present study reported high levels of anger, then the findings of the present study (high aggression and projection for all conditions) are consistent with that position.

Third, the high correlations between aggression and projection scores may also be caused by a confounding of

the two concepts in the scoring systems. In other words, if aggression is the defining feature of projection, then it is unclear how to distinguish defensive projection from simple aggression. It was the point of the present study to attempt to distinguish between the two concepts. The high correlations between aggression and projection scores obtained in the present study suggest the possibility that the concepts of anger and projection are being confounded and raise a question about the discriminant validity of the Defense Mechanism Manual scoring system. Results of the present study were inconclusive with respect to this point, which requires further investigation.

Hypothesis 6 stated that the correlations between projection and aggression scores would be significantly greater for the anger than for the simulated-anger group. For men, the results were contrary to the prediction, that is, the correlations of the anger group were significantly lower than the simulated-anger group. For women, the correlations were the same for both conditions.

Clark (1952), in a study of the expression of sexual motivation on the TAT found that for an aroused group (those shown pictures of nude women) there was less sexual imagery present in stories than for a non-aroused group. Because this finding was assumed to be caused by the guilt or anxiety raised by the pictures, a second study was done using alcoholic beverages as a disinhibitor prior to

viewing the pictures. In the second experiment the expression of sexual imagery was greater in the aroused group than in the non-aroused group. For the present study, the instructions for the simulated-anger group "to pretend to be angry" were intended to make it socially acceptable to express anger in the stories, that is to disinhibit the expression of aggression in stories. The intent was to increase aggression scores for the simulated-anger condition, without increasing the projection scores. The findings of the present study indicate that this did not occur.

The Clark (1952) study also found a curvilinear relationship between the expression of manifest and symbolic sexual imagery. That is, those who expressed high or low levels of manifest sexual imagery, expressed high levels of symbolic sexual imagery in their stories. If the aggression scores of the present study, could be broken down into manifest and symbolic expressions of aggression, it would be interesting to see if the relationship between manifest and symbolic aggressive imagery displayed a similar pattern to that of manifest and symbolic sexual imagery. It would also be valuable to see how the use of manifest or symbolic expressions of aggression relate to the expression of projection in TAT stories. This could be a direction for future research.

Issues Regarding Replication of the Cramer Study

There are a number of possible reasons why this study failed to replicate the Cramer (1991a) findings. First, those in the anger group may not have been made sufficiently angry. The higher levels of anger of the control group, of the present study, may reflect that subjects may have been angry simply at having to fulfill a requirement to participate.

It is also possible that the greater anonymity that was afforded by group testing in the present research, as opposed to Cramer's (1991a) study in which subjects were tested individually, allowed subjects to feel somewhat more free to report anger on the post-experimental questionnaire. The greater anonymity of group testing may also have diffused the impact of the criticism, or may have allowed the subjects to express aggression directly, which may have lessened the need to use projection. In either case, the higher anger scores for subjects at the beginning of the present study, as well as the higher anger scores in the control condition, may have made it more difficult to obtain significant differences between conditions.

The level of anger that was reported by the subjects in the control group on the post-experimental questionnaire was also reflected in the scores obtained from the TAT stories. A visual comparison of mean scores

for aggression (Table 19), reveals a pattern of aggression scores in the TAT similar to that reported by the subjects on the post-experimental questionnaire. Specifically, that subjects in the present study were angry prior to the criticism, and that the control group had a relatively high level of anger all the way through the experiment. The differences in aggression and projection scores between the Cramer (1991a) and the present study can be seen in Table 19, which shows that for the Cramer study, there is a difference between pre- and post-criticism scores for aggression and projection and, for the present study, the pre- and post-criticism scores are very similar.

Table 19

Comparison of Aggression and Projection Means for the Cramer and the Present Study

	Cramer		Present Study	
	Pre-Criticism	Post-Criticism	Pre-Criticism	Post-Criticism
Control				
Aggression	.38	.56	2.16	2.66
Projection	1.04	.91	1.59	1.43
Anger				
Aggression	.28	1.53	2.42	2.79
Projection	.91	2.01	1.82	1.79
Simulated Anger				
Aggression	*	*	2.13	2.88
Projection			1.35	1.58

* Cramer's study did not have this condition.

It could be argued that even if the subjects were angry at the beginning of the experiment, the intervention should have made them significantly angrier. This may not have occurred for the following reasons. First, it is possible that the presentation of the insults was not done in a sufficiently powerful manner to increase anger. However, this did not appear to be the case according to the verbal responses of subjects during debriefing of the data collection groups conducted under anger conditions. While the response of a few of these groups was relatively subdued, the majority consisted of subjects who gave some spoken indication during debriefing that they had in fact become angry. The possibility remains, however, that it may have been that only a few of the subjects in the group became angrier and that they were the more verbal during the debriefing, giving the impression that the group was angrier than it actually was.

Second, it may be that the attitude of subjects who are required to participate in a subject pool may be different than the attitude of subjects who volunteer and are paid for their efforts, as was done in Cramer's (1991a) study.

Third, it may also be possible that the expectation of criticism as part of a college experience or education in general may have taught the students strategies to handle negative criticism without significant increases in

level of anger. The author was told by two students in different anger groups that she was not quite as nasty as some of their regular instructors. Also given the large number of foreign speaking students, there may be cultural differences in the expectation of criticism as part of learning, or different levels of acceptance of the expression of anger. Moreover, Cramer's (1991a) subjects who were drawn from a private college, may have come from a higher socioeconomic class, and may have been better students less accustomed to being told their performance was inadequate. This type of student may be more susceptible to criticism. In summary, the results of this study are inconclusive regarding the validity of the Defense Mechanism Manual.

An Alternative Interpretation of the Data

Based on the finding that the anger and simulated anger groups reported more anger than the control group, on the post-experimental questionnaire, an alternative interpretation of the findings of the present study is possible. Based on the increase in self-reported anger of the anger group, differences would be expected between the control and experimental conditions, at the very least for the TAT aggression scores. In the previous discussion, the finding of no differences among conditions for either aggression or projection has been attributed to (a) possible differences in subjects between the Cramer and

present study, and (b) lack of a strong enough effect of the anger intervention. However, it is also possible that the scoring systems may be inadequate.

1. There are large differences in raw scores for aggression and projection between the present study and Cramer's (1991a) study (see Table 19). The aggression and projection scoring systems may not have been presented with sufficient clarity to exactly replicate Cramer's scoring.

2. The systems themselves may be at fault, in that the construct validity of the scoring systems may be inadequate. The failure of the systems to distinguish between groups that show some differences in reported anger, could mean that the aggression and projection systems are not measuring what they purport to measure.

3. Alternatively, a comparison between self-reported anger and TAT aggression may not be appropriate. McClelland, Koestner and Weinberger (1989) suggest that significant correlations are not usually found between motives which are self-attributed (e.g., from self-report questionnaires) and those which are implicit (e.g., from TAT stories). They suggest that it is unwarranted to reject either type of measure, but rather more appropriate to consider each approach to be measuring different classes of behavior, which respond to different types of stimuli.

Problems with an alternative interpretation. First, although the initial MANOVA showed significantly more self-reported anger in the experimental groups, than in the control group, this finding was reduced to a trend when the scores were statistically adjusted because of problems with homogeneity of variance. This means that there was evidence that the anger manipulation did have an effect, but the effect was simply not strong enough.

The second problem is that the anger and simulated-anger conditions both reported more anger than the control condition. The simulated-anger group had no reason to be angry. Their high anger scores may be inaccurate, that is, they may be due to their complying with a perceived expectation to "continue" their pretense of anger past the story writing and into the questionnaire part of the experiment. However, if this was the case, the hypothesis regarding projection still could have been supported. That is, the projection scores should have been lower in the simulated-anger group than the anger group, if they were not truly angry, but only attempting to comply with the task.

If the high self-reported anger scores, for the simulated-anger group are accurate, they may be due to the process by which subjects generated anger. If students pretended to be angry by drawing on a past experience, they may have tapped into an affective state in which they

did in fact feel the angry affect connected to the memory. If this occurred, the finding that they matched the anger group, not only in reported anger, but in aggression and projection scores as well, would be explained. That is, higher states of anger, whether from current experience or from memory, could produce high aggression and high projection scores.

Thirdly, the possible scores for self-reported anger ranged from zero to seven. However, the range of self-reported anger obtained was only .5 (women, average for control condition) to 2.0 (men, average for anger condition). The raw scores are small considering the possible range, indicating that the intervention may not have been as strong as planned.

Although they are equivocal, the findings of the present research suggest possible problems with the discriminant validity of the DMM projection scoring system and raise questions about Cramer's assumptions regarding the relationship of aggression scores to projection scores.

Directions for Future Research

The reliance on indices of hostility in TAT stories as determinants of the presence of projection is somewhat limiting. If defensive projection is used to protect the self from some unacceptable internal state, then other internal states should be included, such as unacceptable

sexual impulses.

The acceptability or unacceptability of internal impulses is for the most part culturally defined. There may also be different feelings that are considered unacceptable within different cultures. For example, in cultures in which there are very strong gender roles, feelings of the desire for power for women, or maternal feelings for men, may be so unacceptable that they would need to be projected. A scoring system that relies heavily on hostility could miss signs of defensive projection of a wide range of unacceptable feelings. Studying cultural or group specific forms of projection may be another way to understand the defensive process of projection.

Appendix A

Modified TAT Pictures

The following are the two pictures, 8GF and 18GF which were modified for the Cramer (1991) study.

8GF

18GF

Appendix B

Aggression Scoring Systems

Hall and Van de Castle (1966)

1. Covert feeling of hostility or anger without any overt expression of aggression.

2. Aggression displayed through verbal or expressive activity. Included are such activities as one character yelling or swearing at another or when a character criticizes or scowls at another.

3. This subclass covers all situations where there is an attempt by one character to reject, exploit, control, or verbally coerce another character. Such activity may be expressed through dismissals, demands, refusals, disobedience, or any other type of negativistic or deceitful behavior.

4. An aggressive act in which a serious accusation or verbal threat of harm is made against a character.

5. An aggressive act which involves the theft or destruction of possessions belonging to a character.

6. An aggressive act which involves a character being chased, captured, confined, or physically coerced into performing some act.

7. An aggressive act which involves an attempt to physically harm a character. The attempt may be carried out through personal assault or through use of a weapon. Threatening a character with a weapon is also included in this subclass.

8. An aggressive act which results in the death of a character.

Hostility Scale of Whitman, Pierce, Maas, and Baldridge (1961)

6. Human: Death or death threat by stabbing, shooting, pushing, striking, hit by car, drowning, illness, warfare, animal attack, mutilation, drive to insanity, violence.

5. Non Human: Equivalent destruction of animate or inanimate objects other than human.

4. Human: Injury or injury threat by fight, accident, illness, abandonment, rendered helpless, robbed.

3. Non Human: Equivalent injury of animate or inanimate objects other than human

2. Human: Discomfort or discomfort threat by minor difficulty, hurt, annoyance, failure, inappropriate behavior.

1. Non Human: Equivalent discomfort of animate or inanimate objects other than human.

Appendix C

Subject Sign-Up Forms

Subject sign-up forms used at Brooklyn College, City Universty of New York, and Queens College, City University of New York.

Title Creative Imagination in College Students

Location _____ Duration One Hour

Experimenter Jeanne Clancy Phone Number 718 531-3517

Restrictions 7 Men and 7 Women per group

Description IN SMALL GROUPS, SUBJECTS WILL WRITE BRIEF CREATIVE STORIES TO 6 PICTURES AND FILL OUT A BRIEF QUESTIONNAIRE.

ONE NAME PER LINE

DAY / DATE / TIME	STUDENT'S NAME (PRINT IN INK)	PHONE (OPTIONAL)
	MEN	
	WOMEN	

Appendix D

Consent Form

This research project is part of a doctoral dissertation which involves the study of creative imagination. You will be asked to make up stories about some pictures that I will show you and then answer a brief questionnaire. The duration of your participation will be about 45 minutes.

This study is not expected to expose subjects to adverse conditions outside the range of the experiences of the average college student. Benefits from participation should be an increase in your understanding of psychological research and an increase in your knowledge about this topic area.

I understand that my participation is entirely voluntary and that I may withdraw from the study at any time.

I understand that all data are anonymous. The information that is collected may be kept by the researcher and used in publications. This is done with the understanding that my anonymity and confidentiality will be preserved.

After writing the stories and filling out the questionnaire, I understand that there will be time to ask the experimenter questions regarding the study. I also understand that I can request a summary of the findings of this study. If I have any further questions about the study, I can contact the researcher, Jeanne Clancy, at (phone number), or the head of the College subject pool (name and phone number). If I have any questions concerning my rights as a participant in this study, I can call the Office of Sponsored Research, City University of New York at (212) 642-2059.

My signature below indicates that I have read and understood this consent form and I agree to participate in this study.

Participant's Signature _____

Participant's Name (print) _____

Researcher's Signature _____

Researcher's Name Jeanne Clancy, M.A.

Date: _____

Appendix E

Initial Instructions

I will show you some pictures, and your task will be to make up as dramatic a story as you can for each. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking; and then give the outcome. You will have four minutes to write each story. Turn to the first picture. (Adapted for group instruction from Murray, 1943, p. 3)

Appendix F

Debriefing for Subjects in the Anger Condition

Now that we have finished, I can tell you more about the purpose of this research. The main purpose has been to study how anger affects the kinds of stories people write. There were three different conditions in this study. In the neutral condition, participants like yourselves wrote stories to the same six pictures but no comments were made about the quality of their stories. In the anger condition, which was your group, derogatory remarks were made about the quality of the stories. My purpose was to make you angry in order to find out how anger affects story writing. In the simulated or pretend anger condition, participants were asked to imagine that I made negative comments as they wrote their stories.

This research is a study of how defense mechanisms work. The main hypothesis is that persons in the anger condition who cannot express their anger directly toward the experimenter will tend to write stories in which aggression is expressed either directly or indirectly. For example, the main character in a story may harm someone else, or if aggression is projected, someone else may harm the main character. The defense mechanism that I am studying is projection.

Now you can see why I was not able to explain the full purpose of the research before you wrote your stories. If I had told you that I was going to insult your story writing ability beforehand, it would have had no effect.

In psychological research it is sometimes permitted to use deception so long as any false ideas are set straight as soon as possible.

In a moment I am going to give you a chance to ask me questions about the research, but first I am going to be more specific about the false feedback I gave you about your stories. Each time, after I collected your stories, I pretended to read them. Then I told you the stories were not very good. In fact, I never read your stories at all. I was looking at another set of pages that look like yours. The comments I made were preplanned I was reading from a script. They had nothing to do with your stories. For example, after your second story I said: "I've been reading through these stories, and honestly, they are about the worst I've ever read. Could you try to write some better ones?"

It is important for you to understand that these comments

could not have referred to your stories because I never even read your stories. Presumably the stories written by your group are like those written by any other group--no better and no worse.

I want to ask you now if you understand that I was giving you false feedback about your stories.

(ASK STUDENTS TO RESPOND.)

Do you understand my reason for giving the false feedback?

(ASK STUDENTS TO RESPOND.)

Now, you understand that I was simply reading from a script and you understand the reason for the deception.

Even so, some psychologists feel that it is possible that in some corner of your mind you may still feel that you are not very good at story writing. Let me give you an example how the effects of false feedback may linger. Imagine that you are a math teacher and that you give your students a test on a new type of math problem. Now suppose that without even looking at the students' answers you write PASS on half the papers and FAIL on the other half. Students who received a pass would probably draw upon memories of other successes and conclude that they were good at solving this type of math problem. Those who were told that they failed might recall other failures and conclude that they were not good in this area. Such feelings might influence their future expectations. The group that passed may expect future success, and those that failed might expect future failure with similar problems. Your next step would be to tell the students that the PASS or FAIL grade had nothing to do with their test performance. Even though they know the grades were false, some of them may continue to believe that they are good or bad in that particular skill. They may tend to maintain the conclusions that they drew when they first received their false grades.

Fortunately, research has shown that your awareness of this possibility is one way to help remove the effects of false feedback.

Another way to reduce the effects of false feedback is to consider the opposite of the insults given. For example, take a moment to think of something about your stories that was good. (PAUSE). Was there a particular story theme that you liked? (PAUSE) Did you think that there was an interesting character in one or more of your stories? (PAUSE)

How would you have felt if after the second story I said <u>these were the most interesting stories that I had read</u> and that your writing was way ahead of the average college student?

Let me ask you now if you understand that you should be alert to the possibility that your conclusions may tend to persevere?

(ALLOW STUDENTS A CHANCE TO RESPOND.)

In summary, this was a study of defensive projection. The derogatory comments that I made were intended to make you angry. However, those comments were completely unrelated to your stories. If I had told you the full purpose of the study at the beginning I would not have been able to get your true reactions. Finally, it is important for you to be alert to the possibility that your conclusions may be affected by the false feedback so that you can adjust your expectations about your ability to write stories.

(At this point subjects will be given an opportunity to ask questions about the research, and offered the option of receiving a summary of findings.)

I hope that this study will contribute toward knowledge about defense mechanisms and I appreciate your assistance in doing this. I hope that it has been an interesting experience for you. I thank you very much for your participation.

REFERENCES

Atkinson, J. (1958). Thematic apperceptive measurement of motives within the context of a theory of motivation. In J. W. Atkinson (Ed.), Motives in fantasy, action and society (pp. 596-616). Princeton, NJ: Van Nostrand.

Bellak, L. (1944). The concept of projection: An experimental investigation and study of the concept. Psychiatry, 7, 353-370.

Bellak, L. (1950). On the problems of the concept of projection. In L. Abt, & L. Bellak, (Eds.), Projective Psychology. New York: Knopf.

Bellak, L. (1975). The Thematic Apperception Test. The Children's Apperception Test and the Senior Apperception Technique in clinical use (3rd ed.). New York: Grune and Stratton.

Clark, R. A. (1952). The projective measurement of experimentally induced levels of sexual motivation. Journal of Experimental Psychology, 44, 391-399.

Cohen, J. (1988). Statistical Power Analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cramer, P. (1979). Defense mechanisms in adolescence. Developmental Psychology, 15, 476-477.

Cramer, P. (1987). The development of defense mechanisms. Journal of Personality, 55(4), 597-614.

Cramer, P. (1991a). Anger and the use of defense mechanisms in college students. Journal of Personality, 59(1), 39-55.

Cramer, P. (1991b). The development of defense mechanism: Theory, research, and assessment. New York: Springer-Verlag.

Cramer, P. (1996). Storytelling, narrative, and the Thematic Apperception Test. New York: Guilford Press.

Cramer, P. & Carter, T. (1978). The relationship between sexual identification and the use of defense mechanisms. Journal of Personality Assessment, 42, 63-73.

Dollinger, S. & Cramer, P. (1990). Children's defensive responses and emotional upset following a disaster: A projective assessment. Journal of Personality Assessment, 54, 116-127.

Ehrenreich, J. H. (1990). Quantitative studies of responses elicited by selected TAT cards. Psychological Reports, 67, 15-18.

Fenichel, O. (1945). The psychoanalytic theory of neurosis. New York: Norton.

Freud, A. (1946). The ego and the mechanisms of defense. New York: International Universities Press.

Gleser, G. & Ihilevich, D. (1969). An objective instrument for measuring defense mechanisms. Journal of Consulting and Clinical Psychology, 33, 51-60.

Gottschalk, L., Gleser, G., Springer, K., Kaplan, S., Shanon, J., & Ross, W. (1960). Effects of perphenazine on verbal behavior patterns. American Medical Association Archives of General Psychiatry, 2, 632-639.

Gottschalk, L. & Gleser, G. (1969). The measurement of psychological states through the content analysis of verbal material. Berkeley: University of California Press.

Hall, C., & Van de Castle, R. (1966). The content analysis of dreams. New York: Appleton-Century-Crofts.

Haworth, M. (1963). A schedule for the analysis of CAT responses. Journal of Projective Techniques, 27, 181-184.

Haworth, M. (1966). The CAT: Facts about fantasy. New York: Grune & Stratton.

Hibbard, S., Farmer, L., Wells, C., Difillipo, E., Barry, W., Korman, R., & Sloan, P. (1994). Validation of Cramer's Defense Mechanism Manual for the TAT. Journal of Personality Assessment, 63, 197-210.

Jacobson, A., Beardslee, W., Hauser, S., Noam, G., & Powers, S. (1986). An approach to evaluating adolescent ego defense mechanisms using clinical interviews. In G. Vaillant, (Ed.) Empirical studies of ego mechanisms of defense. Washington, D.C.: American Psychiatric Press.

Juni, S. (1980). Classical projection: A critique of experimental methodologies. Genetic Psychology Monographs, 101, 119-146.

Lepper, M., Ross,L., & Lau, R. (1986). Persistence of Inaccurate beliefs about the self: Perseverance effects in the classroom. Journal of Personality and Social Psychology, 50, 482-491.

Lord, C., Lepper, M., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgement. Journal of Personality and Social Psychology, 6, 1231-1243.

McClelland, D., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? Psychological Review, 96, 690-702

McClelland, D., Atkinson, J., Clark, R., & Lowell, E. (1953). The achievement motive. New York: Appleton-Century-Crofts.

Messick, S. (1978). Personality measurement and college performance. In D. Jackson & S. Messick (Eds.), Problems in human assessment (pp. 834-845). Huntington, New York: Robert E. Krieger Publishing Co.

Murray, H. (1938). Explorations in personality. New York: Oxford University Press.

Murray, H. (1943). Thematic Apperception Test manual. Cambridge, MA: Harvard University Press.

Peterson, C. A. & Schiling, K. M. (1983). Card Pull in projective testing. Journal of Personality Assessment, 47(3), 265-275.

Saul, L., & Sheppard, E. (1956). An attempt to quantify emotional forces using manifest dreams: a preliminary study. Journal of American Psychoanalytic Association, 4, 486-502.

Sherwood, G. G. 1981. Self-serving biases in person perception: A reexamination of projection as a mechanism of defense. Psychological Bulletin, 90, 445-459.

Smith, C. P., Feld, S. C. & Franz, C. E. (1992). Methodological considerations: steps in research employing content analysis systems. In C.P Smith (Ed.), Motivation and personalty: Handbook of thematic content analysis (pp. 515-536). New York: Cambridge University Press.

Vaillant, G. (1971). Theoretical hierarchy of adaptive ego mechanisms. Archives of General Psychiatry, 24, 107-118.

Walker, E. L., & Atkinson, J. W. (1958). The expression of fear-related motivation in thematic apperception as a function of proximity to an atomic explosion. In J. W. Atkinson (Ed.), Motives in fantasy, action and society (pp. 143-159). Princeton, NJ: Van Nostrand.

Weisskopf-Joelson, E., Asher, E., Albrecht, K., & Hoffman, M. (1957). An experimental investigation of "label-avoidance" as a manifestation of repression. Journal of Projective Techniques, 21, 88-93.

Whitman, R., Pierce, C., Maas, J. & Baldridge, B. (1961). Drugs and dreams: II Imipramine and prochlorperazine. Comprehensive Psychiatry, 2, 219-226.