Empirical finding on the external validity of patient assault vignettes

James Carifio*, a, Marilyn Lanza b

*University of Massachusetts at Lowell, University Avenue, Lowell, MA 01854, USA
 bEdith Nourse Rogers Hospital, Bedford, MA 01730, USA

Received 17 May 1995; accepted 27 June 1995

Abstract

We had the nurses who were victims of actual assaults (n = 29) use the "blame" scales we have used in our last two written vignette studies to assign blame for their assault, and then we had three judges use the same scales to assign blame for the assault based on a written description of the assault by the victim. No significant difference in mean blame levels were found between victims and the average ratings for the three judges, or between victims, judges and the response levels obtained in our prior two studies. The mean 'blame' scores observed in written assault vignettes, therefore, are relatively and approximately the same as those that are, or will be observed in actual assaults.

Keywords: Blame scale; Empirical findings; Vignettes

1. Introduction

Patient assault (the act of physically attacking or restraining another person with part of a person's body or an object) has been increasingly documented as a serious problem. Studies have shown that assault is increasing across types of institutions (e.g. Carmel and Hunter, 1989), is still vastly under-reported (e.g. Brizer, 1988), causes serious post-traumatic stress reactions for the staff victims of assault (e.g. Lanza and Kayne, 1990), and is costly (e.g. Hunter and Carmel, 1992). Violence in the work lives of nurses is silenced and denied, and blaming the victim is still an all-too-common response (Chin, 1986).

Investigating assaultive behavior is a difficult task for a variety of reasons. The biggest obstacle is the simple fact that researchers cannot control or plan for their occurrence in order to study them in a systematic and controlled fashion. As a result, experimental manipulation of the independent variable, a control group, and random as-
signment of subjects (all requirements of an experimental design) are not possible (Polit and Hungler, 1991). If the researcher chooses to investigate the actual assault situation, external validity will be good; however, internal validity will be weak (Wood and Haber, 1986).

If the researcher chooses a simulation approach (Kelley and McGrath, 1986), internal validity should be very good (Lanza, 1986). In a simulation, the assault situation can be standardized. Everyone responds to the same stimulus. More control over extraneous variables is achieved (Topf, 1990) and the researcher can systematically manipulate the variables of interest. Subjects can be randomly assigned to treatment groups on the basis of a particular variable. This particular feature is extremely useful when there are insufficient numbers of a particular trait (such as male nurses) in the clinic setting to permit random assignment (Topf, 1990).

One type of simulation approach is written assault vignettes. An assault vignette is a carefully constructed description of an assault situation where all of the relevant key features are given and systematically varied in a set of such vignettes to study particular responses and reactions. We have developed and validated a set of six such assault vignettes (see below) and we have shown that the assault vignettes are not only reliable and valid, but that they also produce rich findings about various aspects of assault situations and people’s resulting reactions (see Carifio and Lanza, 1991, 1992).

One of the major problems with using written vignettes (i.e. a simulation) is that the assault stimulus is artificial. Subjects are asked to imagine how they would respond. The concern is always and continually that the actual assault situation might yield quite different results (Marques, 1991). We chose to take advantage of the strengths of the simulation approach when studying assaultive behavior. The success of our prior research with written vignettes provided the impetus for the present study. The purpose of this study was to assess to what degree the use of patient assault vignettes generalizes to actual patient assaults of nurses.

2. Patient assault vignettes

Simulations are increasingly being used in nursing research, and a number of other professional training programs such as those used with teachers, managers, clinicians and healthcare providers. Westfall et al. (1986) used written vignettes and Tanner (1987) used video simulation to study nursing inferences within the framework of information processing theory. Simulations have also been used to study victims’ reaction to violence such as rape (Janoff-Bulman and Frieze, 1983; Marques, 1991). Marques (1991) described, in elaboration, the development of audiotape vignettes for independent judges to assess. We have been using written assault vignettes to assess the assignment of blame for the assault of nurses (and other healthcare professionals) by patients for several years.

In their most current form, our patient assault vignettes describe an incident between a nurse (female or male) and a male patient. The nurse and the patient are having a discussion about the patient’s weekend pass. The discussion takes place in the day room of a psychiatric hospital, and specific conversation and behaviors of the patient and nurses are given (see Carifio and Lanza, 1991 for actual vignettes and details). Communication breaks down and the nurse is either mildly or severely physically assaulted or verbally abused (the control condition). There are, therefore, six [6] vignettes which constitute a 2 × 3 (gender of the nurse victim by severity of the incident with the patient) design.

Each of the six vignettes is followed by 14 items assessing the assignment of blame (attribution of causality) for the incident described in the vignette. The first item is a direct measure of blame attribution. It asks the respondent to select who is responsible for the incident described in the vignette. The remaining 13 items are indirect measures of causal attribution (blame), which judge the degree to which the nurse was responsible for the incident described.

3. Vignette validation

We have validated our six vignettes in a num-
ber of studies (e.g. Carifio and Lanza, 1991, 1994; Lanza and Carifio, 1992a, 1992b). Content validity was assessed by identifying important assault-related factors from the literature and the experience of clinical experts, and using these factors as 'blueprints' for constructing our vignettes. On two occasions, a panel of 12 experts have assessed the quality of our 14-item scale, the plausibility of our vignettes, and have validated the kind and severity of the incident described in each vignette (see Lanza and Carifio, 1992a,b). Using the Campbell and Fiske (1959) paradigm and design, panelists reached similar converging conclusions using different methods to classify the kind and severity of the incidents described in each vignette. The ability of panelists to discriminate vignette versions using a modified version of the Yudofsky overt aggression scale (Yudofsky et al., 1986) corroborated our classification system for the vignettes.

Correlations ($n = 61$) between the direct and indirect causal attribution (blame) scores ranged from 0.26 to 0.48. The concurrent validity coefficient between blame (causal attribution) scores and age ranged from 0.19 to 0.30 (Lanza and Carifio, 1992b).

Correlations between scale items and total scores ranged from 0.40 to 0.89, and the test-retest reliabilities ranged from 0.86 to 0.89. The Cronbach alpha coefficient for the 13 indirect causal attribution items ranged from 0.83 to 0.91 over several studies. Factor analysis found two underlying but correlated factors for the 13 items that accounted for 62% of the variance (see Lanza and Carifio, 1992b). These two factors were personal attribution (48%) and blame of others (14%). The validity evidence for our six vignettes has been excellent.

4. Initial findings

Responses to our six vignettes were obtained from 66 practising nurses who were randomly assigned to a single (one) vignette. In this study, both female and male nurses were blamed as much for the incident that occurred with the patient in the control (verbal abuse) vignette as they were in the severe assault vignette. However, they were blamed more for the incident that occurred in the mild assault vignette than they were in either the control or severe assault vignette. Female nurses, interestingly, were blamed more than male nurses for the incident in the control and severe assault vignette, but not in the mild assault vignette. These complex findings fit the non-linear elliptic umbilic model of catastrophe theory (Zeeman, 1976), and not the linear models and interpretations of these variables currently in the literature.

A closer look at the literature on aggression and assault (see Lanza and Carifio, 1991) found a wide variety of evidence to support non-linear dynamics and non-linear patterns of blame attribution. Using catastrophe theory as a general model and perspective of non-linear dynamics and behavior, a specific theoretical model was generated to explain our findings of professional nurses being blamed most for incidents involving mild assaults. The specific details of this non-linear, 'kinky', or 'twisted' theoretical model, blaming pattern, and general dynamic is given below.

5. Non-linear dynamics

Responding to a mild assault is typically seen as within the range of usual professional functioning for psychiatric nurses, whether male or female. Typical nurse behaviors related to a mild assault situation include assessing a patient's aggressive potential, making predictions about the likelihood of assaultive behavior, and intervening in actual assaults. According to professional guidelines, male and female nurses are held equally accountable for their performance in such situations. Consequently, gender bias disappears in this mild assault condition, and the amount of blame assigned to the nurse professional increases to its highest levels. In the 'extremes' (i.e. verbal abuse and severe assault), which are 'outside' of the professional span of control and thus professional expectation, gender bias asserts itself strongly (as the influence of professionalism is close to nil) and female nurses are blamed more severely than male nurses even though both are
blamed significantly less than in the mild assault conditions. Therefore, it may be that severity of aggression is a surrogate for degree of professional control or behavior for which a nurse professional can reasonably be expected to be accountable.

Professional behavior standards and gender bias are two contradictory factors influencing the attribution of blame. Contradictory factors influencing a dependent variable is the heart of catastrophe theory and non-linear dynamics and what puts the ‘kink’ or ‘twist’ in the data and the dynamics. Carmel and Hunter (1989) and Ryan and Poster (1989) cite evidence to support the idea that there is a gender-based discrepancy between the expectations and actual practices of nurses. Both sets of investigators found that policies and training programs state that all staff, regardless of gender, are believed to be equally capable within broad limits of participating in the containment of violent behavior. In actual practice, however, male staff members may become more actively involved in containing violent behavior, and in fact, female staff members may be selectively excused. Our data may reflect the gender-bias disparity between theoretic (professional) expectations and the actual practices of nurses in assault situations. Simply stated, the gender of the nurse-victim may affect attribution more in extreme situations than in ‘routine,’ everyday professional interactions.

6. Purpose

This study sought to answer the question, ‘To what degree do the causal attribution (‘blame’) scores obtained from carefully constructed written vignettes of assault incidents (i.e. the simulation of reality) agree with and reflect the results that would be obtained from victims and observers of actual assaults in situ (i.e. reality?)’. This external validity question is the outstanding and most frequent question in the literature for both vignette studies to assign blame for their assault. Then we had three judges use the same scale to assign blame for the assault based on a written description of the assault by the victim. Our intention was not only to see if the three judges agreed with the victims, but to see if the responses of the victims and the judges had the same mean levels and non-linear response patterns as in our two previous studies.

7. Pilot study

We initially carried out a pilot study (n = 16) to determine the degree to which responses to our written vignettes predicted responses to actual assault situations. Using a Pearson correlation coefficient, the relationship between responses to the vignettes and response to actual assaults was $r = +0.52 (P < 0.05)$. A fairly strong relationship was suggested in this pilot study. While it was encouraging to see a good correlation between causal attribution (blame) scores for the vignettes (i.e. simulation) and the real world clinical situation and context, methodological improvements were required before the results could be considered more than tentative. Based upon our experience with the pilot study, we re-designed the present study.

8. Methodology

In this study, RN’s who had been assaulted were asked to rate the actual assault using the Causal Attribution Rating Scale. Staff who had been the target of assaults were identified using the nursing ‘24-h report.’ This report is a daily log completed by each shift which records significant incidents and findings. This log is a far superior method of recording assaults than commonly exists in most hospitals (see Lanza and Campbell, 1991). In all, 29 assaults that occurred at one veterans hospital in eastern Massachusetts over a 4-month period constitute the sample for this study. This hospital had 264 nurses who could make entries into the ‘24-h report’ for each unit in the hospital.

Each assault victim who agreed to participate in this study was asked to write an account of the
assault that included:

(1) location of the incident;
(2) a description of the patient's behavior;
(3) a description of their own behavior;
(4) an account of the conversation between the patient and victim;
(5) a description of the physical characteristics of the patient: approximate height, weight, sex and whether of small, medium or large bone structure;
(6) a description of the victim's own physical characteristics, including height, weight, and size of bone structure; and
(7) an account of the immediate outcomes of this incident for both the patient and victim, including site of injury, degree of injury, and subsequent behaviors.

In all, 14 identified assault victims did not agree to participate in the present study. The 29 assault victims who did agree to participate in this study were asked to use the above information when describing their assault so that all of the victim generated descriptions (i.e., vignettes) would contain the same salient information. Also, the above list is the same outline used to devise the vignettes that were validated in our previous studies. Having the victims use the above outline, therefore, helped to ensure that the salient features of the 'vignettes' generated by the victims would be similar to the salient features of our vignettes for consistency.

After developing a written description of their assault, victims were then asked to rate the assault incident using (1) the Causal Attribution Scale (CAS), (2) the Overt Aggression Scale (OAS), and (3) the Global Rating for Severity of Assault (GRSA). In addition, they completed the Sociodemographic Questionnaire and the Just World Scale.

Three judges reviewed the written descriptions of the incident and independently rated each using the Overt Aggression Scale, the Global Rating of Assault Severity Scale, and then completed the Causal Attribution Rating Scale for the incident. All judges had at least 15 years psychiatric nursing experience, had been assault victims, and were accustomed to working with assaultive patients. One of these judges was a male nursing student. The other two judges were female RNs who spent 10 or more years as nurse managers. Judges were not asked to 'reconcile' their ratings or classification of the assault incidents as a baseline on the variance that would occur between judges was desired. These instruments and procedures are the same as those we used to evaluate the construct validity of the vignettes and to conduct the prior studies using the validated vignettes.

9. Just World Scale

Rubin and Peplau (1973, 1975) developed the Just World Scale to examine the relatively enduring individual differences in the way people perceive others as deserving their fates in a wide range of situations. The respondent is asked to indicate his degree of agreement or disagreement on a six-point continuum rated as 'just' or 'unjust'. Scores are based on total item scores reflecting belief in a just world.

In two separate studies, there was a wide distribution of total scores on the scale, ranging from total rejection to qualified acceptance of the just world ideology. The scale had high internal consistency (coefficient alpha or KR-20 equal to 0.80 and 0.81) and predictive validity was evidenced in several studies (e.g., Rubin and Peplau, 1973, 1975). The psychometric data suggest that in spite of the broad spectrum of contents sampled, the scale taps an underlying belief that can be meaningfully viewed as a single attitudinal continuum (Rubin and Peplau, 1973, 1975).

10. Overt Aggression Scale

The Overt Aggression Scale (OAS) cited earlier is divided into four categories: verbal aggression, physical aggression against objects, physical aggression against self, and physical aggression against other people (Yudofsky et al., 1986). For this study, only categories of verbal aggression and physical aggression against people were used. Within each category, four well described observable behaviors that are exemplars of the category
are arranged in order of increasing severity of exemplification. The specific items which describe aggressive behavior are rated as present or absent. The intraclass correlation coefficients were greater than 0.75 for most items (Spitzer, Fleiss, and Endicott, 1978). The rationale for using the OAS is that it is one of the few existing validated scales used for the clinical assessment of aggressive behavior. In fact, the OAS more accurately reflects the occurrence of all types of aggressive behavior than is reported in hospital records and incident reports (Silver and Yudofsky, 1987).

11. Results and discussion

The ages of the 29 assault victims in the sample ranged from 22 to 59 years, with an average age of 44.2 years and a median age of 43 years. Years of work experience ranged from 1 year to 35 years with a mean of 16.2 years and a median of 14.5 years. Forty percent (40%) of the nurses had associate degrees in nursing, whereas 42% had BS degrees in nursing. The remaining 14% of the sample had advanced degrees in nursing. Fifty-eight percent (58%) of the sample were staff nurses while the remaining 41% were head nurses or nursing directors. Twenty-six (26) of the assault victims were female and three were male.

Only two of the 29 victims had not been previously assaulted on the job. Fourteen (14) nurse victims had been assaulted once to three times and 13 victims had been assaulted seven or more times. Nine (9) victims had been slightly injured in previous assaults, while six had been moderately injured and another six were fairly to severely injured in previous assaults. Twelve (12) of the 29 victims had moderate to fairly severe emotional reactions to their prior assaults. This sample of assault victims, therefore, was fairly heterogeneous and representative on a number of relevant factors.

Only seven of the 29 assault victims globally rated their current assault as being severe in nature. In terms of the three judges who rated victim descriptions of their current assault, judge 2 globally rated six of the 29 assaults as being severe, with judges 3 and 1 globally rating five and one of the assaults as being severe, respectively. Judges 2 and 3 agreed on five of the seven victim rated cases of severe assault. The one case rated as severe by judge 1 was not rated as severe by the victim or the other judges.

It should be noted that the low frequencies of globally judged severe assaults in the sample creates several analytical problems. In retrospect, therefore, it may have been better to have had two categories of mild assault (somewhat and mild) and two categories of severe assault (somewhat and severe) for research purposes. The remaining 22 incidents were classified as mild assaults. There were no 'verbal abuse only' (i.e. control) incidents in the sample.

The total Yudofsky Overt Aggression Scale score generated by the victim correlated with the victim's overall global classification of the assault incident as being mild or severe at $r = +0.42$ ($P < 0.05$). As the global classification of the incident is a dichotomy in this sample and the two Yudofsky subscales used have only eight items (four non-physical abuse and four physical aggression), this correlation coefficient is quite good concurrent validity evidence for the correctness and meaningfulness of each victim's global classification of the severity of her or his assault incident. The correlation between each victim's total Yudofsky scale score and total causal attribution scale ('blame') score was $r = +0.43$ ($P < 0.05$), which is good concurrent evidence for the validity and meaningfulness of victim's causal attribution scale scores. There were no significant differences (repeated measures ANOVA; see below) between judges and victims on their (average) total Yudofsky scores for incidents or their overall global classification of assault incidents. Again, this good concurrent evidence for the validity and meaningfulness of the global assault classifications and causal attribution scores and cross-validates the results found for the victims.

The average causal attribution scale ('blame') score for the assault victims was 26.5 with a standard deviation of 6.5. Victims' ratings of their own assaults ranged from 13 to 38. None of the background variables significantly predicted a victim's assessment of her or his own assault experience, except for the prior assault history item indicating the number of days out of work to
recover from a previous assault which correlated to causal attribution scores at \( r = -0.54 \) \((P < 0.01)\). Therefore, the six assault victims who had previously been out of work 1 or more days due to an assault, blamed themselves less for their current assault than the 23 victims who had previously not been assaulted severely enough to be out of work. These latter victims blamed themselves more for their current assault.

This finding of professionals who have a prior history of fairly severe to severe assault experiences (or a large number of prior assaults) blaming themselves (or other assault victims) less for the assault occurrence both supports and cross-validates the same finding in our last study (see Carifio and Lanza, 1994). The finding that the other background variables on which data were collected do not predict causal attribution scores also cross-validates previous findings. In our prior studies, these findings were for subjects rating vignettes which portrayed verbal abuse, mild or severe assaults of nurses. In this instance, the findings are for actual assault victims rating their own actual assault. Although not ‘exactly and perfectly’ comparable, this similarity of findings indicates that the results obtained with written vignettes are similar to and have good correspondence to actual assault situations.

Table 1 presents the results of a repeated measure ANOVA of the victims causal attribution (‘blame’) ratings of their assault and those of the three judges. A repeated measure ANOVA was done even though there was not a true repeated measure in the data so that the subjects factor could be used to assess inter-judge agreement (see Kerlinger, 1985) and the raters (‘treatment’) factor could be used to assess the equivalence of the mean levels of blame assigned to the victim for the assaults for this particular inter-rater agreement question and analysis.

As can be seen from Table 1, there was both a subject effect and a rater effect at the \( P < 0.001 \) level of significance. Judges rated (some of) the same assaults differently from each other (the subjects effect). Two of the judges, on the average, blamed the victims significantly more for their assaults than the victims blamed themselves, and one judge blamed the victims significantly less (judge 1). It should be noted here that one of the two judges who blamed the victims significantly more for their assaults was male and one was female. Neither of these judges had a high prior assault history, whereas judge 1, who was female, had a high prior assault history.

Table 2 presents the means and standard deviations for the assault victim’s causal attribution scores and the average of the three judges scores. A \( t \)-test between these two means, as the victims and the judges (average) causal attribution scores are independent samples, was not significant \((t = 1.6, \text {d.f.} = 48, P > 0.05)\). Averaging the judges’ ratings, therefore, both reduces the variances in the judges ratings and clarifies the findings.

Table 3 presents a summary of repeated mea-

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Repeated measures ANOVA on mean causal attribution (‘blame’) scores for assault victims and judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raters</td>
<td>n</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>Victims</td>
<td>29</td>
</tr>
<tr>
<td>Judge 1</td>
<td>29</td>
</tr>
<tr>
<td>Judge 2</td>
<td>29</td>
</tr>
<tr>
<td>Judge 3</td>
<td>29</td>
</tr>
<tr>
<td>Source</td>
<td>d.f.</td>
</tr>
<tr>
<td>Subjects</td>
<td>28</td>
</tr>
<tr>
<td>Raters</td>
<td>3</td>
</tr>
<tr>
<td>Error</td>
<td>84</td>
</tr>
</tbody>
</table>
Means and standard deviations for causal attribution ('blame') scores for assault victims and the average of the three judges

<table>
<thead>
<tr>
<th>Raters</th>
<th>n</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victims</td>
<td>29</td>
<td>26.5</td>
<td>6.5</td>
</tr>
<tr>
<td>Average of three judges</td>
<td>29</td>
<td>28.8</td>
<td>4.3</td>
</tr>
</tbody>
</table>

$t = 1.6, \text{ d.f.} = 48, P > 0.05$

As can be seen in Table 3, both victims and judges rated victims to be more to blame for mild assaults than for severe assaults. This finding is consistent with numerous prior findings and studies. It should be noted here, however, that this particular effect is not 'simply linear.' The effect is, in fact, non-linear and an instance of catastrophe theory. Judges blame victims as much for verbal abuse as they do for severe assault with victims being blamed most for mild assaults. Furthermore, female nurse victims are blamed more than male nurse victims for verbal abuse and severe assault but not for mild assault (see Table 5 for details).

We have found these results in two studies (see Carifio and Lanza, 1991, 1992), and have proposed a theory to account for them which focuses on the interaction between what is considered to be one's professional span of control and sexism. As previously stated, this theory says that for events that are considered to be within one's professional span of control, sexism has no effect and all professionals are held to high and equal accountability. However, for incidents that fall outside of one's professional span of control (what each professional should 'handle'), sexism will have an effect (see Carifio and Lanza, 1992 for further details).

$t$-tests for independent samples calculated on the mean differences given in Table 3 between victims and the average of the three judges revealed no significant differences between the victims and the average of the three judges on the assignment of blame levels for mild assault ($t = 0.47, \text{ d.f.} = 42, P < 0.05$), or severe assault ($t = 1.2, \text{ d.f.} = 12, P < 0.05$). Although the average of the three judges' blame scores tended
Intercorrelations between victims' and judges' ratings of assaults

Table 4

<table>
<thead>
<tr>
<th></th>
<th>Victim</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson product moment correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victim</td>
<td>1.00</td>
<td>0.25</td>
<td>0.48</td>
<td>0.50</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td></td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Spearman rank order correlations

<table>
<thead>
<tr>
<th></th>
<th>Victim</th>
<th>Judge 1</th>
<th>Judge 2</th>
<th>Judge 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victim</td>
<td>1.00</td>
<td>0.26</td>
<td>0.61</td>
<td>0.54</td>
</tr>
<tr>
<td>Judge 1</td>
<td>1.00</td>
<td>0.42</td>
<td></td>
<td>0.64</td>
</tr>
<tr>
<td>Judge 2</td>
<td>1.00</td>
<td></td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>Judge 3</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

to be higher than those of the victims for both mild and severe assaults, the judges' (average) assignment of blame from reading the victims' written description (i.e. vignette) of her or his assault statistically were the same as the victims. Therefore, this finding is direct evidence of the similarity of findings between (victim generated) vignettes (i.e. simulated reality) and actual assault incidents (i.e. the 'real world'). As previously stated, the question of the similarity of responses to 'simulated realities' and the 'real thing' has been a longstanding question in the literature.

In order to understand the degree to which the assault victims and judges rated assaults similarly and differently, a table of inter-rater correlation coefficients was generated (see Table 4). As can be seen from Table 4, judge 1 agreed with the victim least \( r = +0.25, P > 0.05 \), whereas judges 2 and 3 had significant but moderate rating agreements with the victim \( r = +0.48 \) and \( r = +0.50 \), respectively. The three judges also only moderately agreed with themselves as can be seen from Table 4.

The average of the three judges ratings correlated with the victims ratings at \( r = +0.54 \). The average of the three judges rating, therefore, correlated slightly better than any of the judges individually.

A close examination of the data on a case by case basis revealed that judge 1 widely disagreed (i.e. a difference of 8-12 scale points) with the assault victim's rating in seven of 29 cases. Given the nature of the Pearson product moment correlation and the small size of this sample, these wide disagreements are enough to radically reduce the size of the inter-rater correlation coefficient. This same wide disagreement result was observed between the other judges' ratings and the victims' ratings as well as between the three judges themselves.

This type of case level disagreement is not uncommon between multiple raters, and we did not make the judges 'reconcile' their ratings, which is common practice. We wanted to assess the nature of the discrepancies that would occur to have a baseline on this factor. When 'discrepant cases' were removed, all inter-rater correlations were greater than \( r = +0.70 \). Therefore, given the size of the sample and the fact that the judges were rating victims' written descriptions of their assaults (i.e. victim generated vignettes), the agreement between the victim's assessment of blame for the assault and an independent panel of three judges' assessment for blame for the assault was reasonably good. A larger sample and item level analyses is needed to more definitively answer this agreement question. However, we believe that the results for this sample are both good and highly promising.

As previously stated, the central question this study was designed to answer was, 'To what degree do the causal attribution ('blame') scores
obtained from carefully constructed written vignettes of assault incidents ranging from verbal abuse to severe assault (i.e., the simulation of reality) agree with and reflect the results that would be obtained from victims and observers of actual assaults in situ (i.e., reality). This question is the outstanding and most often posed critical question in the literature on both vignettes and assaults.

To answer this question, we had the victims of actual assaults use the causal attribution scale used also in our last two vignette studies to attribute blame for their assault. Then the three judges used the same scale to assign blame for the assault based on the victim’s written description of the assault. This written description is a victim-generated vignette of the assault incident. Judges rated this victim-generated assault vignette as there was no way to have them observe the actual assault itself. Therefore, this design comes as close to assessing the degree of agreement between causal attribution of blame for actual assault in situ and the causal attribution of blame for assaults in the victim-generated vignettes. As stated above, there was no significant difference between victims’ attribution of blame for their mild and severe assaults and the average of the three judges’ assignment of blame. Given this point, the question that arises is, ‘How similar are the mean blame levels observed in this study and those in our prior two studies which used carefully written vignettes to be rated by judges using the same scale used by victims and judges in this study?’

Table 5 presents the mean response levels by vignette and actual assault type for the present study and the previous two studies that were conducted using vignettes. As can be seen in

<table>
<thead>
<tr>
<th>Study</th>
<th>Severity of incident in vignette (verbal abuse) Control</th>
<th>Mild Assault</th>
<th>Severe Assault</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td><strong>First study</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female nurse*</td>
<td>9</td>
<td>35.2</td>
<td>9.4</td>
</tr>
<tr>
<td>Male nurse*</td>
<td>8</td>
<td>29.6</td>
<td>11.0</td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>32.6</td>
<td>10.3</td>
</tr>
<tr>
<td><strong>Second study</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female nurse*</td>
<td>45</td>
<td>28.4</td>
<td>7.2</td>
</tr>
<tr>
<td>Average of two studies</td>
<td>62</td>
<td>29.2</td>
<td>8.7</td>
</tr>
<tr>
<td><strong>This study</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Victims</td>
<td>22</td>
<td>27.1</td>
<td>6.6</td>
</tr>
<tr>
<td>Judge 1</td>
<td>22</td>
<td>25.5</td>
<td>4.9</td>
</tr>
<tr>
<td>Judge 2</td>
<td>22</td>
<td>31.1</td>
<td>3.7</td>
</tr>
<tr>
<td>Judge 3</td>
<td>22</td>
<td>30.2</td>
<td>6.0</td>
</tr>
<tr>
<td>Average of three judges</td>
<td>22</td>
<td>28.9</td>
<td>4.9</td>
</tr>
</tbody>
</table>

*Gender of the assault victim in the vignette
Table 5, there are no significant differences between the mean blame scores for mild assaults on female nurses in our second prior study ($n = 45$) and the average rating of the three judges for the mild assaults in this study ($t = 0.25$, d.f. = 65, $P > 0.05$), or the mean response levels of the actual mild assault victims ($t = 1.28$, d.f. = 65, $P > 0.05$). This same result was observed for severe assaults. There are no significant differences between the mean blame scores for severe assaults on female nurses in our second prior study ($n = 45$) and the average rating of the three judges for the severe assaults in this study ($t = 0.48$, d.f. = 50, $P > 0.05$), or the mean response levels of the actual mild assault victims ($t = 1.05$, d.f. = 50, $P > 0.05$). It should be noted that the mean levels given in Table 5 for the second study are the response levels of experienced nurses ($n = 45$) so that like elements are being compared to like elements between studies.

The mean blame levels (for experienced nurses) between the first study (Carifio and Lanza, 1991) and this study as well as the first study and the second study (Carifio and Lanza, 1992) are significantly different for both mild and severe assaults, although the relationships between the means is correct. The problem with comparing the first study to this one and the second one is that the sample size for the mild assault condition in the first study is somewhat low (comparatively), the variances are larger, and nothing was known about the prior assault histories of subjects in this first study. It will be recalled that both the second study and this study found that subjects with 'high' prior assault histories tend to blame victims significantly less than subjects without a 'high' prior assault history. Also, there was a somewhat high subject loss in the first study which may have inflated it's means. Given these factors, using the average of the first two studies is the better estimate as opposed to simply dismissing the first study or just going with the second, even though it is a somewhat better and more representative study than the first.

When the average of the first two studies are used, no significant differences are observed between the first two study averages, victims and the average of the 3 judges causal attribution ('blame') scores. The difference between the two study average for the mild assault vignettes and the victims responses to actual mild assaults, however, was close to a significant difference at the 0.10 level ($t = 2.0$, df = 81).

12. Conclusions

Given these findings, it would seem warranted to tentatively conclude that given similar prior assault histories of the subjects (raters or victims), the mean causal attribution ('blame') score levels observed in carefully constructed assault vignettes (i.e. simulated reality) are relatively and approximately the same as those that are or will be observed in actual assaults (i.e. the real world). Subjects' mean responses to carefully and systematically constructed assault vignettes, therefore, are similar to and predictive of their mean level responses to actual assaults, as victims or judges, given roughly similar prior assault histories.

References