

The predictive validity of typical and maximal personality measures in self-reports and peer reports

ROBERT C. KLESGES, HUGH MCGINLEY, GREGORY J. JURKOVIC,
and THOMAS J. MORGAN

University of Wyoming, Laramie, Wyoming 82071

Recent investigations have suggested that personality measures based on a maximal format are superior to those based on a typical format. The results of the present study suggest that peer reports and self-reports of a subject's maximal expression of aggression are more reliable predictors of the subject's laboratory behavior than are their reports of the subject's typical expression of aggression.

Recently, trait approaches to personality theory and assessment have been seriously questioned. Extensive arguments have been expressed by Mischel (1968, 1971, 1973), who contends that situational factors are more influential than personality traits in predicting criterion behaviors. In support of his hypothesis, Mischel cites numerous cases where correlations between responses made in different situations by the same person are consistently low (.24 to .34). Mischel has challenged long-standing assumptions about the nature of personality and, while his conclusions have not been universally accepted (e.g., Alker, 1972; Bowers, 1973), Mischel's arguments have stimulated investigations into how the predictive validity of self-report measures might be enhanced (Turner, 1978; Willerman, Turner, & Peterson, 1976).

In an attempt to increase the predictive validity of self-report personality measures, Willerman et al. (1976) hypothesized that a maximal performance format is superior to the more traditional typical performance format for predicting specific behavior. A typical performance question, such as, "Are you friendly?" forces the individual to subjectively weight and average all relevant behavior that he or she considers to be important to the "trait of friendliness." The result of this weighing and averaging is that averaging errors, ambiguity, and response biases tend to reduce the predictive validity of the typical performance item. In addition, as Willerman, Rosen, Shoup, and Malina

As of August 1979, the senior author will be at the Department of Psychiatry and the Behavioral Sciences, University of Southern California School of Medicine, 2025 Zonal Avenue, Room 630, Mudd Building, Los Angeles, California 90033. Reprint requests can be sent either to the senior author at the above address or to Hugh McGinley, Department of Psychology, University of Wyoming, Laramie, Wyoming 82071. The authors would like to thank Lee Willerman, Robert G. Turner, and Mark L. Pantle for their unselfish assistance and guidance in the experiment. Gregory Jurkovic is now at Georgia State University.

(Note 1) point out, when answering a typical performance item, an individual must average both the frequency of a trait and its intensity. Thus, an individual who is fairly friendly all the time might rate herself or himself the same as an individual who is not usually friendly but who is exceedingly friendly on rare occasions. In contrast to measures of typical performance, measures of maximal performance require only that the subject focus on the extreme of a trait, such as friendliness, and determine what he or she can do or is capable of doing (Mischel, 1973). The frequency with which the behavior occurs is not considered.

Willerman et al. (1976) set out to construct maximal performance tests of personality that could be validated against both typical and maximal criteria by correlating subjects' self-reports of typical and maximal performances of anger and elation with typical and maximal performance laboratory behavior assessments of the subjects' aggression and elation. They found that maximal self-reports, in general, tended to outpredict typical self-reports when the predictor variables were either typical or maximal laboratory behaviors. Other investigations (Turner, 1978; Willerman et al., Note 1) have found results consistent with those reported by Willerman et al. (1976): Maximal self-reports are generally better predictors of laboratory behavior than are typical self-reports.

The studies demonstrating superiority of maximal reports have been limited to self-reports, and it has not been determined whether the superiority of maximal reports will generalize from the laboratory to peer reports. This extension is especially important, since peer reports are considered to be a more independent measure of a subject's behavior than self-reports (Cronbach, 1970) and are often used as a criterion to which subject behavior is compared (e.g., Norman, 1969; Nunnally, 1978).

In concert with previous research, the first general hypothesis of the present study was that the subjects'

reports of their maximal expressions of aggression would correlate higher with both the subjects' maximal and typical aggression in the laboratory than would self-reports of their typically expressed aggression. The second general hypothesis was that peer reports of the subjects' maximal expressions of aggression, as compared to their typically expressed aggression, would be more reliable predictors of the subjects' typical and maximal expressions of aggression in the laboratory. It holds that if the data support both of these general hypotheses, then the peer reports of the subjects' maximal expressions of aggression should correlate higher with the subjects' typical and maximal self-reports of aggression than with the peer reports of the subjects' typical expressions of aggression (Cronbach, 1955).

METHOD

Subjects

One-hundred students (69 females and 31 males) from introductory psychology classes served as subjects. In addition, friends of the subjects served as peer raters (62 females and 38 males). Subjects received extra credit in their classes for participation.

Procedure

Subjects were tested twice, first in their respective classes and then in the laboratory. In the initial assessment, the Willerman et al. (1976) format was used, where the subjects individually indicated their typical and maximal expressions of aggression by writing illustrative stories. For the typical story requests, they were told, "Imagine that your friend has just made you extremely angry. Describe in a short paragraph what you normally or typically would do and say in expressing your anger." For the maximal story request, they were told, "Now this part is a test of ability. Imagine your friend has made you extremely angry. Describe in one short paragraph what you are capable of doing and saying if you maximally express your anger." The subjects were given as much time as they needed to complete the stories. The order in which the story instructions were given was counterbalanced. The subjects were also given a self-consciousness inventory (Fenigstein, Scheier, & Buss, 1975), which was not utilized for the purposes of this study. After the subjects finished the stories and self-consciousness inventories, they signed up for a time to return with a friend of either sex who knew them fairly well.

A subject and her or his friend were met by the experimenter when they appeared 5 to 27 days later for the second testing. The friend was left in one of the laboratory rooms while the subject was taken to a small room that contained a one-way vision mirror and a 127-cm Bobo doll. When in the room, the experimenter read the following instructions: "In this experiment, we are interested in your ability to express your emotions. The experiment will consist of two parts. First, I will leave and when I knock on the window, I want you to imagine your friend you brought with you today has made you very angry. Imagine this Bobo doll [pointing] is your friend. I want you to act out what you would typically or normally do and say in expressing your anger. Then when you are finished, indicate to me that you are done. Then, I will knock on the window again. This time, I again want you to imagine your friend has made you very angry. I want you to act out what you are capable of doing and saying in expressing your anger, in other words, we are interested in your maximal expression of anger."

The experimenter made sure the subject understood the instructions and then he left the room. After 1 min, he, by

knocking on the window, signaled the subject to start his or her typical performance of aggression. After the subject reported that she or he was done with the performance, the experimenter waited for 30 sec and then knocked again, this time signaling the subject to act out her or his maximal performance of aggression. Two judges, separated by a partition, independently judged the subject's performance. The judges had no knowledge of the subject's previous typical-maximal self-reports. This order, typical first and then maximal, was followed throughout the laboratory performances due to previous research (Willerman et al., 1976) and pilot testing that revealed it is apparently difficult for subjects to act out maximal expression of anger and then to act out typical expression of anger. After the subject signaled that he or she was finished, the experimenter returned to debrief the subject.

While the subject participated in the experiment, the subject's friend wrote illustrative stories of how he or she thought the subject would react in typical and maximal expressions of anger. For the typical story requests, the subject's friend was told, "Imagine that you have made your friend (the one you came with today) very angry. Describe in a short paragraph what he or she would typically or normally do and say in expressing his or her anger." For the maximal story requests, they were told, "Now this part is a test of ability. Imagine that you have made your friend very angry. Describe in a short paragraph what he or she is capable of doing and saying if she or he maximally expressed his or her anger." The peers were allotted ample time to complete their stories. The order in which the story instructions appeared was counterbalanced.

The self-reports, laboratory assessments, and peer reports were scored independently utilizing the Willerman et al. (1976) scale of aggression, a 7-point scale with three parts: a physical expression of anger, a verbal expression of anger, and an overall expression of anger. The latter is computed by averaging the first two. Data analyses were of the overall scores. The ratings from the peer reports and the subjects' self-reports of the subjects' expressions of aggression served as the predictors of the subjects' expressions of anger in the laboratory. Also, as a test of consistency, ratings derived from the subjects' self-reports were viewed as a dependent variable for the ratings derived from the peer reports. The interrater reliability for the judges' scoring of the typical self-report stories was .86; for the maximal self-reports, it was .88. The interrater reliabilities for the judges' scoring of the typical and maximal peer reports were .86 and .91, respectively; for the typical and maximal laboratory assessments, the reliabilities were .86 and .95, respectively. Retesting was conducted for 56 subjects 7 weeks later. The test-retest reliabilities were .41 for the typical self-reports, and .66 for maximal self-reports. Test-retest reliabilities of the peer reports and laboratory assessments were not available.

RESULTS

An analysis of variance was conducted in order to check that maximal expressions of aggression resulted in higher ratings of aggression than typical expressions of aggression. The 2 by 3 by 2 analysis of variance (male, female by laboratory, self, and peer ratings by typical and maximal expressions of aggression) revealed a significant main effect for the typical-maximal expression of aggression factor, indicating that the instructions were effective [$F(1,294) = 366.50, p < .001$]. There was also a significant main effect for the source of the aggression rating [$F(2,294) = 4.11, p < .05$]. Paired comparisons revealed that the ratings derived from the peer reports were higher than both the ratings

from the self-reports and the laboratory ratings, while there was no significant difference between the latter [$t(199) = 2.21, p < .05$, for the least-different comparison]. This effect, however, did not significantly interact with the other factors. There were no significant effects for the sex of the subject factor or for any of the interactions (all $F_s < 1$). The mean ratings for the typical and maximal ratings of aggression by source of rating and by sex of the subject who was rated are presented in Table 1.

The intercorrelations between the maximal-typical variables are listed in Table 2. Since the analysis of variance showed a significant main effect for neither sex nor sex-related interactions, the data for female and male subjects were combined.

The analyses support the first hypothesis. Using a test for the difference between dependent correlations (Bruning & Kintz, 1977), ratings made from self-reports of maximal aggression outpredicted ratings derived from self-reports of typical aggression when the dependent variable was either typical aggression in the laboratory [$t(97) = 2.28, p < .05$] or maximal aggression in the laboratory [$t(97) = 6.19, p < .001$].

The second hypothesis was supported also. Peer reports of maximal aggression correlated more reliably with the two laboratory measures of aggression than did the peer reports of typical aggression [$t(97) = 2.03, p < .05$, when typical laboratory aggression was the dependent variable; $t(97) = 3.10, p < .01$, when maximal laboratory aggression was the dependent variable].

Given these results, it would be expected that the correlations between peer reports of maximal aggression and self-reports of either typical or maximal aggression would be higher than the correlations between these measures and peer reports of typical aggression. Unexpectedly, neither the typical nor the maximal peer reports correlated very highly with the subjects' self-reports of their typical aggression. The respective correlation coefficients were .18 and .17; the difference between these coefficients is, obviously, not significant. However, the expected relationship was evidenced with the subjects' self-reports of maximal aggression [$t(97) = 3.73, p < .01$].

An inspection of Table 2 shows that not only did the peer reports and self-reports of maximal aggression outpredict their respective typical reports, but the maximal reports more reliably predicted maximal expressions of aggression in the laboratory than they predicted typical expressions of aggression in the laboratory [$t(97) = 2.37, p < .05$, for peer data; $t(97) = 3.62, p < .001$, for self data]. Given these results, one would expect that the relationship between peer ratings of maximal aggression and self-ratings of maximal aggression would be stronger than the relationship between peer ratings of maximal aggression and self-ratings of typical aggression. It was [$t(97) = 4.48, p < .001$].

DISCUSSION

Consistent with previous findings (Turner, 1978; Willerman et al., 1976; Willerman et al., Note 1), the results of the present study suggest that self-report measures of aggression that are based on a maximal format are superior to measures that rely on a typical format when predicting either typical or maximal aggressive behaviors that are expressed in the laboratory. In addition, the results suggest that this superiority of the maximal format, in its prediction of laboratory behavior, holds for peer reports of a subject's aggressive behavior as well.

If the laboratory situation that was used in the study can be said to be analogous to a normal interpersonal situation, then both peer reports and self-reports about what level of aggression a person is maximally capable of expressing, as compared to what the person is typically capable of expressing, would be better predictors of the aggressive behaviors the person would normally express. In many interpersonal situations, however, it might be more important to be able to predict the maximum level of aggression that a person will express. The data from the study suggest that this maximal level of aggression would be best predicted by peer reports and self-reports of the person's maximal expression of aggressions as well.

Although the results of the study clearly support the use of peer reports and self-reports of maximal behaviors in predicting a person's actual behavior, the study did not experimentally address the question of why this phenomenon occurs.

One explanation as to why a maximal format is superior to a typical one is that self-reports and peer reports of maximal behavior are less ambiguous, require less averaging of behavior, are easier to rate, and force raters to deal with the intensity of the behavior rather than the frequency of the behavior (Willerman et al., Note 1).

Another explanation for the superiority of maximal self-reports and peer reports may be their increased reliability. In this and two other studies (Turner, 1978; Willerman et al.,

Table 1
Means and Standard Deviations of Aggression Ratings for the Source of Ratings, Sex of Subject, and the Total Sample

Variable	Males			Females			Total		
	Mean	SD	n	Mean	SD	n	Mean	SD	n
ML	3.02	1.2	31	2.70	1.1	69	2.80	1.1	100
TL	1.80	.7	31	1.78	.6	69	1.78	.6	100
MS	3.19	1.1	31	3.00	1.1	69	3.05	1.1	100
TS	1.78	.7	31	1.83	.8	69	1.81	.7	100
MP	3.38	1.1	38	3.17	1.2	62	3.22	1.2	100
TP	2.12	.9	38	1.98	.9	62	2.02	.9	100

Note—The differences between maximal (M) and typical (T) ratings are significant in every case ($p < .01$). L = laboratory, S = self, P = peer.

Table 2
Intercorrelations of Aggression Ratings for Maximal (M) and Typical (T) Variables

	TL	ML	TS	MS	TP
ML	.59†				
TS	.18*	.09			
MS	.45†	.69†	.14		
TP	.19*	.30**	.18*	.29**	
MP	.38†	.56†	.17*	.63†	.50†

Note—For all correlations, $n = 100$. L = laboratory, S = self, P = peer. * $p < .05$. ** $p < .01$. † $p < .001$.

1976), the absolute values of the reliability coefficients were higher for the maximal reports and maximal laboratory assessments than for the typical self-reports and typical laboratory assessments. Even though the differences between the coefficients were not significant, higher coefficients were found for the maximal reports. Additionally, in the present study, the absolute values of the test-retest reliability coefficients were higher for the maximal assessments. If this trend is valid, then it may be that subjects and peers are able to more validly report what they will do in the maximal situation, or it is possible that more intense aggressive behaviors are easier for subjects and peers to reliably rate. A less subjectively scored assessment procedure for rating aggressive behavior is currently being investigated that will hopefully eliminate the differences in scorer reliability so this question can be more adequately addressed.

Whatever the explanation may be for this phenomenon, it is quite clear that the maximal format for peer reports and self-reports more reliably predicts behavior than does the more commonly used typical format. Because of this, the phenomenon warrants careful consideration and further exploration.

REFERENCE NOTE

1. Willerman, L., Rosen, J., Shoup, R., & Malina, R. M. *A comparison of maximal and typical performance measures in the prediction of self-reported delinquency*. Unpublished manuscript, 1978. (Available from Lee Willerman, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.)

REFERENCES

- ALKER, H. A. Is personality situationally specific or intrapsychically consistent? *Journal of Personality*, 1972, **40**, 1-16.
- BOWERS, K. Situationism in psychology: An analysis and a critique. *Psychological Review*, 1973, **80**, 307-336.
- BRUNING, J. L., & KINTZ, B. L. *Computational handbook of statistics*. Glenview, Ill: Scott, Foresman, 1977.
- CRONBACH, L. Processes affecting scores on "understanding others" and "assumed similarity." *Psychological Bulletin*, 1955, **52**, 177-193.
- CRONBACH, L. *Essentials of psychological testing*. New York: Harper & Row, 1970.
- FENIGSTEIN, A., SCHEIER, M., & BUSS, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 1975, **43**, 522-527.
- MISCHEL, W. *Personality and assessment*. New York: Wiley, 1968.
- MISCHEL, W. *Introduction to personality*. New York: Holt, Rinehart, & Winston, 1971.
- MISCHEL, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, **80**, 252-282.
- NORMAN, W. "To see ourselves as others see us!": Relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. *Multivariate Behavioral Research*, 1969, **4**, 417-443.
- NUNNALLY, J. C. *Psychometric theory*. New York: McGraw-Hill, 1978.
- TURNER, R. G. Consistency, self-consciousness, and the predictive validity of typical and maximal personality measures. *Journal of Research in Personality*, 1978, **12**, 117-132.
- WILLERMAN, L., TURNER, R., & PETERSON, M. A comparison of the predictive validity of typical and maximal personality measures. *Journal of Research in Personality*, 1976, **10**, 482-492.

(Received for publication March 26, 1979.)