

*A set of ratings by others was obtained in conjunction with a collection of self-report data on subjective quality of life. Respondents and two or three "relevant others" nominated by them rated the respondents' satisfaction with aspects of their life. The correlations between respondent-respondent, respondent-other, and other-other ratings provide estimates of the convergent and discriminant validity of the self-ratings and ratings by others. The results of these analyses provide evidence for the external validity of self-ratings in the quality of life area. More importantly, they illustrate the possibility of using ratings by other as a validation criterion for self-ratings in general. The validity of ratings by others who vary in their relationship with their respondents is also explored, and suggestions for further uses of ratings by others are provided.*

## VALIDATION OF SELF-REPORT MEASURES USING RATINGS BY OTHERS

RICK CRANDALL  
*University of Michigan*

**f**ew objective criteria exist against which to measure the validity of self-report measures. Using self-report measures of quality of life as an example, this paper describes the application of one approach (ratings by others) which can be easily used as a validation criterion for self-report measures.

In order to measure subjective aspects of social change, the basic reliability and validity of self-report measures need to be demonstrated. Measurement in this area has tended to rely on face validity. Since there are many sources of bias in self-reports (e.g., Crowne and Marlowe, 1964), it is important to demonstrate external validity for self-reports even when face validity is very high. To date, however, research which has validated self-reports against a criterion has been restricted to a few areas

*AUTHOR'S NOTE: Acknowledgments of support for this research and the author's current address are given in Note 1.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 4 No. 3, February 1976  
© 1976 Sage Publications, Inc.

where objective information was available (e.g., Kain & Quigley, 1972; Kish & Lansing, 1954; Weaver & Swanson, 1974).

It is particularly difficult to find objective criteria (i.e., behaviors) against which to validate self-report measures of quality of life. Objective measures of quality of life consist of both demographic data and descriptions of physical aspects of life. Race, income, number of bathrooms, crime rates, and population density are examples of such objective measures (Executive Office of the President, 1973). Yet these objective measures cannot have the relevance and power of direct reporting of feelings about life conditions.

For self-ratings of life quality such as satisfaction, relevant objective criteria are not available. While actual salary may relate to satisfaction with salary, there could be many reasons why a person making \$10,000 per year might be more satisfied with his income than a person making \$20,000 per year.

Ratings by others who know respondents well, often called peer-ratings, have proven to be useful as a criterion for the measurement of personality traits (e.g., Norman, 1963; Hamilton, 1971) and for the prediction of task success in military and industrial settings (e.g., Amir et al., 1970; Hollander, 1954; Mayfield, 1972; Mouton et al., 1955). In the past, work with ratings by others has proceeded relatively independently in the personality, military, and industrial areas. However, the three complement and confirm each other and there is no reason that with appropriate modifications, the techniques for peer-ratings developed in these areas should not be applicable to other self-report areas such as quality of life. Relevant others (relatives, acquaintances, friends) represent an untapped source of validity information. Others might be expected to have knowledge of both a respondent's concrete life circumstances and subjective attitudes. They can form their judgements through independent observation of the conditions affecting the respondent and through direct and indirect indicants of the respondent's attitudes and responses to objective conditions.

## METHOD

Respondents were 222 adults living in the Toledo, Ohio area. They were solicited by an existing interview staff in Toledo. The staff was allowed to obtain acquaintances and up to two members from the same family. However, they also had to fill quotas for age, sex and race. The group obtained was not random; however, it had considerable heterogeneity and was not too unlike a 1972 probability sample of Toledo collected as part of another project.<sup>2</sup> Table 1 summarizes the data on the two groups. The largest difference between the groups reflects the fact that this group had somewhat more education than the probability sample.

The questionnaires were administered to groups in several sessions at centrally located churches during July, 1973. Respondents were paid \$25 to answer a 640-item questionnaire about their subjective quality of life which took from one and one-half hours to four hours to finish. The questionnaire included the respondents' ratings of their feelings about their: house, job, family, neighborhood, spare time, national government, fun, standard of living, freedom and independence, beauty and attractiveness, freedom from bother, safety, accomplishments, acceptance by others, and life as a whole.<sup>3</sup> Similar forms of some items were asked (i.e., "How do you feel about your job?" and "How do you feel about the work you do on your job—the work itself?").<sup>4</sup> At the end of the questionnaire, respondents were asked to provide three names and phone numbers of persons who knew them well enough to make the ratings. A brief explanation given to respondents was:

We would like to contact two or three people who know you pretty well and ask them about a dozen short questions . . . as to how *you* feel about . . . your house or apartment . . . neighborhood . . . freedom from annoyance, and the like.

Some people were unable to list complete addresses at this time, and when necessary, respondents were later sent a note asking

TABLE 1  
Demographics of Toledo Respondents:  
1972 Probability Sample and 1973 Sample

Characteristic	Present Sample (1973)*	Probability Sample (1972)**
Age	$\bar{X}=41.48$	$\bar{X}=45.52$
Sex		
Male	44.1%	39.4%
Female	55.9	60.6
Race		
Black	12.4%	13.1%
White	86.4	86.0
Other	1.2	0.9
Marital Status		
Married	70.3%	63.4%
Separated	1.8	2.9
Divorced	4.1	9.5
Widowed	6.8	12.4
Never married	16.9	11.8
Education		
8th grade or less	4.2%	14.3%
Some high school	11.4	22.5
Completed high school	30.6	35.3
Some college	25.6	18.6
Completed college	28.3	9.3
Job Status		
Employed	52.3%	60.1%
Housewife	18.6	23.0%
Retired	9.5	11.0
Student	5.5	2.3
Other or no response	14.1	3.6

\*N = 222  
\*\*N = 346

them for more information. Sixteen respondents (7%) listed no useable raters.

Potential raters were contacted by mail and the purpose of the study was explained in a cover letter. Confidentiality was stressed. A total of 549 potential raters with apparently usable addresses were contacted. Of these, 318 were contacted once, 222 twice, and 9 three times. A third request was sent only if none of the potential raters had responded for a given respondent.

Raters were asked to estimate how a respondent felt about 14 areas of life and about life as a whole. With appropriate rewording, the raters were given the same items and 7-point scale<sup>5</sup> that the respondents had completed. The scale has been used extensively in several national surveys. It is reproduced in Andrews and Withey (1974) and appears to be the most reliable and valid of several scales tested.

#### ANALYSES

The validity of both the self-ratings and ratings by others can be examined by intercorrelating them in several ways. The obtained correlations provide approximations of the values necessary for the convergent-discriminant matrix proposed by Campbell and Fiske (1959). Overall there are two sources of ratings, or methods: self-ratings and ratings by others. There are fifteen areas of life, or traits, which are rated by respondents and others who know them. For some items answered by respondents there are similar forms which can be used as a conservative estimate of the validity of the respondent's ratings. For most respondents there are two ratings by others who know them which can be used to estimate the validity of the ratings by others.

The basic approach is fairly straightforward. The correlations between the self-ratings and the ratings by others on a scale measuring one area of quality of life represent convergent validity—the extent to which two methods measure the same

underlying construct. The correlations between self-ratings on one scale and ratings by others on other scales represent discriminant validity. For self-ratings, the convergent correlations give one kind of reliability estimate. The discriminant correlations represent the effect of method variance if the rated variables are conceptually independent. (For these data the discriminant correlations will be inflated because the variables are conceptually related.) For the ratings by others, the convergent correlations reflect the agreement among raters and the discriminant correlations represent the methods effect. For each set of ratings the convergent correlations should be higher than the discriminant correlations. Ideally the convergent values from each set of ratings should also be higher than the discriminant values of the other sets. Similar patterns of discriminant correlations for both measurement methods help confirm that both methods are measuring the same construct. In summary, the correlations of self-ratings with the average rating by others on each of 15 parallel items represent the heteromethod homotrait matrices.<sup>6</sup> The intercorrelations of the eight pairs of similar self-ratings represent one monomethod heterotrait matrix. The intercorrelations of the 15 ratings from two independent raters for each respondent represent the other monomethod heterotrait matrix.

The intercorrelations of both the self-ratings and the ratings by others provide conservative estimates of convergent validity compared to those derived from the heteromethod matrix. For the self-ratings this is due to the use of nonidentical scales. For the ratings by others this is because the "rating by others" score used for the correlations with the self-ratings consisted of an average of 2.3 relevant others. This provided a more stable score than that estimated from the intercorrelations of two *single* relevant others used to estimate the convergent validity of the other's ratings. The discriminant validity estimates are also conservative because most of the "traits" are conceptually related.

## RESULTS

Eighty-four percent (461) of the potential raters returned usable forms. This is a much better response rate than those normally elicited by general surveys of the population, including cases where the respondents have had previous contact with an organization (Blumberg et al., 1974). Direct reference to their friend is probably responsible for the high response rate. The obtained responses provided: one or more ratings for over 90% of the original sample of respondents: one rater for 32 respondents, two raters for 75 respondents, and three raters for 93 respondents. There were 22 respondent for whom no ratings were obtained. Of these, 16 had listed no usable raters.

Figure 1 summarizes the intercorrelations<sup>7</sup> between the self-ratings and ratings by others on the fifteen scales. The self-other intercorrelations generally show good convergent validity, as can be seen in Table 2. The average correlation in the convergent diagonal is .33. The average discriminant correlations are only .13. As might be expected, the greatest convergence between self-ratings and ratings by others tend to be on aspects of life such as house, job, and neighborhood. Raters might be expected to have some direct knowledge of these areas. Convergence is also reached on rather abstract ratings such as "How he/she feels about the amount of beauty in his/her world" ( $r = .25$ ), and "How he/she feels about life as a whole" ( $r = .37$ ). With the exception of "Freedom from bother and annoyance," each item demonstrates significant convergent validity and almost all show significant discriminant validity as well.<sup>8</sup>

The self-ratings and ratings by others matrices also show significant convergent and discriminant validity. The self-ratings show considerably higher convergence than the ratings by others and are essentially conservative short-term test-retest measures. Interestingly, the ratio of  $r^2$  of the convergent correlation compared to  $r^2$  of the discriminant correlations is very similar in all three submatrices (self-others, .18; self-self, .18; others-others, .21). The only weakness in the overall picture is created by the average self-rating discriminant correlation of

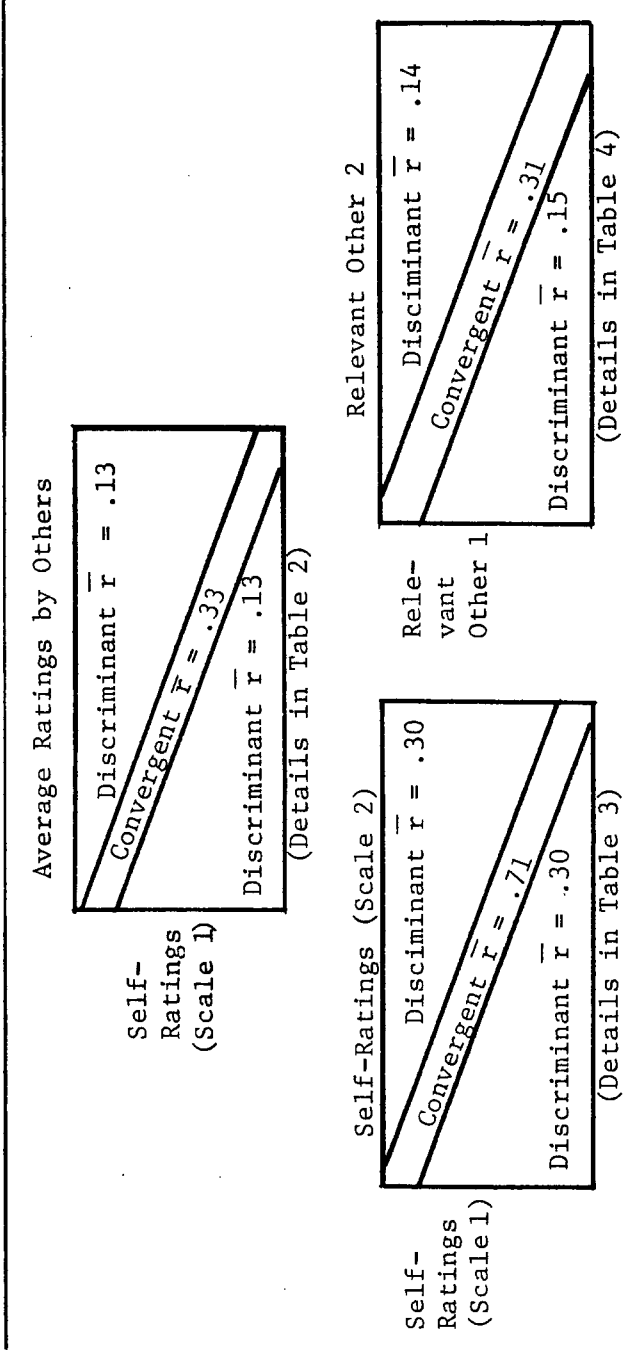


Figure 1: Summary of Average Convergent and Discriminant Validity of Self-Ratings and Ratings by Others for Measures of Life Satisfaction



TABLE 2  
Intercorrelations of Self-Ratings and Ratings by Others on Aspects of Life Satisfaction\*

Self-ratings	Ratings by Others																Average discriminant value	Median discriminant value
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O			
A House	.37	.07	.09	.36	.17	.09	.15	.29	.06	.09	.09	.05	-.07	.08	.08	.13	.09	
B Job	.11	.54	.09	.13	.17	.21	.11	.15	.09	.30	.18	.11	.43	.23	.16	.18	.16	
C Family	.01	.00	.35	-.04	-.10	-.08	.08	.08	.10	.05	.10	.05	-.04	.00	.14	.06	.07	
D Neighborhood	.24	.15	.11	.54	.09	.19	.15	.17	-.06	.11	.18	.15	-.04	.11	.09	.13	.13	
E Spare Time	-.00	-.09	.13	.02	.30	.08	.25	.11	.19	.24	.15	.13	.02	.11	.18	.12	.12	
F Nat'l. Gov't.	.05	-.10	-.04	.17	.07	.32	.13	.06	-.05	.07	.14	.15	-.15	-.06	-.02	.09	.07	
G Fun	.10	.01	.12	.10	.26	.24	.34	.21	.22	.29	.20	.20	.02	.10	.27	.17	.19	
H Std. of Living	.23	.08	.16	.30	.28	.02	.25	.33	.17	.20	.18	.08	.03	.10	.17	.16	.17	
I Freedom	.15	.09	.08	.07	.20	.09	.25	.21	.27	.23	.19	.06	.12	.05	.15	.14	.14	
J Beauty	.09	.08	.08	.06	.25	.08	.21	.18	.13	.25	.15	-.01	.08	.24	.19	.13	.11	
K Free from bother	.03	.03	-.03	.13	.12	.12	.09	.05	.12	.10	.09	.08	-.07	-.02	.05	.07	.07	
L Safety	.05	.07	.07	.33	.19	-.10	.22	.12	.01	.22	.19	.33	.02	.08	.19	.13	.11	
M Accomplishments	.14	.15	.14	.05	.26	.14	.23	.16	.18	.29	.22	.09	.36	.28	.32	.19	.17	
N Acct. by Others	.00	.13	.07	-.00	.14	.01	.13	.01	.06	.25	.19	.07	.08	.25	.16	.09	.08	
N <sub>1</sub> Acct. by Others	.03	.04	.02	-.04	.14	.06	.06	.04	.05	.12	.14	.06	.07	.24	.16	.07	.06	
Q Life as a whole	.09	.06	.27	.06	.18	.11	.27	.25	.29	.29	.23	.17	.19	.21	.37	.19	.20	
Average discriminant value**	.09	.08	.10	.13	.18	.10	.17	.14	.12	.19	.17	.10	.10	.11	.16	.13	.14	
Median discriminant value	.09	.08	.09	.07	.17	.09	.15	.15	.10	.22	.18	.08	.07	.10	.16	.12	.10	

\* See Footnote 3 for further elaboration of these aspects.

\*\* On this and subsequent tables averages are done using Z scores. As a conservative measure the absolute values of the correlations are used for the averages.

TABLE 3  
Intercorrelations of Self-Ratings (Form<sub>1</sub>) and Self-Ratings (Form<sub>2</sub>)  
on Aspects of Life Satisfaction\*

Self-ratings (Form <sub>1</sub> )	Self-Ratings Form <sub>2</sub>										Average discriminant value	Median discriminant value
	A	B	C	D	F	G	J	O				
A House	.80	.31	.29	.44	.20	.37	.38	.32	.33			
B Job	.21	.74	-.09	.30	.16	.27	.28	.32	.24			
C Family	.17	.13	.63	.04	.07	.31	.14	.40	.18			
D Neighborhood	.56	.27	.06	.60	.22	.30	.36	.27	.30			
F Nat'l. Gov't.	.20	.12	.11	.29	.73	.30	.22	.21	.21			
G Fun	.32	.33	.35	.38	.34	.83	.57	.62	.42			
J Beauty	.32	.33	.24	.39	.24	.51	.67	.52	.37			
O Life as a whole	.33	.27	.54	.22	.14	.46	.46	.64	.35			
Average discriminant value	.31	.25	.25	.30	.20	.36	.35	.39	.30	.30		
Median discriminant value	.32	.27	.24	.30	.20	.31	.36	.32	.32	.30		

\*See Footnote 3 for further elaboration of these aspects.

TABLE 4  
Intercorrelations of Ratings by Relevant Other 1 and Relevant Other 2  
on Aspects of Life Satisfaction\*

Ratings by Relevant Other 1	Ratings by Relevant Other 2															Average discriminant value	Median discriminant value
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
A House	.43	.33	.06	.35	.14	.27	.32	.37	.29	.06	.12	-.02	.15	.03	.07	.18	.15
B Job	.05	.45	-.04	-.01	-.13	-.06	-.02	.11	.02	-.10	-.01	.01	.22	.01	.06	.06	.05
C Family	.32	.29	.43	.28	-.04	.10	.10	.24	.19	.02	.17	.11	.29	.18	.27	.19	.19
D Neighborhood	.14	.13	-.01	.39	.06	.03	.11	.18	.09	-.05	.16	.20	-.06	.02	.01	.09	.08
E Spare Time	.04	.24	.20	.02	.22	.08	.21	.35	.24	.08	.07	.09	.21	.19	.22	.16	.20
F National Gov't.	.07	.15	.09	.14	.17	.24	.13	.08	.04	.08	.07	-.02	.10	-.02	-.04	.09	.08
G Fun	.19	.34	.32	.06	.08	.10	.27	.39	.34	.13	.14	.05	.35	.19	.29	.21	.19
H Stnd. of Living	.19	.20	.05	.22	.03	.07	.13	.42	.23	.07	-.03	.05	.19	-.04	.12	.12	.10
I Freedom	.09	.26	.26	.05	.05	.05	.13	.30	.14	.12	.03	.14	.29	.03	.27	.15	.13
J Beauty	.12	.27	.16	.22	.14	.12	.09	.10	.17	.09	.09	-.05	.28	.20	.29	.16	.15
K Free from Bother	.18	.26	.20	.17	.02	.03	.14	.21	.11	.04	.18	-.00	.23	.09	.20	.13	.14
L Safety	.12	.14	.08	.10	-.05	.13	.11	.13	.02	.12	.07	.37	-.01	.13	.13	.10	.12
M Accomplishments	.12	.43	.19	.02	.05	.08	.16	.26	.24	.12	.09	.15	.33	.21	.34	.18	.16
N Accpt. by Others	.16	.12	.08	.05	.07	.07	.11	.16	.13	-.05	.13	.11	.12	.21	.22	.11	.12
O Life as a Whole	.26	.35	.24	.25	.12	.23	.18	.27	.22	.09	.42	.16	.25	.27	.36	.24	.25
Average discriminant value	.15	.25	.14	.14	.08	.10	.14	.23	.17	.08	.11	.08	.20	.12	.18	.14	.14
Median discriminant value	.15	.26	.13	.12	.07	.08	.13	.23	.18	.08	.09	.07	.22	.11	.21	.16	.13

\*See Footnote 3 for further elaboration of these aspects.

.30. This value is high compared to the convergent values of the other-other and self-other matrices of .31 and .33. It should be remembered that for the reasons previously noted, each component in these comparisons is a conservative estimate of the true value.

Because of the lack of previous data in this area, the effects of sex of respondents and of the number of relevant others making ratings were investigated. Columns A and B of Table 5 show the convergence of self-ratings and ratings by others for men and women. The average convergence is identical. The largest difference is on the acceptance-by-others items, but no general pattern is evident.

The number of relevant others to be used as a validity criterion for the self-ratings is an important practical problem. As the number of raters increases, the stability of the index increases, but the cost of collection also increases. It was expected that the level of convergence between self-ratings and ratings by others would increase as relevant others were added to the rater index. This must be true if each set of other's ratings is equal in "true" variance while the error variances are independent. Theoretically, added raters will asymptotically increase reliability. However, error variances of ratings by others are probably not independent, so after some point added raters make a negative contribution. By comparing the convergence with self-ratings of ratings by one, two, and three others, an estimate of the value of collecting further ratings can be obtained.

Columns C, D, and E of Table 5 show the unexpected result of this analysis. The group for whom only one rater was available showed somewhat greater average convergence than the two- and three-rater groups. The effect is not large and may be random. However, if people who only provided one rater also provided a different type of rater, this difference could suggest how to maximize the investment when collecting ratings by others by focusing on certain people. As a first check to confirm that the convergent values from the "only one rater" group are unusual, convergent values were computed singly for

the first, second, and third raters provided by each respondent. When the group of 32 cases for whom only one rating was available (Col. C) is excluded, the average convergence from these three other one-rater sets (columns F, G, and H of Table 5) is slightly lower than that from two raters. This result had been expected in the "only one rater" group and suggests that the raters in that group were different.

Only a limited amount of information was available about the raters. From their names and addresses, two factors distinguished the raters of respondents for whom only one rating was available from the raters provided by other respondents. Twenty-six percent of the raters had the same last name as the respondent, and 26% lived on the same street or even at the same address. The comparative figures for the remaining respondents were 19% same last name, and 15% same street. Neighbors and relatives may know the respondents better than other raters, which should increase the accuracy of their ratings (see, Hjelle, 1968). In order to test whether these differences could be systematically changing the convergence level for the "only one rater" group, two new analyses were run. Using the group of raters for respondents who provided two or more raters, all raters with the same name as their respondents were run for the 15 items. A comparable analysis was done for all neighbors who rated other respondents. Columns I and J of Table 5 show the results of these analyses. Relatives show greater convergence with respondents than an average single rater. These results could account for the higher self-other convergence of the one-rater-only group. The only item for which neighbors appear to be better than average raters is neighborhood.

#### DISCUSSION

The convergence between self-ratings and ratings by others of the respondents' quality of life presents a hopeful picture. While the values are not as high as we would like, the level of conver-

TABLE 5  
Convergence Between Ratings by Others and Self-Ratings for Different Groups\*

	Column: A		B	C	D	E	F	G	H	I	J
	Men	Women	Only one rater	Two raters	Three raters	First rater**	Second rater**	Third rater**	Relatives**	Neighbors**	
A House	.28	.42	.38	.29	.48	.28	.43	.24	.61	.10	
B Job	.55	.52	.48	.28	.66	.49	.39	.45	.64	.14	
C Family	.24	.41	.41	.32	.35	.30	.30	.20	.46	.14	
D Neighborhood	.47	.59	.47	.47	.59	.45	.37	.51	.31	.53	
E Spare Time	.29	.29	.52	.13	.28	.26	.02	.28	.40	.11	
F Nat'l. Gov't.	.32	.32	.46	.31	.18	.26	.20	.12	.10	.32	
G Fun	.40	.31	.49	.31	.21	.27	.12	.23	.42	.28	
H Stnd. of Living	.39	.28	.49	.13	.34	.23	.26	.16	.24	.14	
I Freedom	.15	.32	.23	.12	.42	.26	.11	.31	.40	.29	
J Beauty	.30	.20	-.06	.39	.33	.35	.18	.17	.25	.12	
K Free from Bother	.09	.08	.12	.02	.11	.05	.08	.01	.12	.18	
L Safety	.32	.32	.18	.37	.38	.38	.25	.24	.46	.44	
M Accomplishments	.28	.43	.39	.32	.35	.23	.26	.31	.45	.46	
N Acct. by Others	.37	.17	.37	.16	.20	.17	.15	.10	.21	.01	
N <sub>1</sub> Acct. by Others	.38	.16	.29	.13	.23	.15	.16	.10	--	--	
O Life as a Whole	.32	.39	.46	.34	.34	.27	.25	.26	.22	.22	
Mean r	.33	.33	.37	.26	.35	.28	.22	.24	.36	.24	

\* See Footnote 3 for further elaboration of the items.

\*\* Excluding respondents for whom only one peer was available.

gence is higher than that typically obtained with other criteria such as demographic characteristics. For instance, self-reported income, sex, race, age, education, and stage of family cycle are related to self-reported satisfaction with life (using etas) from .03 to .20 with a median predictability of only .09 (Andrews and Withey, 1974: 20). In this analysis peer ratings of aspects of life related to self-reported satisfaction with life using the more conservative  $r$  from .06 to .37 with a median predictability of .21. Even more telling, self-reported value of home correlated with satisfaction with home  $r = .39$  and self-reported family income correlated with satisfaction with family income  $r = .13$ . In contrast to these correlations which are from the same data source, the convergence obtained here between self-ratings and ratings by others as an independent data source looks large indeed. This convergence gives us evidence that respondent self-ratings are not merely arbitrary ratings produced for our benefit, but measure attitudes and feelings of which important others in the respondent's environment are aware. The ratings also exhibit discriminant validity. Since many of the rated areas of satisfaction should show positive intercorrelations, the obtained estimates of discriminant validity are conservative.

Several possible problems could exist in the ratings by others. Contamination could occur. In our instructions to raters we emphasized the importance of their opinion and of confidentiality. Since the rater's form was mailed separately and neither the rater nor respondent benefits from biasing the answers it probably did not occur. Another factor suggesting that this could have been, at best, a minor factor is the great variability in the levels of the convergent correlations. They ranged from under .10 to over .50. Any amount of direct cheating would have raised the lowest correlation above that level.

Another more subtle problem is projection by the raters. If they are experiencing some of the same life conditions as their respondent, then reports of their own feelings could inflate their accuracy. The best way to rule this out is to collect raters' perception of *their own feelings* and compare these to

their estimates for others as a measure of discriminant validity. This would be a valuable addition to future methodological studies. To some extent raters' projection could also act as a main effect which would not influence the level of the correlations.

There are several ways to test the value of the ratings by others and to collect further data. For instance, increasing the reliability of the ratings can improve the convergence with self-ratings. This study suggested that three raters are better than two or one. Further increases in the number of raters could prove useful though expensive. Since one name relative (often a spouse) was as good as three unselected raters, several selected raters may greatly increase the levels of convergence. The correlations with self-ratings are also limited by the unreliability of the self-ratings. For a few items multiple self-ratings were available. Summing the self-ratings of satisfaction with house on four different scales increased the correlation with the ratings by others to .47 from the .37 obtained with only one self-rating. In a parallel analysis for life as a whole, no increase was obtained.

Other evidence that the ratings by others and the self-ratings may be more valid than would be estimated by their convergence with respondent ratings comes from the fact that the two groups are not really rating the same thing. The addition of respondents' ratings of how they think others will rate them and more concrete explanations focusing others' rating on how they think respondents will answer (possibly taking into account any distortions the respondent may provide) would allow greater convergence between self- and other-ratings. This may seem unnecessarily elaborate; however, in the personality measurement area, Norman (1969) obtained self-ratings, peer-ratings, and in addition, ratings by respondents of how they thought others would rate them. Using factor scores, he found that self-ratings and "self as others see you" ratings had an average correlation of .81 on five-trait scales. While self-ratings correlated an average of .41 with peer-ratings, "self as others see you" ratings



correlated an average of .50 with peer-ratings for a sample of Peace Corp trainees. These findings strongly suggest that people are aware that others will rate them differently than they rate themselves and they can predict the direction of the discrepancies. Similarly, the raters may be able to discriminate between how the respondent truly feels and how he/she will answer the question. For further discussions of approaches to estimating rater bias and how to collect ratings by others, see Klimoski and London (1974), Kaufman and Johnson (1974), and Norman and Goldberg (1966).

While the technique of ratings by others has not been used in national surveys, its application should present no problems. Respondents would provide the names and addresses of appropriate raters who could then be contacted by mail. In home interviews, the problem of occasional incomplete addresses, which was met here, should disappear. Other variations may increase the already excellent response rate. For instance respondents could be asked to address and sign a standardized cover letter asking their raters to cooperate.

Interviewers are another source of data easily available to survey researchers. At the end of one national survey sample of about 1400, collected in May 1973,<sup>9</sup> interviewers rated the respondent's satisfaction with life. The correlation of the interviewers' ratings with the average of two similar self-ratings by the respondent was .57. While memory of respondents' previous answers cannot be ruled out in this case, ratings could also be obtained before the respondents answered relevant questions. Such first impressions may be more informative than might be expected. Gormly and Edelberg (1974) found that group ratings by undergraduate judges of people's aggressiveness, based only on seeing two people enter a room and sit down, significantly predicted the ratings of a group of peers who lived with the people in fraternities. Ratings further increased in predictiveness after observing the people in dyadic discussions. Similarly, any interaction between interviewer and respondent should further increase the validity of interviewer ratings.

Further uses of ratings by others should encourage respondents to provide relatives as raters and attempt to specify other

groups of raters such as best friends who might also be more knowledgeable raters. The relationship between the respondent and raters should be noted (i.e., neighbor, co-worker, relative, and the like), and other characteristics of the raters besides their relationship to the respondent (e.g., sex) should be investigated. Ratings might also be improved by greater specification of question content and relevant comparison groups. Respondents could be encouraged to suggest raters who best know their state of mind about the areas of greatest interest to the investigators and who would be cooperative.

Another way to approach this type of data involves confirmatory factor analysis or structural equation models such as path analysis (Alwin, 1974). The intent of this paper was to discuss the value of rating by others as a validity criterion rather than to explore different modes of analysis for these ratings. However, estimates using these methods were gathered as part of another study focusing on different methods of self-ratings (Andrews and Crandall, 1976). Joreskog's (1973) computer program was used to derive estimates of validity, and methods effects from structural equation models. The results of those analyses suggest that using these procedures, about 15% of the variance of others' ratings accurately measures how the individuals really feel, and about 30% of the variance of others' ratings is methods effects such as bias or false opinions. The majority of the variance cannot be accounted for. Using some of the procedures discussed here the accountable valid variance of ratings by others could probably be increased.

Self-reports often have a face validity which cannot be shared by any other source of data. Yet it is important to use external criterion to assess the meaning of self-reports. The results of this study suggest that ratings by others can be used as an external criterion for the validation of self-report measures such as those used in the quality of life area. Variations are discussed which could allow further work to proceed without great cost and could add further support to the validity of both the method of ratings by others and of related self-reports. Further work needs to be done to determine the cost of collecting

ratings by others compared to their practical value in improving item and scale construction in various self-report areas.

### NOTES

1. The author is now at the Department of Leisure Studies at the University of Illinois, Champaign, Illinois 61820. Data collection was supported by NSF grant GS3322 to Drs. Frank M. Andrews and Stephen B. Whitey. Their help and that of Marita DiLorenzi, Elizabeth Taylor, and William Murphy is appreciated. Thanks to Frank Andrews for detailed comments on early version of the manuscript. Thanks to Carolynn Crandall and Manuel London for comments on a later version. Further information on the larger project is available from Drs. Andrews and Withey, Institute for Social Research, University of Michigan.

2. These data were collected by Drs. Withey, Kline and Robinson, Institute for Social Research, University of Michigan.

3. The specific questions were:

How do you feel about . . .

- A. Your *house* or apartment?
- B. Your *job*?
- C. Your own *family* life—Your husband or wife, your marriage, and your children, if any?
- D. Your particular *neighborhood* as a place to live?
- E. The way you spend your *spare time*, your non-working activities?
- F. What our *national government* is doing?
- G. How much *fun* you are having?
- H. Your *standard of living*--the things you have like housing, car, furniture, recreation, and the like?
- I. Your independence or *freedom*--the chance you have to do what you want?
- J. The amount of *beauty* and attractiveness in your day to day life?
- K. The *freedom* you have *from* being *bothered* and annoyed?
- L. Your *safety*?
- M. What you are *accomplishing* in your life?
- N. How much you are *accepted* and included *by others*?
- N<sub>1</sub>. How much you are *accepted* and included *by others*?
- O. Your life as a whole, and How do you feel about your life as a whole?

4. Below are the parallel forms of the items listed in Note 3. Each item is listed next to the letter of the item in Note 3 which it parallels.

- A. Check the box next to the words which best describe how you feel about your house or apartment?
- B. How do you feel about the work you do on your job--the work itself?
- C. How do you feel about . . . your marriage?
- D. How do you feel about . . . this community as a place to live?

- F. How do you feel about . . . what our national government is doing about the economy—jobs, prices, profits?
- G. How do you feel about . . . the amount of fun and enjoyment you have?
- J. How do you feel about . . . the chance you have to enjoy pleasant or beautiful things?
- O. (Both items are listed in Note 3.)

Because these pairs of items are not identical, the correlations between them provide a conservative estimate of the convergent validity of the self-ratings. Two identical self-ratings of feelings about acceptance by others were included as another reliability check. The life as a whole self-rating was an average of two self-ratings.

5. The categories are: delighted, pleased, mostly satisfied, mixed (about equally satisfied and dissatisfied), mostly dissatisfied, unhappy, and terrible. Respondents could also answer: neutral (neither satisfied nor dissatisfied), I never thought about it, or does not apply to me.

6. There were 168 cases where two sets of peer ratings were available for the same respondent. For individual rating scales there were from 112 to 144 pairs of ratings available.

7. All averages are computed using z scores.

8. Correlations are based on from 135 to 200 observations. With samples of this size, correlations of .22 are significant at the .01 level and differences between correlations of about .17 (i.e., difference between convergent and discriminant levels) are significant at the .05 level.

9. These data were collected by the ISR project headed by Drs. Frank M. Andrews and Stephen B. Withey.

## REFERENCES

- ALWIN, D. (1974) "Approaches to the interpretation of relationships in the multi-trait-multimethod matrix," in H. Costner (ed.) *Sociological Methodology 1973-1974*. San Francisco: Jossey-Bass.
- AMIR, Y., Y. KOVARSKY, and S. SHARAN (1970) "Peer nominations as a predictor of multistage promotions in a ramified organization." *J. of Applied Psychology* 54: 462-469.
- ANDREWS, F. M. and R. CRANALL (1976) "The validity of measures of self-reported well-being." *Social Indicators Research*.
- ANDREWS, F. M. and S. B. WITHEY (1974) "Developing measures of perceived life quality: results from several national surveys." *Social Indicators Research* 1: 1-26.
- BLUMBERG, H., C. FULLER, and A. P. HARE (1974) "Response rates in postal surveys." *Public Opinion Q.* 38: 113-123.
- CAMPBELL, D. T. and D. W. FISKE (1959) "Convergent and discriminant validation by the multitrait-multimethod matrix." *Psych. Bull.* 56: 81-105.
- CROWNE, D. and D. MARLOWE (1964) *The Approval Motive*. New York: Wiley.
- Executive Office of the President: Office of Management and Budget (1973) *Social Indicators, 1973*. Washington, D.C.: Government Printing Office.

- GORMLY, J. and W. EDELBERG (1974) "Validity in personality trait attribution." *Amer. Psychologist* 29: 189-193.
- HAMILTON, D. (1971) "A comparative study of five methods of assessing self-esteem, dominance, and dogmatism." *Educ. and Psych. Measurement* 31: 441-452.
- HJELLE, L. A. (1968) "Accuracy of Personality and Social Judgments as Functions of Familiarity." *Psych. Reports* 22: 311-319.
- HOLLANDER, E. P. (1954) "Buddy ratings: military research and industry implications." *Personnel Psychology* 7: 385-393.
- JORESOKG, K. G. (1973) "A general method for estimating a linear structural equation system," in A. S. Goldberger and O. D. Duncan (eds.) *Structural Equation Models in the Social Sciences*. New York: Siminar.
- KAIN, J. and J. QUIGLEY (1972) "Note on owner's estimate of housing value." *J. of the Amer. Stat. Assn.* 67: 803-806.
- KAUFMAN, G. G. and J. C. JOHNSON (1974) "Scaling peer-ratings: an examination of the differential validities of positive and negative nominations." *J. of Applied Psychology* 59: 302-306.
- KISH, L. and J. LANSING (1954) "Response errors in estimating the value of homes." *J. of the Amer. Stat. Assn.* 49: 520-538.
- KLIMOSKI, R. J. and M. LONDON (1974) "Role of the rater in performance appraisal." *J. of Applied Psychology* 59: 445-451.
- MAYFIELD, E. C. (1972) "Value of peer nominations in predicting life insurance sales performance." *J. of Applied Psychology* 56: 319-323.
- MOUTON, J. S., R. R. BLAKE, and B. FRUCHTER (1955) "The validity of sociometric responses." *Sociometry* 18: 181-206.
- NORMAN, W. T. (1969) "'To see ourselves as others sees us!': relations among self-perceptions, peer-perceptions, and expected peer-perceptions of personality attributes." *Multivariate Behavioral Research* 4: 417-443.
- (1963) "Toward an adequate taxonomy of personality attributes: replicated factor structure in peer nomination personality ratings." *J. of Abnormal and Social Psychology* 66: 574-583.
- and L. R. GOLDBERG "Raters, ratees, and randomness in personality structure." *J. of Personality and Social Psychology* 4: 681-691.
- WEAVER, C. and C. SWANSON (1974) "Validity of reported date of birth, salary and seniority." *Public Opinion Q.* 38: 69-80.