The neglect of the consequences of the Extreme Response Style (ERS) phenomenon in cross-cultural research is documented through a selective literature review. Consideration of ERS differences permitted derivation of three logical consequences that pertain to (a) differences in group means, (b) the internal consistency of measures and the correlation among measures, and (c) results of dimensional analyses. Comparison of U.S. and Korean student samples on a 130-item questionnaire revealed stronger ERS in the U.S. sample. Furthermore, the predictions derived from ERS differences were confirmed, thus demonstrating the credibility of the ERS argument. General implications of ERS for cross-cultural research are discussed.

EXTREME RESPONSE STYLE IN CROSS-CULTURAL RESEARCH
A Reminder

KI-TAEK CHUN and JOHN B. CAMPBELL
Institute for Social Research
University of Michigan

JONG HAE YOO
Yonsei University (Seoul, Korea)

By extreme response style (ERS) we mean the tendency to endorse extreme response categories in a multiple response category format, such as "Strongly agree/disagree" rather than less extreme responses like "Somewhat agree/disagree" or "Mildly agree/disagree." One might say it is the kind of difference between an exploding pronouncement "I could kill you for that" and a muted reaction "I don't appreciate

AUTHORS' NOTE: This study was supported in part by a grant from the International Liaison Committee for Research on Korea and by the National Science Foundation's Capability-Building Grant GI-29904 to the Institute for Social Research, University of Michigan. The paper has benefited from critical readings by Charles F. Cannell, John R. P. French, Jr., Robert D. Caplan, and S. Richard Pinneau, Jr.

Journal of Cross-Cultural Psychology, Vol. 5 No. 4, December 1974

that." Cross-cultural differences in ERS would be of interest either as a reflection of cultural differences on substantive dimensions or for its implications for the methodology of cross-cultural research, or both. This paper points out the inattention to the phenomenon, and particularly to the consequences, of ERS in cross-cultural research, examines the methodological implications of ERS, and reports a finding of dramatic, though serendipitous, ERS differences between U.S. and Korean samples. It proposes that cross-cultural studies routinely provide information on ERS.

THE NEGLECT OF ERS IN CROSS-CULTURAL RESEARCH

In one of those early papers that touched off the response set controversy, Cronbach recognized ERS as a major response set, and he went so far as to suggest a reduction of Likert-type scales to a two-choice format as a means of controlling its effects (1946: 488; see also Cronbach, 1950: 22 ff.). Although it is less visible than such response sets as acquiescence and social desirability, interest in ERS has persisted through the 1960s to the present. As the recent reviews by Hamilton (1968) and O'Donovan (1965) show, there is now a fair amount of literature on the topic. The main themes of this literature, however, have little to do with cross-cultural differences and their consequences.[1]

To be sure, Triandis and Triandis (1962) discussed ERS differences between U.S. and Greek student samples, and major reviews have noted those instances in which cross-cultural differences in ERS were found. For example, Hamilton's review (1968) refers to Brengelmann's comparison (1959) between German and English samples, and Triandis, Malpass, and Davidson (1972) mention the study by Zax and Takahashi (1967) comparing U.S. and Japanese samples. However, the implications of ERS are not systematically pursued in these studies.[2] One misses any explicit

discussion of ERS effects in the much-needed book on cross-cultural research methods by Brislin, Lonner, and Thorndike (1973).[3] An exception to this trend is the recent book by Triandis (1972), who does recognize ERS as a problem in cross-cultural research. However, he suggests the use of correlational techniques as a means of controlling ERS effects (1972: 53-54), and it will be pointed out subsequently that the correlational technique itself is affected by the ERS effects.

THE CONSEQUENCES OF NEGLECTING ERS

Although the substantive meaning of ERS needs to be clarified, we take the atheoretical position of viewing ERS as a mere empirical regularity whose methodological implications merit examination. Several consequences might arise from the failure to consider ERS in cross-cultural research. First, the ERS differences can produce differences in group means that would be uninterpretable. Second, they may affect the level of item intercorrelations within a measure, thus spuriously raising or lowering the index of internal consistency. Similarly, the magnitude of correlations between measures may be affected by ERS. Finally, the ERS differences could influence the outcome of analyses of underlying dimensions (i.e., factor or cluster analyses).

First, consider the simplest case where there is only one item in the measure. If the groups under comparison are both homogeneous in the direction of answering (i.e., everyone giving either affirmative or negative answers), then the group with a stronger ERS would have a higher score than the group with a weaker ERS (or lower score depending upon the direction of scoring). If the scale consists of several items that are all phrased and scored in one direction, the difference between the groups would be the sum of such differences contributed by individual items. If a measure has an equal

number of positively phrased and negatively phrased items, the inflated and deflated components (contributed by the ERS effects of positive and negative items, respectively) may at best cancel out, producing no group differences. If the two components do not cancel out, the observed difference will be essentially uninterpretable.[4] If the ERS tendency is indeed different for positive and negative items, as suggested by Crandall (1973), mere balancing of the number of positive and negative items would be insufficient to cancel out the ERS effects. This confounding in group means could also arise if there are sex differences in ERS (see Light, Zax, and Gardiner, 1965; Zax and Takahashi, 1967), and if the groups under comparison are not matched in terms of the proportion of males and females. Thus, one consequence of neglecting ERS, when the direction of answering is uniform or nearly so, is to obtain differences in the group mean scores that confound genuine differences with ERS effects.

Second, to the extent that subjects in each of the comparison groups give heterogeneous or directionally non-uniform responses (i.e., both agreeing and disagreeing), the group with a stronger ERS would give rise to a wider spread of responses on each item. That is to say, the group whose members display stronger ERS would have a larger variance on individual items than the weaker-ERS group. This restriction in the range of scores on individual items will reduce the magnitude of correlations among items for the weaker-ERS group. The group with a weaker ERS will thus have a lower level of item intercorrelations, resulting in a lower internal consistency. Since the internal consistency of a measure can be affected by the strength of ERS, the adequacy of a measure across comparison groups needs to be evaluated in connection with the possible ERS effects. Furthermore, because of the wider spread of responses on individual items, the total scores of a measure will be more dispersed for the stronger-ERS group (i.e., a larger variance) than for the weaker-ERS group. Due to the restricted range

of total scores, the intercorrelations among measures will be necessarily lower for the weaker-ERS group. To the extent that samples under comparison have differential ERS, it will be useful to apply statistical formulae which adjust for the restricted range in construct validational or correlational analyses. The difference between the corrected and uncorrected correlations would indicate the strength of ERS effects. Thus, when the direction of answering is heterogeneous, ERS could affect both the internal consistency of individual measures and the correlations among measures.[5]

Finally, since ERS could affect the size of item intercorrelations, it is capable of introducing complications in factor or cluster analysis. As a simple case, assume that the matrix of item intercorrelations for one group is uniformly increased or decreased. Such a uniform change, which is conceivable only for those measures that are unidimensional and high in their internal consistency, will merely result in a differential saliency of factors. On the other hand, serious complications could arise when only a subset of the intercorrelation matrix is affected, or even when subsets are differentially affected. Such nonuniform change in a correlation matrix may occur for a variety of reasons. For example, the direction of phrasing (i.e., positive versus negative) might interact with ERS such that the endorsement of extreme responses might be operative only for either positively phrased or negatively phrased items. The content of items may interact with ERS and exaggerated expression may be found only for the culturally-approved behavior (i.e., those items with high social desirability). The "modesty" norm could make ERS pertinent especially to the evaluative items. The norm of moderation or reticence might affect only those items that are expressive of emotions and feelings. Thus, when subsets of a correlation matrix are differentially affected by ERS, one's ability to compare the dimensional similarity of a measure across groups is seriously impaired. It seems necessary then that any conclusion of dimensional dissimilarity

across cultures ought to be weighed against the possibility of ERS effects on subsets of the correlation matrix.

At this point, it would be useful to clarify the context in which the argument of ERS effects is applicable. First, we assume that the relevance of the foregoing ERS effects is obvious when comparison is on individual variances directly across groups or cultures which differ on ERS. Second, we note that the alleged ERS effects are still a problem even if one first determines a relationship among variables within one group (or culture) and then compares it to the corresponding relationship obtained from another group. Since both the nature and strength of relationships among variables can be affected by ERS, comparison of relationships across groups without consideration of ERS effects is inappropriate.

Our examination of ERS effects in the restricted context of a multiple-response questionnaire is not meant to imply that ERS is idiosyncratic to the format of multiple-response categories. We believe that ERS is a general phenomenon that would be equally operative with an open-ended or interview format. With such nonstandardized research instruments, the feasibility of quantifying ERS differences and assessing their consequences will be problematic. However, investigators will do well to be sensitized to the potential, intrusive effects of ERS.

METHOD

PROCEDURE

As part of a larger study, a 130-item questionnaire dealing with various facets of interpersonal and sociopolitical trust was administered to two student samples, one at the University of Michigan, Ann Arbor (n = 187), and the other at the Yonsei University, Seoul, Korea (n = 204). The

questionnaire used a five-point Likert format with response categories of (1) "Strongly agree," (2) "Mildly agree," (3) "Agree and disagree equally," (4) "Mildly disagree," and (5) "Strongly disagree." The mean and standard deviation of responses were computed for each item separately for the U.S. and Korean samples. Since the standard deviation of a given item reflects the degree of extreme responses for that item, we have adopted the cross-sample difference in standard deviations of individual items as a measure of the differential degree of ERS between the two samples.

RESULTS

Differences in ERS. The range of the standard deviations of individual items, i.e., item standard deviations (hereafter called ISDs), was from .87 to 1.45 (M = 1.123, SD = .127) for the U.S. sample, and likewise .58 to 1.17 (M = .861, SD = .107) for the Korean sample. Considering the 130 items as a whole, the U.S. sample has a larger mean ISD ($t = 17.9$, $p < .001$). This difference in the two ISD means is readily visible in Figure 1. The picture remains essentially the same even when standard deviations are compared at the level of individual items. Except for three ties and two reversals (these items did not appear to share any common characteristics), the ISD was always larger for the U.S. sample (i.e., for 125 out of 130 items). Thus the trend is unmistakable, and the evidence overwhelming: the U.S. sample shows larger ISDs, indicating a stronger ERS.

To examine whether this observed difference in ERS interacts with the direction of item phrasing, the mean of the ISD differences for the positively-phrased items (M = .236, SD = .147) was compared with that for the negatively-phrased items (M = .286, SD = .139). The difference was significant ($t = 2.01$, $p < .05$) which suggests an interactive relationship between ERS differences and the direction of

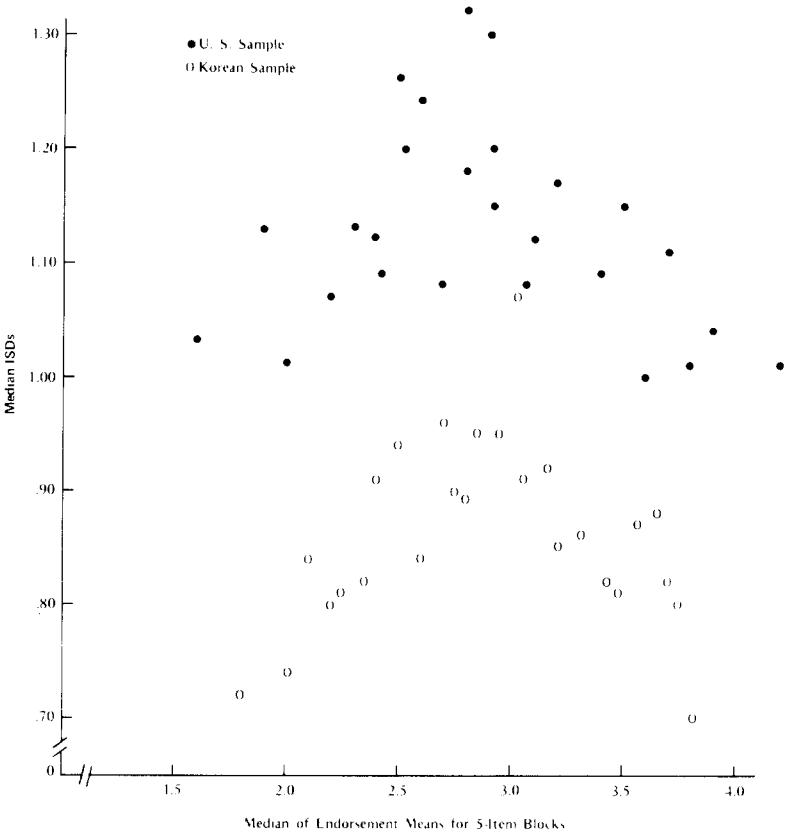

Figure 1: SCATTERPLOT OF MEDIAN ISDs AND ENDORSEMENT MEANS FOR BLOCKS OF FIVE CONSECUTIVE ITEMS ORDERED BY SAMPLE ENDORSEMENT MEANS

phrasing. Upon closer examination we find that the negatively-phrased items have a higher mean ISD (M = 1.162, SD = .123) than the positively-phrased items (M = 1.080, SD = .118) in the U.S. sample ($t = 3.816$, $p < .01$), while the positive and negative items show comparable mean values in the Korean sample (.846 versus .874, $t = 1.515$, $p > .10$). We conclude that ERS is more pronounced in the U.S. sample and that the interaction between ERS and the direction of

phrasing is present only in the U.S. sample. It should be noted, however, that since our analyses are based on ERS indices computed from the total sample without regard to sex, the conclusions may not be the same when males and females are considered separately.

Reliabilities of ERS. Table 1 lists for each sample split-half reliabilities (uncorrected for length) for the extreme positive responses (i.e., "Strongly agree"), extreme negatives (i.e., "Strongly disagree"), and combined extremes, separately for males, females, and the combined sample. For example, the value of .69 listed for U.S. males and extreme positives (A) is the correlation between the frequency of "Strongly agree" responses to the odd-numbered, positively-phrased items and the corresponding frequency of "Strongly agree" responses to the even-numbered, positively phrased items in the U.S. sample. While sex differences are virtually nonexistent in the U.S. sample, reliabilities in the Korean sample tend to be slightly higher for the males, although the differences seem insignificant. (See Hamilton, 1968: 195, for a review of the conflicting evidence regarding sex differences in ERS.)

TABLE 1
Reliabilities[a] of ERS by Culture, Sex, and Direction of Phrasing

Phrasing	Males	Females	Males and Females
United States			
Extreme positives (A)	.69	.69	.68
Extreme negatives (B)	.72	.72	.72
Extreme responses (A+B)	.86	.88	.87
n	71	116	187
Korea			
Extreme positives (A)	.89	.87	.88
Extreme negatives (B)	.81	.76	.78
Extreme responses (A+B)	.93	.92	.92
n	112	92	204

a. Odd-even split, uncorrected for length.

The reliabilities of ERS are uniformly higher for the Korean sample. That is to say, although Korean students as a whole endorse extreme responses less frequently than their American counterparts, their extreme responses (when made) are more reliable and internally consistent. This point is not to be confused with our earlier statement that U.S. students display stronger ERS, which results in a higher reliability of measures for the U.S. student sample.

The cross-sample difference in reliabilities for the extreme positives (.88 versus .68) is greater than the corresponding difference for the extreme negatives (.78 versus .72). This direction-specific group difference suggests that, in comparison to U.S. students, Korean students are more consistent in their extreme positive responses than in their extreme negative responses.

The reliabilities of the U.S. sample are comparable to those values compiled in Hamilton (1968: 193, Table 2), that is, .88 for our extreme responses versus his median value of .82, .72 for our extreme negatives versus his median value of .71, and .68 for our extreme positives versus his median value of .75. The fact that our sample of Ann Arbor students resembles, at least in one respect, students on other U.S. campuses seems to indicate that although our sample is not a representative, probability sample, the findings from the Ann Arbor students may be more generalizable than the sampling scheme alone would warrant. The observed differences between the two samples, then, may well reflect a genuine cross-cultural dissimilarity rather than a peculiarity due to sampling biases.

Consequences of ERS. Our questionnaire included two standardized measures, the 25-item Rotter Interpersonal Trust Scale (Rotter, 1967) and the 14-item Trustworthiness subscale from Wrightsman's Philosophies of Human Nature Scale (Wrightsman, 1964). For each of these two measures, Table 2 lists for both U.S. and Korean samples the mean

TABLE 2
Mean Item-Intercorrelation, Coefficient Alpha, Mean, and Standard Deviation for Rotter's and Wrightsman's Scales and their Positive and Negative Subscales

Sample	Total Scale				Positive Items				Negative Items				Correlation of Positive and Negative Items
	$\bar{r}_{ij}$	α	$\bar{X}$	σ	$\bar{r}_{ij}$	α	$\bar{X}$	σ	$\bar{r}_{ij}$	α	$\bar{X}$	σ	
Rotter Interpersonal Trust Scale[a]					Positive Items (11)				Negative Items (13)				
United States	.114	.754	64.79	10.18	.117	.594	31.98	5.18	.133	.666	32.81	6.51	.51
Korea	.063	.618	65.81	6.96	.070	.451	31.70	3.91	.097	.582	34.12	4.82	.26
Wrightsman Trustworthiness Scale					Positive Items (7)				Negative Items (7)				
United States	.214	.792	41.90	7.92	.242	.691	22.65	4.26	.231	.677	19.25	4.68	.57
Korea	.177	.751	41.36	6.11	.190	.621	21.43	3.38	.227	.673	19.93	3.75	.47

a. Since one of the 25 Rotter items ("Most salesmen are honest in describing their products") was not included in the Korean questionnaire, all analyses are based on 24 items.

item-intercorrelation, coefficient alpha, mean score, standard deviation of the total scale as well as the positive and the negative items, and the correlation between the positive and negative items.

The mean scores on these two measures are virtually identical for the two samples. This finding is not surprising since there is an approximately equal number of positive and negative items in each of these measures.

Consistent with our earlier finding of a stronger ERS in the U.S. sample, the mean item-intercorrelation is higher for the U.S. sample in both the Rotter (.114 versus .063) and Wrightsman (.214 versus .177) scales. As is required by this difference in the mean item-intercorrelations, in the U.S. sample we observe larger values of coefficient alpha (.754 versus .618 for the Rotter and .792 versus .751 for the Wrightsman scale) and standard deviation (10.18 versus 6.96 for the Rotter and 7.92 versus 6.11 for the Wrightsman scale) for the total scales. This cross-sample difference in coefficient alpha and standard deviation prevails for both the positive and the negative items. Furthermore, the correlation between the positive and the negative items is larger for the U.S. than the Korean sample in both the Rotter scale (.51 versus .26, $t = 2.91$, $p < .01$) and the Wrightsman scale (.57 versus .47, $t = 1.35$, $p \cong .18$; this failure to reach statistical significance may be attributable to a relatively smaller number of items in the Wrightsman scale). Recall that a larger correlation was predicted for the stronger-ERS group from considerations of ERS effects. If one views the positive and the negative items as measures, the observed finding is quite consistent with our prediction. In addition, when the statistical correction for restricted range is applied in the Korean sample, the cross-sample difference in the correlations of the positive and negative Rotter items becomes insignificant (.56 versus .44, $t = .89$). This shift from significant to insignificant findings demonstrates the need to correct for ERS effects.

SUMMARY

At an early phase of data analysis we made a serendipitous discovery of striking differences in response extremity between U.S. and Korean samples. We proceeded to systematize our study of ERS, first, by establishing the phenomenon of ERS within our data set, second, by logically deriving the consequences of ERS effects, and then by examining whether the predictions derived from considerations of ERS were sustained. As presented in Table 2, our findings are quite consistent with the predictions. This confirmation renders support to our argument that the phenomenon of ERS and its consequences merit greater attention than accorded in the past.

COMMENTS

Based upon our demonstration of the consequences of ERS, we propose that reports of cross-cultural research routinely provide information on ERS so that the reported data can be interpreted appropriately and misleading conclusions may be avoided. The minimal information helpful for a reader would be knowledge of whether one sample shows a stronger ERS than the other sample under comparison. If the cross-sample difference prevails over all or most of the items, it would be reasonable to attribute the difference to a generic cross-cultural difference in response style. On the other hand, if the ERS difference is present only for a subset of items, it would be useful to know whether that subset is identifiable in terms of an underlying content dimension or the direction of item phrasing.

Our primary concern in this paper has been to underscore the need to examine ERS as a basic methodological issue in cross-cultural research. At this point, two closing notes might help place our concern in a larger perspective. First, the format of multiple response categories will need to be

compared to alternate formats (e.g., trichotomous, dichotomous, open-ended, and so on) in terms of its susceptibility to ERS effects. Second, a theoretical framework helpful for understanding ERS as a phenomenon in its own right is needed. We have in mind a social psychological delineation of the process of responding that articulates the underlying parameters. Hopefully, consideration of these parameters would permit predictions of the degree of ERS for different cultural groups.

NOTES

1. These themes center on such issues as: (a) the temporal stability of ERS (Merrens, 1970; also extensively reviewed by Hamilton, 1968), (b) the generalizability of ERS across different tests (Hamilton, 1968) and across different sense modalities (Merrens, 1970), (c) sex differences in ERS (Crandall, 1973; Light et al., 1965; Zax and Takahashi, 1967), (d) ERS as indicative of psychopathology (O'Donovan, 1965), (e) personality correlates of ERS (Hamilton, 1968), and (f) the partitioning of the total variance into those components attributable to the direction of response (e.g., yes-no, agree-disagree, and so forth) and to the degree of extremeness (e.g., very, somewhat, a little, and so on) (Peabody, 1962).

2. As a further check on this observation of inattention, we examined all the papers published in the four 1972 issues of the *Journal of Cross-Cultural Psychology*. Judging from the written accounts of methods, procedures, instruments, and so on, we concluded that there are three papers in which ERS is clearly involved, and another three papers which seem to be related to ERS (insufficient description of instruments does not permit a definite statement). ERS is not mentioned in any of these six studies, nor have we detected any indication that the authors of these papers might have been aware of ERS-related issues.

Those researchers who are active in cross-cultural research seem to be personally aware of the ERS phenomenon. For example, according to Suzuki (1973), the Japanese descendants in Hawaii, compared to their counterparts in Japan, are less constrained in their expression, talk more freely about their feelings and opinions, and often use a more vivid language. In short, one might infer that the Japanese Hawaiians exhibit a stronger ERS than the Japanese mainlanders. (Reports of comparative surveys on the Japanese national character between Hawaii and mainland residents, on which this observation is based, are available from Tatsuzo Suzuki, Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-Ku, Tokyo, Japan.)

3. But see a brief discussion on p. 205 of Brislin et al.–Ed.

4. We are aware of the argument that comparison of the means of Likert scales is illegitimate in a strict sense. Our point, however, is that ERS could affect such a comparison regardless of its legitimacy.

5. It follows from the discussion that the use of correlations as a means of controlling for ERS effects, suggested by Triandis (1972: 53-54), will be ineffective.

REFERENCES

BRENGELMANN, J. C. (1959) "Differences in questionnaire responses between English and German nationals." Acta Psychologica 16: 339-355.

BRISLIN, R. W., W. J. LONNER, and R. M. THORNDIKE (1973) Cross-cultural Research Methods. New York: John Wiley.

CRANDALL, J. E. (1973) "Sex differences in extreme response style: differences in frequency of use of extreme positive and negative ratings." J. of Social Psychology 89: 281-293.

CRONBACH, L. J. (1950) "Further evidence on response sets and test design." Educ. and Psych. Measurement 10: 3-31.

--- (1946) "Response sets and test validity." Educ. and Psych. Measurement 6: 475-494.

HAMILTON, D. L. (1968) "Personality attributes associated with extreme response style." Psych. Bull. 69: 192-203.

LIGHT, C. S., M. ZAX, and D. H. GARDINER (1965) "Relationship of age, sex, and intelligence level to extreme response style." J. of Personality and Social Psychology 2: 907-909.

MERRENS, M. (1970) "Generality and stability of extreme response style." Psych. Reports 27: 802.

O'DONOVAN, D. (1965) "Rating extremity: pathology or meaningfulness?" Psych. Rev. 72: 358-372.

PEABODY, D. (1962) "Two components in bipolar scales: direction and extremeness." Psych. Rev. 69: 65-73.

ROTTER, J. B. (1967) "A new scale for the measurement of interpersonal trust." J. of Personality 35: 651-665.

SUZUKI, T. (1973) Personal communication.

TRIANDIS, H. C. (1972) The Analysis of Subjective Culture. New York: John Wiley.

--- and L. M. TRIANDIS (1962) "A cross-cultural study of social distance." Psych. Monographs 76, 21.

TRIANDIS, H. C., R. S. MALPASS, and A. R. DAVIDSON (1972) "Cross-cultural psychology," pp. 1-84 in B. J. Siegel (ed.) Biennial Review of Anthropology: 1971. Palo Alto, Calif.: Stanford Univ. Press.

WRIGHTSMAN, L. S. (1964) "Measurement of philosophies of human nature." Psych. Reports 14: 743-751.

ZAX, M. and S. TAKAHASHI (1967) "Cultural influences on response style: comparisons of Japanese and American college students." J. of Social Psychology 71: 3-10.

Ki-Taek Chun was born in Korea and received both his M.A. (in statistics) and Ph.D. (in psychology) from the University of California at Berkeley. He is a Study Director in the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106. His research interests include reification of psychological constructs, epistemology of cross-cultural research, and personality assessment. His current research deals with the conceptual and methodological problems in the measurement of trust.

John B. Campbell was born in the United States and received his B.A. from the College of Wooster, in Wooster, Ohio. He is currently a Ph.D. candidate in psychology at the University of Michigan, where he works as a teaching fellow, and was formerly an Assistant Study Director at the Institute for Social Research. His research interests center on interpersonal trust, personality assessment, and Person x Situation interaction.

Jong Hae Yoo was born in Korea and received his Ph.D. in political science from the University of Michigan. He is Professor and Associate Dean of the Graduate School of Public Administration, Yonsei University, Seoul, Korea. He was formerly an assistant professor in the Department of Political Science at Eastern Michigan University in Ypsilanti. His main research interests include children's political socialization process and theories in social change and comparative public administration. He co-authored Development Administration *(Pummunsa, Seoul, Korea, 1973).*