

THE STRUCTURE OF PGC MORALE SCALE IN AMERICAN AND JAPANESE AGED: A FURTHER NOTE

ABSTRACT. This study involves a further replication of cross-cultural comparison of the structure of the Philadelphia Geriatric Center Morale Scale (PGCMS). Using Japanese and American data sets, the present research replicates and extends the findings reported by Liang et al. (1987). In particular, the earlier findings that four PGCMS items behave differently in two cultures are replicated. The present study yields two additional observations. First, the invariance in the PGCMS can now be extended beyond the urban elderly residents studied by Liang et al. (1987) to the entire aged population in the U.S. and Japan. Second, this comparability is robust despite the elimination of correlated measurement errors from the earlier specifications and when several exogenous variables are controlled. Further, the impact of selected demographic variables on the PGCMS was evaluated. In addition, qualitative data from in-depth interviews provide further insights concerning the cultural differences in the expression of well-being.

Key Words: comparative, cross-cultural, invariance, measurement model, well-being

INTRODUCTION

It has long been recognized that cross-cultural differences in health and well-being reflect not only genuine variation in the incidence and prevalence of morbidity, disability, and mortality, but also the social processes by which these data were generated. These processes may be far removed from the biological reality. A major assumption underlying the comparative study of health is that beliefs about health, diseases, and medical treatment are the product of a particular historical, social, and cultural setting (Lock 1987). For instance, the concept of depression has been suggested as unique to European and North American cultures since equivalent conceptual terms have not been found among many non-European groups (Marsella et al. 1985). Thus, the subjective experience of depression, its behavioral manifestations, and the social responses invoked by it may vary across cultures. Similar observations have been made of other concepts such as symptoms, illness (Angel and Thoits 1987), and subjective well-being (Liang et al. 1987).

As suggested by Angel and Thoits (1987), there are two traditions in the study of physical and emotional illness. The epidemiological tradition seeks to understand the etiology, progress, and consequences of diseases in different populations. The ethnographic tradition, which is largely anthropological and sociological, attempts to understand the illness-labeling and help-seeking process. What is lacking in the epidemiological approach to the study of the impact of culture on illness is the recognition of the actor's perspective, the cognitive structure that mediates the definition of sickness and health.

A major research issue in comparative studies therefore has been the equiv-

alence of measures. Whereas it may be less complicated when only objective physical measures are employed; this is a particularly serious issue when self-reported data are involved. Cross-cultural comparisons in health and well-being are further complicated by the fact that different indicators are often used to measure the same concept. The implications of such practices have not been adequately ascertained. The investigator must understand the similarities and differences between measures of a given variable in different cultures in order to make a meaningful comparison.

This paper describes a further assessment of the cross-cultural comparability in a widely used measure of subjective well-being, the Philadelphia Geriatric Center Morale Scale (PGCMS), between the American and Japanese elderly. As reported by Liang and his associates (1987), while the 15-item version of PGCMS initially derived from the U.S. data did not fit the Japanese data well, a version containing 11 of the original items was found to fit both the American and Japanese data adequately. No major differences in the factorial structure of the 11-item PGCMS were uncovered.

The study by Liang and his associates (1987) can be improved in several respects. First, 345 Japanese elderly persons who resided in Tokyo were compared with 490 American aged adults from large urban areas with a population of one million or more in the United States. The generalizability of the findings beyond the elderly American and Japanese in large urban communities remains to be established. According to Cook and Campbell (1979), external validity is concerned with generalizability: To what populations, settings, treatment variables, and measurement variables can a given observation be generalized? As a powerful approach to extend external validity, distinct and divergent replication may be designed specifically to submit theories and observations to the most difficult test of falsification (Cook and Campbell 1979; Kish 1987). This is illustrated by the practice in disciplines such as physics and chemistry where little attention is given to representativeness but great concern is accorded to repeatability by independent researchers. External validity is enhanced more by multiple replications across different times, settings, respondents, and investigators. An analysis involving samples representative of national populations of elderly people in these two countries affords an opportunity to extend the validity of prior findings reported by Liang and his associates.

Second, the comparisons by Liang et al. (1987) were made between the American and Japanese elderly persons of somewhat different age ranges, and the closest approximation to the U.S. retirement age (i.e., 65) in Japan was assumed to be age 60. Although this assumption does have some justification, an alternative comparison may involve identical age ranges. Such a comparison can provide additional evidence concerning the equivalence of PGCMS.

Third, the specifications proposed by Liang and associates (1987) contained five pairs of correlated measurement errors which are difficult to explain theoretically. In view of the fact that the specification of these correlated measurement errors only contributed to a moderate degree of improvement in

the goodness of fit, a model without correlated measurement errors may be a more appropriate basis for comparison.

Fourth, Liang et al. (1987) focused exclusively upon the internal structure of PGCMS within and between the American and Japanese elderly. The hypothesis that the observed cross-cultural differences in the measurement specifications are a function of population heterogeneity was not examined. If differences in the measurement model parameters can be attributed to characteristics of the two populations under studies, then these differences should disappear once these variables are explicitly incorporated into the analyses. Consequently, by explicitly considering the effects of certain relevant exogenous variables, the assessment of equivalence can be made under a more "comparable" circumstance.

Fifth, the data were collected at different times. Whereas the data on the American aged were gathered in 1968, the data in Japan were collected in 1983. Given the substantial turn-over in the elderly population, the possible impact of different period and cohort is difficult to assess. As suggested by Glenn (1977) and Labouvie-Vief (1985), a substantial amount of variation between different cohorts may be attributed to differences in education and other characteristics. Accordingly, even when data from the comparable cohorts are not available, the confounding effect of cohort differences may be partially addressed when population heterogeneity including education and health are controlled.

The present study addresses the first four of the five limitations described in the above. It involves the comparison of data from national samples of aged individuals in the United States and Japan. The comparisons are made with the same age groups in terms of models without correlated measurement errors. In addition, quantitative comparisons of the internal structure of the PGCMS are made when the impact of several exogenous variables were taken into account. Through the statistical control of population heterogeneity, the confounding effect of cohort differences is partially addressed.

DATA AND METHODS

Data for this research came from the 1968 National Senior Citizens (NSC) Survey (Schooler 1979) and the 1987 National Survey of Japanese Elderly (NSJE). The NSC survey involved an area probability sample of 3,996 non-institutionalized people 65 years old and over living in the continental United States in 1968. The survey focused on characteristics of residential environments, social relations, health, and knowledge and use of services and facilities.

NSJE was undertaken in November, 1987, through a stratified two-stage probability sample survey of 2,200 non-institutionalized older adults aged 60 and over in Japan. Names of potential respondents were obtained from the Resident Registries, which contain an up-to-date listing of the name, address, sex, and age of all individuals in Japan. A total of 2,200 interviews were completed successfully from the list of 3,288 names. Excluding 88 persons who died or were hospitalized during the survey, and 99 persons who were never

contacted due to a change in residence or an unknown address, the overall response rate was 71 percent (2,200/3,086).

To maximize the comparability between the American and Japanese data, the present research was primarily based on data from a subsample of 1,506 Japanese respondents who were 65 years old or older at the time of survey. To fully replicate the analyses done by Liang et al. (1987), a comparison involving the entire Japanese sample and the American data was also made.

Given that the NSC survey is the only national data set in the U.S. which contains the PGCMS, it was used in this replication. Although this may introduce certain confounding effects, it still remains that the NSC and NSJE are the two most comparable data sets in existence for the comparative analysis of PGCMS. A major assumption underlying the present research is that the structure of PGCMS in the American aged remains stable over the 20 year period. Despite the fact that this assumption is not directly testable in this study, several studies conducted in the mid and late 1970s reported a similar factorial structure of PGCMS, lending some support to stability assumption (McCulloch 1989; Liang and Lawrence 1986).

The original English version of PGCMS was translated into Japanese, and was pretested several times before it was finalized. The Japanese version was subsequently translated back into English. The accuracy of translation was ensured in that there was no discrepancy of substantive importance between the two versions.

Data analyses were undertaken by using LISREL 7 (Jöreskog and Sörbom 1988) and were replicated in four randomly divided subsamples of the NSC and NSJE surveys. The adequacy of the size of these subsamples to detect significant effects must be evaluated with reference to the proposed models. As suggested by Bentler (1989), the ratio between sample size and the number of free parameters could be as low as 5 to 1, depending on the distributional assumptions such as multinormality or being distribution-free. In the 15-item and 11-item measurement model specifications, we hypothesize 33 and 25 free parameters, respectively, that are significantly different from zero. Given that the size of the subsamples range from 271 to 321 for Japan and 956 to 964 for the U.S., these subsamples are more than adequate for the evaluation of the proposed models.

To assess whether the Japanese elderly understand the survey questions and interpret them within the same framework as the investigators, in-depth interviews of a small number of Japanese and American elderly persons were conducted respectively in July 1989 and July 1990. In particular, 23 elderly Japanese across Japan were selected in terms of their age (i.e., under 75 vs. 75 and over), sex, and type of community (i.e., urban versus rural). Given their high correlations with age and sex, marital status and living arrangement were not used as strata. The American sample consisted of 28 elderly residents in Ann Arbor, Michigan. They were recruited through newspaper advertisement and local senior citizen centers, conforming as close as possible to the characteristics of the 23 Japanese respondents in terms of age and sex. For both countries, each

respondent was asked all of the questions contained in the interview schedule used for the NSJE. In addition, a follow-up probe was carried out for 36 out of the 101 questions. With reference to the PGCMS, 8 out of the total 15 items were selected for probe. The findings from these in-depth interviews will also be used to illuminate the meanings of the quantitative survey results.

FINDINGS

A. Descriptive Analysis

Table I presents descriptive statistics of the social demographic characteristics and the PGCMS items for the total American sample (N=3,996) and the subsample of those aged 65 and over in the Japanese sample (N=1,506). The American sample differed significantly from the Japanese sample in a number of characteristics. The American elderly respondents were slightly older, were more likely to be female, and appeared to be somewhat better educated. Whereas a higher proportion of the Japanese elderly seemed to consider their health status to be good or better than the American elderly, they were less likely to view their health as better relative to their peers and their previous health status. Significant differences were found between the American and Japanese elderly in 11 out of the 15 PGCMS items. Given that the cross-cultural equivalence of these items has not been fully established, caution must be exercised in interpreting the meanings of these differences.

B. Measurement Equivalence

The assessment of comparability was undertaken first by evaluating the construct validity of the proposed measurement model for the PGCMS in each of the two cultures. A cross-cultural comparison is meaningful only when there exists a model which fits both cultures well. Second, measurement equivalence was examined by testing a series of nested hypotheses involving equivalence constraints on one or more matrices constituting the factorial structure.

We started with a model (M_1) previously developed by Liang and Bollen (1983), which consists of three dimensions: agitation (η_1), dissatisfaction (η_2), and attitude toward one's own aging (η_3). They involve, respectively, six, four, and five indicators. In addition, a second-order factor, subjective well-being (ξ_1), is assumed to influence these three dimensions directly. Finally, five pairs of correlated measurement errors were hypothesized. This model was empirically evaluated and supported by analyzing data from the NSC survey (Liang and Bollen 1983). As previously stated, in the present analysis, only models containing independent measurement errors were examined. This was based on the recognition that the pattern of correlated measurement errors previously specified lacks strong theoretical justification, and, at the same time, provides only moderate improvement in the goodness-of-fit.

The measurement model of the 15 item PGC Morale Scale with independent

TABLE I
Descriptive statistics for items in the analyses for U.S. and Japanese samples

Item	United States (N=3996)	Japan (N=1506)
Age		
Mean	73.13***	72.51
SD	6.29	5.50
Sex		
% Female	60.2	55.9
Education		
% High School Graduate equivalent or more (a)	27.9	19.9
Health		
Satisfaction with health (b)		
% very or completely satisfied	--	68.1
Present health status (c)		
% good or better	51.6	80.8
Health compared to others your own age		
% better	50.0***	44.4
Health compared to 1 year ago		
% better	14.2***	6.4
Health compared to when 60 years old (d)		
% better	10.8	--
Agitation		
Things bother me more		
% yes	34.4***	20.2
Worry so much can't sleep		
% yes	29.9	28.8
Afraid of things		
% yes	12.9***	21.1
Get mad more than I used to		
% yes	14.6	13.3
Take things hard		
% yes	25.1***	41.0
Get upset easily		
% yes	27.6**	22.8

TABLE I (Continued)

Dissatisfaction		
Life isn't worth living		
% yes	18.3***	13.9
A lot to be sad about		
% yes	22.6***	12.5
Life is hard for me		
% yes	22.4***	61.8
Satisfied with life today		
% yes	85.0	84.2
Attitudes about own aging		
Things get worse as one ages		
% yes	50.3***	27.3
Have as much pep as last year		
% yes	54.1***	62.9
As one ages, he/she gets less useful		
% yes	56.0	56.2
Things better/worse than I thought		
% better	12.5***	19.0
As happy as when younger		
% yes	55.0***	64.5

Note: To conserve space, all items are presented in compressed form. Significance tests on the proportional difference of the U.S. and Japan samples are noted by asterisks (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). (a) For Japan, this includes persons with 11 or more years of education. Given that the two items are not completely comparable, no statistical test was performed. (b) This item is in the Japanese data only. (c) The range for this item in the U.S. data is 1–4 while in the Japanese data it is 1–5. Given the difference in range, no statistical test was performed. (d) This item is in the U.S. data only.

measurement error terms (M_2) is provided in Figure 1. Several measures were applied in assessing the overall fit of the proposed model. With 87 degrees of freedom, the likelihood ratios in all four American subsamples vary from 256.20 to 310.65, indicating a probability of less than 0.001. As the likelihood ratios range from 155.10 to 207.69, the probabilities of the fitting function associated with M_2 are near zero in all four Japanese subsamples. These reflect a poor fit of M_2 to both of the American and Japanese data.

Given that the likelihood ratio is very sensitive to sample size, a more appropriate fit index would be one that is not or is less affected by the sample size. Two such measures are the Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit Index (AGFI) devised by Jöreskog and Sörbom (1988). The GFIs and AGFIs associated with M_2 vary from 0.939 to 0.965 in the American samples, and 0.887 to 0.930 in the Japanese samples. These are considered to be of reasonable goodness-of-fit by most standards. Two other measures of fit assessed are the Bentler-Bonett (1980) normed fit index (Δ_1) and the Bollen (1989) non-normed fit measure (Δ_2). Whereas Δ_1 s range from 0.901 to 0.914 in the American samples, they vary from 0.716 to 0.801, reflecting a less satisfac-

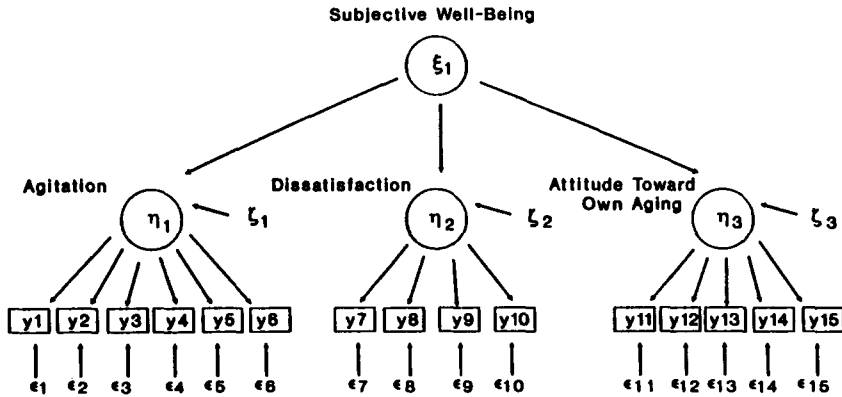


Fig. 1 Measurement model of 15-item PGC morale scale.

tory fit in the Japanese samples. Similar conclusions can be derived from the Δ_2 s associated with the American and Japanese samples. In summary, the results regarding the goodness-of-fit of M_2 are somewhat mixed. Although M_2 fit the American data well, its fit to the Japanese data is somewhat marginal but acceptable. Given that M_2 was developed on the basis of the NSC survey data (Liang and Bollen 1983), this is not surprising.

The parameter estimates differ across the American and Japanese data in several ways. Specifically, four items showed lower standardized factor loadings (i.e., less than 0.4) in the Japanese subsamples. One item, "I am afraid of a lot of things" exhibited a factorial complexity of greater than one, whereas the other three indicators displayed factor loadings of different magnitudes in comparison with those in the American sample. The item, "I get mad more than I used to," exhibited a greater unstandardized loading than its counterparts in the U.S. subsamples. Two other items, "Life is hard for me most of the time," and "Things are better/worse than I thought" showed relatively smaller unstandardized factor loadings in the Japanese sample than those in the American sample.

These observations were subsequently supported by simultaneous factor analysis in that the hypothesis of equivalence in the first-order factor loadings across the American and Japanese samples was rejected in all four subsamples. In addition, differences in parameter estimates were statistically significant in two or more subsamples for all four items except one, "I get mad more than I used to." Regarding this item, the differences in the unstandardized loadings were in the order of 0.11 to 0.16 but were statistically not significant. Moreover,

TABLE II
 Goodness-of-fit indices of models for the 11-item PGC morale scale (M₃)
 with various equivalence constraints

Models	Likelihood Ratios (a)		Incremental χ^2 's		Bentler-Bonett Normed Index $\Delta 1$	Bollen Normed Index $\Delta 2$	GFI	JP	
	χ^2	df	χ^2	df				US	JP
1. Null Model									
Sample 1 ($n(\text{US})=958; n(\text{JP})=318$)	2499.69	110					0.587	0.662	
Sample 2 ($n(\text{US})=957; n(\text{JP})=321$)	2990.57	110					0.542	0.576	
Sample 3 ($n(\text{US})=956; n(\text{JP})=317$)	2587.47	110					0.580	0.687	
Sample 4 ($n(\text{US})=964; n(\text{JP})=271$)	2560.06	110					0.560	0.725	
2. Model with no equivalence constraints									
Sample 1	236.47	82			0.905	0.936	0.970	0.956	
Sample 2	295.09	82			0.901	0.927	0.960	0.952	
Sample 3	268.16	82			0.896	0.926	0.967	0.950	
Sample 4	243.22	82			0.905	0.935	0.966	0.959	
3. Model with equivalence constraints on Λ_y									
Sample 1	244.53	90	8.06	8	0.902	0.936	0.970	0.952	
Sample 2	309.23	90	14.14	8	0.897	0.924	0.959	0.947	
Sample 3	282.42	90	14.26	8	0.891	0.923	0.967	0.943	
Sample 4	249.56	90	6.34	8	0.903	0.938	0.966	0.955	
4. Model with equivalence constraints on Λ_y, Γ									
Sample 1	245.38	92	0.85	2	0.902	0.936	0.970	0.952	
Sample 2	311.84	92	2.61	2	0.896	0.924	0.959	0.945	
Sample 3	292.27	92	9.85*	2	0.887	0.920	0.966	0.939	
Sample 4	250.67	92	1.11	2	0.902	0.936	0.966	0.954	

TABLE II (Continued)

5. Model with equivalence constraints on $\Lambda_y, \Gamma, \theta\epsilon$									
Sample 1	315.17	103	69.79***	11	0.874	0.911	0.967	0.924	
Sample 2	376.14	103	64.30***	11	0.874	0.905	0.957	0.919	
Sample 3	354.37	103	62.10***	11	0.863	0.899	0.964	0.916	
Sample 4	306.42	103	55.75***	11	0.880	0.917	0.964	0.926	
6. Model with equivalence constraints on $\Lambda_y, \Gamma, \theta\epsilon, \Psi$									
Sample 1	335.10	106	19.93***	3	0.866	0.904	0.966	0.920	
Sample 2	383.58	106	7.44	3	0.872	0.904	0.956	0.917	
Sample 3	374.89	106	20.52***	3	0.855	0.892	0.963	0.908	
Sample 4	331.95	106	25.53***	3	0.870	0.908	0.962	0.917	
7. Model with equivalence constraints on all matrices									
Sample 1	348.68	107	13.58***	1	0.861	0.899	0.966	0.912	
Sample 2	385.52	107	1.94	1	0.871	0.903	0.956	0.916	
Sample 3	396.70	107	21.81***	1	0.847	0.883	0.962	0.898	
Sample 4	355.68	107	23.73***	1	0.861	0.899	0.962	0.904	

(a) all likelihood ratios are significant at the 0.001 level.
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE III
Standardized maximum likelihood estimates for the 11-item American and Japanese PGC model (M_3)

Parameters	U.S.				Japan			
	S1 N=958	S2 N=957	S3 N=956	S4 N=964	S1 N=318	S2 N=321	S3 N=317	S4 N=217
Variance of the second-order factor	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Second-order factor loadings (γ_s)								
η_1 : Agitation (a)	0.889	0.885	0.798	0.874	0.676	0.741	0.554	0.662
η_2 : Dissatisfaction	0.890	0.903	0.897	0.830	1.006	0.940	0.848	1.187
η_3 : Attitudes	0.774	0.770	0.739	0.788	0.747	0.912	0.939	0.707
First-order factor loadings (λ_s)								
η_1 : Agitation								
y1: Little things bother me (a)	0.589	0.556	0.568	0.625	0.667	0.625	0.679	0.573
y2: Worry so much can't sleep	0.604	0.599	0.619	0.594	0.658	0.607	0.575	0.596
y3: I take things hard	0.615	0.656	0.640	0.672	0.451	0.433	0.570	0.494
y4: I get upset easily	0.635	0.653	0.684	0.673	0.631	0.651	0.517	0.605
η_2 : Dissatisfaction								
y5: Life isn't worth living (a)	0.562	0.594	0.591	0.576	0.516	0.710	0.616	0.360
y6: A lot to be sad about	0.579	0.645	0.615	0.615	0.470	0.616	0.476	0.450
y7: How satisfied with life	0.467	0.549	0.564	0.489	0.475	0.521	0.386	0.350
η_3 : Attitudes toward aging								
y8: Things get worse (a)	0.625	0.635	0.582	0.641	0.566	0.535	0.656	0.436
y9: I have as much pep	0.533	0.592	0.526	0.527	0.518	0.540	0.459	0.520
y10: As one ages, he/she gets less useful	0.505	0.543	0.529	0.586	0.496	0.473	0.411	0.475
y11: I am as happy as when I was younger	0.510	0.485	0.525	0.456	0.417	0.522	0.416	0.478

TABLE III (Continued)

Error variances in equations										
$\psi(1,1)$	0.210	0.217	0.363	0.236	0.542	0.450	0.693	0.562		
$\psi(2,2)$	0.208	0.184	0.195	0.311	-0.012+	0.116+	0.280+	-0.409+		
$\psi(3,3)$	0.400	0.407	0.454	0.378	0.442	0.169+	0.117+	0.500		
Measurement error variances in observed indicators										
$\theta\epsilon(1,1)$	0.653	0.691	0.677	0.609	0.555	0.610	0.539	0.672		
$\theta\epsilon(2,2)$	0.635	0.641	0.617	0.648	0.567	0.632	0.670	0.645		
$\theta\epsilon(3,3)$	0.622	0.570	0.591	0.549	0.797	0.812	0.675	0.756		
$\theta\epsilon(4,4)$	0.597	0.573	0.533	0.547	0.602	0.576	0.733	0.633		
$\theta\epsilon(5,5)$	0.684	0.647	0.651	0.668	0.734	0.496	0.621	0.870		
$\theta\epsilon(6,6)$	0.665	0.584	0.622	0.621	0.779	0.621	0.774	0.798		
$\theta\epsilon(7,7)$	0.782	0.699	0.682	0.761	0.774	0.728	0.851	0.877		
$\theta\epsilon(8,8)$	0.610	0.597	0.661	0.589	0.680	0.714	0.569	0.810		
$\theta\epsilon(9,9)$	0.716	0.650	0.724	0.722	0.732	0.708	0.789	0.730		
$\theta\epsilon(10,10)$	0.745	0.705	0.721	0.656	0.754	0.777	0.831	0.775		
$\theta\epsilon(11,11)$	0.740	0.765	0.725	0.792	0.826	0.728	0.827	0.772		

Note: To conserve space, item wordings have been compressed. All estimates are significant at the 0.05 level except those marked by a plus (+) sign. (a) Parameter value is constrained to 1.0.

this item had a standardized factor loading of 0.25 or smaller in three out of four Japanese subsamples. The above findings replicate those reported by Liang et al. (1987) and further reinforce the conclusion that these four items behave differently across cultures.

A revised model (M_3) was subsequently specified by excluding the four noncomparable items. This 11-item model fit both the American and Japanese data sets quite satisfactorily (GFIs and AGFIs range from 0.919 to 0.970), and was used as the basis of the comparison. According to Table II, there are significant American-Japanese differences in the measurement error variances ($\Theta_{\epsilon s}$) and possibly in the residual error variances (ψ_s), and variance of the second-order factor (ϕ). Specifically, models involving equivalence constraints on the first and second-order factor loadings exhibit quite good fit, while the fit tends to deteriorate when additional constraints are imposed. However, such differences are more apparent in the incremental chi-squares, Δ_1s , and Δ_1s but less clear in terms of GFIs and AGFIs.

To gauge the differences in parameter differences, confidence intervals around the differences in the measurement error variances were formed. Five measurement error variances were identified as being nonequivalent, including those associated with the following items: (a) "Little things bother me"; (b) "I take things hard"; (c) "Life isn't worth living"; (d) "I have a lot to be sad about"; and (e) "I was as happy as when I was younger." In view of the fact that nonequivalence in measurement errors can be easily adjusted, it is reasonable to conclude that there is no major cross-cultural difference in the structure of the 11-item version PGCMS between the American and Japanese aged. Standardized maximum likelihood estimates associated with M_3 are found in Table III.

As an additional check, the above analyses were also run employing the age cut-offs that were used in Liang et al. (1987) i.e., 60 years of age or greater for the Japanese and 65 years of age or greater for the Americans. The results from these analyses were basically identical to those using identical age cut-offs.

C. Analysis Involving Exogenous Variables

In the following, the impact of several exogenous variables including age, sex, education, and physical health status on the PGCMS was examined. More importantly, to further evaluate the extent that different cohorts may differ in terms of age and sex composition, education, and health status, the equivalence analyses were rerun after adding these variables to the model as exogenous variables.

Figure 2 presents a path diagram which describes the linkages between PGCMS and exogenous variables such as age, sex, education, and self-rated health (M_4). It assumes that the effects of all exogenous variables on the first-order factors are mediated by the second-order factor (M_4). To simplify the presentation, observable indicators and measurement errors are omitted. The arrows signifying the effects of the exogenous variables on dimensions of the PGCMS are denoted as gammas (γ_s).

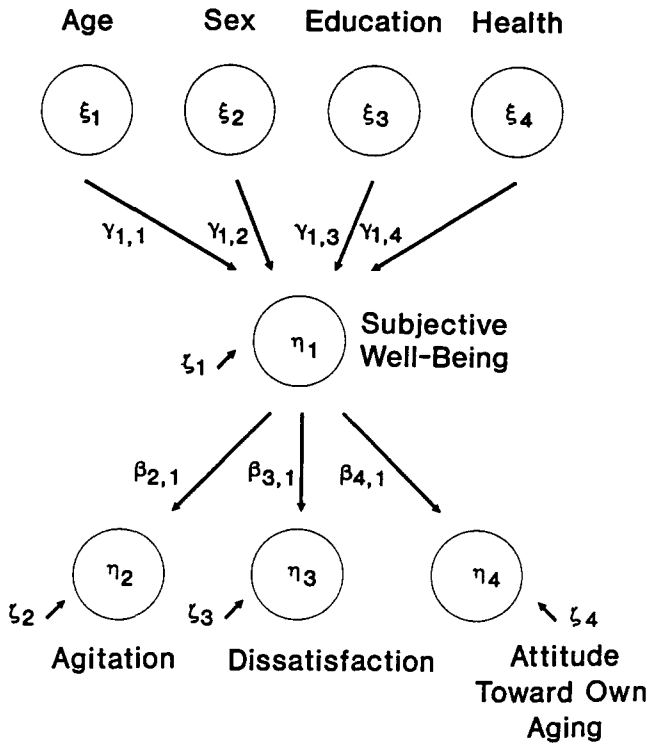


Fig. 2. Second-order factor PGC model with exogenous variables (M_g).

Before an in-depth assessment, a brief description of the operationalization of the exogenous variables and the coding of the PGCMS items is in order. In both American and Japanese data sets, age is a continuous variable and sex is a dichotomous variable with males coded 0 and females coded 1. Education in the Japanese sample is measured in actual years of school (range 0 through 17), while in the American sample it reflects ordered categories ranging from 0 (none) to 7 (finished college or graduate work). Because age, sex, and education were measured by single-indicator measures, factor loading and measurement error coefficients associated with them were fixed according to known reliabilities (i.e., 0.90, 0.95, and 0.60 respectively) derived from the 1970 U.S. Census (U.S. Bureau of the Census 1975). Given that the authors have no information concerning these reliabilities in Japan, the same reliabilities were assumed for the American and Japanese data. This seems to be a reasonable resolution in view of the consequence of ignoring measurement errors.

Self-rated health is comprised of 4 items in each data set with all items coded

TABLE IV
The unstandardized direct and total effects of the exogenous variables on morale and the first order factors and measures of fit for M_4 in the American and Japanese samples (a)

	Japan							
	U.S.				Japan			
	Age	Sex	Education	Health	Age	Sex	Education	Health
Direct effect on second order factor								
Subjective well-being								
Sample 1	-0.016*	-0.006	0.024**	-0.162***	-0.021*	-0.007	-0.014	-0.124***
Sample 2	-0.013	-0.026***	0.033***	-0.201***	-0.018	-0.009	0.016	-0.036***
Sample 3	-0.023***	-0.019**	-0.021*	-0.162***	-0.009	0.003	-0.039**	-0.084***
Sample 4	-0.024**	-0.008	0.032	-0.162***	-0.022**	0.002	0.017*	-0.066***
Total effects on first-order factors								
Agitation								
Sample 1	-0.016*	-0.006	0.024**	-0.162***	-0.021*	-0.007	-0.014	-0.124***
Sample 2	-0.013	-0.026***	0.033***	-0.201***	-0.018	-0.009	0.016	-0.136***
Sample 3	-0.023***	-0.019**	0.021**	-0.162***	-0.009	0.003	0.039**	-0.084***
Sample 4	-0.24*	-0.008	0.032**	-0.162***	-0.022**	0.002	0.017*	-0.066***
Dissatisfaction								
Sample 1	-0.012*	-0.004	0.017*	-0.119***	-0.016*	-0.005	-0.010	-0.094***
Sample 2	-0.011	-0.022***	0.028**	-0.167***	-0.022	-0.011	0.019	-0.162***
Sample 3	-0.021**	-0.018**	0.020**	-0.151***	-0.009	0.003	0.039**	-0.085***
Sample 4	-0.016**	-0.005	0.022**	-0.112***	-0.024**	0.002	0.019*	-0.074***
Attitudes toward own aging								
Sample 1	-0.025*	-0.010	0.037**	-0.253***	-0.030*	-0.010	-0.020	-0.178***
Sample 2	-0.018	-0.036***	0.048***	-0.277***	-0.024	-0.012	0.012	-0.181***
Sample 3	-0.036***	-0.030**	0.034**	-0.258***	-0.018	0.007	0.076**	-0.164***
Sample 4	-0.037***	-0.012	0.050***	-0.252***	-0.052**	0.004	0.041*	-0.160***

TABLE IV (Continued)

Measures of fit	U.S.				Japan			
	S1	S2	S3	S4	S1	S2	S3	S4
	χ^2 (df=125)	539.20	589.06	529.73	552.39	286.73	319.82	297.09
GFI	0.938	0.931	0.938	0.938	0.912	0.899	0.908	0.915
AGFI	0.916	0.906	0.916	0.916	0.879	0.862	0.875	0.884
Δ_1	0.846	0.846	0.851	0.854	0.780	0.804	0.763	0.789
Δ_2	0.877	0.874	0.882	0.883	0.863	0.871	0.848	0.891

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. (a) Sample sizes for the U.S. data are 958, 957, 956, and 964 for subsamples 1 through 4, respectively. For the Japanese data, sample sizes are 318, 321, 317, and 271, for subsamples 1 through 4, respectively.

such that a high score reflects poorer health. Only two items are directly comparable: "health compared to others your own age" and "health compared to a year ago" ("better", "same", or "worse"). A third item representing global (present) health status" is in both data sets but with different categories. In the American data, this item has four categories ranging from "very good" (1) to "poor" (4), while in the Japanese data it has five categories ranging from "excellent" (1) to "poor" (5). The fourth item in the U.S. data is a rating of "health now as compared to when you were 60" ("better", "same", or "worse"). Finally, the fourth item in the Japanese data deals with "satisfaction with health" and ranges from "completely satisfied" (1) to "not at all satisfied" (5).

Turning to the PGCMS items themselves, all items were recoded such that a higher score reflects a more positive outlook. More specifically, higher scores on items imply either less agitation, more satisfaction, or a positive attitude toward one own aging, depending on which factor the item loads.

Table IV presents the measures of overall fit and unstandardized estimates of the total effects of the exogenous variables on the first and second-order factors associated with M_4 . The results pertaining to the U.S. samples will be described first. Age, education, and health have significant effects on subjective well-being in at least three out of four subsamples, whereas findings concerning the effect of sex are inconclusive. Substantively, young-old, better educated, and healthy respondents tend to have higher psychological well-being. Similar conclusions tend to hold for the total effects of exogenous variables on the first-order factors. These results are largely replicated in the Japanese subsamples. The significant effect of health and the lack of sex differences are also sustained by the Japanese data. The effects of age and education are only partially replicated in that only two out of the four subsamples yielded statistically significant results.

The equivalence findings regarding the PGCMS items remained the same even when age, sex, education, and health status were statistically controlled. These results constitute additional evidence that our findings are not an artifact of population heterogeneity. They can also be viewed as an indication that the cohort differences between the American and Japanese respondents are unlikely to account for our observations.

D. In-depth Interviews

To better understand the observed cross-cultural differences uncovered in the four PGCMS items, probed responses made of these items were evaluated. These probed items were included as part of two studies, involving 28 in-depth interviews with Americans and 23 interviews with the Japanese elderly. As a whole, both the American and Japanese respondents reported difficulties in responding to the PGCMS items because they are so vague and general in nature. The Japanese respondents seemed to experience much more difficulty. Whereas the proportions of "don't know" in the American sample ranged from 0.9 to 4.45 percent (median = 1.3%), the percentages of "don't know" in the

Japanese sample varies from 1.7 to 17.6 percent (median = 3.4%). This is probably due to the fact that PGCMS was originally developed in English, and some of the concepts are more alien to the Japanese. This is also consistent with the hypothesis that individuals inherit from their cultures structured vocabularies which limit the possibilities for the interpretation of physical and psychological states (Angel and Thoits 1987). The unstructured responses associated with these four items are described in the following.

Two of the four items that behaved differently in the American and Japanese data were indicators of agitation, "I am afraid of a lot of things," and "I get mad more than I used to." The majority of the American respondents did not endorse the statement, "I am afraid of a lot of things." Among these who did, financial problems, poor health, and crime were mentioned as major concerns. Approximately one half of the Japanese respondents indicated that they were anxious about a lot of things, particularly regarding their own health. In addition, health conditions of their significant others and financial difficulties were mentioned. Confronted by the question, "Do you get angry more often than before?", the majority of American and Japanese respondents said no. Among those who said yes, interpersonal problems, personality, and poor health were reported as reasons for getting angry.

As an indicator of the factor of dissatisfaction, the item "Life is hard for me most of the time" also exhibited different characteristics among the American and Japanese elderly. Only a minority of the American respondents considered life hard most of the time, because of health, finances, and interpersonal problems. In contrast, the majority of Japanese respondents agreed that life is hard most of the time. Deteriorating health, financial problems, interpersonal relations, and the lack of family support were cited as the primary reasons for considering that life is hard. Hardship experienced during young adulthood and World War II was also referenced.

The PGCMS dimension of attitude toward one's own aging consists of five items representing an assessment of the conditions of one's old age, with the emphasis on the contrast between one's current situation and the situation when one was young. One indicator of this factor, "As I get older, things are better/worse than I thought," exhibited different properties between the American and Japanese elderly. About half of the American and Japanese respondents felt that life would get worse as they get older. The major reasons given by the American and Japanese respondents were similar, including retirement, financial difficulties, and health conditions. Conversely, the reasons for considering life would get better in old age include financial security, good physical health, satisfying family life, and adequate social support.

Although there are some differences in the proportions of American and Japanese respondents endorsing the probed items, the framework used by both groups seemed largely similar. One's own health, health of significant others, finances, and interpersonal and marital relations were cited by the American and Japanese respondents as the primary reasons for their responses. However, the attributions of the Americans tended to be more diffuse, including personality

characteristics, fear of crime, and specific phobias in addition to health, finance, and social relations.

The respondents were also probed of four other PGCMS items which were demonstrated to be equivalent between the American and Japanese elderly. These included the following items: (a) Do things keep getting worse as you get older?; (b) Are you as happy now as you were when you were young?; (c) How satisfied are you with your life today?; and (d) Do you get upset easily? There was a remarkable similarity in terms of the reasons suggested as the basis of the answers given by the American and Japanese respondents, thus suggesting that these items were interpreted and understood equally well by these two groups.

Based on the above observations, one may conclude that the American and Japanese elderly seem to have similar understanding of these four PGCMS items and interpret these items within the same framework. This is true although significant differences in frequency distributions were observed in 11 out of the 15 PGCMS items. Differences in frequency distributions must not be confused with those in correlations. The relative frequencies of individual scale items may or may not reflect differences in the structure of their interrelationships. However, the different patterns of factor loadings associated with the four items cannot be attributed to different interpretations by the American and Japanese respondents.

DISCUSSION AND CONCLUSIONS

A major objective of this research was to further examine the cross-cultural comparability of the PGCMS between the Japanese and American elderly. In contrast to the often small and more limited samples used by previous investigators, this research incorporated data obtained from national probability samples of the aged in the U.S. and Japan. The measurement models examined in the present study differed from those analyzed by Liang and his associates (1987) in that no correlated measurement errors were assumed. This study also improved upon previous research by comparing the elderly of the same age range and controlling population heterogeneity. Through in-depth interviews, qualitative data were obtained to provide additional assurance of the accuracy of the translation and that the PGCMS items were meaningful to the Japanese elderly respondents.

The findings from the present research lend further support to the results reported by Liang and his associates (1987). Although the 15-item model of the PGCMS initially derived from the American data manifested acceptable overall fit to the Japanese data, several parameter estimates showed undesirable properties. A revised formulation consisting of 11 PGCMS items exhibited an acceptable fit to both data sets in terms of the overall model fit and measures of component fit. Invariance was demonstrated in terms of the first and second-order factor loadings but not in the measurement error variances. In view of the fact that nonequivalence in measurement errors can be easily adjusted, it is reasonable to conclude that there is no major cross-cultural differences in the

structure of the 11-item version of PGCMS between the American and Japanese aged. Accordingly, the present study provides two additional observations. First, the invariance in the PGCMS can now be extended beyond the urban elderly residents studied by Liang et al. (1987) to the entire aged population in the U.S. and Japan. Second, this finding is also robust despite the elimination of correlated measurement errors from the earlier specifications and the inclusion of exogenous variables.

There has been little research on the differences in the manifestations of subjective well-being between European and non-European cultures in general and between Americans and Japanese in particular. The vast majority of reported observations are concerned with the presence or absence of certain symptoms or complaints instead of the patterns of relationships among them (Marsella et al. 1985). For instance, it has been suggested that the traditional Japanese culture strictly suppresses the expression of anger, particularly among older and presumably more mature persons (Matsumoto et al. 1988). The relative frequencies of such complaints or symptoms may or may not reflect differences in the structure of their interrelationships. In order to obtain a better understanding of these differences, a more in-depth examination of the meanings associated with these items within different cultural contexts is warranted.

What are the research implications of the above observations? First, much more attention needs to be directed to the meanings of various scale items. This is true even when the accuracy of translation appears to have been assured. To obtain a more in-depth understanding of individual stimuli, qualitative strategies should be consistently employed in conjunction with more conventional quantitative methods. Techniques such as random probe (Schuman 1966), focus group, and in-depth interview (McCracken 1988) should be seriously considered, as they will provide culture-specific information valuable for the interpretation of quantitative results and for the generation of new hypotheses. Second, in making quantitative comparisons across different cultures, one should always distinguish the culture-general items from those unique to each culture. Comparisons across different cultures should probably be made only in terms of culture-general items, whereas the different relationships exhibited by the culture-specific components need to be taken into account. The present research represents a step towards these directions.

It should be noted that the Japanese data on the PGCMS were collected within the conceptual framework specified by the American investigators who developed this scale. There are at least two aspects toward which efforts of further elaboration may be directed. First, additional dimensions of subjective well-being may be identified. Based on qualitative and quantitative data collected from American adults, Ryff (1989a, 1989b) reported that certain key aspects of positive functioning have not been included in the existing measures. These involve positive relations with others, autonomy, purpose in life, and personal growth. Second, efforts should be made to explore new dimensions as well as new indicators of psychological well-being within the Japanese culture.

Conceivably, psychological well-being can be defined and operationalized in Japan without making any reference to the PGCMS. The Japanese structure of well-being may differ significantly from that derived from the American culture. Lock (1980) reported the Japanese health beliefs and practices which are quite different from those in the United States. For example, the abdomen, *hara*, in Japanese, has long occupied the place corresponding to that of the heart in western culture. It is commonly believed that if the stomach and spleen were treated therapeutically, then all the organs of the body would automatically be harmonized. The abdomen is also considered the origin of human emotions. Many everyday terms related to the expression of emotion are framed with reference to *hara*. Accordingly, the Japanese conception of well-being may differ significantly from that derived from the American culture. Measures of well-being can be developed in Japan, translated into English, and administered in the United States. This will provide another avenue for assessing the cross-cultural equivalence in well-being.

In addition to the non-comparable items, cross-cultural differences were uncovered in measurement error variances associated with five items. Such differences in measurement error variance can be addressed by employing structural equation modeling of latent variables (Bollen 1989), and therefore, should not be an obstacle in cross-cultural quantitative assessment. Research concerning cross-cultural differences in measurement error is scant. The extensive literature on survey methodology has indicated that measurement error is generally due to (a) effects of interviewers on the respondents' answers; (b) error due to respondents; (c) weaknesses in the wording of survey questionnaires; and (d) the effects of the mode of data collection (Groves 1989). In the present research, efforts were made to apply comparable methodologies in terms of the formats of questionnaires and mode of data collection. However, further systematic research in this regard is clearly needed.

This research also provides a comparative assessment of the equivalence of PGCMS by explicitly taking into account the effects of age, sex, education, and health status. The equivalence of the 11-item PGCMS model remained valid when population heterogeneity was adjusted. Furthermore, the exogenous variables exhibited essentially the same effects on PGCMS in both American and Japanese older people, thus yielding some evidence concerning the causal equivalence of the 11-item version of the PGCMS.

Given the data in the U.S. and Japan were collected almost 20 years apart, the validity of our findings are predicated on the assumption that the same structure of PGCMS can be replicated or the effects of cohort and period do not make a significant difference. Although this is an issue that is not resolved in the literature, and it certainly cannot be resolved with the data available, there is some support for the replicability of the structure of the PGCMS over time. A 1977 longitudinal study entitled, "The Impact of Retirement in Aging and Adaptation," conducted by Robert Atchley in (Atchley 1982) included a 12-item version of the PGCMS. Data were collected via a mail questionnaire sent to all respondents 50 years of age and older living in Oxford Township, Ohio, in July

1975. Liang and Lawrence's (1986) analyses of the 12-item PGCMS included in Atchley's study showed that Liang's earlier measurement model was supported. In another analysis by McCulloch (1989) of a two-wave panel study (from 1976 to 1986) of over 400 rural elderly in the southeastern United States, the hierarchical structure of PGCMS was also replicated.

Finally, this study must be viewed as an initial step toward the accomplishment of a substantial research agenda. In view of the current state of comparative research on aging, such an agenda may be outlined as the following. First, the collection of comparable data and the assessment of conceptual and metric equivalence of key variables deserve the highest priority. Without resolving these concerns, direct quantitative comparisons are difficult and probably unjustified. Once equivalence in measurement is demonstrated, cross-cultural similarities and differences in causal linkages can be properly examined in that substantive explanations for the observed differences can be sought without being confounded by measurement non-equivalence (Markides, Liang, and Jackson 1990). To accomplish this objective, further explication of the theories of comparative aging and considerably more empirical research are essential.

ACKNOWLEDGEMENTS

This is a revised version of a paper presented at the 42nd Annual Meeting of the Gerontological Society of America, November 17–22, 1989, Minneapolis, Minnesota. Support for this research was provided under Grant R37 AG06643 by the National Institute on Aging and matching funds contributed by the Tokyo Metropolitan Institute of Gerontology. Data on the American aged came from the 1968 National Senior Citizens' Survey conducted by Kermit Schooler. This data set was made available by the Inter-University Consortium for Social and Political Research. Neither the original collector of the data nor the consortium bears any responsibility for the analysis and interpretation presented here. Gina Jay, Neal Krause, and George Myers have contributed many useful comments to earlier versions of this paper. The assistance provided by Cathy Fegan and Naoko Muramatsu is gratefully acknowledged. A full-length report containing the details of the analysis is available upon request.

REFERENCES

- Angel, R., and P. Thoits 1987 The Impact of Culture on the Cognitive Structure of Illness. *Culture, Medicine, and Psychiatry* 11:465–494.
- Atchley, R. 1982 The Process of Retirement: Comparing Women and Men. In *Women's Retirement*. M. Szinovacz, ed. Pp. 153–168. Beverly Hills: Sage Publications.
- Bentler, P.M. 1989 EQS: Structural Equations Program Manual. Los Angeles: BMDP Statistical Software, Inc.
- Bentler, P.M., and D.G. Bonnet 1980 Significance Tests and Goodness-of-Fit in the Analysis of Covariance Structures. *Psychological Bulletin* 88:588–606.
- Bollen, K.A. 1989 *Structural Equations with Latent Variables*. New York: Wiley.

- Cook, T.D., and D.T. Campbell 1979 *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally.
- Glenn, N.D. 1977 *Cohort Analysis*. Beverly Hills: Sage Publications.
- Groves, R.M. 1989 *Survey Errors and Survey Costs*. New York: Wiley.
- Jöreskog, K.G. and D. Sörbom 1988 *LISREL 7: A Guide to the Program and Applications*. Chicago, IL: SPSS Inc.
- Kish, L. 1987 *Statistical Design for Research*. New York: Wiley.
- Labouvie-Vief, G. 1985 *Intelligence and Cognition*. In *Handbook of Psychology of Aging* J.E. Birren, and K.W. Schaie, eds. Pp. 500–530. New York: Van Nostrand Reinhold Company.
- Liang, J., and K.A. Bollen 1983 *The Structure of the Philadelphia Geriatric Center Morale Scale: A Reinterpretation*. *Journal of Gerontology* 40:468–477.
- Liang J., and R.H. Lawrence 1986 *Change and Stability in Two Dimensions of the Philadelphia Geriatric Center Morale Scale*. Paper presented at the 39th Annual Scientific Meeting of the Gerontological Society of America, Chicago, Illinois.
- Liang, J., H. Asano, K.A. Bollen, E.F. Kahana, and D. Maeda 1987 *An American-Japanese Comparison of the Structure of the Philadelphia Geriatric Center Morale Scale*. *Journal of Gerontology* 42:37–43.
- Lock, M.M. 1980 *East Asian Medicine in Urban Japan*. Berkeley, CA: University of California Press.
- Lock, M.M. 1987 *Introduction: Health and Medical Care as Cultural and Social Phenomena*. In *Health, Illness, and Medical Care in Japan*. E. Norebeck and M. Lock, eds. Pp. 1–23. Honolulu, HA: University of Hawaii Press.
- Markides, K., J. Liang, and J. Jackson 1990 *Race, Ethnicity, and Aging: Theoretical and Methodological Issues*. In *Handbook of Aging and Social Sciences*, 3rd ed. R. Binstock and L.K. George, eds. Pp. 112–129. New York: Academic Press.
- Marsella, A.J., N. Sartorius, A. Jablensky, and F. Fenton 1985 *Cross-Cultural Studies of Depressive Disorders: An Overview*. In *Culture and Depression*. A. Kleinman and B. Good, eds. Pp. 299–324. Berkeley: University of California Press.
- Matsumoto, D., T. Kudoh, K. Schever, and H. Wallbott 1988. *Antecedents of and Reactions to Emotions in the United States and Japan*. *Journal of Cross-Cultural Psychology* 19:267–286.
- McCracken, G. David 1988 *The Long Interview*. *Qualitative Research Methods*, Vol. 13. Newbury Park, CA: Sage Publications.
- McCulloch, B.J. 1989 *A Longitudinal Investigation of the Factor Structure of Well-Being as Measured by the Philadelphia Geriatric Center Morale Scale*. *Dissertation Abstracts International* 50(6-A):1768.
- Ryff, C.D. 1989a *In the Eye of the Beholder: Views of Psychological Well-Being Among Middle-Aged and Older Adults*. *Psychology and Aging* 4:195–210.
- Ryff, C.D. 1989b *Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being*. *Journal of Personality and Social Psychology* 57:1069–1081.
- Schooler, K.K. 1979 *National Senior Citizen Survey, 1968*. (ICPSR 7626). Ann Arbor, MI: Inter-University Consortium for Political and Social Research.
- Schuman, H. 1966 *The Random Probe: A Technique for Evaluating the Validity of Closed Questions*. *American Sociological Review* 31:218–222.
- U.S. Bureau of Census 1975 *1970 Census of Population and Housing. Evaluation and Research Program, Vol. 11: Accuracy of Data for Selected Population Characteristics as Measured by the 1970 CPS-Census Match*. PHC(E)-11. Washington, DC: Government Printing Office.

Jersey Liang
The Institute of Gerontology and School of Public Health
The University of Michigan
Ann Harbor, MI 48109, U.S.A.

Joan Bennett
The Institute of Gerontology
The University of Michigan
Ann Harbor, MI 48109, U.S.A.

Hiroko Akiyama
School of Social Work
The University of Michigan
Ann Harbor, MI 48109, U.S.A.

Daisaku Maeda
Department of Sociology
The Tokyo Metropolitan Institute of Gerontology
Tokyo, Japan