INT'L. J. OF PUB. ADMIN., 15(7), 1469-1492 (1992)

THE RIGOR OF QWL EVALUATIONS OVER TIME: EVIDENCE OF A MODIFIED POSITIVE-FINDINGS BIAS

Robert T. Golembiewski
Department of Political Science
University of Georgia
Athens, GA 30602
and
Ben-chu Sun
Graduate Program in Public Administration
National Chengchi University
Taiwan, R.O.C.

ABSTRACT

The present design focuses on possible changes over time in the rigor of QWL evaluations, which might support the interpretation of a positive-findings bias (PFB) in a limited range of the study's 22-year period of observation. Other analyses show no major PFB in QWL evaluative studies considered as an aggregate. This analysis adds a longitudinal perspective to these analyses, and finds evidence of PFB in part of the observational period.

Recent detailed analysis of Quality of Working Life (QWL) evaluative studies, N = 231, supports three major conclusions concerning this popular family of organizational interventions. Overall, success rates are substantial--indeed, formidable (1), (2). Moreover, although QWL applications in business enjoy a higher success rate, the record of

public-sector applications does not lag very far behind (3), (4). In addition, considering the entire batch of studies, only small proportions of the variance in success rates can be attributed to differences in rigor (5). No case can be made for a robust positive-findings bias, then, which proposes that QWL success varies inversely with the degree of rigor of evaluative studies. The attractive results of QWL applications cannot be attributed to wimpish methodology.

These important conclusions require a significant caveat, however. The 231 evaluative studies were performed during an interval of more than two decades, and the results summarized above consequently can be faulted in neglecting trends over time. Specifically, one might argue that in the early years of QWL research boosterism overwhelmed methodological scruples and no positive-findings bias would show then because of the limited range of differences in rigor. However, one might continue the argument, more recent years might have raised the methodological and design rigor in evaluative studies, which would depress success rates.

This study tests for possible longitudinal effects in two basic ways. First, and more elementally, success rates will be assessed during 5-year intervals from 1965 through early 1987, the end of the present period of observation. Second, a detailed analysis will search for variations over time between methodological/design rigor and success rates of individual studies. This will permit an estimate of the degree to which attractive

overall QWL success rates can be attributed to incaution or slow learning concerning methodological strictures.

These two basic analytic foci require numerous operational details, by way of introduction. Immediate attention goes to a formidable but unavoidable array of measurement conventions. Subsequently, attention will be directed at QWL success rates over time, and also at the positive-findings bias as a possible explanation for those success rates.

FIVE MEASUREMENT DETAILS

Five emphases related to measurement require up-front treatment: the search for QWL applications; estimates of QWL outcomes; several ways of determining rigor; an overview of the reliability of the numerous coding decisions underlying this research; and introducing the statistical methods used to assess the degree to which a positive-findings bias explains the attractiveness of QWL outcomes, over time. Details are available elsewhere (6).

1. Sketching the Search Process. This program of research sought a comprehensive panel of QWL applications, although Sun (7) isolates over three dozen of them, all of which report substantially-positive results, overall. However, these findings are questioned because of the small sample size of the typical survey of applications, which clusters around 50-60 cases, on average.

This QWL panel includes 231 applications, and results from a determined search for published and

unpublished sources during the interval 1965-1987. The search for published QWL applications takes a conventional, three-pronged approach. In sum:

- 38 bibliographies and survey studies are isolated;
- nearly 100 journals are searched for the full interval: they are almost always in English, but are published in numerous in countries--including the U.S., Canada, Great Britain, India, and Australia;
- about 100 books on QWL and related matters provide both reports of applications as well as citations.

In addition, special attention goes to seeking unpublished studies--in-house reports, dissertations or theses, and the like--on the general theory that published works might over-represent successful applications. The effort in this connection includes two basic thrusts:

- Dissertation Abstracts are systematically reviewed, and available conference proceedings are searched--Academy of Management, OD Network, ASTD, and so on
- letters to approximately 100 persons associated with QWL seeking unpublished materials or citations to published work.

Although by far the largest ever assembled (8), the present panel of 231 applications is not ideal. Consider one central particular. The search process for unpublished sources generates a bit less than 13 percent of all panel entries, which certainly does not proportionately represent unpublished reports. However, the panel's published and unpublished applications have similar success rates (9).

Longitudinal effects in this panel of 231 QWL evaluative studies are assessed by focusing on four 5-year intervals. Each case is assigned to an interval on the basis of the date of first publication, or the date of an unpublished study or report. A few 1987 cases--during the early part of which year the search was closed--are included in the 1981-86 interval. In sum, this distribution of QWL applications exists over time:

	Cases
1965-70	18
1971-75	81
1976-80	68
1981-86	64

A few cases may have appeared pre-1965, but they fall outside our observational interval.

2. Estimating QWL Outcomes. This program of research also seeks to improve on the available literature in a second significant particular. That is, most QWL surveys emphasize "soft" data--changes in attitudes, self-reports about heightened collaboration, and so on. Not only are "soft" data often seen as suspect and somehow inferior to "hard" data, but they also leave survey studies open to the charge that they deal only with "expectation effects" or "Hawthorne effects". In the common view, these involve ephemeral rather than "real" consequences.

Consequently, the criterion for admission to the QWL panel is that each study must provide "hard" or objective data--dollars saved, units produced, turnover experienced, and so on. As fate would have it, all studies in the panel also measure "soft"

effects. This research distinguishes over three dozen separate variables assessing outcomes, about equally divided between "hard" and "soft" variables.

Operationally, the outcomes of each of the 231 QWL applications are assessed twice by three independent raters. The raters were advanced doctoral students in the behavioral and management sciences, and made two assessments of each application--a Hard-Criteria assessment of effects, as well as a Global assessment that takes into additional account "soft" changes in worksite features. For each assessment, raters assign each case to one of four categories:

I highly positive and intended effects
II balance of positive and intended effects
III no appreciable effects
IV negative effects

Conservative conventions govern assessing QWL outcomes, as two rules-of-thumb establish. For rating I, we require that a majority or more of all pre- vs. post-test changes reported not only fall in the intended direction, but also that a majority or more of all changes either attain statistical significance ($P \leq .05$) or are 10 percent or greater than at pre-test. Rating II then goes to those cases whose effects fall mostly or entirely in the intended direction but are not frequently enough such large changes as to receive a I rating. Rating IV goes to any case with a substantial proportion of unintended effects, as well as to all cases generating large contrary effects. Cases rated III have few or no large changes associated with them.

Rater disagreements were treated conservatively. As sub-section 4 below establishes, raters agreed

about almost all assessments--about 9 times out of 10, on average. When the raters still disagreed after discussion, a case was assigned to the lower or lowest of the outcome categories in contention.

3. Measuring Rigor. The critical literature often alleges, with solid reason, that apparently-successful QWL applications merely reflect a degree of rigor incapable of assessing the unattractive things really going on. Few QWL surveys evaluate rigor, and the few exceptions typically rely on a single estimate.

Hence this program of research uses three estimates of rigor, and they relate not only to methodology and design (M/D) but also to the degree to which each study minimizes threats to internal and external validity (I/EV). Although the second M/D measure is intended to materially improve on the other, in this study the two measures correlate .91 and hence share over four-fifths of their variance. Nonetheless, analysis involves all three measures, which are briefly introduced.

Terpstra's M/D Score

Terpstra (10) provides the first evaluation of rigor of studies in planned change. He employs six dimensions, each rated 0 or 1, with an 0-6 range of scores indicating least/most rigor. To conserve space, the six Terpstra dimensions basically are included in the list immediately below of Woodman and Wayne's enhanced version of the original scale. See dimensions 1, 2, 3 restricted to true control groups, 4, 5, and 8.

Woodman and Wayne's M/D Score

Terpstra's (11) formulation was criticized as too limited (12), and this encouraged Woodman and Wayne (13) to develop a 9-dimension estimate of methodology and design rigor. Their version has been slightly modified for present purposes (14), and it assigns 0 or 1 scores for:

1. Sampling strategy: 1 indicates a full census or a representative sampling strategy like random sampling, stratified sampling, or cluster sampling. A nonrepresentative sampling strategy or an unspecified sampling plan is coded 0.
2. Sample size: 1 is assigned when $N > 30$, 0 when $N \leq 30$.
3. Control or comparison groups: 1 indicates the presence of such a group, and 0 indicates its absence.
4. Random assignment utilization: 1 represents use of random assignment, and 0 its absence.
5. Measurement strategy: 1 indicates longitudinal measurement, or measures taken at two or more times; 0 indicates a cross-sectional or one-time measurement.
6. Reliability and validity of measures employed: 1 represents reliability of measures $\geq .6$ with some evidence of validity; 0 is assigned for no reliabilities reported or reliabilities $< .6$ with no validity evidence.
7. Criteria for dependent variables(s): 1 equals presence of some objective criteria; 0 stands for perceptual data only.
8. Significance level: 1 indicates a probability of type I error $\leq .05$; 0 indicates $p > .05$.

9. Use of a statistical analysis procedure: 1 is yes, and 0 is no.

Morrison's I/EV Score

Following Morrison's (15) adaptation of Campbell and Stanley (16), an estimate of internal and external validity provides additional perspective on the rigor of the 231 QWL evaluations. "External validity" refers to the generalizability of results to other settings; and "internal validity" refers to the likelihood that the relationship between an independent and dependent variable actually exists.

Morrison's I/EV score varies from 0 to 12, with 1/0 assignments depending on whether or not each of the threats described below is reduced. Details are available elsewhere (17), but dealing with the eight threats to internal validity involves eliminating or reducing these possible effects: non-treatment factors; aging or development; sensitivity to testing instruments; measurement instability of the instruments; regression to the true mean; non-random selection; experimental mortality over the period of observation; and interaction of selection and maturation.

In addition, Morrison's I/EV score assesses each design's management of four threats to external validity: reactive or interacting effects of testing; interaction of selection effects and the experimental variables; generalizability to other settings; and multi-treatment interference.

4. Assessing Coders' Reliabilities. Evidence implies a substantial reliability in the judgments

by three independent observers. In sum, inter-observer reliabilities average in the low-to-mid .90's for the six measures employed here--two measure of outcomes, and three estimates of rigor. In addition, all differences in ratings were discussed after the reliabilities were determined. Any emergent consensual rating was adopted. If agreement was not achieved, a case got assigned the lower or lowest rating in contention.

5. Statistical Methods. All associations of outcomes with rigor are tested by one-way analysis of variance, for each of four 5-year intervals. In addition, each case that ANOVA shows to include non-random variance ($P \leq .05$) is subjected to an analysis of all possible paired-comparisons to assess both the magnitude as well as the direction of all differences in rigor and outcomes. The Least Significant Difference test, or LSD, as modified for unequal sub-sample sizes, is used for this purpose.

Perfect support for the positive-findings bias requires meeting stringent conditions. Thus each rigor and outcome association would have to show statistically-significant variance by ANOVA. Moreover, every possible pair of rigor and outcome combinations would be in a direction consistent with the positive-findings bias. Specifically, this means that as rigor increases, the outcomes deteriorate for all 5-year intervals. In addition, substantial proportions of variance have to be explained, as estimated by eta^2. Arbitrarily, we define "substantial" as 10-15 percent or more.
Finally, each paired-comparison would involve a difference that is "large" as well as in the

appropriate direction--that is, the LSD test, modified, would reveal statistically-significant differences between all pairs, for each of the four 5-year intervals, were perfect support of the positive-findings bias to exist.

SUCCESS RATES OVER TIME

Success rates do not vary appreciably over time, as Table 1 reflects. All intervals show very much the same proportions of positive effects, clustering tightly around 90 percent in Categories I and II. Post-1971 studies do isolate a few more cases coded in outcome Categories III and IV.

Note that a similar data array for Global Assessment, which takes into account both soft as well as hard criteria, reflects much the same pattern as Table 1. Over time, Category I outcomes drop off a bit in later studies, but more Category II cases appear. In both Table 1 and its unpublished Global counterpart, for all intervals, very close to 90 percent of the cases fall in Categories I plus II. In fact, the arithmetic difference for the eight comparisons averages only a bit over 2.5 percentage points.

RIGOR SCORES OVER TIME

As Table 2 helps show, the rigor of QWL evaluations has increased over time, but significantly so only when the earliest 5-year interval is compared to later ones. All three ANOVAs in Table 2 surpass the .05 level; all 18 paired-comparisons indicate increases over time for each of the three estimates of rigor; and nearly 28

TABLE 1

QWL Success Rates by 5-Year Intervals, 1965-1986, Hard-Criteria Assessment

	(Categories of QWL Outcomes)							
	I		II		III		IV	
Intervals	Highly Positive and Intended Effects		Definite Balance of Positive and Intended Effects		No Apprec-iable Effect		Negative Effects	
	N	%	N	%	N	%	N	%
1965-70	13	72.2	5	27.8	0	0	0	0
1971-75	44	68.8	14	21.9	1	1.6	5	7.8
1976-80	57	70.4	15	18.5	9	11.1	0	0
1981-86	46	67.8	16	23.5	3	4.4	3	4.4
Totals	160		50		13		8	

percent of the paired-comparisons attain the .05 level, by LSD Modified, which far exceeds the proportion attributable to chance alone. However, all significant differences in rigor involve the 1965-70 interval, which in 5 of 18 cases were significantly lower than in later intervals. No significant differences exist between rigor estimates in the three more recent intervals of time. Moreover, eta^2 indicates that about 7 percent of the variance is explained, on average.

TABLE 2

Rigor Scores by 5-Year Intervals

	Means		
Interval	Terpstra M/D Score	Woodman & Wayne M/D Score	Morrison I/E Score
1965-70	2.2	2.6	3.4
1971-75	2.5	3.0	4.2
1976-80	2.8	3.6	4.9
1981-86	3.2	4.1	5.2
ANOVA F =	3.86	4.78	9.91
P =	.01	.003	.0001
eta^2 =	.048	.059	.116

POSITIVE-FINDINGS BIAS OVER TIME

Two basic steps are involved in this test of the positive findings bias (PFB) hypothesis. Thus a single run of the analysis will be illustrated first, and then all runs will be summarized.

1. Illustrative Test of PFB Hypothesis. Table 3 illustrates the approach used here to test the PFB hypothesis. In sum, ANOVA reveals a non-random association that is quite robust. Specifically, as eta^2 indicates, one quarter of the variance is accounted for. The direction of this association is also regular in all six paired-comparisons of

TABLE 3

Woodman and Wayne's M/D Scores and QWL Outcomes, Global Assessment, 1976-80, ANOVA

	M/D Scores			ANOVA	
	N	Means	S.D.	F	Prob.
				8.35	.0001
QWL Outcomes					
I. Highly Positive and Intended Effects	50	2.92	1.79		
				eta^2 = .25	
II. Definite Balance of Positive and Intended Effects	24	4.58	1.53		
III. No Appreciable Effects	4	5.25	.96		
IV. Negative Effects	3	5.67	.58		

outcomes. That is, the rigor of evaluative studies is lowest for outcome I, and increases progressively for each of the other three categories of outcomes.

In addition, the Least Significant Difference test--as modified for unequal sub-sample sizes--permits a specific estimate of the significance of

the differences between all 6 possible pairs of outcome and rigor scores. In the case of Table 3, 3 of the 6 paired-comparisons attain or surpass the .05 level. This far exceeds the proportion of such differences that chance alone would generate.

In sum, Table 3 provides virtually perfect support for the positive-findings bias, considering only Woodman and Wayne's measure of methodology/ design rigor, and considering only the global assessment of outcomes. Perfect support for the PFB hypothesis would meet four criteria:

- the overall ANOVA would achieve or surpass P = .05;
- eta^2 would indicate substantial variance is accounted for;
- all paired-comparisons of rigor and outcomes would show that outcomes become less attractive as rigor increases; and
- the LSD test would show that all differences in paired-comparisons achieve statistical significance as well as fall in the expected direction.

The analysis associated with Table 3 fails to meet only the last of these criteria, and even in that case provides substantial support for the PFB hypothesis.

2. Summary of All Tests of PFB Hypothesis. In all, 47 runs in addition to the one illustrated in Table 3 are employed in the full test of whether the positive-finding bias constitutes a robust explanation of the success rates of the present large panel of QWL evaluative studies. In introductory sum, the 48 separate runs are required to encompass:

- two measures of outcomes: hard-criteria and global assessments
- three estimates of the rigor of methodology and research design
- two ways of arraying outcomes: by four categories; and by collapsing III and IV to test whether their small sub-sample sizes generate unstable patterns of association
- four 5-year time intervals: 1965-70, 1971-1975, 1976-80, and 1981-86.

Table 4 summarizes the 48 separate runs. Note that the number of possible paired-comparisons will vary from run to run because of two factors. Thus a few cells have no entries, and these are distributed throughout the four 5-year intervals. In addition, analyses involving four outcome categories contain a maximum of 6 paired-comparisons of rigor scores, while those analyses which collapse categories III and IV require only three paired-comparisons.

Overall, Table 4 falls far short of the four criteria detailed above for perfect support of the PFB. Thus less than a third of all ANOVAs achieve $P \leq .05$. Moreover, although eta^2 for the minority of significant cases indicates that about 16.1 percent of the variance in rigor and outcomes is accounted for, over two-thirds of the cases reflect random differences only. In addition, less than 6 in 10 of the paired-comparisons of rigor and outcomes fall in a direction consistent with the PFB hypothesis, and only a bit over 12 percent of all pairs attain statistical significance and also fall in the expected direction.

TABLE 4

Overview of 48 Tests of PFB Hypothesis, Summaries of Paired-Comparisons by LSD

Statistically Significant ANOVAS	In PFB Direction	In PFB Direction and Statistically Significant	In Direction Opposite PFB	In Direction Opposite PFB, and Statistically Significant
15 of 48, or 31.3%	106 of 177 or 59.9%	22 of 177, or 12.4%	71 of 177, or 40.1%	0 of 177, or 0%

3. One Possibility of Modified PFB. Despite this clear overall record of failure to support the PFB hypothesis, the data also contain some intriguing hints that one cannot simply discard it. Consider those 15 of 48 cases which showed overall non-random variance, and which are summarized in Table 5. Eighty percent of those 15 cases involve the interval 1976-80.

What do Table 5 and the prominence of 1976-80 entries suggest? Three points make the present argument. First, for the 15 selected cases, Table 5 provides substantial support for a positive findings bias. On average, over 9 in 10 rigor and outcome pairs fall in the PFB direction, and well over a third of the differences between pairs achieve $P \leq .05$. Few contrary cases exist, moreover, and all of those reflect random differences only.

Second, the support for PFB bias appears to be quite similar for both of the two modes of assessing QWL outcomes (see Table 5).

Third, 80 percent of the cases in Table 5 relate to the interval 1976-80, and none to either 1965-70 or 1981-86. Most notably, all 22 of the statistically-significant cases in the PFB direction in the entire panel involve the 15 applications considered in Table 5. See also Table 4. Directly, if only for the 15 cases with non-random variance, Table 5 implies that differences in rigor show a marked and inverse association with QWL outcomes during 1976-80. This association does not characterize the three other 5-year intervals--two earlier intervals and one later than 1976-80. Some

TABLE 5

Summary, Sub-Sample of 15 QWL Studies

Assessment Mode	% Variance Accounted for (eta^2)	In PFB Direction	In PFB Direction and Statistically Significant	In Direction Opposite PFB	In Direction Opposite PFB, and Statistically Significant
Hard Criteria, 7 cases	19.5	20 of 21, or 95.2%	7 of 21, or 33.3%	1 of 21, or 4.8%	0%
Global, 8 cases	12.3	32 of 36, or 88.9%	15 of 36, or 41.7%	4 of 36, or 11.1%	0%

small cell sizes are involved, to be sure, but it does not seem extravagant to suggest that Table 5 implies a modified PFB when increases in rigor are most associated with lower QWL outcomes during 1976-80.

It constitutes a reach, albeit an interesting one, to propose that three phases may characterize the historical evolution of QWL applications. In the earliest phase, the rigor of evaluative studies has little association with outcomes. This might be due to well-known factors--the "magic island" atmosphere associated with breakthroughs, special care in selecting targets, and perhaps especially super-commitment by skillful employees and change-agents.

In an intervening phase, rigor might show up in a modified positive-findings bias. In effect, the novelty has worn off and reputations can be made by rigorous research designs that challenge the no-longer-new wisdom. The resulting research either rejects the once-new initiative, or highlights experience and knowledge for fine-tuning applications.

In QWL research, only a small minority of applications from the middle observational periods --specifically, 12 of the 81 cases during 1976-80, and 3 of 64 cases from 1971-75--reflect marked support for a modified PFB. The high and persisting success rates, along with general increases over time in rigor, imply that experience and theory are being satisfactorily absorbed in practice.

Next comes a third phase of QWL applications, roughly dated from the later 1970s. Increases in

rigor continue and success rates remain high, but research activity drops and PFB effects no longer characterize even a small minority of cases. These trends suggest a period of effective fine-tuning, as when journey-men become more astute in adapting applications to run-of-the-mill sites, develop a larger inventory of skills and attitudes for coping, and gain an understanding of institutional supports to shore-up problematic applications. In this third stage, the positive-findings bias again becomes diluted, and the incidence of research drops because of substantial confidence about what can be done, when, how, and with what consequences. The rate of applications increases in this scenario.

CONCLUSIONS

In sum, analysis does not reveal a positive-findings bias over 20-plus years of observing QWL applications. However, one interpretation of the present data suggests that such a hypothesis does apply--but only in diluted form, and for a limited portion of the present observational time-frame. This interpretation is offered tentatively, of course.

Three other points require attention, largely as caveats. First, the present data batch suffers from an attenuated range of scores in two senses, and each (or both together) can influence the present results. Thus QWL applications generate few cases assignable to the two least-attractive categories of outcomes. Moreover, the distribution of rigor scores is perhaps most appropriately summarized as biased toward the low-to-moderate

reaches of all three of the measures. Conceivably, studies with very high rigor scores might lead to different conclusions than the present ones.

Second, in at least one sense, the strength and direction of associations in this study are understated, if anything. Other analysis (18) finds in public-sector applications no evidence of PFB, while modest support for the PFB hypothesis exists in business applications. The panel analyzed here contains applications from both sectors, which probably operates to dilute the strength of the present associations. The small number of public-sector applications in this panel (N = 44) prevents a detailed test of this surmise.

Third, although the most comprehensive QWL panel available, the 231 cases are not necessarily ideal. For example, less than 15 percent of the cases come from unpublished sources, and this may imply higher success rates due to the widely-assumed preference of editors to publish "positive results." Perhaps this is the case, but such a bias probably would diminish over the long observational period in this program of research. In the present panel, in any case, the success rates of published versus unpublished sources differ only in minor and apparently-random ways (19).

REFERENCES

(1) Golembiewski, Robert T. and Sun, Ben-chu. "Enriching Work and Empowering Employees." Paper presented at Annual Meeting, American Society for Public Administration, Miami, FL., 1989A.

(2) Golembiewski, Robert T. and Sun, Ben-chu. "Positive Findings Bias in QWL Research: A

Comparison of Public and Business Sectors." Public Productivity Review 13 1989B: 145-155.

(3) Golembiewski, Robert T. and Sun, Ben-chu. op.cit., 1989B.

(4) Golembiewski, Robert T. and Sun, Ben-chu. "Positive Findings Bias in QWL Studies." Journal of Management 16 1990A: 39-46.

(5) Golembiewski, Robert T. and Sun, Ben-chu. "Positive Findings Bias in QWL Studies." Journal of Management 16 1990A: 39-46.

(6) Sun, Ben-chu. Quality of Working Life Programs. Unpublished doctoral dissertation, University of Georgia, Athens, GA., 1988.

(7) Sun, Ben-chu. op. cit., 1988.

(8) Sun, Ben-chu. op. cit., 1988.

(9) Golembiewski, Robert T. and Sun, Ben-chu. op. cit., 1990A.

(10) Terpstra, D.E. "Relationship between Methodological Rigor and Reported Outcomes in Organizational Development Evaluation Research." Journal of Applied Psychology 66 1981: 541-543.

(11) Terpstra, D.E. op. cit., 1981.

(12) Bullock, R.J. and Svyantek, D.J. "Positive-Findings Bias in Positive-Findings Bias Research." In K.H. Chung (Ed.), Proceedings: 221-224. Annual Meeting, Academy of Management, Dallas, TX., 1983.

(13) Woodman, R.W. and Wayne, S.J. "An Investigation of Positive-Findings Bias in Evaluation of Organization Development Interventions." Academy of Management Journal 28 1985: 889-913.

(14) Sun, Ben-chu. op. cit., 1988.

(15) Morrison, P. "Evaluation in OD." Group & Organization Studies 3 1978: 42-70.

(16) Campbell, D.T. and Stanley, J.C. Experimental and Quasi-Experimental Designs for Research. Boston, MA.: Houghton Mifflin, 1963.

(17) Sun, Ben-chu. op. cit., 1988.

(18) Golembiewski, Robert T. and Sun, Ben-chu. op. cit., 1989B.

(19) Golembiewski, Robert T. and Sun, Ben-chu. op. cit., 1989B.