愛知淑徳大学大学院論文集―グローバル・コミュニケーション学部篇―　第 4 号　2020.3 1-16

The Measurement Equivalence of Organizational Commitment between the United States and Japan: A Re-analysis of Lincoln and Kalleberg's Data

Tetsushi Fujimoto and Naotaka Watanabe

Abstract

The purpose of this study was to provide a psycho-metrical discretion to the debate on the comparability of organizational commitment among American and Japanese workers in terms of Lincoln and Kalleberg's (1990) study. To tackle with this issue, we employed Item Response Theory (IRT) for detecting the existence of item biases between English and Japanese version of questionnaire. In addition to the original data obtained in Indiana, U.S. (n=4,567) and in Atsugi, Japan (n=3,735), Aichi, Japan data (n=1,025) was newly collected for detecting intra-language as well as inter-language differential item functioning (DIF) of the questionnaires. Through the item level of re-analysis of the data, following facts were found; 1) There is an inter-language bias in four out of six items between original English version and translated Japanese version, 2) Regarding two items in which no bias was detected, there were no statistically significant differences between American and Japanese workers in the average levels of organizational commitment, 3) There is intra-language bias in only one out of six items between two samples collected in Japan. Discussions were addressed about that we need caution when examining the result from Lincoln and Kalleberg's study. That is why it is possible that their result was in part due to translation bias in the measurement tool used in their study.

Keyword

Organizational commitment, Re-analysis, Item Response Theory, Language translation, Differential item functioning

Introduction

A few years ago, several organization scholars debated on the comparability of organizational commitment among American and Japanese workers (Besser 1993; Cole, Kalleberg, and Lincoln 1993). One of the central issues in the debate was Lincoln and Kalleberg's (1985; 1990) interpretation of their finding that the average level of commitment among Japanese workers is no

higher than that among American workers. Lincoln and Kalleberg (1990) discredited this finding that Americans are more committed, claiming that commitment is indeed greater in the Japanese sample than in the US sample, once adjusted for the influence of job satisfaction. Besser (1993) criticized Lincoln and Kalleberg's conclusion, arguing that the evidence of US-Japan commitment gap favoring American workers is real because Americans are more committed than are Japanese in terms of work attitudes (e.g., loyalty), but the Japanese are more committed than are Americans in terms of work behaviors (e.g., low turnover rates). She argued that given the different political-economic and labor market structures in Japan and the United States, attitudinal and behavioral commitment ought to be separately considered when comparing worker orientations in the two countries.

These views are two alternative explanations for why Japanese workers report in attitudinal surveys lower average levels of commitment to their employing organizations than do American workers. Notice, however, that these explanations hold one key assumption in common: there was no measurement bias in organizational commitment in Lincoln and Kalleberg's research.

Our study begins by questioning this assumption. We ask: "Was cross-national measurement equivalence of organizational commitment truly maintained in Lincoln and Kalleberg's study?" Measurement equivalence exists "when individuals with equal standing on the trait measured by the test but sampled from different subpopulations have equal expected observed test scores" (Drasgow, 1987). We wonder if language translation of commitment items had any impact on the level of Japanese commitment. While Lincoln and Kalleberg do not elaborate on this point, we see it as a critical issue because there remains a possibility that the level of commitment reported by Japanese workers was low because the wordings of the scale items in Japanese inhibited Japanese respondents from giving positive responses to the questionnaire statements. Lincoln and Kalleberg defend their measurement strategies by stressing how deliberately they prepared their Japanese questionnaires, translating the items into Japanese, and then back-translating them into English, on numerous occasions to iron out conceptual inequivalence. They also argue that the factor structure c:i organizational commitment is identical in Japanese and American samples, based on confirmatory factor analysis (CFA) of measurement models using (see Lincoln and Kalleberg 1990, p.76 for detail).

We are not fully convinced with Lincoln and Kalleberg's justifications for the measurement equivalence of organizational commitment. Because translation processes necessarily introduces the problem of measurement inequivalence (Ellis, 1989), we believe that the issue of US-Japan commitment gap deserves a more careful and closer scrutiny. At least, some researchers regard

qualitative methods, such as back-translation, as merely a minimum requirement of language translation (Hulin 1987; Ellis 1989). Generally in a confirmatory factor analytic approach, translation equivalence can be detected only in a scale level, but not in an item level. We cannot detect sophisticated wording differences of each item by using CFA.

Our temptation to examine measurement equivalence of commitment also derives from the fact that Lincoln and Kalleberg never reported in detail how their commitment scale statements were translated into Japanese. We contend that even a simplest eyeball comparison of item statements in Japanese and English languages could provide meaningful clues to whether there were no between-country disparities in wordings and their meanings. While such comparisons of survey questions remain intuitive, screening of literal inexactness in the two versions can provide an important ground for questioning the assumed translation equivalence in each item.

Thus, our study examines the US-Japan commitment gap from a perspective of cross-cultural item translation equivalence. We investigate the possibility that the lower levels of commitment for Japanese workers may have been caused by the way questionnaire items were translated into Japanese. We re-analyze Lincoln and Kalleberg’s data using statistical methods based on item response theory (IRT) to evaluate the measurement equivalence of organizational commitment in English and Japanese. We also analyze data from two subsamples of Japanese employees to see whether measurement equivalence is maintained across two separate samples drawn from the same cultural setting. If our results suggest that the commitment items in Lincoln and Kalleberg's study maintain translation equivalence between Japanese and American samples, the meaningfulness of the debate between Cole, Kalleberg, and Lincoln and Besser would be reinforced. If, however, the results indicate that translation equivalence is not maintained between the Japanese and American samples, while confirming equivalence between the two Japanese subsamples, we may be obliged to reconsider the whole debate about organizational commitment disparities between Japanese and American workforce.

The Commitment Scale in Japanese

Lincoln and Kalleberg's commitment scale consisted of six items in total; five items were drawn from the original fifteen-item Porter scale and one item was their original one (see Lincoln and Kalleberg, 1990, pp. 66-67 for detail). Lincoln and Kalleberg version of the Porter scale (referred to as "Porter scale" or “the commitment scale" in the rest of this paper)
in English is as follows:

(1) "I am willing to work harder than I have to in order to help this company succeed."

(2) "My values and the values of this company are quite similar."

(3) "I am proud to work for this company."

(4) "I feel very little loyalty to this company."

(5) "I would take any job in order to continue working for this company."

(6) "I would tum down another job for more pay in order to stay with this company."

These attitudinal statements were translated into Japanese by Lincoln and Kalleberg's research collaborators (Lincoln and Kalleberg. 1985. p.743). The corresponding Japanese statements were as follows (cf. Hanada, 1988):

(1)“Kono kaisha wo seikou saseru tamedeareba futsuu ijou ni doryoku surukoto wo itowanai.”

(2)“Kono kaisha no sosshikifuudo ya shafuu wa watashikojin no kachikan ni pittari atteiru.”

(3)“Kono kaisha no iclii-in de aru to iukoto ni hokori wo motteiru.”

(4)“Kono kaisha ni taishite chuusei-shin to ittamono wa mochiawasete inai.”

(5)“Kono kaisha de hataraki tsuzukeru tamedeareba donoyouna shigotomo hikiukeru.”

(6)“Kono kaisha ni nokoru tamedeareba hokano kaisa kara yoritakai kyuuyo wo dasaretemo sore wo kotowaru.”

We simply compared these statements in the two languages, in order to assess their intuitive comparability. Among the above listed items, the first three statements appeared to include phrases that could possibly convey different meanings in the two languages. We briefly summarize potential problems with the three statements.

(1) Working harder than I have to.

In the first item (“I am willing to work harder than I have to in order to help this company succeed.”), the phrase “to work harder than I have to” was translated into Japanese as “futsuu ijyou ni doryoku surukoto,” which literally means “to make efforts more thanaverage.” A potential problem with this translation lies in the ambiguity in the wording “average”. While the English phrase "more than I have to" seems to connote "beyond the level of my work assignment stipulated by job description," the Japanese phrase "more than average" does not clearly indicate to what levels of workload it refers.

The vast majority of Japanese organizations attempt to diffuse worker responsibilities by not employing job descriptions (Cole, 1979; Lincoln, Hanada, and Olson, 1981; Wood, 1989). Also, it is well known that Japanese workers in general work longer hours than do American workers (Kawahito, 1990). Furthermore, Japanese national opinion polls repeatedly show that diligence and hard work are attributes ranked high in how the Japanese perceive themselves (Cole, 1979; Lincoln and Kalleberg, 1990). Given these facts, Japanese workers could perceive their "average" levels of workload and job responsibility to be already at the ceiling level. If this is the case, the chances that

Japanese respondents endorse this statement may be low, compared with their American counterparts.

(2)The values of this company.

Next, consider the second statement ("My values and the values of this company are quite similar."). In this item, the phrase “values of this company” was translated into Japanese as, "soshikifuudo ya shafuu," which is equivalent to "organizational climate and culture." Still another phrase in this item "quite similar" was translated as "pittari atte iru,” which literally means "perfect fit." Therefore, this item approximates to a Japanese statement that reads "The organizational climate/culture of this company perfectly fits my personal values."

A first problem with this translated statement is that organizational climate/culture and company values do not necessarily refer to the same things. Values are generally broad, nonspecific feelings, such as of good and evil (Hofstede et al., 1990). Corporate values often represent the values of founders and significant leaders of the corporation (Peters and Waterman, 1992). Hofstede et al. (1990) argue that such values of corporate founders and leaders could be translated into organization's culture through shared perceptions of daily practices (i.e., conventions, customs, mores, rules, traditions etc.) among the bulk of the organization's members. With shared practices, values of corporate heroes could permeate into organizational members as culture. Thus, organizational culture may encompass broader daily practices in the company than the company values might connote. Therefore, while the English statement seems to imply the match between a respondent's personal values and the values of corporate founders and leaders, the Japanese statement appears to ask whether one's personal values are perturbed by the practices in broader organizational life.

A second problem is the difference in intensity conveyed by the expressions "quite similar" and "perfect fit" Obviously, the expression "perfect fit" has significantly stronger nuance than being "quite similar." It appears that even when two things are quite similar, it still leaves a small room for them to be different When one thing perfectly fits another, however, it seems to leave little room for difference. Therefore, with this seemingly broader and stronger statement in Japanese, reactions of Japanese respondents to this question could become less positive than those of American respondents.

(3) Working for this company

In the third statement ("I am proud to work for this company."), the phrase "to work for this company" was translated into Japanese as "ichi-in de aru to iukoto", which literally means "to be a

member of (this company)". Phrased this way, the Japanese statement seems to carry an image of worker membership in the corporate community, reflecting the conventional view that Japanese organizations take household structures (Nakane, 1970). On the other hand, the English statement seems to connote a contractual nature of employment relationships. Therefore, there appears to be a subtle difference between the two versions in what a respondent is asked to be proud of. Note, however, that in the fifth item ("I would take any job in order to continue working for this company") the phrase "to continue working for" was not translated as "to continue being a member of (this company)". Instead, the phrase was literally translated as "to continue working for" in Japanese.

Although the above comparisons mostly remain intuitive, they imply that the ways the Porter scale items were translated into Japanese language may have yielded some between-country disparities in the underlying meanings of questions. One striking fact we must note about these three statements is that among six items from the Porter scale used in Lincoln and Kalleberg's study, the cross-cultural differences in average scores reached statistical significance ($p<.001$) only in these three items (see Lincoln and Kalleberg, 1990, p.75). This suggests a possibility that the commitment gap between Japanese and American workers might have emerged because Japanese respondents were guided to score low on the items by the way the questionnaire statements were phrased.

These backgrounds provide a rationale for our analysis of translation equivalence in the commitment scale, based on a more sophisticated statistical method. In the next section, we briefly describe the data and review item response theory.

Methodology

Data　The data used in this study come from the U.S.-Japan comparative study conducted by James Lincoln and Arne Kalleberg between 1981 and 1983 (for details, see Lincoln and Kalleberg, 1990.) Lincoln and Kalleberg gathered questionnaire data from samples of managers, supervisors, and workers employed in a variety of functional departments of the plants. The effective sample size was 4,567 in the U.S. and 3,735 in Japan. The plants studied were located in central Indiana (mostly the Indianapolis metropolitan area) and in the Atsugi area of Kanagawa prefecture.

In the Indiana survey, the plant access rate was 35%, yielding usable data from 52 plants. The samples in both countries include plants of well-known multinational corporations as well as local private companies. Questionnaires were randomly administered to a sample of full-time employees, primarily from production departments in the plants. The questionnaire data collection in Indiana yielded high response rate (65 percent), while the overall response rate from the Atsugi sample was higher (78 percent). The Japanese and the U.S. worker samples were reasonably similar.

We also used Japanese data from a separate worker attitude study conducted in Aichi prefecture in 1993 (N = 1025) (see Aichi Department of Labor, 1994 for detail). In the Aichi study, a revised version of the Porter scale that consists of fifteen items was used to tap organizational commitment. We compared five Porter scale items commonly employed in Aichi and Atsugi studies to tap commitment, in order to determine whether measurement equivalence was maintained for the two subsamples of Japanese workers.

Item Response Theory Item response theory (IRT) is a measurement theory that has been applied in the analysis of translated tests to detect measurement inequivalence and discover cultural and language differences (Hambleton and Swaminathan, 1985; Hulin et al., 1983; Lord, 1980). IRT has been applied to different types of translated tests to assess the measurement equivalence of the original and translated versions (Candell and Hulin, 1987; Hulin, 1987; Hulin, Drasgow, and Komocar, 1982; Hulin and Mayer, 1986; Van de Vijver and Leung, 1977).

IRT makes an assumption that an individual's response to a set of test items can be accounted for by defining characteristics called latent trait (Lord and Novick, 1968). The theory holds that a relationship exists between the latent trait measured by the test and an individual's response to a particular item in the test (Ellis, 1989). Mathematical functions of IRT models relate the probability of response occurring to an item, to the trait measured by the item. This mathematical function is referred to as an item characteristic curve (ICC), a plot of the probability of giving a positive response against a latent trait level.

An ICC represents the regression of item scores on the latent trait, and therefore given an ICC for an item, the probability of a particular response for any given latent trait level can be assessed. An empirical ICC, in general, is a S-shape curve rather than a linear function. When the ICC for an original and a translated test items are statistically identical, this suggests that the original and translated versions are equivalent When the ICCs are not statistically identical, the translated item is said to contain item bias, which is recently referred to as differential item functioning (DIF; Holland and Thayer, 1988; Thissen, Steinberg and Wainer, 1988). A test item displays DIF when individuals who have the same latent trait, but who come from different groups, do not have the same probability of giving a response in the same direction (Hulin et al., 1983).

The IRT model used in this study was a binary response, two-parameter logistic model. The mathematical form of this model is as follows:

$$p_i(\theta) = \frac{1}{1 + e^{-Da_i(\theta - b_i)}}$$

$P_i(\theta)$ is the probability of a positive response to an item i among respondents with a latent trait θ. The a, or discrimination, parameter shows the steepness of the ICC and is proportional to the ICC's slope at the point of inflection. The b parameter, referring to item popularity in attitude measurement and item difficulty in ability measurement, shows the location of the ICC along the latent trait theta continuum. The D is a scaling constant normally set equal to 1.702. Theta (θ) refers to the latent trait estimated by a set of items constitute the scale. While there are several methods for estimating IRT model parameters (Baker, 1992), in this study we employed joint maximum likelihood estimation procedure, using LOGIST 5 program.

Analytic Procedure　The following procedure was adopted for detecting DIF in the commitment scale. First, we conducted factor analysis for each data set to examine the unidimensionality of the scale; i.e., scale unidimensionality is an indispensable condition for IRT analysis. Second, the data from Indiana and Atsugi were compared based on the following steps:

(1) Conducting t-test for each item on the scale using data from Indiana and Atsugi.
(2) Estimating parameters a, b, and θ by joint maximum likelihood; i.e., two parameter logistic model was adopted.
(3) Conducting t-test for each item on the scale, for Indiana and Atsugi samples, using estimated latent trait values, θ.
(4) Equating item parameters a and b.
(5) Depicting the ICCs of each item for Indiana and Atsugi samples, using equated item parameters.
(6) Detecting DIF of each item by area method.
(7) Conducting t-test for the scale, constructed with non-biased (non-DIF) items.

Third, the data from Atsugi and Aichi were compared, following the above steps (2), (4), (5), and (6). Since the original data were obtained in a Likert-type response scale, we transformed the original 7 and 5 point interval data to 0-1 binary data when estimating the item and latent trait parameters. Namely, values 1 and 0 were assigned to indicate positive and negative responses respectively in each item.

Results

We first conducted a factor analysis for the commitment scale separately for the three data sets from Indiana, Atsugi, and Aichi, in order to confirm the usability of these data for an IRT analysis. Is revealed that factor loading was high for the primary factor in each data set, indicating that the scale

maintained a unidimensionality of latent trait θ for all these data. This suggests that these data satisfied a prerequisite for an IRT analysis that a unidimensionality of latent trait be maintained.

In order to re-confirm Lincoln and Kalleberg's finding that the average level of commitment was higher for Americans than for Japanese, we compared mean scores of six items on the scale between Indiana and Atsugi samples. As the results in Table 1 show, in item numbers 1, 3, and 4,

Table 1. Comparisons of Porter Scale Items for American and Japanese Respondents on t-Tests: Means and Standard Deviations.

	American		Japanese		
Items	Mean	(SD)	Mean	(SD)	Difference
ITEM#l "I am willing to work harder than I have to in order to help this company succeed." (5=strongly agree)	3.91	(.89)	3.44	(.98)	.47 *
ITEM#2 "I would take any job in order to continue working for this company." (5 = strongly agree)	3.12	(1.14)	3.45	(1.06)	-.3.3
ITEM#3 "My values and the values of this company are quite similar." (5=strongly agree)	3.15	(1.06)	2.68	(.95)	.47
ITEM#4 "I am to work for this company." (5 =strongly agree)	3.70	(.94)	3.51	(1.02)	.19 *
ITEM#5 "I would turn down another job for more pay in order to stay with this company." (5 =strongly agree)	2.71	(1.17)	2.68	(1.08)	.03

ITEM#6	3.45	(1.13)	3.40	(1.03)	.05
“I feel very little loyalty to this company.”(5=strongly disagree)					

* p<.05

American respondents scored higher than their Japanese counterparts. This is the same result as that reported by Lincoln and Kalleberg in their original study. We also compared a scale-level average scores for American and Japanese respondents, and the results similarly indicated that Americans were more committed than Japanese. Since average scores and standard deviations for these six items are exactly the same as those reported by Lincoln and Kalleberg (1990, p.75), we confirmed that the data we used in this study were the same as the original data in Lincoln and Kalleberg's study.

We also estimated latent trait values of commitment using joint maximum likelihood estimation (JML) for the Indiana and Atsugi samples. While we tested the difference in average values of latent trait between the two countries, we found that the U.S.-Japan difference was statistically significant Thus, this result, along with the disparity in scale― level scores, reinforced the prior finding that Americans are more committed to their organizations than their Japanese counterparts.

The next step of our analysis was to explore a cross-cultural item bias in the commitment scale. First, we equated the estimated item parameters (Crocker and Algina, 1986). Secondly for chi-square testing, we divided the θ continuum from -4 to +4 into 41 equally distanced intervals, and we computed the probability of a positive response (i.e., the height of item characteristic curve, ICC) for both English and Japanese versions of each item. Table 2 presents estimated item parameters, equated item parameters, and results of chi-square tests. The results show that chi-square scores for item numbers 1 through 4 reached statistical significance ($df = 40$, $p<.05$), whereas chi-squares for item numbers 5 and 6 were not statistically significant. Thus, in four out of six items that consist the commitment scale, DIF was detected between Indiana and Atsugi samples.

Table 2. Item Parameters and Chi-Square Results for Indiana and Atsugi Data

	Indiana		Atsugi		
Item#	*b* parameter	*a* parameter	*b* parameter	*a* parameter	Chi-Square
1	-1.775	.526	-.583	1.476	
			-1.020	**1.306**	252.4 *
2	.539	.587	.187	.533	
			-.148	**.489**	55.9 *
3	.343	1.740	1.553	.676	
			1.394	**.598**	106.7 *
4	-.558	2.906	-.694	1.640	
			-1.146	**1.451**	186.7 *
5	1.550	.575	1.499	.591	
			1.333	**.62.5**	1.0
6	-.617	.429	-.503	.625	
			-.930	**.553**	6.4

* df = 40, p<.05

Scores in bold face represent equated item parameters.

Figures 1 and 2 present item characteristic curves (ICC) for item number 3, in which we detected

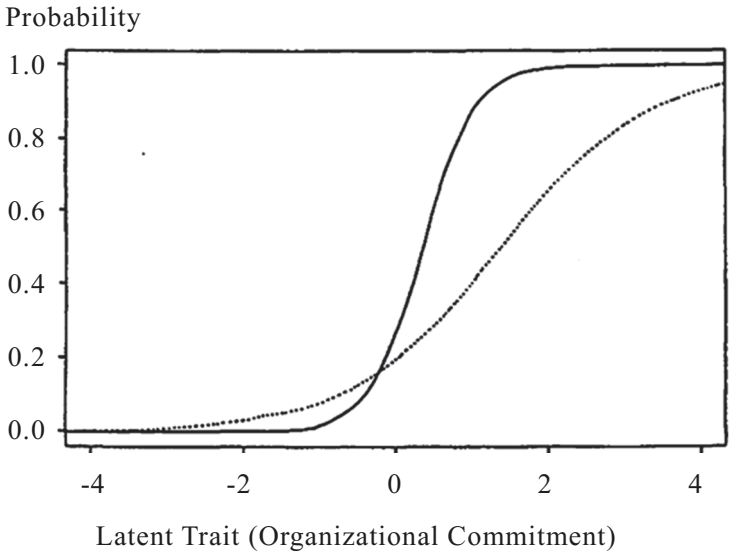

Figure 1. Item Characteristic Curves for the Item # 3 for Indiana and Atsugi Samples
(Solid Line = Indiana, Broken Line = Atsugi)

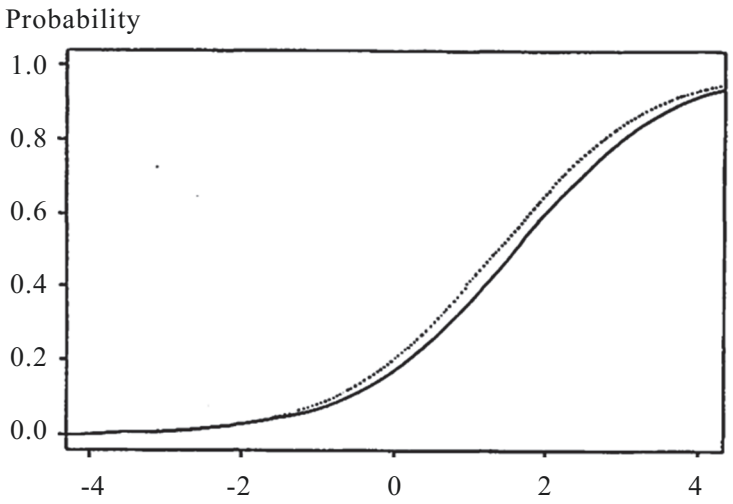

Figure 2.Item Characteristic Curves for the Item# 5 for Indiana and Atsugi Samples
(Solid Line = Indiana, Broken Line = Atsugi)

DIF, and those for item number 6, in which we detected no DIF. As the ICCs for these two items indicate, when an item contains measurement bias, as in item number 3, curves for two versions take different shapes. In contrast, when an item is free of bias, as in item number 6, the shapes of curves resemble each other.

Note that cross-cultural differences in average scores of commitment for item numbers 2, 5 and 6 were not statistically significant, as shown in Table 1 (also reported by Lincoln and Kalleberg). Our findings show that two items in which Lincoln and Kalleberg found no U.S.-Japan mean differences were not affected by measurement bias. Reversely put, three items that Lincoln and Kalleberg did find U.S.-Japan mean differences were indeed contaminated by measurement bias.

Finally, we compared the data from Atsugi and Aichi subsamples to examine DIF among Japanese respondents. For this step, we compared item numbers I through 4 and 6 in Table 1, which were adopted from the original Porter scale. As in the previous step, we first equated the estimated item parameters, then we compared the shapes of ICCs for the two subsamples. Table 3 reports estimated item parameters, equated item parameters, and chi-square results obtained from the analysis. As the results show, while chi-square for item number 6 reached statistical significance, chi-square from the remaining 4 items were not statistically significant Therefore, we detected
differential item functioning between Atsugi and Aichi samples only in one item of the Japanese version of Porter scale.

Table 3. Item Parameters and Chi-Square Results for Atsugi and Aichi Data

	Atsugi		Aichi		
Item#	*b* parameter	*a* parameter	*b* parameter	*a* parameter	Chi-Square
1	-.554	1.443	-.227	.788	
	-.417	**1.244**			29.6
2	.347	.499	-.904	.738	
	.629	**.430**			50.3
3	3.303	.323	3.876	.189	
	4.058	**.278**			35.1
4	-.673	1.675	-.268	1.384	
	-.555	**1.444**			17.1
6	-.438	.602	.997	.887	
	-.239	**.519**			264.9*

* df = 40, $p<.05$

Scores in bold face represent equated item parameters.

Summary and Conclusion

The findings in our re-analysis of Lincoln and Kalleberg's U.S.-Japan comparative data are summarized as follows:

(1) Using statistical methods based on item response theory, we detected a cross-cultural measurement bias (differential item functioning) in *four out of six* items that constitute the Porter organizational commitment scale used in Lincoln and Kalleberg's study.

(2) In three out of the four Lincoln and Kalleberg's commitment scale items in which we detected differential item functioning, the average levels of commitment were higher for Americans than for their Japanese counterparts. In contrast, for the two items in which no bias was detected, there were no statistically significant differences in the average levels of commitment between American and Japanese workers.

(3) Our analysis of the data from two subsamples of Japanese workers showed that in general measurement equivalence of commitment scale items was maintained. In one item, however, differential item functioning was detected even among Japanese respondents.

These findings warn us that we need caution when examining the result from Lincoln and Kalleberg's research (1985; 1990) that the average level of organizational commitment for American workers is higher than that for Japanese workers. Given our findings, it is possible that their result

was in part due to translation bias in the measurement tool used in their study. It may be that the level of commitment reported by Japanese workers was low because the wordings of the measurement tool in Japanese language inhibited Japanese respondents from giving positive responses to the questionnaire statements.

Lincoln and Kalleberg were aware of the possibility that measurement of organizational commitment was not equivalent between American and Japanese respondents in their study. They employed a confirmatory factor analytic strategy to validate their use of English and Japanese versions of commitment scale to tap commitment in the two countries, and their conclusion based on the analysis was that there was no translation bias in their measurement tools. Note, however, that even though confirmatory factor analysis may be appropriate to detect a measurement bias in a scale level, it is not necessarily so in an item level. Our study reveals that even when the factor structure of commitment is identical between American and Japanese samples, as Lincoln and Kalleberg stress, this does not necessarily mean that the problem of measurement inequivalence for each component item is ruled out. Thus, the present study opened up a possibility that at least four items on Lincoln and Kalleberg's version of commitment scale measured different things in the U.S. and Japan. This means that the prior debate about the cross-national disparities of organizational commitment between American and Japanese workforce needs to be reconsidered.

Ambiguity, however, remains in our research. We did not expect to detect a differential item functioning in any item that consisted the Japanese version of Porter scale between Atsugi and Aichi subsamples. Hulin (1987) suggests that when a differential item functioning is detected, it may be due to (1) a problem in item translation itself, and/or (2) the possibility that the item has different meanings for two groups of respondents. Since the same five items on the Japanese version of Porter scale, phrased exactly alike, were used to tap commitment for respondents from Atsugi and Aichi in Japan, the possibility of translation problem is unlikely. Yet, we have no solid explanation so far why the Porter scale item (item #6） in Japanese could convey different meanings for the two subsamples of Japanese workers.

Notes

This paper was originally prepared for presentation at the 1998 Annual Meetings of the American Sociological Association, San Francisco, CA. The content of this paper was partially published in chapters of “Soshiki Shinri Sokuteiron” (1999) and “Soshiki Shinri Testing no Kagaku” (2015) Tokyo: Hakuto Shobo in Japanese language. Upon the readers’ request for seeing the English version, we hereby publish the original document as it is.

References

Aichi Department of Labor. 1994. "Kinrousha Kyouiku Kunren Jittai Chousa Kekka Houkokusho." (Research report on the present state of worker training in Aichi) Aichi Prefectural Government

Besser, Terry L. 1993. “The Commitment of Japanese Workers and U.S. Workers: A Reassessment of the Literature.” *American Sociological Review* 58: 873-881.

Candell, Gregory L., and Charles L Hulin. 1987. "Cross-Language and Cross-Cultural Comparisons in Scale Translations: Independent Sources of Information about Item Nonequivalence." *Journal of Cross-Cultural Psychology* 17: 417-440.

Cole, Robert E. 1979. *Work, Mobility, and Participation: A Comparative Study of American and Japanese Industry*. Los Angeles, CA: University of California Press.

Cole, Robert E., Arne L. Kalleberg, and James R. Lincoln. 1993. "Assessing Commitment in the United States and Japan: A Comment on Besser." *American Sociological Review* 58:882-885.

Lincoln, James R., Mitsuyo Hanada, and Jon Olson. 1981. "Cultural Orientations and Individual Reactions to Organizations: A Study of Employees of Japanese-Owned Finns.” *Administrative Science Quarterly* 26: 93-115.

Lincoln, James R., and Arne L Kalleberg. 1985. "Work Organization and Workforce Commitment: A Study of Plants and Employees in the U.S. and Japan." *American Sociological Review* 50: 738-60.

_____ 1990. *Culture, Control, and Commitment: A Study of Work Organization and Work Attitudes in the United States and Japan.* New York: Cambridge University Press.

Drasgow, Fritz. 1987. "Study of Measurement Bias of Two Standardized Psychological Tests." *Journal of Applied Psychology* 72: 19-29.

Ellis, Barbara. 1989. "Differential Item Functioning: Implications for Test Translations." *Journal of Applied Psychology* 74: 912-921.

Hambleton, Ronald K, and Hariharan Swaminathan. 1985. *Item Response Theory: Principles and Applications.* Boston: Kluwer-Nijhoff.

Hanada, Mitsuyo. 1988. "*Soshiki Shakaika to Soshiki Komittomento.*" (Organizational Socialization and Organizational Commitment) In Mitsuru Wakabayashi and Toshihiro Matsubara (eds.) Soshiki Shinrigaku (Organizational Psychology) pp. 58-81. Tokyo: Fukumura Shuppan.

Hofstede, Geert, Bram Neuijen, Denise Daval Ohayv, and Geert Sanders. 1990. "Measuring Organizational Cultures: A Qualitative and Quantitative Study across Twenty Cases. *Administrative Science Quarterly* 35: 286-316.

Holland, Paul W., and Dorothy T. Thayer. 1988. "Differential Item Performance and the

Mantel-Haenszel Procedure." In Howard Wainer and Henry I. Braun (eds.) *Test Validity* pp. 129-145. Hillsdale, NJ: Erlbaum.

Hulin, Charles L. 1987. "A Psychometric Theory of Evaluations of Item and Scale Translations: Fidelity across Languages. *Journal of Cross-Cultural Psychology* 18: 115-142.

Hulin, Charles, Fritz Drasgow, and John Komocar. 1982. "Applications of Item Response Theory to Analysis of Scale Translations." *Journal of Applied Psychology* 67: 818-825.

Hulin, Charles, Fritz Drasgow, and Charles L. Parsons. 1983. *Item Response Theory: Applications to Psychological Measurement.* Homewood, IL: Dow Jones Irwin. Hulin, Charles and Laura J. Mayer. "Psychometric Equivalence of Translation of the Job Descriptive Index into Hebrew." *Journal of Applied Psychology* 71: 83-94.

Kawahito, Hiroshi. 1991. "Death and the Corporate Warrior." *Japan Quarterly*. April-June: 149-l57.

Lord, Frederic. 1980. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Erlbaum.

Nalcane, Chie. 1970. *Japanese Society*. Berkeley, CA: University of　California Press. Peters, Thomas J., and Richard H. Waterman. 1982. *In Search of Excellence: Lessons from America Best-Run Companies.* New York: Harper & Row.

Thissen, David, Lynne Steinberg and Howard Wainer. 1988. "Use of Item Response Theory in the Study of Group Differences in Trace Lines." In Howard Wainer and Henry I. Braun (eds.) *Test Validity* pp. 147-170. Hillsdale, NJ: Erlbaum.

Van de Vijver, Fons, and Kwok Leung. 1997. *Methods and Data Analysisfor Cross-Cultural Research.* Thousand Oaks, CA: Sage.

Wood, Stephen. 1989. “The Japanese Management Model:Tacit Skills in Shop Floor Participation.” *Work and Occupations* 16: 446-460.