# Challenges of Agreement Scales in Cross-national Surveys: From the Aspect of East Asia

Kuniaki Shishido
Noriko Iwai
Tokio Yasuda

1. Introduction
2. Global Development of Cross-national Surveys
3. Scales of attitude-related questions used in cross-national surveys
4. Issues on Agreement Scales
  4-1. The Number of Choices, Inclusion or Exclusion of a Mid-point
  4-2. Adverbial Expressions in Response Choices and Their Translations
5. Coping with Problems of Agreement Scale
6. Conclusion

## 1. Introduction

Cross-national surveys, which had been conducted mostly in Western countries, have expanded to countries and regions all over the world. In recent years, a number of cross-national surveys in Asia have been actively implemented. The aim of cross-national surveys are to clarify commonalities and differences across cultures, but it becomes more and more difficult to conduct credible surveys with an increase of participating countries and regions and a broadening of cultural difference. Many researchers (e.g., Harkness 1999; Johnson 1998; Jowell 1998; Hayashi and Hayashi 1995) have pointed out that assuring equivalence in concepts and verbal expressions among different cultures is the most important and the most difficult methodological task.

This article focuses on importance and difficulty in establishing culturally sensitive agreement scale. An agreement scale is one of the response scales to measure the degree of agreeableness toward a certain opinion, and is one of the most frequently used response scales to measure attitudes. Because slight differences in question statements and response alternatives fluctuate attitude scores more sensitively than demographic and behavioral scores, attitude measurement is most likely to be a point of dispute in

cross-national surveys. In fact, many studies indicate that designs of response alternatives directly affect response distribution (e.g., Smith 2004; Smith 2008; Presser and Schuman 1980; Bishop 1987; Onodera 2002; Manabe 2000; Manabe 2001).

Measurement equivalence across cultures is critical in conducting cross-national surveys. Johnson (1998) pointed out that cross-cultural equivalence can be classified into procedural equivalence (equivalence of measures and procedures) and interpretive equivalence (equivalence of substantial meanings) and these two kinds of equivalence should be ensured in every phase of the cross-cultural project. Focusing on designing of agreement scale, this article regards the two types of equivalences as follows. Procedural equivalence is achieved when unified agreement scale with the same number of choices and response statements is used in all participating countries and regions. Interpretive equivalence, on the other hand, cannot be achieved by only using the same agreement scale. At least, well-suited translation is essential in order to enhance interpretive equivalence. Besides, it is important to consider patterns of responses unique to each country or region, and design agreement scales which enable proper measurement of attitudes in all countries and regions.

This article investigates several problems on agreement scales and their solutions, not from Western but East Asian viewpoint, especially by introducing agreement scales used in cross-national surveys in Japan.

## 2. Global Development of Cross-national Surveys

A few decades ago, cross-national surveys had been conducted mainly in Western countries, but now it has been expanding worldwide. One of the representative cross-national surveys is the World Values Survey (WVS); it started with 12 countries and regions in 1981, most of which belong to Western countries. In 2004, however, as many as over 80 countries and regions participated in the WVS, including countries and regions in Asia, Africa and Latin America. The International Social Survey Programme (ISSP) launched in West Germany, USA, UK and Australia in 1984, and the number of countries in 2006 survey amounted to 43, including non-Western countries and regions.

In Asia, besides these Western-origin surveys, a number of cross-national surveys have emerged since the beginning of a new millennium. Such examples are the East Asia Value Survey (EAVS), the East Asia Barometer (EAB), the AsiaBarometer (AB) and the East Asian Social Survey (EASS) [1]. Minato (2008) pointed out that although these surveys have different objectives and frameworks, they share two research focuses in common: commonalities and differences among Asian countries and regions, and change of Asian features accompanying the advance of globalization.

Because many existing cross-national survey projects have been led by Western coun-

tries, such projects have employed response scales that are generally used in Western cultures (i.e., a 5-point scale labeling from "strongly agree" to "strongly disagree") for non-Western countries and regions. In recent years, however, cross-national surveys have emerged which do not give priority to comparison within Western countries and regions, and it has become necessary to develop response scales suitable for precise measurement of non-Western societies. Applying response scales generally used in the West to non-Western countries and regions without critical reexamination is fraught with danger of overlooking features of non-Western societies.

## 3. Scales of attitude-related questions used in cross-national surveys

In this section, we examine scales of attitude-related questions practically used in cross-national surveys. Specifically, we refer the ISSP and the WVS as worldwide cross-national surveys, and the EAVS, the EAB, and the AB as cross-national surveys with Asian focus. We also examine Japanese General Social Surveys (JGSS) which implements repeated cross-sectional surveys with a view of comparison with GSS in the USA and of cross-national comparison within East Asia. In order to capture overall characteristics of response scales used in recent cross-national surveys, we classified attitude-related questions based on the type of response scales (Table 1).

Scales can be classified by several ways (Smith 2004; Smith 2008). Bipolar scales that place agreement and disagreement as both ends in symmetrical manner and unipolar scales that place either agreement or disagreement on only one end. Scales are also classified into numerical scales and verbal scales. In the former, choices are shown in numbers and only both ends of them have verbal labels. In the latter, each of choices has a verbal label.

Response scales listed in the six surveys have two features in common: (1) most of the scales are verbal and bipolar scales, and (2) few scales have more than five response alternatives.

There are differences among the six surveys in whether or not a mid-point is included. The ISSP, mainly conducted in Western countries and regions, uses 5-point scales with a mid-point in a large part of its questionnaire. On the other hand, the WVS adopts 2-point or 4-point scales without a mid-point. Response scales used in surveys with Asian focus,

1） The East Asia Value Survey was conducted by the Institute of Statistical Mathematics in Japan. The East Asia Barometer was headquartered by National Taiwan University, and the AsiaBarometer by Takashi Inoguchi at Chuo University in Japan. The East Asian Social Survey is being implemented jointly by JGSS Research Center at Osaka University of Commerce (Japan), Sungkyunkwan University Survey Research Center (South Korea), Institute of Sociology at Academia Sinica (Taiwan) and Department of Sociology at the Renmin University of China and Survey Research Center at the Hong Kong University of Science and Technology (China).

Table 1 Type of Response Scales for Attitude-related Questions Used in Cross-national surveys

| | ISSP-2002 | | WVS-2000 | | EAVS-2002 | | EABS-2003 | | ABS-2003 | | JGSS-2001 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. of questions | % | No. of questions | % | No. of questions | % | No. of questions | % | No. of questions | % | No. of questions | % |
| 2-point scale | 0 | 0% | 33 | 24% | 23 | 32% | 3 | 3% | 3 | 3% | 4 | 4% |
| 3-point scale | 5 | 14% | 15 | 11% | 13 | 18% | 1 | 1% | 10 | 11% | 35 | 32% |
| 4-point scale | 0 | 0% | 54 | 39% | 26 | 36% | 70 | 69% | 35 | 38% | 33 | 30% |
| 5-point scale | 26 | 74% | 17 | 12% | 3 | 4% | 22 | 22% | 44 | 48% | 36 | 33% |
| 7-point scale | 3 | 9% | 0 | 0% | 7 | 10% | 0 | 0% | 0 | 0% | 1 | 1% |
| 10-point scale | 1 | 3% | 20 | 14% | 0 | 0% | 5 | 5% | 0 | 0% | 1 | 1% |
| 11-point scale | 0 | 0% | 0 | 0% | 0 | 0% | 1 | 1% | 0 | 0% | 0 | 0% |
| Verbal Scale | 34 | 97% | 119 | 86% | 65 | 90% | 96 | 94% | 92 | 100% | 94 | 85% |
| Numerical Scale | 1 | 3% | 20 | 14% | 7 | 10% | 6 | 6% | 0 | 0% | 16 | 15% |
| Unipolar Scale | 5 | 14% | 4 | 3% | 11 | 15% | 3 | 3% | 9 | 10% | 19 | 17% |
| Bipolar Scale | 30 | 86% | 135 | 97% | 61 | 85% | 99 | 97% | 83 | 90% | 91 | 83% |
| Scale with a mid-point | 29 | 83% | 29 | 21% | 15 | 21% | 23 | 23% | 45 | 49% | 53 | 48% |
| Scale without a mid-point | 6 | 17% | 110 | 79% | 57 | 79% | 79 | 77% | 47 | 51% | 57 | 52% |
| Total | 35 | 100% | 139 | 100% | 72 | 100% | 102 | 100% | 92 | 100% | 110 | 100% |

such as the EAVS, are mostly those without a mid-point.

## 4. Issues on Agreement Scales

Next, we would like to clarify issues on agreement scales. This article deals with two main issues on agreement scales, which are closely related with each other. The first issue is the desirable number of choices in response scales and whether or not a mid-point is included in scales (Bishop 1987; Bishop 1990; Presser and Schuman 1980; Si and Cullen 1998; Payne 1951; Alwin 2008; Cox 2008; Krosnick and Berent 1993; Krosnick et al. 2008; Smith 1997; Converse and Presser 1986). The second issue is how adverbial expressions should be set and how they should be translated (Onodera 2002; Manabe 2000; Manabe 2001; Mohler et al. 1998; Harkness et al. 2004; Behling and Law 2000).

### 4-1. The Number of Choices, Inclusion or Exclusion of a Mid-point

Table 2 shows agreement scales by different scale points in the six surveys. The table also shows number of question items using each of the scales. All response scales are bipolar and verbal scales. Krosnick and Berent (1993) suggested that reliability of verbal scale is higher than numerical scale. Verbal scales are advantageous, because they clarify the meanings of each scale point and reduce respondents' burden (Krosnick et al., 2008).

Table 2 Agreement Scales by Different Scale Points Used in Cross-national Surveys

| | ISSP-2002 | | WVS-2000 | | EAVS-2002 | | EABS-2003 | | ABS-2003 | | JGSS-2001 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No. of questions | % | No. of questions | % | No. of questions | % | No. of questions | % | No. of questions | % | No. of questions | % |
| 2-point scale | 0 | 0% | 3 | 8% | 3 | 21% | 0 | 0% | 0 | 0% | 2 | 8% |
| 3-point scale | 0 | 0% | 2 | 5% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| 4-point scale | 0 | 0% | 17 | 45% | 11 | 79% | 41 | 100% | 0 | 0% | 19 | 79% |
| 5-point scale | 25 | 100% | 16 | 42% | 0 | 0% | 0 | 0% | 7 | 100% | 3 | 13% |
| Total | 25 | 100% | 38 | 100% | 14 | 100% | 41 | 100% | 7 | 100% | 24 | 100% |

All the agreement scales in the ISSP is a 5-point scale with a mid-point. The WVS uses both 4-point scales without a mid-point and 5-point scales with a mid-point. 4-point scales are also frequently used in cross-national surveys within Asia with the exception of the ABS. These tendencies expose an issue in response scales whether a 5-point scale with a mid-point or a 4-point scale without a mid-point should be adopted.

There has been a dispute in regards to which scale should be adopted. Smith (1997) discussed that the bipolar scale including a mid-point has smaller risk of misleading judgment of respondents, and therefore desirable for cross-national comparison than the unipolar scale or scales without a mid-point. Klopfer (1980) and Krosnick et al. (2008) also suggested that offering a midpoint is desirable because omitting the middle alternative lead respondents to randomly select one of the moderate scale points closest to where a midpoint would appear. On the other hand, Converse and Presser (1986) suggested that middle alternative should not be explicitly provided because providing a midpoint leads losing information about direction in which people lean.

Si and Cullen (1998) investigated whether questionnaires using explicit midpoints produce different results for different cultural groups. They found that East Asian people (China, Japan, Hong Kong) have a greater tendency than Western people (United States, Germany, United Kingdom) to choose middle response categories when offered an explicit midpoint response categories. Besides, East Asian people are less likely than Western people to select either end-point categories. They explained the central tendency of East Asian people from cultural factors. For example, Confucian teaching regarding "the middle way" of East Asia tells its followers that the most important thing they can do is to keep themselves from extremes. Si and Cullen (1998) suggested that when surveys use questionnaires with multiple-rating scale points, researchers do so under the condition of needing to obtain results from respondents with obvious yes-no tendencies, using questionnaire items with even-numbered response categories without explicit midpoint should be more effective in measuring and exposing respondents' real opinions in Asian nation.

Many surveys in Asia commonly avoid a mid-point in response scales. For example, most social surveys in Japan use 4-point scales without a mid-point. The major reason

why a mid-point is avoided in Japan is that the Japanese are less likely to express pros and cons clearly, and their answers tend to concentrate on a mid-point (Hayashi and Hayashi, 1995). Furthermore, response patterns yielded by "neither agree nor disagree" as a mid-point response are similar to the response pattern yielded by "can't choose" (Nakao, 2000). Otani et al. (1999) also argue that the inclusion of response alternatives, such as "neither agree nor disagree," "can't choose" or "it depends" will make it difficult to interpret the results. All most of the Japanese select one of the moderate scale point even if a middle point is omitted (Shishido et al. 2009). Taken together, a scale without a mid-point is favorable in surveys in Japan (Iwai 2003; Sugita and Iwai 2003).

### 4-2. Adverbial Expressions in Response Choices and Their Translations

Here we focus attention on response statement of each response alternative in agreement scales, and examine their adverbial expressions and translations. Especially, we describe response statements in Japan, where translation is apt to be a major concern. As mentioned above, the Japanese do not prefer intensive adverbial expressions, and their responses tend to be centralized. This tendency should be taken into consideration in designing response scales of cross-national surveys.

Distribution of responses varies considerably even within the same 5-point scales, depending on whether they consist of "Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree" or "Agree / Somewhat agree / Neither agree nor disagree / Somewhat disagree / Disagree". Many surveys in Japan use scales starting from "Agree" without adverbs, such as "absolutely", "completely", and "strongly" because in Japan extremely few respondents select choices with intensive adverbial expressions like "strongly" (Shimazaki and Otake, 2000). However, agreement scales including "Strongly agree" are widely used in cross-national surveys.

Next, let us consider translation of response statements. In the case of surveys conducted in Japan, the same response statement in English is translated into different Japanese. Although English statement is the same, each cross-national survey uses different Japanese translations. For example, "Strongly agree" has been translated into Tsuyoku sansei (*strongly approve of*), Mattaku sonotouridato omou (*absolutely I think so*), or Sou omou (*I think so*). Each local research agency translates English statements based on their own experience and judgment, and there is no consensus among them on how to translate each of the specific choices.

Such discrepancy in translation exists in other countries and regions as well as in Japan. Here we examine how the same question statement in the WVS and the ISSP is asked in six countries, namely France, Japan, the Philippines, Sweden, Taiwan and USA. The question in focus is asking the degree of agreeableness of an opinion "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work". A response scale used in the WVS is a 4-point scale such as

"Strongly agree / Agree / Disagree / Strongly disagree". On the other hand, a response scale used in the ISSP is a 5-point scale such as "Strongly agree / Agree / Neither agree nor disagree / Disagree / Strongly disagree".

It should be noted that agreement scales in the two surveys differ in the number of response choices as well as translation. The ISSP uses a 5-point scale with a mid-point, while the WVS uses a 4-point scale without a mid-point. Although we have to take into consideration the difference in the number of choices, it is obvious, as will be shown below, that difference in translation causes more significant effect than difference in number of choices.

Translation of response statements in six countries is shown in table 3. Statements are quoted directly from target language questionnaires used actually in each country and region. Translations are completely consistent in French version of questionnaire, but are more or less different in questionnaires used in other countries. Especially, considerable difference can be seen in Japanese questionnaires.

Table 3 Response Scales for a Given Question used in the ISSP and the WVS

| | Japan | Taiwan | Philippine |
|---|---|---|---|
| ISSP | 1 そう思う | 1 非常同意 | 1 Lubos na sang-ayon |
| | 2 どちらかといえばそう思う | 2 同意 | 2 Sang-ayon |
| | 3 どちらともいえない | 3 無意見 | 3 Hindai tiyak |
| | 4 どちらかといえばそう思わない | 4 不同意 | 4 Hindai sang-ayon |
| | 5 そう思わない | 5 非常不同意 | 5 Lubos na hindai sang-ayon |
| WVS | 1 強く賛成 | 1 很賛成 | 1 Talagang sumasang-ayon |
| | 2 賛成 | 2 賛成 | 2 Sumasang-ayon |
| | 3 反対 | 3 不賛成 | 3 Di sumasang-ayon |
| | 4 強く反対 | 4 很不賛成 | 4 O talagang di sumasang-ayon |

| | USA | France | Sweden |
|---|---|---|---|
| ISSP | 1 Strongly Agree | 1 Tout à fait d'accod | 1 Instämmer starkt |
| | 2 Agree | 2 Plutôt d'accod | 2 Instämmer |
| | 3 Neither agree nor disagree | 3 Ni d'accod ni pas d'accod | 3 Varken instämmer eller tar avstånd |
| | 4 Disagree | 4 Plutôt pas d'accod | 4 Tar avstånd |
| | 5 Strongly Disagree | 5 Pas du tout d'accod | 5 Tar starkt avstånd |
| WVS | 1 Agree strongly | 1 Tout à fait d'accod | 1 Instämmer Helt |
| | 2 Agree | 2 Plutôt d'accod | 2 Instämmer Delvis |
| | 3 Disagree | 3 Plutôt pas d'accod | 3 Tar avstånd Delvis |
| | 4 Disagree strongly | 4 Pas du tout d'accod | 4 Tar avstånd Helt |

These differences in translation led to wide divergence in distribution of responses between the two surveys [2)]. Since the target populations in the two surveys are different, here we focus on respondents aged 20 through 89, who are included in both surveys (figure 1).

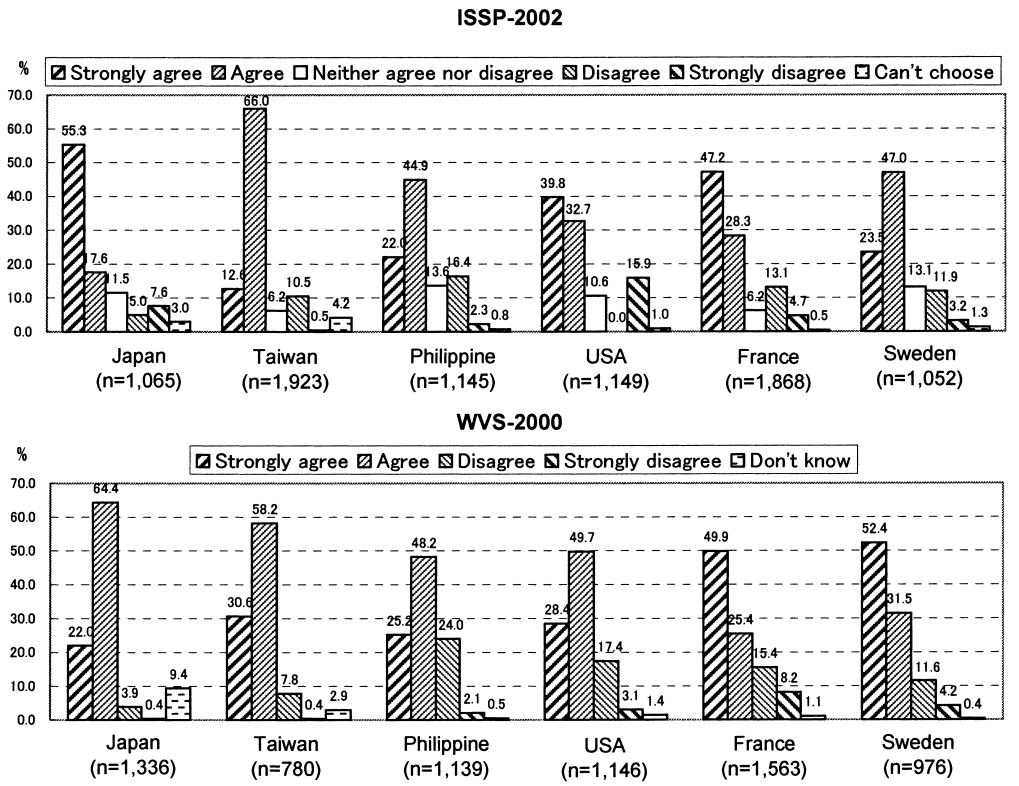


Figure 1 Response Distributions: ISSP-2002 and WVS-2000 for 6 Countries

There is a considerable difference in the translation of response statements between the two surveys in Japan, and the distributions in the two surveys are also different. The result of the ISSP indicates that Japanese respondents are the most radical, to be precise, the most affirmative to working mothers among six countries. However, result of the WVS suggests that the Japanese show ambiguous attitudes without expressing strong agreement. This huge difference in the results from two surveys may stem from translation in their response scales. Japanese local agency of the ISSP does not translate the adverb, "strongly", in response statement, whereas Japanese local agency of the WVS does. Also, Japanese local agency of the ISSP translates "Agree/Disagree" into "I think so (*Sou omou*)/I do not think so (*Sou omowanai*)", whereas Japanese local agency of the WVS translates "Agree/Disagree" into "Favor (*Sansei*)/Oppose (*Hantai*)". Distributions are also divergent in countries where there are slight differences in translation between the two

2） There are also differences between the two surveys other than scales. For example, survey mode of the ISSP-2002 is the interview survey in Japan, Taiwan and Philippine, the self-administered survey in USA, the telephone survey in France, and combination of the mail survey and the telephone survey in Sweden. On the other hand, survey mode of the WVS-2000, the mail survey in Japan, the interview survey in other five countries and regions. However, the differences of their survey modes between the two surveys do not seem a sole cause of the discrepancy of the distributions.

surveys. In France, however, there is no difference in translation and results of the two surveys are similar.

Obviously, one of the obstacles in cross-national surveys is translation of questionnaires. Currently a guideline is being developed to overcome the obstacle (Harkness et al. 2004; Behling and Law 2000). Although researchers are aware that there are specific problems in translation of not only questions but also response statements, the problem is that it is sometimes hard to find culturally sensitive translation.

The real problem pointed out in this article is that there are plural well-suited translations for a single response statement. And slight differences among these translations have a great influence on the distribution of responses. This problem cannot be detected through conventional procedure of back translation. Even if response choices are back translated, there is still a large chance to make several translations for the same English response statement across cross-national survey projects.

## 5. Coping with Problems of Agreement Scale

This article examined current conditions and problems of agreement scales in cross-national surveys. The discussion so far can be summarized into the two points:

(1) In Western countries 5-point scales with a mid-point are generally used, while 4-point scales without a mid-point are common in cross-national surveys in East Asia. There is a dispute in regards to which scale should be adopted.
(2) Intensive adverbial expressions such as "strongly" is commonly used in agreement scales of cross-national surveys. However, translations of these adverbs varied depending on research agencies. As a result, different surveys show inconsistent features of the same country.

This section examines potential solutions of these two problems. First, it is necessary to check target language questionnaires as well as source language questionnaires and see how choices are translated, when analyzing data of cross-national surveys. It is dangerous to analyze the data by only seeing "value labels" of distributed dataset. In many cases meanings of the response choices are shown in English, and it is impossible to understand inexplicable difference of distribution by seeing only value labels in English, as we observed in the case of the ISSP and the WVS.

Second, it is better not to offer the middle point in response scales if the direction in which people are leaning on the issue is the type of information wanted (Presser and Schuman 1980; Payne 1951). If the middle point should be kept in response scales because of some realistic reasons, 7-point scales may be desirable for cross-national surveys,

especially when Asian countries will be included in analysis. In countries like Japan, responses are likely to be congested in a mid-point. However, such situation may be improved by adopting a 7-point scale rather than a 5-point scale. A 7-point scale is considered to measure degree of agreement and disagreement of respondents more properly, because both strong and weak (e.g. "Somewhat agree") adverbial expression can be contained in the scale. Previous studies in Western countries showed that increasing the number of response choices alleviated the congestion to a mid-point (Mangione 1995; Groves et al. 2004). Cox (2008) also pointed out that overuse of the neutral category by respondents can generally be avoided by providing them with an adequate number of reasonable response alternatives. Furthermore, increasing the number of choices, especially seven plus or minus two, improves reliability and validity (Krosnick and Fabrigar 1997; Groves et al. 2004; Schwarz and Hippler 2004; Lehmann and Hulbert 2008; Masters 2008; Alwin 2008). The authors conducted an experimental survey in Higashiosaka city in 2005 with split-ballot method [3]. As shown in Table 4, the 7-point scale, compared with the 5-point scale, restrained responses from centralizing.

Table 4 The Number of Responses to the Middle Alternative out of 18 Items

| | 5-point scale | 7-point scale |
|---|---|---|
| 0-4 | 21 (*15.4%*) | 73 (*26.0%*) |
| 5-9 | 75 (*55.1%*) | 150 (*53.4%*) |
| 10-14 | 35 (*25.7%*) | 55 (*19.6%*) |
| 15-18 | 5 ( *3.7%*) | 3 ( *1.1%*) |
| Total | 136 (*100%*) | 281 (*100%*) |

$X^2 = 9.43,\ df = 3,\ p < .05$

Third, countries and regions participating in cross-national survey projects should make an effort to examine intensiveness of adverbial expressions and to enhance equivalence of the expressions within participant countries and regions. As for adverbial expressions in agreement scales, an approach of Research into Methodology of Intercultural Surveys (MINTS) project by ZUMA in Manheim, Germany and National Opinion Research Center (NORC) at the University of Chicago is of reference. This project examined equivalence of agreement scale based on "direct rating approach" which quantifies and measures the impression of the respondents from the wording of response statements. Using a value between 0 and 20, the respondents were asked to score the strength of 28 response statements, which used adverbs to express various shades of agreement or disagreement ranging from "completely disagree" to "completely agree" (Mohler et al. 1998).

3） The survey was conducted in July 2005 through mailing method. Target respondents were 750 adults aged 20 through 89, chosen with two-stage random sampling method, and 417 valid responses were collected.

Onodera (2002) reexamined the approach of the MINTS project in Japan and found that, like USA and Germany, variety of adverbs could be sorted into seven stages in the light of intensiveness.

Forth, given the differences in the distribution of responses possibly caused by translation differences in response statements, each English response statement should be translated identically among cross-national surveys in order to compare the results. However, changing translation of response statements faces an enormous hurdle because most of the current cross-national surveys are repeated cross-sectional. In other words, changing translation of response statements makes time-series comparison impossible. Even so, it is desirable to seek a way to integrate translation by sharing information among local research agencies.

## 6. Conclusion

Issues of agreement scale discussed in this article are becoming more and more important, along with worldwide development of cross-national surveys. Of course, the issues in this article are only a part of methodological problems in cross-national surveys. As Jowell (1998) pointed out, researchers should be cognizant of situation and culture of the countries in focus before they conduct comparative analysis and interpret results of the analysis. Furthermore, plenty of fundamental methodological problems exist among countries participating in surveys, such as matching target respondents, sampling method, data collection mode etc. Countries included in cross-national surveys have different social conditions and different practice of surveys. Therefore, there are always considerable disconnect between "desirable" and "possible" methods in cross-national surveys.

Compared with the problems above, designing response scales is relatively feasible and has immeasurable effect. The significance of culturally sensitive design of response choices should be emphasized particularly in cross-national surveys.

Issues of agreement scales discussed in this article should gain more attentions from scholars. Now that cross-national surveys have broadened their coverage to non-Western cultures, and the surveys have included countries and regions with wide variety of culture, efforts to solve the problems pointed out in this article certainly contribute to the development of cross-national surveys.

## References

Alwin, Duane F. 2008. "Feeling Thermometers versus 7-Point Scale: Which Are Better ?" In *Attitude Measurement Volume 2 Designing Direct Measures*, ed. Caroline Roberts and Roger Jowell, 242-261. London: Sage Publications.

Behling, Orlando, and Kenneth S. Law. 2000. *Translating Questionnaires and Other Research Instruments: Problems and Solutions.* London: Sage Publication.

Bishop, George F. 1987. "Experiments with the Middle Response Alternative in Survey Questions." *Public Opinion Quarterly* 51:220-232.

Bishop, George F. 1990. "Issue Involvement and Response Effects in Public Opinion Surveys." *Public Opinion Quarterly* 54:209-218.

Converse, Jean M., and Stanley Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire.* London: Sage Publication.

Cox, Eli P., III. 2008. "The Optimal Number of Response Alternative for a Scale: A Review." In *Attitude Measurement Volume 2 Designing Direct Measures,* ed. Caroline Roberts and Roger Jowell, 208-240. London: Sage Publications.

Groves, Robert, M. 2004, *Survey Errors and Survey Costs.* New York: John Wiley & Sons, Inc.

Groves, Robert M., Fowler J. Floyd, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2004. *Survey Methodology.* New York: John Wiley & Sons, Inc.

Harkness, Janet. 1999. "In Pursuit of Quality: Issues for Cross-national Survey Research." *International Journal of Social Research Methodology* 2(2):125-140.

Harkness, Janet, Beth-Ellen Pennell, and Alisu Schoua-Glusberg. 2004. "Survey Questionnaire Translation and Assessment." *In Methods for Testing and Evaluating Survey Questionnaires, ed.* Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer, 453-473. New York: John Wiley & Sons, Inc.

Hayashi, Chikio, and Fumi Hayashi. 1995. "Kokuminsei no Kokusai Hikaku (Comparative Study of National Character)." *Proceedings of the Institute of Statistical Mathematics* 43: 27-80.

Iwai, Noriko. 2003. "JGSS Project (2) Chosa-hoho to Chosa-komoku (JGSS Project(2) Research Designs and Questionnaires)." *Statistics* Nov: 48-55.

Johnson, Timothy P. 1998. "Approaches to Equivalence in Cross-Cultural and Cross-National Survey Research." *ZUMA-Nachrichten Spezial No3 Cross-Cultural Survey Equivalence:* 1-40.

Jowell, Roger. 1998. "How Comparative is Comparative Research?" *American Behavioral Scientist* 42-2: 168-177.

Klopfer, Frederick J. 1980. "The Middlemost choice on attitude items: Ambivalence, Neutrality, or Uncertainty? *Personality and Social Psychology Bulletin* 6: 97-101.

Krosnick, Jon A., and Leandre R. Fabrigar. 1997. "Designing Rating Scales for Effective Measurement in Surveys." In *Survey Measurement and Process Quality,* ed. Lars E. Lyberg, Paul Biemer, Martin Collins, Edith D. De Leeuw, Cathryn Dippo, Norbert Schwarz , Dennis Trewin, 141-164. New York: John Wiley & Sons, Inc.

Krosnick, Jon A., Charles M. Judd, and Bernd Wittenbrink. 2008. " The Measurement of Attitudes." In *Attitude Measurement Volume 2 Designing Direct Measures,* ed. Caroline Roberts and Roger Jowell, 1-85. London: Sage Publications.

Krosnick, Jon A., and Matthew K. Berent. 1993. "Comparisons of party identification and policy preference: The impact of survey question format." *American Journal of Political Science,*

37, 941-964.

Lehmann, Donald R., and James Hulbert. 2008. "Are Three-Point Scales Always Good Enough ?" In *Attitude Measurement Volume 2 Designing Direct Measures*, ed. Caroline Roberts and Roger Jowell, 194-200. London: Sage Publications.

Manabe, Kazufumi. 2000. "Kokusai Hikaku Chosa ni okeru response scale no Tokasei ni kansuru Kenkyu(1) (A Study of Equivalence of Expressions from response scales Used in Cross-National Survey Research(1))." *Bulltin of the School of Sociology, Kwansei Gakuin University* 86: 107-116.

Manabe, Kazufumi. 2001. "Kokusai Hikaku Chosa ni okeru response scale no Tokasei ni kansuru Kenkyu(2)(A Study of Equivalence of Expressions from response scales Used in Cross-National Survey Research(2))." *Bulltin of the School of Sociology, Kwansei Gakuin University* 89: 107-121.

Mangione, Thomas W. 1995. *Mail Surveys: Improving the Quality*. London: Sage Publications.

Masters, James R. 2008. "The Relationship between Number of Response Categories and Reliability of Likert-Type Questionnaires." In *Attitude Measurement Volume 2 Designing Direct Measures*, ed. Caroline Roberts and Roger Jowell, 201-207. London: Sage Publications.

Minato, Kunio. 2008. "Cross-national Social Survey in East Asia World Values Survey, ISSP, AsiaBarometer, Asian Barometer, East Asia Value Survey and EASS." *JGSS Monographs* 7: 159-193.

Mohler, Peter, Ph, Tom W. Smith, and Janet A. Harkness. 1998. "Respondents' Ratings of Expressions from response scales: A Two-Country, Two-Language investigation on equivalence and translation." *ZUMA-Nachrichten Spezial No3 Cross-Cultural Survey Equivalence:* 159-184.

Nakao, Keiko. 2002. "Sentakushi no Sentaku: Nichi-bei Chosa Data niokeru Shokugyo Kachikan no Hikaku kara erareta Chiken (Selection of Alternatives in Multiple Choice Questions)," *Proceedings: 30th Conference of the Behaviormetric Society of Japan:* 352-353.

Onodera, Noriko. 2002. "Hizyoni to Kanari de kotonaru Kaito: Kokusai Hikaku Chosa ni okeru Sentakushi Hyogen no Kento (Answers Depend on Adverbial Terms Used in Questions: A Review of Expressions of Degree in Choices Used in International Comparative Studies)." *The NHK Monthly Report on Broadcast Research*, January: 62-75.

Otani, Shinsuke, Eiji Kinoshita, Noriaki Goto, Hiroshi Komatsu, and Takeshi Nagano. 1999. *Syakai-Cyosa eno Approach: Ronri to Hoho (Approach to the Social Surveys: Logic and method)*. Tokyo: Minerva Publishing Co. Ltd.

Payne, Stanley. 1951. *The Art of Asking Questions*. Princeton: Princeton University Press.

Presser, Stanley, and Howard Schuman. 1980. "The Measurement of a Middle Position in Attitude Surveys." *Opinion Quarterly* 44:70-85.

Schwarz, Norbert, and Hans-Jurgem Hippler. 2004. "Response Alternatives: The Impact of Their Choice and Presentation Order." In *Measurement Errors in Surveys*, ed. Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 41-56. New York: John Wiley & Sons, Inc.

Shimazaki, Akihiko, and Nobuyuki Otake. 2000. "*Shakaichosa no Zissai: Toukeichosa no Houhou to Deta Bunseki (The Actual Social Survey: Method of Statistical Research and Analysis of Data)*" Tokyo: Gakubunsha.

Shishido, Kuniaki, Noriko Iwai, Tokio Yasuda. 2009. " Designing Response Categories of Agreement Scales for Cross-national Surveys in East Asia: The Approach of the Japanese General Social Surveys " *International Journal of Japanese Sociology* 18:97-111.

Si, Steven X. 1998. "Response Categories and Potential Cultural Bias: Effects of an Explicit Middle Point in Cross-Cultural Surveys." *The International Journal of Organization Analysis* 6(3):218-230.

Smith, Tom W. 1997. "Improving Cross-National Survey Research by Measuring the Intensity of Response Categories." *GSS Cross-National Report* 17. Chicago: National Opinion Research Center, University of Chicago.

Smith, Tom W. 2004. "Developing and Evaluating Cross-National Survey Instruments." In *Methods for Testing and Evaluating Survey Questionnaires*, ed. Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer, 431-452. New York. John Wiley & Sons, Inc.

Smith, Tom W. 2008. "Developing Comparable Questions in Cross-National Surveys." In *Attitude Measurement Volume 4 Expanding the Measurement Horizons*, ed. Caroline Roberts and Roger Jowell, 242-261. London: Sage Publications.

Sugita, Hizuru, and Noriko Iwai. 2003. "JGSS Project (3) Sokutei Shakudo to Sentakushi (JGSS Project (3) Measurements and Wording of Questions)." *Statistics* Dec: 49-56.