

How to Design a Survey Questionnaire by Employing a Weighting Method

Yuji Sato

Graduate School of Policy Science, Mie Chukyo University
1846 Kubo, Matsusaka, Mie, 515-8511 Japan
ysatoh@mie-chukyo-u.ac.jp

Summary: The objective of this study was to examine the effectiveness of multiple-choice method, ranking method, rating method and the application of Analytic Hierarchy Process in measuring human perception. The AHP not only clearly identifies the most important alternative but also the preference for each alternative by each decision maker. Therefore, using AHP to analyze the decision-making process may result in a precise clarification of preference for alternatives. Based on survey researches on social issues, the results offered some evidence that the application of the AHP was superior to traditional questionnaire methods in representing human perceptions.

Keywords: survey questionnaire, multiple-choice, ranking, rating, Feeling Thermometer

1. Introduction

Questionnaire design for survey research, such as public opinion polls, presents one of the biggest challenges for survey researchers in terms of accuracy in measuring respondents' perceptions (Traugott and Lavrakas 2000). Consequently, many ways of asking questions have been proposed and much discussion has been generated. One rating scale—called a Feeling Thermometer (FT, Kabashima 1998)—was extensively used in survey questionnaires, which ranges from 0, the coldest feeling toward alternatives, to 100, the hottest, with 50, being neutral. In surveys, this method asks respondents to express their perceptions by indicating their "temperature" for each alternative for a given question. Although this method helps respondents precisely clarify their judgments for each alternative, consistency among responses to the alternatives is not always satisfactory (Sato 2005).

A more traditional method for measuring respondents' perceptions is the multiple-choice question format, which has been thought to be well suited to questionnaire formatting because respondents find the questions easy to answer and they allow researchers to easily identify the main concerns of the respondents (Jerard 1995; Downing 2004). This method takes two different forms: one is simple multiple-choice (SMC); the other is modified multiple-choice (MMC, Sato 2004). In the SMC format, respondents must choose one from among the given alternatives. The SMC identifies only the most important alternative for each respondent, thus preventing the respondent from expressing his or her preference concerning a selected alternative over the others. Moreover, no information regarding the relationship among the non-selected alternatives is derived (Sato 2004). In the MMC format, respondents have the option of indicating their top-two (or more) alternatives. Since respondents are allowed to express their preferred alternatives, the MMC can be expected to be an effective way to make up for the lack of information incurred by the SMC. Nevertheless, the difference in the degree of importance among the selected alternatives is not clarified, nor is the information concerning non-selected alternatives reflected in the results (Sato 2004). Let us consider the case asking respondents why they are non-partisan, for example. On a question designed in the SMC format, respondents must express their opinion by choosing one from among the reasons given. Respondents with a definite reason regarding the issue could choose one alternative without confusion if they found that it exactly represented their perception. This format could be expected to function quite well for these respondents. On the other hand, it might be that some respondents are non-partisan for no particular reason, while others are non-partisan for complex reasons. The MC would not be an effective format for those respondents.

Another scale that has been applied is the ranking scale used by Ronald Inglehart and Paul Abramson (1993) in their World Values Survey. This method asks respondents to rank all given alternatives in a

question, from the most preferred to the least, thus allowing researchers to identify a respondent's preference order for all alternatives. The problem with this method, however, is that the more alternatives a questionnaire offers, the more difficult it is for the respondent to answer (Inglehart and Abramson 1993). Another drawback to this approach is that it does not allow for ties (Sato 2003). For example, let us consider the case asking executive staff members of a prefectural government who have authority for final budget decisions "which governmental projects should be budgeted with high priority for next year?" On a question designed in the ranking format, respondents must express their opinion by ranking all projects given in the question. Respondents with definite preferences on the issue could rank all the projects without hesitation. On the other hand, some respondents might have no definite preference concerning the issue while others might have many ties in the priority for projects.

One possible option for formatting questionnaires is to apply the AHP, a popular method for decision-making developed by T. Saaty (1977, 1980). Since it was released, many individuals and groups in various fields have used the AHP because of its user-friendly interface for multi-criteria decision-making (Vargas, 1990). In the AHP, data from a decision-maker's judgments, called pairwise comparisons, are aggregated, and the degree of importance of each alternative is quantified in the AHP. This procedure identifies not only the most important alternative but also the preference for all alternatives for each decision-maker (Crawford and Williams, 1985). Using the AHP to analyze the decision-making process, therefore, results in a precise clarification of respondents' preferences for alternatives (Sato 2007).

In this study, we compared the answers to four sets of questions on a particular issue, each formatted in a different way. The first two of the four sets consisted of questions formatted using two different types of the MC method and the AHP; the third set consisted of two pairs of questions formatted using the ranking method and the AHP; the fourth set consisted of two pairs of questions formatted using the rating method and the AHP. We then evaluated the four methods in terms of appropriateness for representing each respondent's perceptions.

First, we focused on the difference of the aggregated ranking of alternatives across all respondents between the MC and AHP formats. The ranking derived from the SMC implies aggregated plurality, while that elicited from the AHP suggests aggregated intensity.¹ Since both rankings reflect the pattern of responses for alternatives, they are likely to produce similar results. In addition, we evaluated the effectiveness of the MMC in terms of its ability to make up for the lack of information incurred by the SMC. Since the MMC is a type of the MC question format that allows respondents to indicate their second-best alternative, it may reflect each respondent's preference for alternatives more precisely than does the SMC. To compare the two methods, three questions—formatted in the SMC, MMC and AHP—on the same issue were posed: each asked about the reasons that respondents were non-partisan. Details of the data set are shown in Sections 2.1 and 2.2.

Second, we compared the preference orders of alternatives across all respondents between the ranking and the AHP formats, and evaluated these two formats in terms of the appropriateness for representing each respondent's perception. The aggregated ranking of alternatives across all respondents derived from the ranking method implies preference order of alternatives for respondents. Similarly, aggregated weight for alternatives across all respondents elicited from the AHP also implies preference order of alternatives for respondents. Since both preference orders reflect the entire trend concerning each alternative of a population, these two methods are likely to produce similar results. To compare the two methods, two different sets of a pair of questions on a particular issue were used: one concerned an abstract issue—the basic concepts of refining governmental program policy; the other related to a concrete issue—governmental projects with high priority. Details of the data set are shown in Section 2.3.

Third, we focused on the difference of the aggregated weight of alternatives across all respondents between the FT and the AHP formats. The weight derived from FT is based on absolute evaluation method, while that elicited from the AHP is based on comparative evaluation method. Since both aggregated weight reflect the pattern of responses for alternatives, they are likely to produce similar

¹ From discussion at the May Conference, University of Michigan, May 2000.

results. In addition, we evaluated the effectiveness of each method in terms of its ability as independent variables in regression analyses. In the above-mentioned comparisons, a data set obtained from a 2004 survey on public opinion was employed. To compare the two methods, two types of questions on a particular issue were employed, each formatted in a different way, one using the FT and the other using the AHP. Each asked about respondents' intention to vote for a party in the next election. Details of the data set are shown in Section 2.4.

2. Outline of surveys

In this study, we compared the three above-mentioned methods—the MC method, the ranking method, rating method, and the AHP method—using survey data. The data sets were obtained from the following four surveys. For the AHP, respondents were asked to respond to a series of redundant pairwise comparisons. We thus needed to take into account the possible inconsistency of a pairwise comparison matrix in analyzing elicited weights (Webber, Apostolou and Hassell, 1997). In this paper, however, we do not cut off those samples exceeding the threshold of C.I., such as 0.15, because the respondents whose C.I. exceeded the threshold are social existences, such as constituencies in elections, too.

2.1. Survey 1 (January 1999)

Survey 1 was carried out in January 1999, one month after the coalition cabinet of the Liberal Democratic Party and the Liberal Party was established in Japan. Respondents were 834 students at a university in Japan. The purpose of the survey was to identify "the political attitude of students when a coalition cabinet was established." The survey included 30 questions. In Q.12, respondents were asked their party identification, and for only the 398 respondents who answered "non-partisan" to the question, the following three sub-questions were posed in three ways to determine why they were non-partisan. The reasons offered were: Too much political realignment; Political apathy; Non-confidence with party and politician; and Corruption of political ethics. Q.13 (hereinafter referred to as Q1S) and Q.29 (Q1M) in Survey 1 were respectively formatted by the SMC method, which requires respondents to choose only one from among four given reasons, and by the MMC format, which gives respondents the option of indicating their second-best alternative. Q.26 (Q1A) was formatted in the AHP system, in which respondents are required to conduct pairwise comparisons across all possible combinations of reasons.

2.2. Survey 2 (April 2001)

Survey 2 was conducted in April 2001, the month when graduating students usually begin to start looking for a job for the next fiscal year in Japan. In this survey, respondents were 323 students of a university in Japan. The intellectual purpose of this survey was to clarify "the main concerns of Japanese graduating students as they begin their job search." The main concerns offered were: Job specifications; Welfare program of the company; Salary; and Place of employment. In this survey, three differently formatted questions were posed, each asking the respondents' main concerns of their job-search activities. The first question (hereinafter referred to as Q2S) was formatted in the SMC, the second question (Q2A) in the AHP and the third question (Q2M) in the MMC. We leave out the details of Qs. Q2S, Q2A and Q2M; suffice it to say that each question had exactly the same format as Qs. Q1S, Q1A and Q1M in Survey 1, respectively.

2.3. Survey 3 (January 2002)

Survey 3 was carried out in January 2002, the month executive staff members of a prefectural government in Japan finalize the preparation of the budget for the following year. Respondents were 35 executive staff members having authority for final budget decisions on governmental projects. The purpose of the survey was to identify "the main concerns of executive staff members in budgeting." The survey included two

issues: one asked about the basic concepts for refining the prefectural administration (abstract issue); the other asked about actual governmental projects with high priority (concrete issue). The first pair of questions (hereinafter referred to as Q3A₁ and Q3R₁) concerned an abstract issue—the basic concepts of refining governmental program policy. They were: *Concept I*—for cultural development of our area; *Concept II*—for safety of our social life; *Concept III*—for environmental preservation of our area; *Concept IV*—for economic growth of our area; and *Concept V*—for enhancement of our area. The respondents were asked, in two ways, which of the basic concepts they thought significant for refining governmental program policy: Q3A₁ was formatted in the AHP and Q3R₁ used the ranking method. Q3A₁ in the Survey 3 was formatted in the AHP system, in which respondents were asked to conduct pairwise comparisons across all possible combinations of the basic concepts. Q3R₁ was offered in the same format as Q.4 (Q3R₂), but the alternatives offered were different.

The second pair of questions (Q3A₂ and Q3R₂) related to a concrete issue—governmental projects with high priority (PHP, for short). They were: *PHP I*—support for school education (for cultural development of our area); *PHP II*—improvements to rivers, mountains and coasts (for safety of our social life); *PHP III*—preservation of water resources (for environmental preservation of our area); *PHP IV*—support for entrepreneurs (for economic growth of our area); and *PHP V*—construction and repair of roads (for enhancement of our area). The respondents were asked in two ways which of the governmental projects should be budgeted with high priority: Q3A₂ was formatted in the AHP and Q3R₂ used the ranking method, which required respondents to rank all the governmental projects in the given question. Q3A₂ was formatted the same as Q3A₁, but the alternatives offered were different.

2.4. Survey 4 (March 2004)

Survey 4 was conducted in March 2004, one month after Japan Self-Defense Forces was dispatched to Iraq. Respondents were lay citizens of a local city in Japan. The purpose of the survey was to identify "the political attitude of citizens when the Self-Defense Force was dispatched to a country in the state of warfare." The sample size was 30; each respondent's political attitude was elicited by interviewing him/her one by one. The survey included 33 questions. In Qs.9 and 17, respondents were asked their intention to vote for a party in the next election. The parties were: Liberal Democratic Party of Japan (LDP); Democratic Party of Japan (DPJ); New Komeito (NK); Japanese Communist Party (JCP); and Social Democratic Party (SDP). Q.9 (hereinafter referred to as Q4W) was formatted in the FT and Q.17 (Q4A) in the AHP. The FT gave Feeling Score for each party, and the AHP gave the weight of each party.

We also asked respondents about their political slant in Q.13, and whether they support Prime Minister and the president of LDP, Junichiro Koizumi, in Q.33. Each of these issues was mutually related to the party electorates would vote for in the House of Councilors election. The outputs from these questions would serve as ideal dependent variables in a regression analysis. Q4W in the Survey 4 was formatted by the FT method, which requires respondents to assign an adequate number between 0 and 100 according to their intention to vote for each party; Q4A was formatted in the AHP system, in which respondents are required to conduct pair-wise comparisons across all possible combinations of parties. Qs.13 and 33 were formatted by the SMC method, which requires respondents to choose one from among the given alternatives.

3. Results

In this section, we analyze responses to the SMC-formatted questions, the MMC-formatted questions and the ranking-formatted questions of Surveys 1, 2, 3 and 4, based on the weight of each alternative elicited from the AHP.

3.1. Comparison between SMC and AHP

First, we focus on the difference of the aggregated ranking of alternatives between the SMC and the AHP;

in particular, the observed ratios of each alternative derived from the SMC and the weight of each alternative elicited from the AHP are compared.

Table 1 is a cross tab summarizing the results of two questions, Q1S and Q1A, from Survey 1. The numbers in each row are the AHP-based weights for each reason aggregated across each response to the SMC. The last row represents the weight of each reason aggregated across all responses. As was done in Table 1, Table 2 summarizes the results of two questions, Q2S and Q2A, from Survey 2. The maximum elements in each row are highlighted.

Table 1: Reason for being non-partisan (SMC and AHP)

SMC (Q1S)	AHP (Q1A): aggregated weight over each answer to SMC				Obs.	Ratio	Ranking
	Apathy	Realignment	Corruption	Non-confidence			
Apathy	0.192	0.350	0.223	0.235	108	38.4%	1
Realignment	0.184	0.171	0.227	0.418	40	14.2%	4
Corruption	0.177	0.199	0.371	0.253	56	19.9%	3
Non-confidence	0.298	0.189	0.248	0.265	77	27.4%	2
Average	0.217	0.250	0.260	0.273	281	100%	
Ranking	4	3	2	1			

Table 2: Main concern of job-search activities (SMC and AHP)

SMC (Q2S)	AHP (Q2A): aggregated weight over each answer to SMC				Obs.	Ratio	Ranking
	Specifications	Welfare	Salary	Place			
Specifications	0.405	0.097	0.178	0.320	83	47.4%	1
Welfare	0.276	0.219	0.184	0.322	28	16.0%	3
Salary	0.356	0.179	0.266	0.199	46	26.3%	2
Place	0.306	0.066	0.087	0.540	18	10.3%	4
Average	0.361	0.134	0.193	0.312	175	100%	
Ranking	1	4	3	2			

As can be seen in Tables 1 and 2, each diagonal element is not always the maximum in each row; that is, an answer on the SMC and the most weighted criterion in the AHP do not necessarily coincide. For example, in Table 1, respondents who answered *Apathy* to the SMC attached their weight mostly to *Realignment* in the AHP, while respondents who answered *Realignment* to the SMC attached their weight mostly to *Non-confidence* in the AHP. Furthermore, the highest ratio of the answer to the SMC was *Apathy*; the most weighted reason in the AHP, on the other hand, was *Non-confidence*. Thus, based on the SMC method, the survey research would conclude that the most important reason respondents were non-partisan is *Apathy* and the next is *Non-confidence*, and so on. In contrast, based on the AHP method, the most important reason is *Non-confidence* and the next is *Corruption*, and so on. Clearly, these two methods yield a different aggregated ranking of alternatives.

The SMC method is superior for two reasons. First, it is easy for respondents to fill out the questionnaire, and second, the main concerns of a respondent can be easily identified. However, this method prevents respondents from expressing their preference for a particular alternative over the others. Furthermore, no information regarding the relationship among non-selected alternatives can be derived. On the other hand, the AHP method makes it possible to reflect the relative importance of alternatives to results, despite the fact that it requires that respondents answer complex questions and thus requires much more time than the SMC. Which of these methods accurately represents respondents' perceptions is still an open question; verification requires more empirical tests. These results imply, however, that the output of the SMC, widely employed in survey research, might conceivably provide erroneous information.

3.2. Comparison between MMC and AHP

Next, we compare the observed ratios for each alternative derived from the MMC and the weight of each alternative elicited from the AHP. The MMC is another type of the MC question format allowing respondents to express their top two alternatives, thus giving them a greater degree of freedom in answering questions. Each respondent's preference for alternatives is likely to be specified more precisely than it is with the SMC. In Surveys 1 and 2, however, almost half the respondents in each survey chose only one alternative in the MMC. Accordingly, we define as "Singular" the group of respondents that

chose only one alternative despite the fact that they were given the option of indicating their second choice in the MMC (Q1M and Q2M), and we define as "Plural" the group of respondents that chose two alternatives. In this section, we first focus on the difference of aggregated ranking of alternatives among the "Singular" and the "Plural" of the MMC and of the AHP.

Table 3 is a cross tab summarizing the results of the two questions, Q1M and Q1A, for those defined as "Singular." Numbers in each row are the AHP-based weights of each reason aggregated across each response to the MMC. The last row represents the weight of each reason aggregated across all responses. Table 5 similarly summarizes the results of the two questions Q2M and Q2A. Tables 4 and 6 are cross tabs summarizing the results concerning those defined as "Plural," where two answers are counted for each respondent. In each table, the maximum elements for each reason are highlighted.

As can be seen in Tables 3 and 5, each diagonal element is not always the maximum in each row; that is, an answer of the "Singular" to the MMC and the most weighted criterion in the AHP do not necessarily coincide. For example, in Table 5, respondents who answered *Specifications* to the MMC attached their weight to *Place* most in the AHP, and respondents who answered *Welfare* to the MMC also attached their weight most to *Place* in the AHP. Furthermore, the highest ratio of the answer to the MMC was *Salary*; however, the most weighted concern in the AHP was *Specifications*. As a result, these two methods yield a different aggregated ranking of alternatives (see Tables 3 and 5).

The fact that the choice in the MMC is more compatible than the SMC with the most weighted alternative in the AHP is likely to result in a rationale that the MMC functioned well for the "Plural." As shown in Table 4, for example, each diagonal element is always the maximum in each row. Table 6, however, shows that the relationship between the choice in the MMC and the most weighted alternative is in a state of chaos. Furthermore, as was seen in Tables 3 and 5, the MMC and the AHP yield a different aggregated ranking of alternatives for both Tables 4 and 6, while the rankings elicited from the AHP are robust despite a change in population (see Tables 1, 3 and 4 from Survey 1 and Tables 2, 5 and 6 from Survey 2).

Lastly, we evaluate the effectiveness of the MMC in making up for the lack of information encountered with the SMC. Since the MMC gives respondents the option of indicating their second-best alternative, whether or not a second-best alternative is chosen may depend on the strength of that alternative vis-à-vis the first alternative. In other words, if the MMC functions well for the "Plural," it could contribute by supplying missing information (e.g., the difference in the degree of importance among selected alternatives and the respondent's perception concerning non-selected ones).

Let w_p^1 and w_p^2 respectively denote respondent p 's maximum and second-maximum element of the eigenvector corresponding to the Frobenius root, then $\Delta w_p \equiv w_p^1 - w_p^2 (\geq 0)$ represents the discrepancy in the maximum weight and the second maximum weight of alternative for respondent p . Therefore, the larger the Δw_p is, the more clearly respondents would distinguish their best alternative from the second-best.

Table 7 summarizes the number of the "Singular" and the "Plural" respondents in Surveys 1 and 2, the table was stratified by the size of Δw_p . As shown in the table, the difference in the distribution of responses between the "Singular" and the "Plural" seems small. Indeed in Chi-square tests, $X^2_{(Q1M)} = 2.979 < X^2_{(6, 0.8)} = 3.070$ and $X^2_{(Q2M)} = 2.108 < X^2_{(6, 0.9)} = 2.204$ hold. These results from the chi-square tests imply that there might not be any difference in Δw_p between the "Singular" and the "Plural"; in other words, whether or not respondents add their second choice in answering the MMC-formatted question may be independent of the discrepancy in the degree of importance between the best and the second-best alternative. Thus, we could conclude that the MMC does not succeed in precisely specifying a respondent's preference of alternatives.

The MMC method, overall, does seem to be a good option for designing questionnaires because it enhances the degree of freedom in answering questions for respondents. Many researchers thus widely employ the MMC in their survey questionnaires. Insofar as our surveys are concerned, however, the MMC might not be effective in filling the information gap between the SMC and the AHP.

Table 3: Reason for being non-partisan (MMC "Singular" answer, and AHP)

MMC (Q1M) "Singular" answer	AHP (Q1A): aggregated weight over each answer to MMC				Obs.	Ratio	Ranking
	Apathy	Realignment	Corruption	Non-confidence			
Apathy	0.281	0.207	0.229	0.283	41	31.3%	2
Realignment	0.169	0.372	0.230	0.229	48	36.6%	1
Corruption	0.154	0.165	0.415	0.266	25	19.1%	3
Non-confidence	0.142	0.191	0.219	0.448	17	13.0%	4
Average	0.198	0.258	0.263	0.281	131	100%	
Ranking	4	3	2	1			

Table 4: Reason being non-partisan (MMC "Plural" answers, and AHP)

MMC (Q1M) "Plural" answer	AHP (Q1A): aggregated weight over each answer to MMC				Obs.	Ratio	Ranking
	Apathy	Realignment	Corruption	Non-confidence			
Apathy	0.275	0.244	0.245	0.236	83	28.0%	1
Realignment	0.223	0.322	0.203	0.251	72	24.3%	3
Corruption	0.199	0.211	0.324	0.266	78	26.4%	2
Non-confidence	0.209	0.216	0.246	0.330	63	21.3%	4
Average	0.228	0.248	0.256	0.267	296	100%	
Ranking	4	3	2	1			

Table 5: Main concern of job-search activities (MMC "Singular" answer, and AHP)

MMC (Q2M) "Singular" answer	AHP (Q2A): aggregated weight over each answer to MMC				Obs.	Ratio	Ranking
	Specifications	Welfare	Salary	Place			
Specifications	0.283	0.079	0.162	0.476	19	23.5%	2
Welfare	0.309	0.215	0.096	0.381	11	13.6%	4
Salary	0.367	0.169	0.314	0.150	38	46.9%	1
Place	0.247	0.140	0.082	0.530	13	16.0%	3
Average	0.320	0.149	0.212	0.319	81	100%	
Ranking	1	4	3	2			

Table 6: Main concern of job-search activities (MMC "Plural" answers, and AHP)

MMC (Q2M) "Plural" answer	AHP (Q2A): aggregated weight over each answer to MMC				Obs.	Ratio	Ranking
	Specifications	Welfare	Salary	Place			
Specifications	0.445	0.136	0.230	0.189	57	32.4%	1
Welfare	0.283	0.225	0.093	0.399	40	22.7%	3
Salary	0.426	0.134	0.286	0.154	45	25.6%	2
Place	0.328	0.196	0.063	0.413	34	19.3%	4
Average	0.381	0.167	0.181	0.271	176	100%	
Ranking	1	4	3	2			

Table 7: Comparison of Δw_p between "Singular" and "Plural"

Δw_p	Q1M (Survey 1)		Q2M (Survey 2)	
	Singular	Plural	Singular	Plural
0	46	45	24	30
(0 , 0.1]	34	35	22	21
(0.1 , 0.2]	16	24	9	13
(0.2 , 0.3]	7	10	6	8
(0.3 , 0.4]	10	14	5	5
(0.4 , 0.5]	6	10	6	5
(0.5 , 1.0]	12	10	9	6

3.3. Comparison between ranking method and AHP

First, we analyze the difference of the preference order for all alternatives concerning the abstract issue, between the ranking method and the AHP. Specifically, the aggregated ranking of each concept for refining governmental program policy derived from the ranking method and the aggregated weight of each concept elicited from the AHP are compared. Table 8 is a cross tab summarizing the results of the two questions, questions Q3A₁ and Q3R₁, from Survey 3. The numbers in the first row are the aggregated ranking of each concept across all responses. The second row represents the AHP-based weights for each concept aggregated across all responses. The last row indicates the change from year 2001 to 2002 in the actual implementation of the budget for each concept, which corresponds to one of the five budgeting categories in the prefecture.

Table 8: Concepts of refining governmental program policy

Q3R ₁ & Q3A ₁	Concept I	Concept II	Concept III	Concept IV	Concept V
Ranking	2.000	1.250	3.500	4.500	3.750
Weight	0.18803	0.36152	0.19979	0.08191	0.16874
Annual change in budget 2001 - 2002	0.355%	-1.62%	1.26%	-0.231%	-1.11%

As shown in Table 8, the difference of the preference order for all alternatives between the ranking method and the AHP seems small. For example, the highest ranked concept in the aggregated ranking derived from the ranking format was *Concept II*; the most weighted concept in the AHP was also *Concept II*. The case is the same for both the lowest concept in the ranking format and the lightest weighted concept in the AHP format: one concept—*Concept IV*—resulted for both. In addition, the line charts produce similar patterns, and the correlation coefficient (CC, for short) between these two is -0.87. Consequently, the preference order of the basic concepts derived from the ranking method coincided almost exactly with that elicited from the AHP. As a result, these two methods would likely produce similar results concerning respondents' preference order of alternatives.

Nevertheless, the preference order of basic concepts for refining governmental program policy did not coincide with the annual change in the actual implementation of the budget. Specifically, the concept thought to be the most significant by executive staff members for refining governmental program policy (*Concept II*) and the category budgeted with the highest priority (corresponding to *Concept III*) were different. Consequently, the preference orders obtained from the ranking method and the AHP did not coincide with the annual change in budget, nor did high-lows of the line charts and those of the bar chart synchronize with each other. Indeed, CCs between the preference orders of concepts and the actual change in budget are -0.44 and 0.32. These results might imply that both the ranking format (Q3R₁) and the AHP format (Q3A₁) did not function well in capturing respondents' perceptions. We would not, however, be able to draw such a conclusion based solely on this survey because the change rates in the budget were so small (from -1.62% to +1.26%) that the correlation between the preference order and the change rate might have been affected by measurement error.

Next, we focus on the difference of the preference order for all alternatives concerning the concrete issue between the ranking method and the AHP. Specifically, the aggregated ranking of each governmental project with high priority obtained from the ranking method and the aggregated weight of each project elicited from the AHP are compared. Table 9 is a cross tab summarizing the results of the two questions, Q3A₂ and Q3R₂. The numbers in the first row are the aggregated ranking of each PHP across all responses. The second row represents the AHP-based weights for each PHP aggregated across all responses. The last row indicates the annual change from year 2001 to 2002 in the actual implementation of the budget for each PHP.

As shown in Table 9, the preference order for all alternatives between the ranking method and the AHP

are different. For example, both the highest PHP in the aggregated ranking obtained from the ranking format and the most weighted PHP in the AHP were *PHP I*, while the lowest PHP in the ranking format (*PHP V*) and the lightest weighted PHP in the AHP format (*PHP III*) were different. Indeed, the CC was equal to -0.32. As a result, the ranking method and the AHP would reason differently.

Table 9: Governmental projects with high priority

Q3R ₂ & Q3A ₂	PHP I	PHP II	PHP III	PHP IV	PHP V
Ranking	2.000	2.750	2.750	3.250	4.250
Weight	0.39234	0.14968	0.08169	0.14725	0.22904
Annual change in budget 2001 - 2002	10.3%	-7.72%	-8.16%	-5.87%	-0.300%

In contrast to the case of the abstract issue, the preference order of PHPs induced from the AHP were positively correlated with the annual change in the actual implementation of the budget for the case of the concrete issue. As can be seen in the Table 9, the preference order elicited from the AHP coincided almost exactly with the annual change in the actual implementation of the budget (CC=0.98), while that generated by the ranking format did not (CC=-0.32). Although these results might have been affected by measurement error, these results imply that each respondent's preference for alternatives elicited from the AHP is likely to be specified more precisely than it is with the ranking method.

3.4. Comparison between rating method and AHP

First, we focus on the difference of the aggregated ranking of alternatives between the FT and the AHP; in particular, the average of Feeling Score for each party obtained from the FT formatted question and that of the weight of each party derived from the AHP formatted question are compared. Since both rankings reflect the entire trend concerning each party of a population, they are likely to produce similar results.

Table 10 summarizes the results of two questions, Q4W and Q4A, from the survey. The numbers in the first and the second rows are the average of Feeling Score and that of the AHP-based weight of each party aggregated across each response to the FT and the AHP, respectively. The last row represents the correlation coefficients between Feeling Score and the weight of each party among all respondents.

As shown in Table 10, Feeling Scores and the weights of the AHP imply that the most and the second most favored party among respondents are DPJ and LDP, respectively; both the FT and the AHP clearly identify the top 2 parties. On the other hand, the ranking for the remaining parties, NK, JCP and SDP, are different, even though the difference of Feeling Score or the weight of each party among those parties is quite small. Indeed, the correlation coefficients between Feeling Score and the weight of each party are high on some level. As a result, answers on the FT and on the AHP do not necessarily coincide, which yields a different aggregated ranking of alternatives. Both the FT and the AHP, however, produced similar results, and they overall seem to specify respondents' preference for parties in the choice of House of Councilors election.

Table 10: Aggregated Feeling Score and weight of each party

Q4W & Q4A	LDP	DPJ	NK	JCP	SDP
Feeling Score	53.6	66.6	19.2	14.4	14.2
Weight	0.34662	0.38969	0.08945	0.08007	0.09417
Correlation Coefficient	0.81257	0.72848	0.77450	0.77443	0.75042

Next, we evaluate the effectiveness of each method in terms of its ability as independent variables in regression analysis; in particular, we formulate four regression models and compare R^2 s. In the evaluations, as independent variables, we employ Feeling Score and not normalized weight of each party except for those of the Social Democratic Party (SDP), because the party was minority group (The number of the member of House of Councilors was; LDP: 116, DPJ: 70, NK: 23, JCP: 20 and SDP: 5, as at the survey was carried out. (March 2004)). Insofar as dependent variables, based on Quantification Theory I, we employ two variables; one is respondent's political slant (Q.13) in Regression Models 1 and 2, and the other is whether respondent supports Prime Minister Koizumi or not (Q.33) in Regression Models 3 and 4. As can be seen in Fig. 3, we offered six alternatives including "Don't Know" answer in Q.13. Among those alternatives, we set "Progressive" or "Slightly progressive" as 0, and "Conservative" or "Slightly conservative" as 1 in the Regression Analysis A. In the same way as for Regression Analysis A, we set "Support Koizumi" as 0, and "Do not support Koizumi" as 1, in Regression Analysis B. Thus, the actual regression models can be formulated as follows.

- Regression Analysis A
 - dependent variables; Political Slant
"Progressive" or "Slightly progressive" = 0, "Conservative" or "Slightly conservative" = 1
 - regression equations;
 - Political Slant = $a_1 + a_2*FS_{LDP} + a_3*FS_{DPJ} + a_4*FS_{NK} + a_5*FS_{JCP}$ (regression model 1)
 - Political Slant = $b_1 + b_2*w_{LDP} + b_3*w_{DPJ} + b_4*w_{NK} + b_5*w_{JCP}$ (regression model 2)
- Regression Analysis B
 - dependent variables; Support Koizumi or not
"Support" = 0, "Do not support" = 1
 - regression equations;
 - Koizumi = $c_1 + c_2*FS_{LDP} + c_3*FS_{DPJ} + c_4*FS_{NK} + c_5*FS_{JCP}$ (regression model 3)
 - Koizumi = $d_1 + d_2*w_{LDP} + d_3*w_{DPJ} + d_4*w_{NK} + d_5*w_{JCP}$ (regression model 4)

Table 11 summarizes the results obtained from Regression Analysis A. As can be seen in the table, the R^2 s of regression models 1 or 2 are not large enough to predict respondents' political slant; respondent's intention for voting does not necessarily correlate with his/her political slant. By employing the weights derived from the AHP, however, R^2 was improved from 0.23755 to 0.42850. Furthermore, focusing on the p-values, the degree of precision of the regression model 2 was higher than that of the regression model 1.

Table 12 summarizes the results obtained from Regression Analysis B. As shown in the table, R^2 s are relatively larger than those of the Regression Analysis A; whether respondent supports Prime Minister Koizumi or not can be predicted on some level based on regression models 3 or 4. By employing the weight derived from the AHP, R^2 was again improved from 0.42293 to 0.51299. Furthermore, focusing on the p-values, the degree of precision of the regression model 4 was higher than that of the regression model 3.

Table 11 : Results of Regression Analysis A (models 1 and 2)

Political slant	model 1 (Feeling Score)		model 2 (weight)	
	coefficient		coefficient	
constant	1.48488	**	1.35099	***
LDP	-1.58819	**	-0.00968	***
DPJ	-0.51939	*	-0.00175	*
NK	-0.45019	*	-0.00330	**
JCP	-1.62317	*	-0.00642	**
N = 25	$R^2 = 0.23755$	SE = 0.48460	$R^2 = 0.42850$	SE = 0.41955
p-value: 0.05 < p <= 0.1: *, 0.01 < p <= 0.05: **, p <= 0.01: ***				

The FT method overall seem to be a good option for designing questionnaires because it assures respondents expressing their preferences for all alternatives, and it enhances the degree of freedom in answering questions for respondents. Therefore, many researchers, especially in political science field, have widely employed the FT in their survey questionnaires. Insofar as our survey is concerned, however, Feeling Scores may not be effective in predicting respondents' preferences. Either in Regression Analysis A or B, R^2 s based on the FT were not large enough to predict respondents' preferences. On the other hand, the results of this section imply that the AHP could quantify respondents' preferences in terms of the distribution of the weight of each alternative; by employing weights of the AHP, R^2 s in regression analyses were improved from the use of Feeling Scores. Since Feeling Score measures the feeling for parties and the weight of the AHP measures the preference for parties, we cannot simply compare these R^2 s and conclude that the weight is the better explainer than Feeling Score. The weights derived from the AHP, however, could be conjectured to function better as independent variables in regression analyses than Feeling Scores.

Table 12: Results of Regression Analysis B (models 3 and 4)

Support Koizumi	model 3 (Feeling Score)		model 4 (weight)	
	coefficient		coefficient	
constant	-0.05558	*	-4.22262	**
LDP	0.00918	**	5.29983	***
DPJ	-0.00148	*	4.09808	**
NK	0.00481	**	6.46932	***
JCP	0.00125	*	9.04207	***
N = 17	$R^2 = 0.42293$ SE = 0.45130		$R^2 = 0.51299$ SE = 0.41459	

p-value: 0.05 < p <= 0.1: *, 0.01 < p <= 0.05: **, p <= 0.01: ***

4. Concluding remarks

Questionnaire design poses one of the biggest challenges for survey researchers because how respondents are asked questions can have a great effect on results. One political scientist² conducting a public opinion poll remarked that different question formats yielded different results, despite the fact that they were asking about the exact same content. Consequently, various ways of eliciting opinions have been proposed and evaluated in order to determine which best represents each respondent's perception. In particular, the multiple-choice method, ranking method, and rating method have often been compared, generating a great deal of discussion.

Among the aforementioned methods, each method has the pros and cons. The multiple-choice method has been most extensively used because of its ease of response and its ease in identifying for the researcher the respondents' main concerns. In addition, multiple-choice provides some formats that enhance a respondent's degree of freedom in answering questions. Meanwhile, no information regarding the non-selected alternatives or the relative importance among selected alternatives can be derived. The ranking method, often considered ideal for designing questionnaires, is also widely used because it allows researchers to easily identify respondents' preference orders. This method, however, does not allow ties for alternatives nor can they represent the degree of importance for each alternative. The rating method, such as the Feeling Thermometer, has also been widely used because of its ease for respondents to answer and its ease in identifying for the researcher the respondents' concerns for all alternatives. On the other hand, this method does not fit to natural way of thinking for human. In contrast, the AHP, a support system for decision-making, can thus be a possible option for formatting questionnaires. This method makes it possible to reflect the relative importance of alternatives to results, despite the fact that it requires that respondents answer complex questions, thus necessitates much more time.

² Chris Achen, University of Michigan, February 2001 (personal communication).

In this study focused on survey research in which the questions that respondents were asked involved issues, we verified the effectiveness of the multiple-choice method, the ranking method, and the rating method by using the weight of alternatives elicited from the AHP. The results were: (1) the simple or modified multiple-choice and the AHP yielded a different aggregated ranking of alternatives, while the AHP-based rankings were robust despite a change in population; (2) whether or not respondents added a second choice in answering the modified multiple-choice formatted questions was irrelevant to the discrepancy in the degree of importance between the best and the second-best alternative; (3) for the abstract issue, the ranking format and the AHP format yielded similar aggregated ranking of alternatives; (4) for the concrete issue, the results from the AHP format coincided almost exactly with the annual change in the actual implementation of the budget, unlike those generated by the ranking method; (5) the Feeling Thermometer method and the AHP yielded a similar aggregated ranking of alternatives; (6) in regression analyses, Feeling Scores may not be effective in predicting respondents' preferences, while the weight derived from the Analytic Hierarchy Process could predict their preferences on some level. These results, insofar as our survey is concerned, provide some evidence that the multiple-choice method, the ranking method, and rating method do not succeed in precisely specifying a respondent's perception for alternatives and thus might not be appropriate for measuring human perception in questionnaires of survey research. The application of the AHP to questionnaire design in survey research, on the other hand, might very well be superior to those traditional methods.

Nevertheless, several issues remain. The first issue is which criteria would be appropriate for evaluating the effectiveness of the questionnaire format. For example in the Survey 3, is the annual change in the actual implementation of a budget really adequate? This issue relates to the criterion evaluating the method: the criterion needs to be sensitive in measuring the correlation between what is in respondents' minds and their actual behavior. As for the theme of the survey, the annual change in budget is still a functional criterion for the evaluation because the change is one of the distinguishing phenomena representing the priority of executive staff members in the government. On the other hand, other criteria for evaluation could be considered: the number of staff members for a particular project, the frequency of being one of the main agenda items in the executive staff meetings, the number of times a particular project is covered in the public relations material of the government, and so on. Further empirical tests are required for verification.

The second issue concerns what the results would be if the number of alternatives in a question were other than we used in the surveys. Would similar results be obtained if the question offered nine alternatives, for instance? In this study, we analyzed the case for just four or five alternatives; investigations of up to at least nine alternatives would therefore be warranted in order to verify what an adequate format for questionnaire design would be.

References

- Downing, M. Steven. (2004) "Reliability: on the Reproducibility of Assessment Data," *Medical Education*, 38-9, 1006-1012.
- Inglehart, R. and Abramson, P. (1993) "Values and Value Change of Five Continents," paper presented at the 1993 Annual Meeting of the American Political Science Association, Washington D.C., September 1-5.
- Jerard, K. (1995) "Writing multiple-choice test items Practical Assessment," *Research and Evaluation*, 4-9.
- Kabashima, I. (1998) "Seiken Koutai to Yuken-shya no Taido Henyou," (in Japanese) Bokutaku-shya.
- Saaty, T. L. (1980) *The Analytic Hierarchy Process*. McGraw-Hill.
- Saaty, T. L. (1994) "Highlights and Critical Points in the Theory and Application of the Analytic Hierarchy Process," *European Journal of Operational Research*, 52, 426-447.
- Sato, Y. (2003) "Comparison between Ranking Method and the Analytic Hierarchy Process in Program

Policy Analysis,” *Proceedings of the Seventh International Symposium on the Analytic Hierarchy Process*, 439-447.

Sato, Y. (2004) “Comparison Between Multiple-choice and Analytic Hierarchy Process: Measuring Human Perception,” *International Transactions in Operational Research*, 11-1, 77-86.

Sato, Y. (2005) “Questionnaire Design for Survey Research: Employing Weighting Method,” *Proceedings of the Eighth International Symposium on the Analytic Hierarchy Process*, ISSN 1556-830X.

Sato, Y. (2007) “Administrative Evaluation and Public Sector Reform: An Analytic Hierarchy Process Approach,” *International Transactions in Operational Research*, 14-5, 445-453.

Traugott, W. Mike and Lavrakas, J. Paul. (2000) “The Voter’s Guide to Election Polls, 2ed.” Chatham House Publishers, Seven Bridges Press, LLC.

Webber, A., Apostolou, B. and Hassell, J. M. (1997) “The Sensitivity of the Analytic Hierarchy Process to Alternative Scale and Cue Presentations,” *European Journal of Operations Research*, 96, 351-362.