

## Method and Technique in Speech Act Research: The Case of Refusal Realization in EFL

Keiko SATO

### Abstract

The purpose of this study is to examine the validity and reliability of the investigation of speech act realization by learners using the study of how recent Japanese college students use English when they refuse requests in different situations, and whether there is some difference according to English proficiency level. In the refusal study, written discourse completion test questionnaire (DCT) was employed as analyzed materials. The questionnaire of eight situations was designed to identify three different influential factors (i.e. degrees of social distance, psychological distance, and offendedness). The reliability of measurements, that is, the questionnaire and the scales of perception on factors was tested, and validity of those measurements was considered.

The analytical framework of the refusal study was decided from Brown and Levinson (1978)'s chart of possible strategies for doing FTAs (p.69). The use of strategies on negative politeness on record and those off record was statistically proceeded by percentage and Chi-square of frequency. As a result, the degree of offence was a decisive factor on whether or not a speech act of refusal was done to both proficiency levels, and the degree of psychological distance was a factor that chose direct or indirect refusal expressions. The results are considered along with reliability and validity.

### INTRODUCTION

Most studies of pragmatic competence, particularly in speech act realization, have employed discourse completion tests (DCTs) as a means of data collection in the quantitative approach. Recently, however, some researchers began to be doubtful of the validity and reliability of such data (Rose, 1992, 1994, 1995; Kasper, 1998; Hinkel, 1997, etc.). They compared two different types of data collecting techniques and attempted to decide which technique is the better measurement, based on their assumptions.

Rose (1992) reported the results from comparing the DCT form with hearer response with the DCT without it. His DCT questionnaire consisted of six request situations involved in two variables: social distance and social dominance. The subjects were students with a variety of background in the freshman composition required course. Their performances were analyzed for six components: alerter, perspective, strategy, downgraders, supportive moves, and length. As a result, there were no significant differences be-

tween the two forms of the DCT discovered. Rose raised a question that “the DCT may not be an adequate instrument for collecting data on hearer-based languages”. He meant, based on Lakoff (1984) and Lebra (1976), that Japanese is a hearer-oriented language and that the Japanese speakers usually leave the understanding of incomplete utterances to the hearers. Although he added the condition of “if Lakoff is correct”, he seems to believe the two researchers’ argument. He concluded that “the DCT, especially one which includes hearer response, may prove insufficient for speech act research on languages which is characterized by hearer-based interaction.”

In Rose (1994) too, the assumption of Japanese language attitude was employed based on Lakoff (1985), Lebra (1976), and Clancy (1986)’s views. Rose quoted from Clancy that “ideal interaction is not one in which the speakers express their wishes and needs adequately and listeners understand and comply, but rather one in which each party understands and anticipates the needs of the other, . . .” and from Lebra that “the speaker does not complete a sentence but leaves it open-ended in such a way that the listener will take it over before the former clearly expresses his will or opinion.” Although this observation is correct in some respect, the Japanese do not always complete their utterances. Open-ended utterances are not always ideal in Japanese, but rather ones that should be corrected by mothers and teachers when the speakers are young. Open-ended utterances are conventionally used as a kind of softener, mitigating device without limiting to requesting in Japanese. It does not mean that the Japanese will use indirect requests frequently.

Rose (1994) compared the DCT to the multiple-choice questionnaire (MCQ). As the coding scheme he used the nine-level scale from most direct to least direct for analysis of requesting strategy based on the Cross Cultural Speech Act Realization Project (CCSARP). It was expected according to the previous assumption on Japanese that Japanese subjects would use more indirect requests or hints. The results of their experiments were far from what he had expected. Their subjects used direct requests more frequently than the subjects of American English native speakers. They ascribed the unexpected results to the method of data-collection, that is, the DCT. They inferred the possibility of the Japanese subjects’ responses without awareness of face-to-face interaction, which inspired them to conduct the second experiment using the MCQ. As a result, there were significant differences in seven situations out of eight between the responses from the DCT and those from the MCQ. They suggested that DCTs may not be a valid means of data collection procedure with non-Western subjects. The two studies, however, have some biases in that: (1) they employed different subjects, though they were Japanese in both studies, (2) while the DCT seems a pseudo-production task, the MCQ can be a perception or judgment task, and (3) the assumption that the Japanese speakers use indirect requests by all means in any cases is taken to be applicable to any case. Which one appears to reflect the reality?

Japanese has a specific honorific system, and it cannot be analyzed only in terms of direct/indirectness measure. It seems to me that directness assumption cannot decide

whether the DCT measure is valid for non-Western speakers. Rather, the effect on the result may be subjects' proficiency of a target language.

Rose & Ono (1995) compared the DCT to MCQ administered to the same Japanese subjects. Rose's previous study (1994) prompted this study in order to address methodological validity in speech act research. After they reviewed previous studies of DCT development and criticisms comprehensively, they pointed out two main weaknesses of Rose (1994): different groups of subjects for the DCT and the MCQ, and unlikely situations in Japan because of translation from American version. Their revised study indicated significant differences between the DCT and the MCQ in most situations, supporting Rose's (1994) results. Nevertheless, they recognized their results did not indicate that previous studies using the DCTs were of no value, but suggested that they need to be treated carefully as their weak conclusion.

Hinkel (1997) also investigated the difference of DCT advice data and MC advice data in terms of four levels of coding: direct, hedged, indirect, and nothing. She found that native speakers of English selected fewer direct or hedged advice than the Chinese subjects did with the MCQ, whereas with the DCT English speakers used more direct and hedged advice than the Chinese learners of English, the result of which is opposite to that of Rose (1994, 1995). In the comparison of the two instrumental procedures, she states the characteristics of the two. "Since making choices from MC selections requires only a fraction of the effort necessary for writing responses to DCTs, the latter may elicit data more representative of planned, rather than the spontaneous discourse characteristics of an actual speech act" (p.19). Responding to the DCT imposes great demands on the linguistic skills of learners' target languages. Which data-collection technique to employ depends on practical problems such as limited research period, number of subjects accessible, research purpose (to investigate speech act production or perception), and the like.

Rintell & Mitchell (1989) attempted to investigate what differences would be found between the results from a written discourse completion test and those from an oral role play task when they are used for data elicitation. Two types of elicitation procedures were administered to four groups: the written DCT to native speakers and foreign students and the spontaneous oral role play to different English speaking students and different foreign students. The language elicited in both forms was very similar in general. In detail, there were subtle differences in length between native and non-native oral data, and in frequency of request strategy use between oral and written data for both native and non-native speakers. They explained the reason for this result by stating that "the discourse completion test is, in a sense, a role-play like the oral one" (p. 270).

Bonikowska (1988) paid attention to the significance of act undone in a specific speech act. She argued that "pragmatics should expand its research interest to include not only the study of how speakers perform speech acts but also the investigation of instances where they decide not to perform them" (p.169). Her questionnaire of complaint was constructed in such a way that the subjects were not asked to play artificial roles.

Since this study took a qualitative approach, the effect of adopting the opting-out factor was not reported. The factor of non-performance of acts should be included in pragmatic research, particularly in face-threatening acts. Hinkel (1997) included opting-out strategy with direct, hedged, and indirect strategies in her advice study.

Cohen & Olshtain (1993) attempted the description of non-native speakers' cognitive process such as assessing, planning, and executing speech act utterances when they produced them. Retrospective verbal report protocols were analyzed in terms of the degree of assessing and planning, the degree of attention to grammar and pronunciation, and the sources for language. They found that there were three types of subjects who could be called metacognizers, avoiders, and pragmatists. They also argued issues of research methodology. Since there are different types of speakers in thought process, if the respondent was not allowed to opt out, the results might be deflected. Forcing unnatural behavior is not always the case in the real world.

Cohen & Olshtain (1994) presented the model of collecting speech act data, the cycle of speech act research methods. They claimed that "investigators would start with the generation of initial hypotheses based on observational data from natural speech in L1 and L2, . . . (t)hen we would elicit simulated speech such as that in role-plays, . . . we might go on to a paper-and-pencil task such as discourse completion . . . (f)inally, we might be advised to validate the findings by means of further naturalistic, observational data" (p.148). Furthermore, they added two more methods: a perception task such as acceptability checks and a verbal report such as role-play interviews. Since each procedure has some degree of weaknesses, triangulation to be adopted may enhance the validity and reliability of data and the interpretation from the results.

This study assesses the validity and reliability of instruments used in our recent research of refusal.<sup>1</sup> In this refusal research, university students' written responses to the DCT with the other party's responses were analyzed. The DCT format includes three perception scales attaching to each situation and some other introspective written questions. The purposes of this study is to evaluate the methods of pragmatic data collection and to revise them if they are available to future use.

#### Research Questions of This Study

1. To what degree is the questionnaire with eight situations valid and reliable?
2. To what degree is the scales of perception on three factors valid and reliable?

### *STUDY OF REFUSAL*

#### *Research Questions of Refusal Study*

1. Is there any difference in the use of pragmatic strategies between Japanese students with higher proficiency level and those with lower proficiency level when refusing requests?

2. Is there any difference in the use of pragmatic strategies according to different situations when refusing requests?
3. Is there any language transfer in the use of pragmatic strategies in the realization of refusal?

### *Participants*

Approximately 200 Japanese university students in different majors participated in the study. They were all from freshmen and sophomores except for a few junior or senior students who had to enroll a comprehensive English again because of their absence or personal convenience. Most of them whose majors were French or English were female students. Since the random selection of subjects was impossible, they came from 6 intact classes I happened to teach in 1996. Their majors included French, English, Business, Education, and Law.

To investigate the relationship between the ability to use appropriate pragmatic strategies and the English proficiency, the structure section of Comprehensive English Language Test (CELT) was administered to the participants, and the test takers were divided into three groups in terms of their scores; the upper, middle, and lower groups. The upper and lower groups were decided to be analyzed. The upper group consisted of top 25 % with the score range from 65 to 55 out of 75 in the full mark and the lower group of bottom 25 % with the range from 43 to 33. The number of each group was 49 for each group.

### *Instrument*

The DCT technique was adopted as a data collection procedures for the quantitative analysis. Eight situations were designed according to the three degree of variables from the speaker's point of view: social distance (higher or lower social rank), psychological distance (close or far distance relationship between the speaker and the addressee), and offense the speaker feels the hearer will have. The three variables have two levels. The situations were as follows:

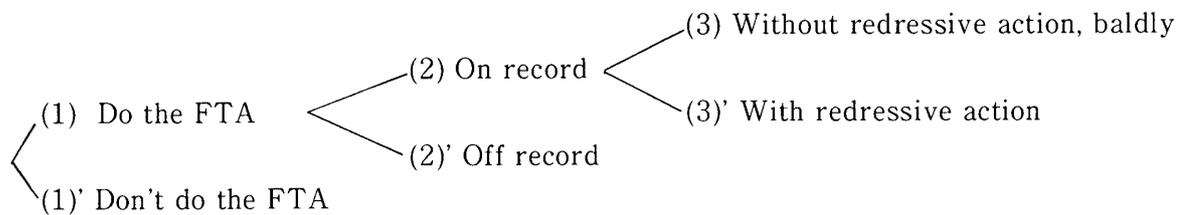
1. Your seminar professor asks you to help the movement from the old office to the new office next Sunday.
2. An unfamiliar professor asks you to help the movement from the old office to the new office next Sunday.
3. Your close friend asks you to help the movement from the old apartment room to the new apartment room.
4. An unfamiliar friend asks you to help the movement from the old apartment room to the new apartment room.
5. Your close friend asks you to lend him/her a pen, though you have only one pen.
6. An unfamiliar friend asks you to lend him/her a pen, though you have only one pen.
7. Your seminar professor asks you to lend him/her a pen, though you have only one pen.

8. An unfamiliar professor asks you to lend him/her a pen, though you have only one pen.

The DCT questionnaire was administered to all the participants, who were asked to choose whether to refuse or accept the request in each situation first, to give verbal response to each request in case of refusing, to mark the point of three perception scales, and to choose the image the responder had out of multiple-choice items when he or she was planning to make a refusing response (for detail in Appendix 2).

*Coding scheme for analysis*

The analytical framework is based on the chart from Brown and Levinson (1987: 69).



(Partially changed from Brown & Levinson)

**Figure 1 Speech Act Realization System**

The choice of (1) or (1)' means that of avoidance strategy or not. The choice of (2) or (2)' indicates whether to use direct or indirect strategy in the head act of refusal. The choice of (3) or (3)' indicates whether to use only direct refusal without any mitigating strategy or to use direct refusal but with one or more mitigating strategies. What kind of strategies were used in the categories of (2)' and (3)' was examined as well as the percentage of frequency in each category the subjects preferred to choose. The coding scheme of strategies in (2)' and (3)' was based on the external modification of request in Faerch and Kasper (1989), from which four possible strategies in refusal were selected. Some researchers call them 'semantic formula' (Ikoma & Shimura, 1993; Beebe & et al, 1990).

External modification functioning as supportive moves

- Grounder (giving reasons for the refusal)
- Apology (e.g., 'I'm sorry')
- Alternative (e.g., 'I can help you another day')
- Desire (e.g., 'I'd like to help you, but . . .')

*Validity of Measurement*

The study of refusal employed the DCT along with perception scales for collecting and analyzing data. Were these instruments valid for the purpose of measuring participants' pragmatic competence? Are they valid instruments for Japanese learners of En-

glish? A number of researchers have recently raised this question as a controversial issue in speech act research. If the study is not valid in the design and the measurement, it is worthless while conducting it any more. Therefore, first of all, it is necessary to examine the validity of measurements of this study: the DCT and perception scales.

### 1. *Situations in the DCT*

The DCT was designed to consist of eight situations, each of which contains either of two levels of three variables: social status of the addressee, psychological distance between the addresser and addressee, and relative degree of the addressee's offense likely caused by refusal.

The degree of likely offense was assumed to be represented by two different request contents. The greater offense would be caused by refusing the request of assistance for moving to a new office or apartment, and the smaller offense would be caused by refusing the request of lending a pen or pencil. The addressee's social status was decided to differ by setting the addressee as a professor or friend. The difference of psychological distance between two interlocutors was specified by adding the word 'close' or 'familiar'. The inclusion of three variables into situations appears simple and clarified. However, one problem was found in the course of analysis. While the researcher took it for granted that a professor was perceived to be in the higher status than a friend, there was little or no difference in the participants' performances between toward a professor and to a friend. Sasaki (forthcoming) commented personally that recent university students have no particular respect for professors but for managers under whom they are working part-time. She changed the addressee from a professor to a manager at the working place in her study of request and refusal. The inspection of participants' assessment of their own perception on roles in situations should be made before questionnaires are designed.

Were the participants – Japanese university students – likely to experience all the eight situations? Is there any possibility to be asked by a professor, especially by an unknown professor to help his/her movement to a new office even if it is on campus? Can it be possible that a Japanese university student ask an unfamiliar friend to help him/her move to a new apartment? The answer will be 'no' definitely. The situations were designed so as to readily analyze the effect of the three variables on performances. All the situations were made under the consideration of a discourse completion test administered to Japanese university students. Two out of eight situations, however, may be inappropriate contents. They might give some threat to the face validity of the refusal study.

Since the prior consideration was on situations for easy access to variables, all the situations may be too simple, not representatives of refusal situations by university students, and lack of comprehensiveness. The participants might have had difficulty in differentiating refusal phrases in situation by situation. In addition, they could predict what the researcher expected. It can be the thread to an internal validity.

## 2. *Perception scales*

Three scales of participants' perception on the three variables - social distance, psychological distance, and offense - were attached to each situation in order to ask participants' own degree of perception and to investigate the relationship between the degree of perception for each situation and the use of pragmatic strategies in refusing. The number of points on scales was problematic. Both extremes and average point cannot be deleted. The problem was how many point should be made in between. Finally, the five-point scales were considered as appropriate because it would be more difficult to decide if there were more than five points. Actually, most of participants had no difficulty in deciding the degree except for a couple of respondents who marked in between 2 and 3 (average).

There seemed to be some threat to construct validity. One of the three scales (familiarity) placed the highest degree at the left end (1) and the lowest (5) at the right end, but the rest two scales (rank and offence) placed the reversed way, that is, the lowest at the left and the highest at the right. This construction might have confused some of the respondents. They might have responded the opposite way they really perceived. Only a few respondents indicated the possibility of their confusion, although it was impossible to judge those cases.

### *Reliability of Measurement*

#### 1. *Situations in the DCT*

What extent to which does the DCT instrument produce consistent, accurate results when it administered under similar conditions? There are some factors of unreliable data assumed: measurement error, fatigue of participants, problems with the data collection environment, participants' lack of familiarity with a particular type of test, etc.

The DCT of refusal was administered in part of usual ninety-minute English class. It was planned to leave the last thirty minutes for students to respond the DCT. Some students can respond quickly to all situations, while some others cannot. Therefore, it was allowed that slow-respondents stayed ten more minutes to engage themselves in the DCT. Although the last part of class period left to the test administration was convenient for them, another factor for triggering unreliable data, that is, participants' fatigue was raised. Some male students looked rushing to respond in short sentences without working on seriously. Their performance might have spoiled the data. It is still a question whether to eliminate such participants from the data.

My participants were not familiar with the DCT type of test. The explanation of how to respond and situational description was made in the participants' language, that is Japanese. The researcher herself administered the DCT to the participants, supervising and checking if there was anyone who did not understand how to do. Therefore, there seemed no problem concerning test taking.

To estimate the reliability of measurement, Cronbach's alpha and split-half reliability was computed for the DCT of the refusal study in terms of situations and resulting use of

strategies. The results are shown in tables 1 and 2 at the appendix C. The reliability alpha was fairly high among situations (.79 for Cronbach alpha), but not so high among strategies (.69 for Cronbach alpha). The reliability among respondents varied in strategy measures. The ID numbers were randomly assigned, and divided into two groups. Since the data were dichotomous, tetrachoric  $r$  (interactive approx) was procedured. The attempted split-half reliability indicated different coefficients between 'grounder'/'apology' and 'alternative'/'desire' strategy use. The former was 6.899 and the latter was .2673. This may means that the former strategies (grounder and apology) were used by almost all the informants and the latter two (alternative and desire) were used by very few students. Even tetrachoric  $r$  cannot compute reliability of the proportions of use or non-use are extreme such as 95% of 0's and 5% of 1's (see for further detail in Appendix C). This study did not examine by test-retest reliability. This issue should be reconsidered for future research.

## 2. *Perception scales*

The three scales were also computed to estimate the reliability by the same reliability procedures as with the DCT. The alpha resulted was relatively low, which may mean that it is natural that individuals' perception is not consistent but of diversity. That is the reason three scales were attached to each situation. The detailed alphas are shown for each scale in Table 2 in Appendix C.

## *Conclusion*

The validity and reliability of measurement were examined quantitatively by statistical procedure in terms of multiple perspectives. The first research question asking "to what degree is the questionnaire with eight situations valid and reliable?" was partially answered affirmatively, but partially negatively. Strategies "grounder" and "apology" were used by almost every respondent. These two strategies were inevitable to use when we refuse requests in every situation. They do not discriminate from person to person, from situation to situation. The other two strategies "alternative" and "desire" were good discriminants, because very few students used these strategies according to situations. Consequently they were reliable.

The second research question asking "to what degree are the scales of perception on three factors valid and reliable?" was not answered perfectly. The reliability coefficients were not so high. This means that respondents' reaction was not consistent but unique in a certain extent.

To what extent do the reliability and validity affect the results? This issue was not investigated in this paper. However, since the validity and reliability of measuring instruments definitely affect the results, it is of great importance to consider it in quantitative research before starting research using a questionnaire. The DCT measurement is recently a greatly controversial issue and further studies and discussions are indispensable.

NOTE:

<sup>i</sup> Unpublished. Presented at the JACET 36th Annual Convention at Waseda University on September 5 in 1997.

REFERENCES

- Beebe, L. M., Takahashi, T. & Uliss-Weltz, R. (1990). Pragmatic Transfer in ESL Refusals. In R. C. Scarcella, E. S. Anderson & S. D. Krashen (Eds.), *Developing Communicative Competence in a Second Language*. New York: Newbury House Publishers.
- Beebe, L. M. & Cummings, M. C. (1996). Natural speech act data versus written questionnaire data: How data collection method affects speech act performance. In S. M. Gass, & J. Neu (Eds.), *Speech Acts Across Cultures: Challenges to Communication in a Second Language*. Mouton de Gruyter.
- Bonikowska, M. P. (1988). The Choice of Opting Out. *Applied Linguistics*, 9/2, 169-181.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and Validity Assessment*. Newbury Park: Sage Publications.
- Cohen, A.D., & Olshtain, E. (1993). The production of Speech Acts by EFL Learners. *TESOL Quarterly*, 27/1, 33-56.
- Cohen, A.D., & Olshtain, E. (1994). Researching the Production of Second-Language Speech Acts. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research Methodology in Second-Language Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Hatch, E. & Lazaraton, A. (1991). *The Research Manual: Design and Statistics for Applied Linguistics*. New York: Newbury House Publishers.
- Hinkel, E. (1997). Appropriateness of Advice: DCT and Multiple Choice Data. *Applied Linguistics*, 18/1, 1-26.
- Ikoma, T. & Shimura, A. (1993). Eigo kara Nihongo eno Pragmatic Transfer: On the Speech Act of Refusal. *Nihongo Kyouiku*, 79, 41-52.
- Kasper, G. (1989). Variation in interlanguage speech act realization. In S. Gass, C. Madden, D. Preston, & L. Selinker (Eds), *Variation in Second Language Acquisition: Discourse and Pragmatics*. Clevedon and Philadelphia: Multilingual Matters.
- Johnston, B., Kasper, G., & Ross, S. (1993, August). Production questionnaires as a method of data collection in interlanguage pragmatics. Paper presented at the 10th AILA Congress. Amsterdam.
- Rintell, E. M. & Mitchell, C. J. (1989). Studying Requests and Apologies: An Inquiry into Method. In S. Blum-Kulka, J. House, and G. Kasper (Eds.) *Cross-cultural Pragmatics: Requests and Apologies*. Norwood, NJ: Ablex.
- Rose, K. R. (1992). Speech acts and questionnaires: The effect of hearer response. *Journal of Pragmatics*, 17, 49-62.
- Rose, K. R. (1994). On the Validity of Discourse Completion Tests in Non-Western Contexts. *Applied Linguistics*, 15/1, 1-11.
- Rose, K. R. & Ono, R. (1995). Eliciting Speech Act Data in Japanese: The Effect of Questionnaire Type. *Language Learning* 45/2, 191-223.
- Sasaki, M. (forthcoming). Investigating EFL Students' Production of Speech Acts: A Comparison of Production Questionnaires and Role Plays. *Journal of Pragmatics*.
- Yokoyama, S. (1993). Nihongo ni okeru Nihonjin no Nihonjin ni taisuru Kotowari to Nihonjin no Americajin ni taisuru Kotowari no Hikaku: Shakai Gengogaku no level deno Foreigner Talk. *Nihongo Kyouiku*, 81, 141-151.

### Appendix A: Situations

1. One day you happened to see a professor on campus. You don't know him/her very well. You were asked by him/her to move to a new office next Sunday. The moving will take a whole day. The day is inconvenient for you for some reason.
2. One day you visited your professor's office to express your thanks to him/her for helping you in securing a position of employment. Thanks to him/her, you have got a post in the company that you have really wanted to get. You were asked by him/her to help him/her move to a new office next Sunday. The moving will take a whole day. The day is inconvenient for you for some reason.
3. One day you took a seat next to a classmate in the cafeteria. You don't know him/her very well. You were asked by him/her to help him/her move to a new apartment next Sunday. The moving will take a whole day. The day is inconvenient for you for some reason.
4. One day you had lunch with your close friend in the cafeteria. You were asked to help him/her move to a new apartment next Sunday. The moving will take a whole day. The day is inconvenient for you for some reason.
5. During a psychology class, a student next to you asked you to lend him/her a pen to take notes with. You don't know him/her very well. You have only one pen. If you lend him/her the pen, you may not take notes.
6. During a psychology class, your close friend asked you to lend him/her a pen to take notes with. You have only one pen. If you lend him/her the pen, you may not take notes.
7. One day you see a professor of your university at a post office nearby. S/he is going to write an address to send a parcel to. You know him/her by sight, but don't know his/her name. You don't take his/her course, either. S/he seems to notice you are a student of his/her university. S/he asks you to lend him/her a pen, but you have only one pen and you are in a hurry to go.
8. During a psychology class, the professor noticed s/he had no pen to check his/her roll book with. S/he asked you to lend him/her a pen through the class, but you have only one pen. If you lend his/her the pen, you may take notes.

### Appendix B: Questionnaire Sample

(Situation 1)

One day you happened to see a professor on campus. You don't know him/her very well. You were asked by him/her to help him/her move to a new office next Sunday. The moving will take a whole day. The day is inconvenient for you for some reason. Would you accept his/her request or refuse it?

A. I would accept it. B. I would refuse it.

In the case of refusing it, what would you say in actual conversation?

Professor: Next Sunday I'm moving to a new building on this campus. I'm looking for someone to help me. Can you help me?

You:

Professor: That's too bad. I was hoping you could help me.



**Table 2: Reliability result of perception scales**

	Cronbach alpha	Standardized alpha	Split half reliability
Familiarity	.4923	.4686	.4430
Rank	.5806	.5844	.6055
Offence	.7143	.6979	.2218
All factors	.4738	.4513	.6489