Statistics Canada Statistique Canada

Methodology Branch

Census & Household Survey Methods Division

Direction de la méthodologie

Division des méthodes de recensement et d'enquêtes ménages

Canada

[redacted]

ON A REPEATED COMPARISON QUESTIONNAIRE TECHNIQUE FOR MEASURING ORDERED CATEGORICAL VARIABLES IN TELEPHONE SURVEYS

Avinash C. Singh

Working Paper No. SSMD 86-003E

FEBRUARY 1986

ABSTRACT

In designing questionnaires for measuring ordered categorical variables, we propose to use the idea of an accuracy check within a single measurement in terms of internal consistency of responses to repeated comparisons (RCs) of the true level with the reference levels. This idea is important in view of cost effectiveness and cutting respondent burden incurred in call backs. A technique termed RCQ (Repeated Comparison Questionnaire) is proposed which consists of a systematic set of brief and simple questions and is optimal in the sense of ensuring high response accuracy. A simple graphical method for recording, consistency check, and scoring of responses is employed. This is useful for on-spot editing of aberrant responses (if any) by consulting with the respondent about doubtful answers. RCQ can also be applied to interval variables whenever their categorization is considered suitable for certain practical reasons. Although the RCQ theory fits naturally with telephone surveys because RCs arise there almost spontaneously and that CATI can be conveniently used for administering the procedure, it is also applicable to non-telephone surveys and is recommended on account of its optimality. Some illustrative examples are presented.

RÉSUMÉ

Pour l'étude de questionnaires destinés à mesurer des variables catégoriques ordonnées, nous proposons une vérification de la précision au sein d'une mesure unique en termes de cohérence interne des réponses à des comparaisons répétées (CR) du niveau vrai au niveaux de référence. Cette idée est importante en raison des économies et de la réduction du fardeau de réponse représenté par les rappels qu'elle permet. Nous proposons une méthode appelée méthode du questionnaire à comparaisons répétées (QCR); elle fait intervenir un ensemble systématique de questions simples et courtes et elle est optimale en ce sens qu'elle garantit une précision de réponse élevée. Elle utilise une méthode graphique simple d'enregistrement, une vérification de la précision et un classement des réponses. Ceci est utile pour la vérification sur le champ des éventuelles réponses aberrantes, car on peut vérifier avec le répondant toute réponse douteuse. Le QCR peut également être utilisé pour les variables à intervalle lorsque cette catégorie est envisagée pour certaines applications pratiques. Bien que la théorie QCR fasse naturellement partie des enquêtes téléphoniques car des CR s'y produisent presque spontanément, et que les ITAO peuvent très bien convenir à la mise en oeuvre de cette méthode, elle est également applicable aux enquêtes non téléphoniques et elle est recommendée en raison de son optimalité. Nous présentons par ailleurs des exemples concrets.

ON A REPEATED COMPARISON QUESTIONNAIRE TECHNIQUE FOR MEASURING ORDERED CATEGORICAL VARIABLES IN TELEPHONE SURVEYS

Avinash C. Singh

1. INTRODUCTION

In telephone surveys with a variable being measured on an ordered categorical scale (consisting of say, s categories), questions involving repeated comparisons of the true level of a respondent with each of the s categories arise naturally due to lack of visual aids for the display of categories. These questions would seek to determine where the individual's true level stands in relation to the given categories. Thus for a single measurement, the number of component questions (or repeated comparisons - RCs for abbreviation) with dichotomous responses is atmost s - 1 i.e. one less the number of categories.

The use of RCs as defined above is desirable on two accounts: (i) they give rise to questions which are brief and simple. This is of obvious importance in telephone surveys. (ii) If all the RCs were administered and responses (in the form of an (s - 1)-vector of data) explicitly recorded, then these data for a single measurement can be screened for accuracy via an internal consistency check of the response vector. This is possible because of the overlapping information about the true level provided by various comparisons with the ordered categories. It may be noted that although it is possible to administer only a subset of (s -1) comparisons for a partial internal consistency check, it would be preferable, in practice, to present all (s - 1) categories for comparison to eliminate any possible bias that might enter otherwise. Clearly, the feature of a built-in accuracy check within a single measurement process would be of great value in reducing respondent burden and cost incurred in call backs (or reinterviewing). Moreover, on spot editing of aberrant responses (if any) can be performed during the course of interview by repeating those questions causing inconsistency.

The problem considered in this paper can be stated as follows. If the scale precision (i.e. s- the number of categories) were too low, then the extent of built-in accuracy check would not be adequate although the task of performing RCs (i.e. the respondent burden) would be very light in view of a small number of RCs. For instance, even if the individual were responding carelessly or in a haphazard manner, the chance for him to come out consistent would be high. On the other hand, if the scale precision were too high, then apart from the problem of heavy respondent burden (or difficult task of performing many RCs), the practical relevance of the built-in accuracy check becomes questionable. This is so because even if the individual were responding in a reasonable rational manner, the chance for him to come out inconsistent would be high due to mere large number of questions and perhaps confusion resulting from it. We therefore, consider the problem of finding an optimal choice of scale precision (s) (or equivalently, an optimal choice of number of repeated comparisons) in order to achieve a high response accuracy (or low measurement error) which we shall define in terms of the internal consistency of response vector. First a framework of optimality in the sense of reducing response error is presented and then a solution is provided which is essentially an optimal compromise between too coarse and too fine ordered categorical scales. The optimal solution in general consists of a family of scales and the choice of a particular member would depend on the associated respondent burden and practical considerations. The optimality framework also provides as a by product a meaningful probabilistic measure for a given choice of scale precision. This, in turn, resolves the usual problem of an arbitrary choice in deciding the number of categories for commonly used ordinal scales.

The key idea used in this paper is the use of repeated comparisons for an internal consistency check for a single measurement whenever the variable is on an ordered categorical scale. This idea has been used before in the literature but in a different context concerning psychological dysfunctional states by Shapiro (1961, 1966) and Phillips (1963, 1977) on Personal Questionnaire Techniques, Blanz and Ghiselli (1972) on Mixed Standard Scales, and Singh and Bilsbury (1982) on Sequential Pair Comparisons. The problem of optimality however has not been addressed before. It may be pointed out that the well known method of paired comparisons (somewhat similar to repeated comparisons

considered here) used in the statistical problem of ranking and selection is for a purpose completely different from the problem dealt with in this paper. In the method of paired comparisons, a pair of stimuli (or objects) is presented to a subject (or a judge) for ranking with respect to an attribute (see e.g. Thurstone 1959; a review by Bock and Jones 1968; David 1963; Bradley 1976; and a recent review by Bradley 1984). The only similarity between the two types of problems is the use of questions involving comparisons. As a matter of fact, the problem considered here complements the ranking problem because an ordinal grading of each stimulus or assigning a comparative ordinal score to a pair of stimuli with respect to a certain attribute are, of course, required while using the method of paired comparisons. The problem concerning the effect of number of ordered categories in rating scales on precision of estimation of scale values studied by Ramsay (1973) is somewhat similar to our problem but considered in the special framework obtained under Thurstone's successive intervals model.

In ordinal scales, the categories can be for a subjectively measured continuous variable which is necessarily described categorically although its all possible values form a continuum. For example, 'lower', 'middle', and 'upper' for socio-economic status; 'strongly disagree', 'somewhat disagree', 'undecided', 'somewhat agree', and 'strongly agree' on some issue in opinion surveys. Such variables commonly arise in social surveys, for measuring attitudes and opinions on various issues and status of various types, business surveys and health surveys. Ordinal categories are also invariably used in describing objectively measured continuous variables such as income, age, and blood pressure by partitioning the underlying continuum in a few class-intervals (or groups) which then form the ordered categories. These categories usually contain sufficient information and are easily interpretable in practice. Besides, categorization of a continuous response variable may be desirable in dealing with sensitive or personal matters because it provides the individual with the confidence of nonidentifiability of the true value and thus controls for possible response set bias or even nonresponse. Moreover, response error due to poor memory is unlikely when measuring a continuous variable over class-intervals or categories. It may also be added that in view of the optimality theory presented in this paper for ordered categorical scales, one can

ensure, through categorization a low response error within a single measurement as given by the extent of the internal consistency check.

The theory is presented in sections 2 to 4. In section 2, a parametric family of partitions of the underlying continuum is proposed which provides a very wide class of ordinal scales for optimality framework. In section 3 we introduce the method of repeated comparisons for an internal consistency check and determination of a category for ordinal scales defined in the previous section. The issue of controlling for bias and chance error by utilizing a suitable number of RCs is discussed. Next in section 4, analogous to statistical testing, accuracy (or performance) measures via the concepts of Type I and II error probabilities are introduced and then a criterion of optimality is defined. The usual measures of accuracy, namely, bias and standard error, are not applicable in the present context because they require the assumption of at least an interval scale of measurement. Sections 5 and 6 contain applications of the proposed theory to questionnaire designs for telephone surveys and some possible applications to non-telephone surveys also (personal interview or self administered) respectively with some illustrative examples. It will be seen that the scoring task (or respondent burden) for optimal ordered categorical scales is simple, internal consistency checks can be quickly performed by means of a graphical device, and inconsistent responses (if any) can be resolved by repeating certain questions which become apparent during the course of performing RCs. With the introduction of CATI (computer assisted telephone interviewing) the proposed technique can be extremely simple with the aid of automation of all the steps involved. In the final section 7, summary is given.

2. A FAMILY Π OF PARTITIONS FOR ORDERED CATEGORICAL SCALES

We assume that for an ordered categorical variable, there exists an underlying continuous variable (i.e., the possible values of the variable being measured form a continuum) and that the ordered categories (like class-intervals) constitute a partition of the continuum. In the case of subjectively measured variables, there is often a practical limitation on the number of easily

understandable categories which is generally 3 to 5 in number. This does not provide an adequate internal consistency check because of an insufficient number of RCs (repeated comparisons). Therefore, there is a need for constructing refinements of a given partition consisting of subsets of initial categories. A suitable class of partitions is given by a family of ordinal scales, to be denoted by Π, and parametrized by two parameters (d, r) as follows.

Let r denote the given number of reference levels $L_1, \ldots, L_r$ on the continuum. These levels are simply the initial points in partition (but not including the two extremities at $\pm \infty$) corresponding to the given set of (r+1) ordered categories. The parameter d is assigned the value of 0 for the initial partition. The intervals between adjacent pairs of reference levels (r-1 in number) are subdivided in powers of 2 in order to obtain successive stages of partitioning corresponding to $d = 1, 2, \ldots$, respectively. Thus, d denotes the degree (or exponent) of the number 2^d of subdivisions of initial categories. We do not subdivide the two end class intervals on either side of the continuum for reasons concerning difficulty in practical interpretation. The partitioning precision (viz, the number s of categories) of a partition in Π is, therefore, given by $2^d(r-1) + 2$. Thus, s is (r+1) for d = 0, 2r for d = 1 4r - 2 for d = 2, and so on.

The commonly used rating scales namely Likert and Analog Scales (see Guilford 1954) also belong to the family Π. The Analog (which provides an interval level of measurement) can be obtained as a limit when r = 2 and d tends to ∞. The Likert (or the usual ordered categorical scale with s categories) is obtained, on the other hand, by setting r equal to s-1 and d at its minimum value of 0. Here the descriptions of first (s-1) ordered categories define the reference levels $L_1, L_2, \ldots, L_{s-1}$ representing certain points on the continuum and the class intervals corresponding to s categories will be defined as $(-\infty, L_1]$, $(L_1, L_2]$, ..., $(L_{s-2}, L_{s-1}]$, (L_{s-1}, ∞). For example, with 3 categories 'lower', 'middle', and 'upper', the three class intervals are $(-\infty$, lower], (lower, middle], and (middle, $\infty)$.

In practice, it may be preferable to substitute the medians of alternate pairs

of reference levels for the initial partitioning points $L_2, \ldots, L_{r-1}$ (i.e. all except the end points L_1 and L_r). We shall denote the alternate pair medians by L_1/L_3, L_2/L_4, ... The RC for the partitioning point L_{i-1}/L_{i+1} may be construed as an easier task for the individual than the RC for the point L_i. The former RC seeks to determine whether the true level is closer to L_{i-1} to L_{i+1} while the latter RC finds if the true level is more than L_i or not (see section 3 for details). With the modified initial points of partition as mentioned above, the resulting ordinal scale will be different if L_i's are not equally spaced. In this paper we will restrict our discussion to the above modified scales. The appropriate changes for the other case would be obvious whenever required.

Figure 1 (a,b,c) illustrates partitions in the family Π when (d,r) is $(0,9)$, $(1,5)$ and $(2,3)$ with $s = 10$ for each case. The lengths of the categories (or class intervals) are unknown in general but are shown equally spaced for convenience. The points of partition can be grouped according to successive stages of subdivision. More specifically, we have

<u>Stage '0'</u> It consists of end reference levels L_1 and L_r and the alternate pair medians L_i/L_{i+2}, $i = 1, \ldots, r-2$.

For $d = 0$, only stage '0' partitioning points are required. For $d = 1$, we also need stage '1' partitioning points.

<u>Stage '1'</u> It consists of adjacent pair medians L_i/L_{i+1}, $i = 1, \ldots, r-1$.

For $d = 2$, we need stage '2' points in addition to those for stages '0' and '1'.

<u>Stage '2'</u> It consists of adjacent pair first and third quartiles to be denoted by $L_i/(L_i/L_{i+1})$ and $(L_i/L_{i+1})/L_{i+1}$ respectively for $i = 1, 2, \ldots, r-1$.

and so on for higher values of d.

It should be noted that for $d = 0$, 1 and any r, the partitioning points are well ordered and so the categories are well defined. However for $d \geq 2$. some regularity conditions concerning relative distances of L_i's are required whenever $r > 2$. For example, for $d = 2$, the condition is that the distance between any adjacent pair exceeds half of the distance for the pair preceding or following it. This condition would seem to be reasonably met in practice although it may not be possible to verify it. In practice, we will not be interested in partitions corresponding to $d > 2$ when $r > 2$ because they lead to nonoptimality (see section 4). It may be pointed out that there will be no need of the above regularity condition if the initial partitioning points L_2, ..., L_{r-1} were not replaced by alternate pair medians.

3. REPEATED COMPARISONS AND ACCURACEY CHECK WITHIN A SINGLE MEASUREMENT

We will describe RCs (repeated comparisons) for measuring on ordered categorical scales Π by considering the three cases corresponding to $d = 0$, 1, 2 respectively. These RCs are somewhat different from those mentioned in the beginning of section 1 because here comparisons involve partitioning points (i.e. category boundaries) rather than categories directly.

Case I ($d = 0$). It can be seen from Fig. 1(a) that for measuring on partitions Π with $d = 0$, the required RCs can be classified in two steps.

Step (i) RC with Alternate Pairs: The individual is asked which of the two levels in the alternate pair (L_{i-1}, L_{i+1}), $i = 2, \ldots, r-1$, is closer to the true level.

Step (ii) RC with End Levels: The individual is asked whether his true level is more or less than L_i ($i = 1, r$). If the individual is at L_i, then we use the following convention. Assign first category C_1 if at L_1 and last category C_s if at L_r.

<u>Case II</u> ($d = 1$). It follows from Fig. 1(b) that for measuring on Π with $d = 1$, one more step of RCs than those for case I is required. The three steps are:

<u>Step (i)</u> <u>RC with Alternate Pairs</u>

<u>Step (ii)</u> <u>RC with Adjacent Pairs</u>: The individual is asked which of the two levels in the adjacent pair (L_i, L_{i+1}), $i = 1, \ldots, r-1$, is closer to the true level.

<u>Step (iii)</u> <u>RC with End Levels</u>

<u>Case III</u> ($d = 2$). From Fig. 1(c) it can be seen that measurement on Π with $d = 2$ can be accomplished in four steps, namely,

<u>Step (i)</u> <u>RC with Alternate Pairs</u>

<u>Step (ii)</u> <u>RC with Adjacent Pairs Using Middle Option</u>: The individual is asked whether his true level is closest to L_i or L_{i+1} or 'the middle in between' for the adjacent pair (L_i, L_{i+1}).

<u>Step (iii)</u> <u>RC with Adjacent Pairs</u> (without middle option).

<u>Step (iv)</u> <u>RC with End Levels</u>.

Note that for $d = 2$, four equal subdivisions between successive reference levels are achieved by administering the adjacent pair RC with and without the 'middle' option. The questions in step (ii) are now trichotomous whereas all the other questions remain dichotomous. When $d \geqslant 3$, the question for RCs are not quite simple as before. For example, with $d = 3$ and $r = 2$, the required RCs can be obtained from the case $d = 2$ and $r = 3$ by regarding the centre (or median L_1/L_2) as a new reference level. As will be seen in section 4, values of d other than 1 and 2 would be rarely needed in practice.

For each fixed (d, r) in Π, let N denote the total number of RCs and s as before denotes the number of categories. We have

i) $d = 0$, $s = r + 1$, and $N = r$
ii) $d = 1$, $s = 2r$, and $N = 2r - 1$
iii) $d = 2$, $s = 4r - 2$ and $N = 3r - 2$
iv) $d = 3$, $s = 8r - 6$ and $N = 6r - 5$.

For arbitrary (d,r), one can calculate N in a similar manner. If we wish s of 10, then the value of N for
$(d, r) = (0, 9)$ is 9 (all dichotomous), for $(d, r) = (2, 3)$ it is 7 (5 dichotomous and 2 trichotomous) and for $(d, r) = (3, 2)$ it is also 7.

If one has confidence in respondent's accuracy, then there is no need for performing accuracy check via internal consistency and consequently, the required number of RCs will be very small. For example, by employing binary search, it is easily seen that for $d = 0$, $r = 9$, a category (or a score) can be selected by using only 3 - 4 RCs; for $d = 1$, $r = 5$, a score can also be obtained in 3 - 4 RCs while for $d = 2$, $r = 3$, only 2 - 3 RCs will be required. Therefore, if the remaining RCs were indeed administered, they would serve as 'replications' in the sense that they would provide overlapping information which can be used for measuring accuracy via a consistency check. Inconsistency can be caused by either bias or chance error in response. For instance, if irrelevant response set and intentional biases were present, then it would be difficult for someone to deliberately falsify the response and yet come out consistent provided there were enough RCs to be performed (to be discussed further in section 4). Similarly, chance error due to task difficulty (if there were quite a few RCs) would give rise to inconsistency. It is also possible to check bias due to order effects and unintentional factors by the internal inconsistency if randomization in the order of presentation is introduced while administering RCs within each step.

4. A CRITERION OF OPTIMALITY FOR THE NUMBER OF CATEGORIES (TYPE I AND II ERROR PROBABILITIES)

As mentioned in section 1, there are two types of error that should be controlled for with the RC method. The problem is analogous to testing statistical hypotheses. Let H_0 denote the hypothesis of 'random' response model in the sense that the respondent assigns equal chance to all possible answers to each question and that the questions are answered independently. We now define for partitions Π,

α = Type I Error Probability

$$= \Pr [\text{ consistent response in N repeated comparisons } / H_0] \tag{4.1}$$

Clearly α depends on (d, r) because N does. It will be seen that α is a non-increasing function of N (or s). Now, let H_1 denote a 'rational' response model defined by assumptions A1 - A3.

> <u>Assumption A1</u>: There exists a number δ (critical number for discrimination) such that whenever s-1 (the number of partitioning points) is $\leq \delta$, there is no error in selecting a category containing the true level (denoted by λ).

It follows from Miller (1956) that δ is $7(\pm 2)$ in view of the human capacity for discrimination. Whenever $(s-1) > \delta$ but $r \leq \delta$, we define a <u>critical degree</u> (d_c) as that value of d for which

$$s(d_c - 1, r) - 1 \leq \delta \text{ but } s(d_c, r) - 1 > \delta \tag{4.2}$$

where $s(d, r) = 2^d (r-1) + 2$. Note that $d_c \geq 1$ because r is assumed to be $\leq \delta$. Now for $r \leq \delta$ and $d = d_c - 1$, it follows from the assumption A1 that the individual can select a class-interval (I_λ, say) containing the true level λ without any error. We next introduce

Assumption A2: (For the case $r \leq \delta$). Within I_λ, the individual further locates λ according to a continuous probability distribution F with support I_λ.

The assumption A2 becomes superfluous if I_λ is one of the end intervals. The third assumption required to complete the definition of H_1 is

Assumption A3: (For the case $r > \delta$). The individual can choose a subset of δ reference levels (say, $L_1^* < \ldots < L_\delta^*$) such that $L_1 = L_1^*$, $L_r = L_\delta^*$ and that he can locate λ in one of the $(\delta+1)$ intervals formed by the points L_i^*'s without any error. Furthermore, the individual chooses a point within the selected error-free interval according to some distribution F as given in (A2). It is also assumed that the number $(r-\delta)$ of L_i's not chosen in the subset are as much as possible evenly interspersed with L_i^*'s.

The assumptions A1 - A3 complete the definition of H_1. Now we define for partitions Π,

$$\beta = \text{Type II Error probability}$$
$$= \Pr\,[\text{an inconsistent response in N repeated comparisons} \mid H_1] \qquad (4.3)$$

The probability β depends on (d, r), δ, I_λ and F. It will be seen that β is a nondecreasing function of N(or s). Let us also define

$$\beta^* = \Pr\,[\text{an incorrect response in N repeated comparisons} \mid H_1] \qquad (4.4)$$

It is easily seen that $\beta \leq \beta^*$ because correct response implies consistency but not vice versa. We have

$$\beta^* = 0 \text{ implies that } \beta = 0 \qquad (4.5)$$

Therefore, $\beta = 0$ whenever $(s-1) \leq \delta$ under A1. It will be seen that it would be unreasonable to restrict choice of (d, r) such that $(s-1) \leq \delta$, because in

these situations α would generally be unacceptably high. It may be remarked that although both β and β^* depend on the unknown λ, we prefer β for our theoretical treatment because of the practical feasibility in checking consistency of a response (versus checking correctness).

4.1 Computation of α

By definition of H_0, it is easily seen that for a given (d, r),

$$\alpha = s/T \tag{4.6}$$

where $s = 2^d(r-1) + 2$ and T is the total number of possible responses to N RCs corresponding to the given (d,r). Table 1 lists α values for various choices of r when d = 0, 1, 2, 3.

Table 1: α values as d and r vary

d	r								General Formula For $\alpha = s/T$
	2	3	4	5	6	7	8	9	
0	.75	.50	.3125	.1875	.1094	.0625	.0352	.0195	$(r+1)/2^r$
1	.50	.1875	.0625	.0195	.00586	.	.	.	$2r/2^{2r-1}$
2	.25	.0347	.00405	.0004	.	.	.	.	$(4r-2)/2^{2r-1}\ 3^{r-1}$
3	.0347	.0004	.	.	.	.	.	.	$(8r-6)/2^{4r-3}\ 3^{2r-2}$

Proposition 4.1 For a given r, there exist α_0 and d_0 such that

$$\alpha \leq \alpha_0 \text{ iff } d \geq d_0 \tag{4.7}$$

To prove, it is enough to show that α is a decreasing function of d. From Table 1, note that given r, both s and T are increasing functions of d but T increases much faster than s does and so s/T is decreasing in d. Hence the result.

Remark 4.1: Given r, the condition $d \geq d_0$ is equivalent to $s \geq s_0$ where $s_0 = 2^{d_0}(r+1) + 2$. Therefore the proposition 4.1 gives a condition of minimum scale precision in order to control α down to α_0. This provides a probabilistic measure (or interpretation) of s which will be helpful in practice when specifying s. For example, in the class Π pf partitions, if we wish $\alpha_0 \leq 3\frac{1}{2}\%$, then we must choose $s \geq 10$ for all values of r. If we wish $\alpha_0 \leq 2\%$, then $s \geq 10$ is still satisfactory provided $r \geq 4$.

Remark 4.2: It may be possible that the desired level α_0 can be achieved by only a subset of N RCs (in other words, by only a 'partial' instead of a 'full' replication). This would have an important practical implication in reducing respondent burden. One can also check the effect of omission of some replication on α by computing the new α. For example, for d = 1, if replications of step (iii) are omitted, than α is changed to $2(r-1)/2^{2r-3}$. Thus for r = 5, α would increase from .0195 to .0625. For d = 2, by omitting step (iv) replications, α is changed to $(4r-1)/2^{2r-3}\,3^{r-1}$. So for r = 3, α would increase from .035 to .111 and for r = 4, it would increase from .004 to .014.

4.2 Computation of β

Although β in general is not known, it is possible to specify necessary and sufficient conditions under which $\beta = 0$ for arbitrary λ.

Proposition 4.2 Let r and δ be given. Let λ denote the unknown true level. We have

a) for $r \leq \delta$,

$$\max_{\lambda} \beta = 0 \text{ iff } d \leq d_0$$

b) $\delta < r < 2\delta$

$$\max_\lambda \beta = 0 \text{ iff } d = 0$$

c) $r \geq 2\delta$,

$$\max_\lambda \beta > 0 \quad \text{for all } d.$$

Proof: (a) Suppose $\max_\lambda \beta = 0$ but $d > d_c$. Then there exists I_λ for some λ such that it contains at least 3 partitioning points (say $y_1 < y_2 < y_3$, $y_2 = (y_1 + y_3)/2$). Now denoting 1-F by $\bar{F}$, we have

$$\max_\lambda \beta \geq 1 - [F(y_2)\, F(y_3) + \bar{F}(y_1)\, \bar{F}(y_2)]$$

$$> 0$$

which leads to a contradiction. Hence the result '$\Rightarrow$'. To show '$\Leftarrow$', note that the error free interval I_λ obtained at $d = d_c - 1$ contains at most one partitioning point at $d = d_c$, thus requiring only the minimum number of one RC for further location of β. The corresponding β would be zero in view of no replication. Hence $\beta = 0$ for every $d \leq d_c$ because it is obviously a nondecreasing function of d.

(b) Suppose $\max_\lambda \beta = 0$ but $d > 0$. At $d = 1$, $s = 2r > 2\delta$. Therefore by (A3) there exists I_λ for some λ such that it contains at least 2 partitioning points, $y_1 < y_2$ (say) then

$$\max_\lambda \beta \geq 1 - [F(y_1)\, F(y_2) + \bar{F}(y_1)] > 0$$

leading to a contradiction. Hence the result '$\Rightarrow$'. To see '$\Leftarrow$', we assume $d = 0$. Now $r < 2\delta$ implies that $s = r+1 \leq 2\delta$. Therefore, by A3, the I_λ for any λ contains atmost one partitioning point and so $\beta = 0$ as in (a).

c) It easily follows from the proof of (b).

Corollary 4.1: Given δ, r and d,

$$\max_{\lambda} \beta = 0 \text{ iff } s \leq 2\delta \tag{4.11}$$

This is a direct consequence of proposition 4.2 and the fact that for $r \leq \delta$,

$$d \leq d_c \text{ iff } s \leq 2\delta \tag{4.12}$$

To see (4.12), note that

$$s(d_c - 1, r) - 1 = 2^{d_c - 1}(r-1) + 1 \leq \delta$$

$$\text{iff } 2^{d_c - 1}(r-1) \leq \delta - 1$$

$$\text{iff } 2^{d_c}(r-1) + 2 \leq 2(\delta-1) + 2 = 2\delta.$$

Remark 4.3: The condition $s \leq 2\delta$ can be interpreted as the condition of maximum scale precision in order to hold β down to zero.

4.3 A Criterion of Optimality

Analogous to statistical testing, the two error probabilities α and β are inversely related because while α is a decreasing function of d, β is a non-decreasing function of d when r is fixed. It is possible to minimize β holding α fixed for a class of partitions in Π. Thus we define a partition in Π to be optimal if for given δ, α_0, r; the value of d is such that $\beta = 0$ while $\alpha \leq \alpha_0$.

Proposition 4.3 Given δ and α_0, the optimal class of partitions in Π satisfying $\alpha \leq \alpha_0$ is given by pairs (d, r) such that

$$s_0 \leq s \leq 2\delta \tag{4.13}$$

Moreover, if r is also given, then the optimal class satisfying $\alpha \le \alpha_0$ is a subset of the previous class and is given by values of d such that

$$d_0 \le d \le d_c \tag{4.14}$$

Proof follows easily from propositions 4.1, 4.2 and corollary 4.1.

Remark 4.4: For the optimal class of partitions, the conditions of minimum ($s \geqslant s_0$) and maximum ($s \le 2\delta$) scale precision must be satisfied. Therefore, if values of δ and α_0 are such that $s_0 > 2\delta$, there will not exist an optimal partition. For instance, with $\alpha_0 = .0195$ (or about 2%), there is no optimal choice of (d,r) when δ is 5 (the most conservative value of δ in view of Miller's result, namely, $5 \le \delta \le 9$). For the proposed technique (see section 5), we will take $\alpha_0 = .0352$ (or about 3½%) and $\delta = 7$ (the median of the range 5 to 9) as working values because apart from these values being reasonably small, the corresponding optimal class does contain various partitions (d, r) of practical interest. The optimal class for $\delta = 7$, $\alpha_0 = .0352$ is given by:

$$\{(d, r) = (3,2), (2,3), (2,4), (1,5), (1,6), (1,7), (0,8), (0,9), (0,10), (0,11), (0,12), (0,13)\} \tag{4.15}$$

Thus, values of d other than 1 and 2 would be rarely needed because r is generally between 3 to 7 in practice.

Remark 4.5: Under the optimality condition (4.13), β^* may be positive although $\beta = 0$. It should be noted that with a rather stringent condition of $s \le \delta + 1$, we will have $\beta^* = 0$ but α would be generally quite high. It then follows that form the optimal class, one should in practice choose (d, r) such that s is as small as possible in order to keep respondent burden minimum possible which in turn would render β^* small. Note that for the optimal class (4.15), s varies from 10 to 14 whenever r is between 3 to 7. From Miller (1956) it is seen that the most liberal choice of δ is 9 and so whenever possible one should restrict s not to exceed 10 while maintaining $\alpha \le \alpha_0$. Thus for the optimal class (4.15), the best choice of s is 10.

5. THE PROPOSED RCQ (REPEATED COMPARISONS QUESTIONNAIRE) TECHNIQUE FOR TELEPHONE SURVEYS

We will describe the proposed RCQ technique by means of two examples when (d, r) is (1,5) and (2,3), both yielding a partitioning precision of 10. The corresponding values of α for the full replication case are respectively .0195 and .0347. These examples typically arise in applications because 3 to 5 reference levels are generally available in practice.

Example 5.1: (d = 1, r = 5)

Consider a hypothetical situation involving 'job satisfaction' as a variable being measured with respect to five reference levels, L_1 to L_5, namely

very dissatisfied		so-so		very satisfied
(L_1)	(L_2)	(L_3)	(L_4)	(L_5)
	moderately dissatisfied		moderately satisfied	

It follows from section 3 that only 3 - 4 RCs (consisting of dichotomous questions will be required for selecting a category out of 10 when no replications are performed. This nonreplicated (or short) version of RCQ for the case (d=1, r=5) consists of the following steps.

STEP I

Question 1: Closer to L_2 or L_4?

Answer: (L_2) moderately dissatisfied ... Go to Question 2
(L_4) moderately satisfied ... Go to Question 3.

Question 2: Closer to L_1 or L_3?

Answer: (L_1) very dissatisfied ... Go to Question 4.
(L_3) so-so. ... Go to Question 5.

Question 3: Closer to L_3 or L_5?

Answer: (L_3) so-so ... Go to Question 6.
(L_5) very satisfied ... Go to Question 7.

STEP II

Question 4: Closer to L_1 or L_2?

Answer: (L_1) very dissatisfied ... Go to Question 8.
(L_2) moderately dissatisfied ... Select category 'C_3'

Question 5: Closer to L_2 or L_3?

Answer: (L_2) moderately dissatisfied ... Select 'C_4'
(L_3) so-so ... Select 'C_5'

Question 6: Closer to L_3 or L_4?

Answer: (L_3) so-so ... Select 'C_6'
(L_4) moderately satisfied ... Select 'C_7'

Question 7: Closer to L_4 or L_5?

Answer: (L_4) moderately satisfied ... Select 'C_8'
(L_5) very satisfied ... Go to Qestion 9.

STEP III

Question 8: Worse than L_1 (very dissatisfied)?

Answer: worse than (or at) L_1 ... select 'C_1'
better than L_1 ... select 'C_2'

Question 9: Better than L_5 (very satisfied)?

Answer: worse than L_5 ... select 'C_9'
better than (or at) L_5 ... select 'C_{10}'

With objectively measured continuous variables such as income, the above RCQ procedure can be used to select from 10 income categories in 3-4 brief and simple questions in terms of five reference levels of income. If all the nine questions are administered, then we will have a fully replicated (or long) version of RCQ. As mentioned in section 3, the RCs in the replicated version should be presented in a random order. This randomization should be both with respect to question number and level position within a question. This can be easily performed with CATI (computer assisted telephone interviewing). It would be preferable to restrict randomization of questions within each step in order to avoid redundancy of certain questions in practice.

With many ordinal categorical variables in a survey, it would probably not be feasible in practice to administer a replicated verstion of RCQ for each variable. A reasonable compromise would be to give the long RCQ to only a small subset of variables interspersed among others. This will provide an accuracy check at certain points of time during the course of interview. A graphical device shown in Figure 2 can be used for recording, quick consistency check and score determination (see Fig. 3(a)) in case consistency was affirmed. In the case of an inconsistent response (see Fig 3(b)), the aberrant responses can be easily detected and the corresponding questions could be repeated for resolution during the same interview.

Figure 2(a) explains the symbols for recording responses. The circled symbols are joined together from left to right. If the horizontal axis of categories is crossed at only one point, then the response will be consistent and the score is given by the category of intersection (see Fig. 3(a) for the score of C_4 for example). An inconsistent response pattern is shown in Fig 3(b) which shows that the possible categories for score are C_4 and C_7. The questions (1), (5), and (6) must be repeated for the sake of resolution of inconsistency.

<u>Example 5.2</u> (d = 2, r = 3)

Consider an ordinal preference scale with 3 reference levels L_1, L_2 and L_3, namely

(L_1) not at all   (L_2) moderate   (L_3) strong

Some other examples of 3 reference levels are: 'Left', 'Centre', and 'right' for political party preference; 'not present', 'possibly present', and 'probably present' in disease disgnosis etc. It follows from section 3 that 2 - 3 RCs (one trichotomous and others dichotomous) will be required for selection among 10 categories when the short (or nonreplicated) version of RCQ is used. It may be noted that although the required number of RCs is less than that for the previous example, not all RCs for the present example require a simple dichotomy in answers. The short RCQ for the case (d = 2, r = 3) consist of the following steps.

STEP I

Question 1: Closer to L_1 or L_3?

Answer: (L_1) not at all ... Go to Question 2
(L_3) strong ... Go to Question 3.

STEP II

Question 2: Closest to L_1 or L_2 or the middle in between?

Answer: (L_1) not at all ... Go to Question 6.
(L_2) moderate ... select 'C_5'
(L_1/L_2) middle ... Go to Question 4.

Question 3: Closest to L_2 or L_3 or the middle in between?

Answer: (L_2) moderate ... select 'C_6'
(L_3) strong ... Go to Question 7.
(L_2/L_3) middle ... Go to Question 5.

STEP III

Question 4: Closer to L_1 or L_2?

Answer: (L_1) not at all ... Select 'C_3'
(L_2) moderate ... Select 'C_4'

Question 5: Closer to L_2 or L_3?

Answer: (L_2) moderate ... Select 'C_7'
(L_3) strong ... Select 'C_8'

STEP IV

Question 6: At L_1 (not at all)?

Answer: More than L_1 ... Select 'C_2'
At L_1 ... Select 'C_1'

Question 7: Less than L_3 (strong)?

Answer: More than (or at) L_3 ... Select 'C_{10}'
Less than L_3 ... Select 'C_9'.

For the long (or fully replicated) version of RCQ, all the seven questions are administered in a (restricted) randomized order as in the previous example. In practice, it would be preferable to perform Step III before Step II for the long RCQ for the case (d = 2) in order to avoid obvious redundancy of certain questions. With this change (Questions 2 and 3 replaced by Questions 4 and 5 respectively and vice versa), Figure 2(b) shows a graphical device for recording, consistency check and scoring of responses. Inconsistent responses, if any, can be resolved as before.

6. APPLICATION OF RCQ TO NON-TELEPHONE SURVEYS

The RCQ technique provides accuracy check within a single measurement at a cost of a little extra effort in the case of telephone surveys because RCs are naturally performed as it is not practical to display (or read out) all the categories simultaneously. However, with non-telephone surveys such as personal interview or self-administered, there is no problem of displaying or presenting all categories at the same time. Even so, it may be considered desirable to use RCQ with non-telephone surveys in view of benefits of having brief and simple questions and an internal consistency check within a single measurement. It may be remarked that the single task of an overall comparison (or rating) of the true level with all the categories (10 or so for example) simultaneously might be quite difficult, perhaps leading to inaccuracy (although it is not possible to check it with a single rating). Thus, RCQ (consisting of several simple tasks) might be preferable over the task of rating eventhough the latter consists of a single task. Moreover, there is a general problem of arbitrariness in the choice of number of categories in rating with ordinal scales, a solution for which can be obtained from optimality considerations of RCQ. Notice that although RCQ is not used explicitly in rating, the use of RCs may be thought to be implicit in any form of rating and so the theory of RCQ may be deemed to be applicable for rating methods.

There may be several versions of RCQ suitable for different types of variables and correponding reference levels in dealing with non-telephone surveys. Figures 5(a,b) and 6(a,b) correspoding to ($d = 1$, $r = 5$) and ($d = 2$, $r = 3$) respectively show possible versions of short and long RCQ which seem appropriate for many situations.

The selection of a category is self explanatory from Figures 4(a) and 5(a). In Fig. 4(b) and 5(b), the response form is similar to the graphical device (figures 2a and b) except that the alternate symbols on either side of the category axis for recording responses are switched. This will give a zig-zag pattern for a consistent response with only one breakspot. The score (or category) can be easily determined from the location of the breakspot. This alternation was made to guard against possible irrelevant response in self-

administered surveys because of the danger of consistent pattern being too obvious. One can, of course, use any other mixing sequence for symbols in order that the consistent response pattern is not too apparent. It will also be preferable to randomize the order of questions for long RCQ as mentioned in section 5. This can be easily incorporated in Figure 4(b) and 5(b) by renumbering the questions according to the given random sequence. If reference levels are long and descriptive which may occur with subjectively measured variables, then levels in Figures 4 and 5 can be presented vertically with one above the other rather than horizontally.

7. SUMMARY

For measuring over ordered categorical scales, a simple technique termed RCQ (Repeated Comparisons Questionnaire) was proposed. RCQ consists of a systematic set of repeated comparisons (RCs) of the true level with one or two reference levels in order to select a category (or score). A theory of optimality was developed under a suitable framework. In the following, the main observations and results are summarized.

(1) RCQ suits telephone surveys very well because RCs are brief and simple and that they arise naturally due to lack of visual aids for displaying categories.

(2) RCQ can provide a built-in accuracy check within a single measurement via internal consistency of responses (generally dichotomous) to component questions or RCs. A simple graphical method is used for recording, consistency check (on spot editing if necessary) and score determination. This as well as randomization in the order of presentation of RCs can be automated with CATI. With the accuracy check, the respondent burden in callbacks or reinterviewing may be reduced. Furthermore various biases and chance error in response can be controlled by the internal consistency check.

(3) If the categories are too many, the respondent burden will of course

be high due to many RCs; while if the categories are too few, the resulting accuracy check will not be adequate. A solution to the problem of finding an optimal choice of number of categories in order to ensure high accuracy while keeping respondent burden minimum possible was provided by RCQ theory. It turns out that generally speaking, number ten is the best choice. Some other choices may also be optimal under certain specific conditions. It may be noted that the number of categories in commonly used ordinal scales is generally arbitrarily fixed from certain practical considerations.

(4) Usually 3 to 5 categories are easily available and meaningful in practice especially with subjectively measured variables. A method based on RCs for making refinements of a given partition was employed in order to enhance the number of class-intervals partitioning the underlying continuum.

(5) With 10 ordinal categories, the number of RCs required in RCQ for consistency check and scoring is 8 or so. Note that in telephone surveys the number of category comparisons (or RCs with categories instead of partitioning points) required simply for scoring over 10 categories varies from 1 to 9. Here the category comparisons are as in usual rating method and therefore they are somewhat different from RCs as defined in section 2. Thus with one reinterview, the number of category comparisons required on the average would be comparable to the number of RCs in employing RCQ for telephone surveys.

(6) If in a survey there are many variables measured on ordinal scales, it may not be considered necessary to perform accuracy check for each but rather for a few and far in between may be sufficient. Thus the number of RCs for selecting a category by RCQ will be reduced with 10 categories, for instance, to 3 or so from 8 or so.

(7) It may be advantageous to apply RCQ to non-telephone surveys also because accuracy check within a single measurement would be available and that the single task of an overall comparison of the true

level with all the categories (10 for example) might be perceived to be more difficult than several (8 or so) simple tasks of RCs. Moreover, formulation of more than a few (4 ± 1) ordered categories for the purpose of rating is generally not easy in practice.

(8) In view of the optimality, RCQ can also be beneficially used with interval variables (objectively measured continuous variables such as income) because they are commonly discretized into a finite number of class-intervals due to certain practical considerations and that the associated loss of information is not generally deemed important.

ACKNOWLEDGEMENT

Support for this research from Statistics Canada is gratefully acknowledged. Thanks are due to M.P. Singh and D. Drew for their helpful suggestions and encouragement and especially to G.H. Choudhry for several useful discussions. I am also thankful to Pat Pariseau for her efficient typing.

REFERENCES

Blanz, F. and Ghiselli, E.E. (1972). The Mixed Standard Scale: A New Rating System. Personnel Psychology, 25, 185-199.

Bock, R.D. and Jones, L.V. (1968). The Measurement and Prediction of Judgment and Choice. Holden-Day, San Francisco.

Bradley, R.A. (1976). Science, Statistics and Paired Comparisons. Biometrics 32, 213-232.

Bradley, R.A. (1984). Paired Comparisons: Some Basic Procedures and Examples. In Handbook of Statistics, Vol. 4, Eds. P.R. Krishnaiah and P.K. Sen, Elsevier Science Publishers, 299-326.

David, H.A. (1963). The Method of Paired Comparisons, Griffin, London.

Guilford, J.P. (1954). Psychometric Methods, 2nd Ed, McGraw-Hill, N.Y.

Miller, G.A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits to Our Capacity for Processing Information. Psychological Review, 63, 81-97.

Phillips, J.P.N. (1963). Scaling and Personal Questionnaires. Nature, 200, 1347-1348.

Phillips, J.P.N. (1977). Generalized Personal Questionnaire Techniques. In the Measurement of Intra-Personal Space by Grid Technique, Ed. P. Slater, Vol. 2, Wiley, London.

Ramsay, J.O. (1973). The effect of number of categories in rating scales on precision of estimation of scale values. Psychometrika. Vol. 38, 4, 513-532.

Shapiro, M.B. (1961). A Method of Measuring Psychological Changes Specific to the Individual Psychiatric Patient. British Journal of Medical Psychology, 34, 151-155.

Shapiro, M.B. (1966). The Single Case in Clinical-Psychological Research. Journal of General Psychology, 74, 3-23.

Singh A.C. and Bilsbury, C.D. (1982). Scaling Subjective Variables by SPC (Sequential Pair Comparisons). Behavioural Psychotherapy, 10, 128-145.

Thurstone, L.L. (1959). The Measurement of Values. University of Chicago Press, Illinois.

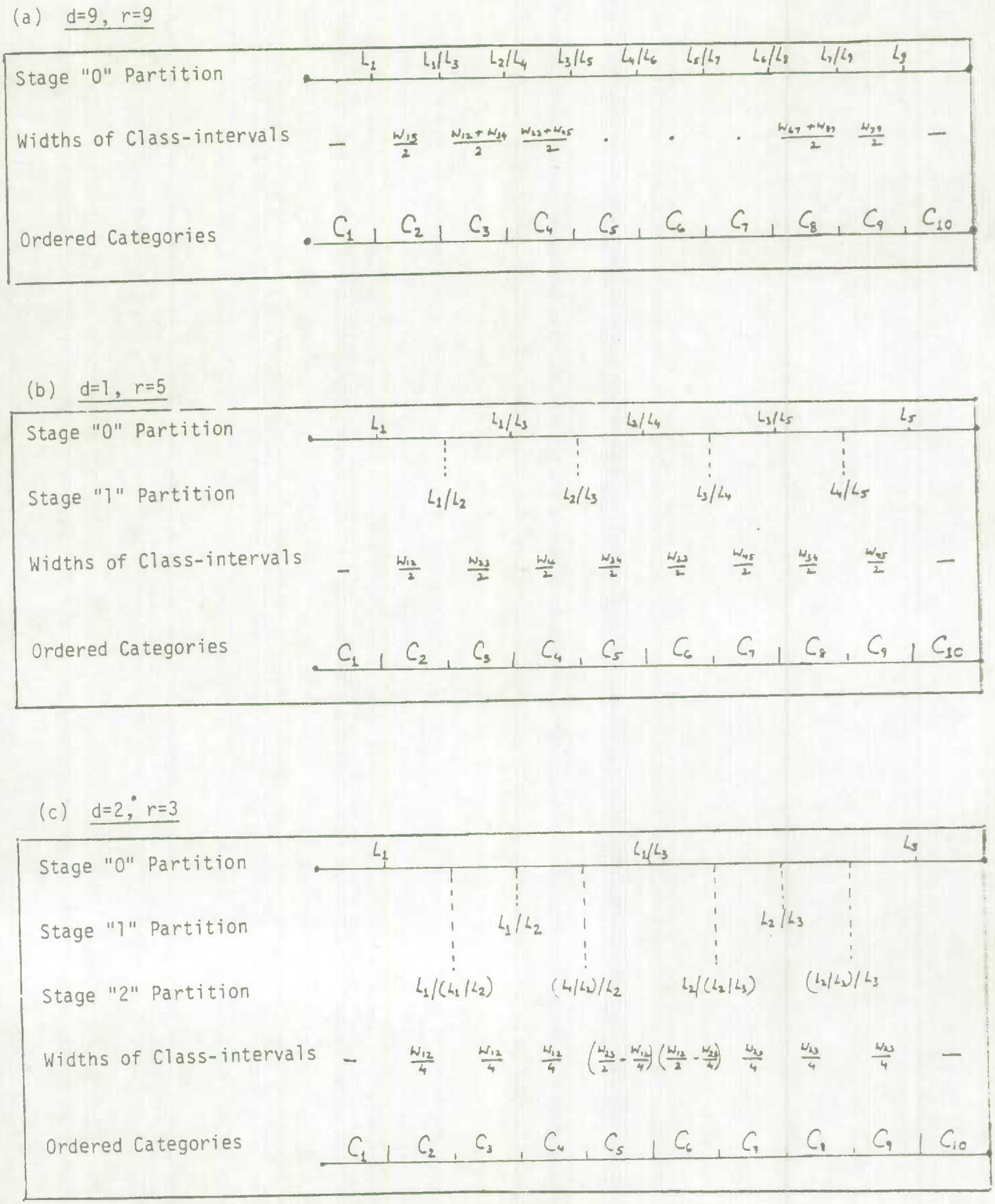
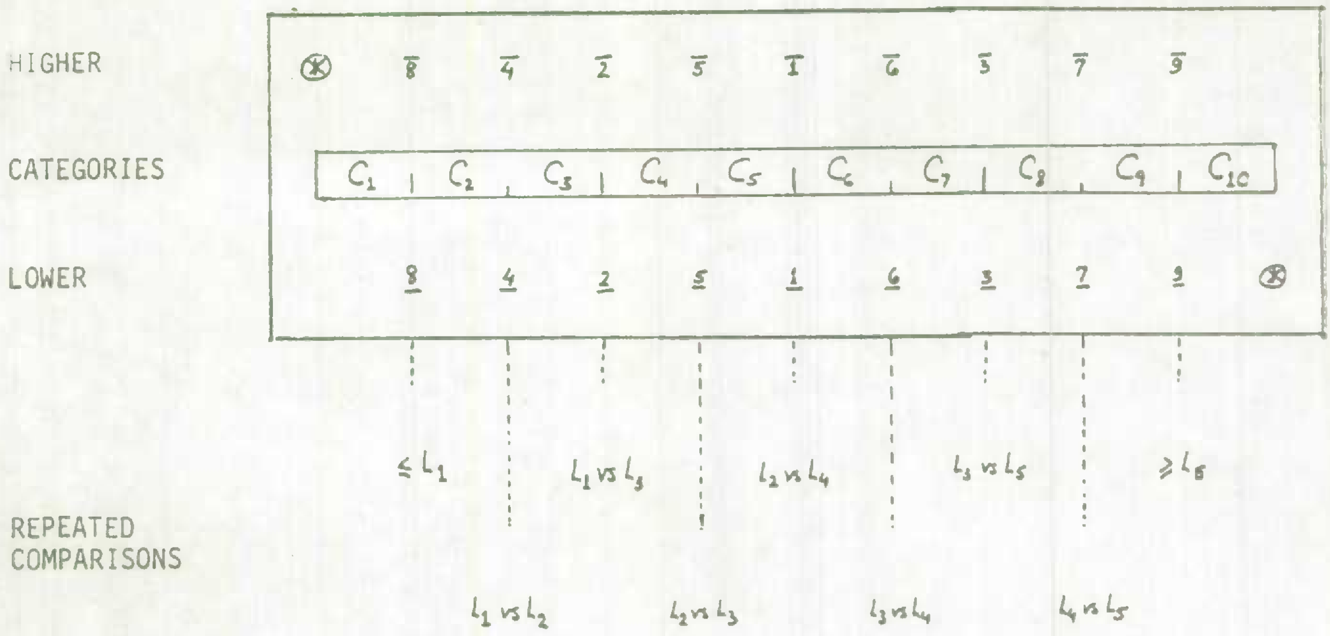


Figure 1. Continuum Partitions in the Family Π with 10 Categories

(The symbol w_{ij} denotes the width of the class-interval (L_i, L_j))

(a) d=1, r=5



(b) d=2, r=3

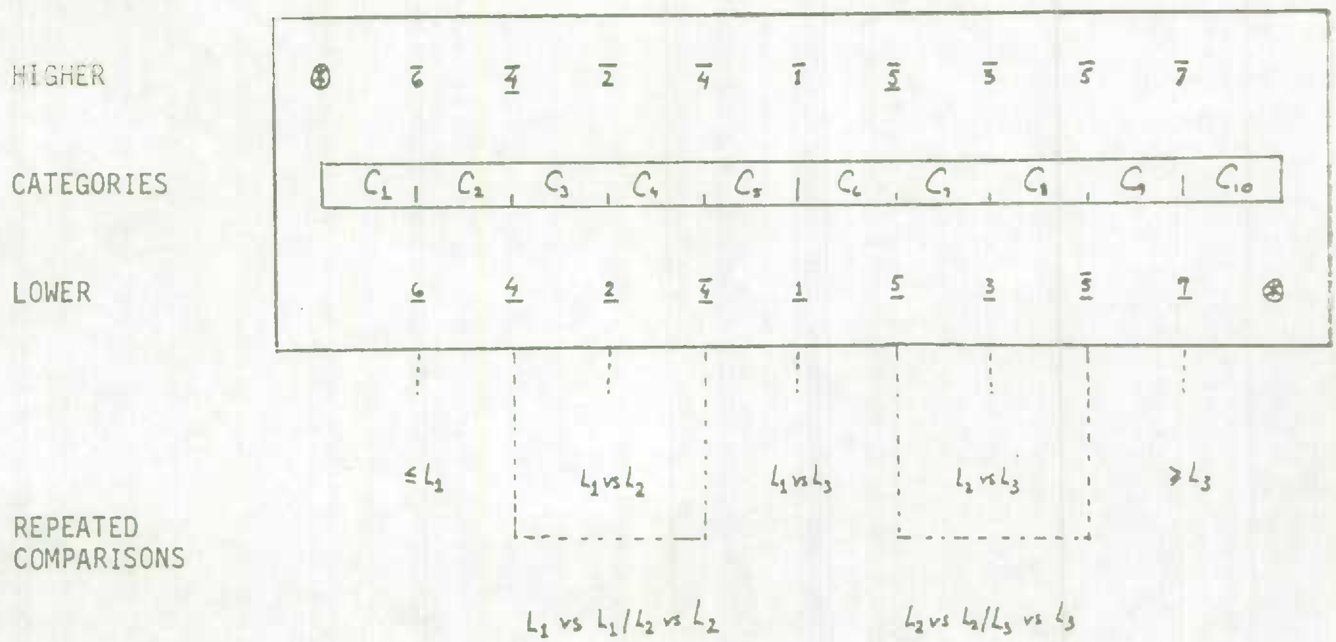
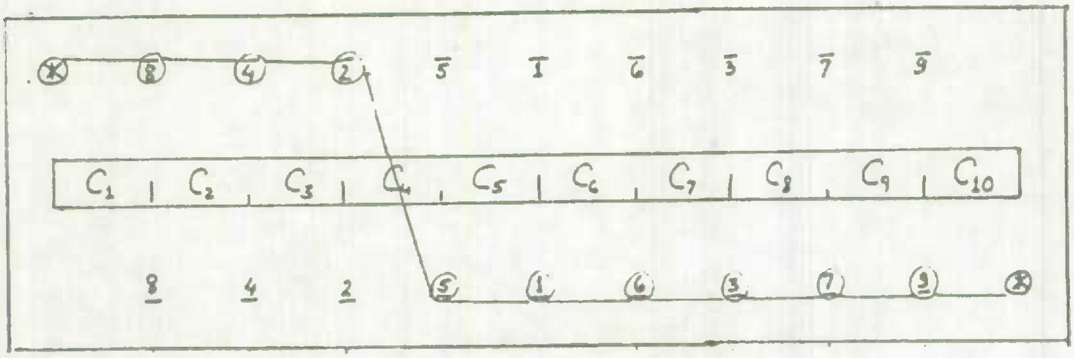


Figure 2 (a) A Graphical Response Form for RCQ (d=1, r=5)

(b) A Graphical Response Form for RCQ (d=2, r=3)

(The symbol $\underline{i}$ or $\bar{i}$ or $\bar{\underline{i}}$ are circled according as response to the i^{th} question corresponds to the lower or the higher or the middle level. The two precircled symbols '*' indicate that the true level is always between the two extremities.)

(a)



(b)

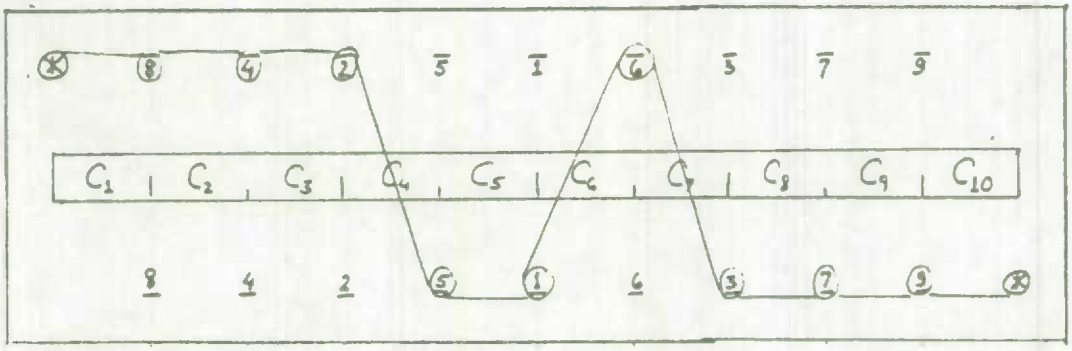
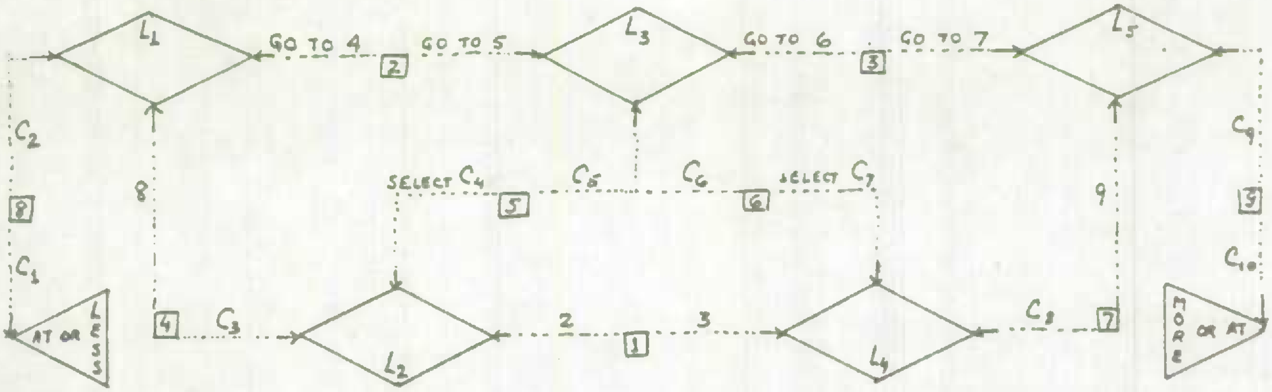


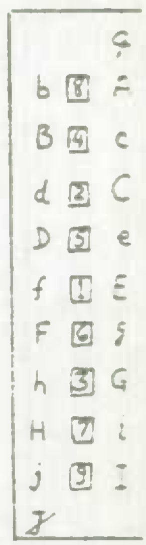
Figure 3 (a) An Illustration of a Consistent Response Pattern and a Score of C_4 for RCQ (d=1, r=5)

(b) An Illustration of an Inconsistent Response Pattern for RCQ (d=1, r=5)

(a) Short RCQ (d=1, r=5)



Response Form



(b) Long RCQ (d=1, r=5)

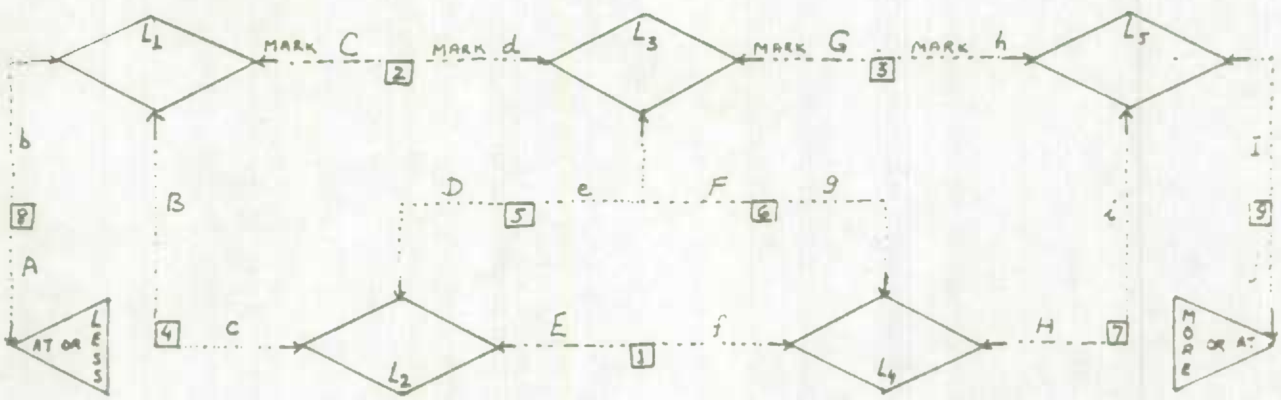
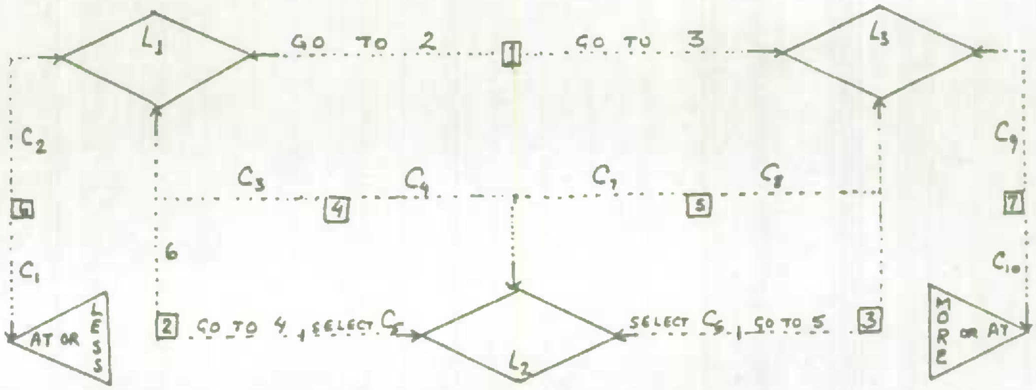


Figure 4 (a) A Flow Chart Version of Short RCQ (d=1, r=5)

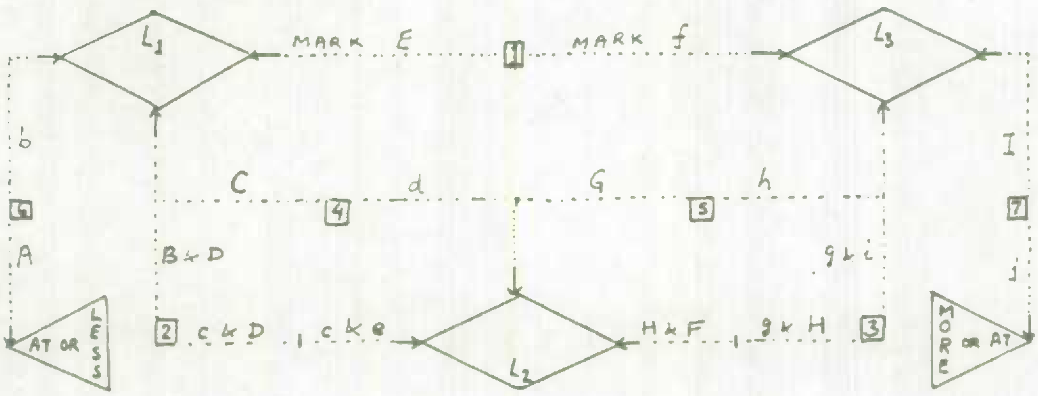
(b) A Flow Chart Version of Long RCQ (d=1, r=5)

(Questions: (1) L_2 vs. L_4, (2) L_1 vs. L_3, (3) L_3 vs. L_5, (4) L_1 vs. L_2, (5) L_2 vs. L_3, (6) L_3 vs. L_4, (7) L_4 vs. L_5, (8) $\leq L_1$, (9) $\geq L_5$)

[a] Short RCQ (d=2, r=3)



[b] Long RCQ (d=2, r=3)



Response Form

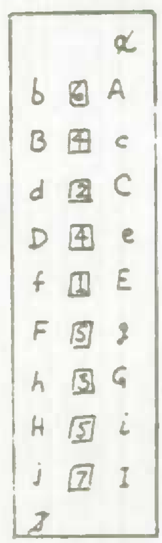


Figure 5 [a] A Flow Chart Version of Short RCQ (d=2, r=3)

[b] A Flow Chart Version of Long RCQ (d=2, r=3)

(Questions: [a] (1) L_1 vs. L_3, (2) L_1 vs. L_1/L_2 vs. L_2, (3) L_2 vs. L_2/L_3 vs. L_3, (4) L_1 vs. L_2, (5) L_2 vs. L_3, (6) $\leq L_1$, (7) $\geq L_3$.

[b] as in [a] except that (2) and (3) are interchanged with (4) and (5) respectively.)