22

Preference ranking: a cautionary tale from Papua New Guinea

William J. Fielding and Janet Riley, with a response from Robert Chambers

Feedback is a forum for discussion in *PLA Notes*. It features articles which raise common concerns or challenges in fieldwork or training, together with a response from another practitioner of participatory approaches. Letters and articles are welcomed for this section, as are your comments on any of the issues raised by ***Feedback***.

• Introduction

Preference ranking is commonly used in participatory research to *understand* local people's preferences and priorities. These can include characteristics, aspects of a new farming method or more general development needs. Ordered lists are an important result of preference ranking, which can be used to guide researchers' future investigations or plan policies.

However, some problematical aspects of ranking and their interpretation have already been raised in *PLA Notes* (See Fielding, Riley & Oyejola, *PLA Notes* 33, October 1998). This article raised the importance of having a sufficient number of respondents in the study before differences in the ranked order of a list of items can be reliably identified. In the current article we discuss some ways in which ranking techniques can be misleading and limiting, and suggest some solutions. One method used to obtain a priority ranking of a list of items is to ask respondents to put each item in rank order; the numbers associated with the ranks are then added across all respondents and the totals are used to obtain an overall ranking from the group.

An example

Consider the following example. Four farmers are asked to rank five characters of a tree species using the numbers 1,2,3,4 and 5 to indicate worst to best respectively. In practice, farmers may be given counters to allocate to the items being ranked. Suppose the following ranks were obtained (see Table 1). The 'TOTAL' column, containing the sum of the numbers associated with the characters, can be used to give the overall preferences of the four farmers. In this exercise the farmers felt that the tree species was most useful as a building material.

Table 1. Example of a preference ranking method

Character	Farmer 1	Farmer 2	Farmer 3	Farmer 4	TOTAL
Fodder	1	2	2	3	8
Thatching	2	3	3	2	10
Straw	3	4	4	1	12
Fire wood	4	5	1	4	14
Building	5	1	5	5	16

Feedback… Feedback… Feedback…

An example using two methods of obtaining preferences

In the example above where five characters were ranked, farmers were asked to allocate the numbers 1,2,3,4 and 5 to each character. This system allows each respondent 15 'votes' to be allocated, but only in a specified way. Another method of obtaining preferences is scoring. Here there is a fixed total number of 'votes' or counters which respondents can allocate as they wish. The literature contains many examples of both ranking and scoring methods. But can the choice of method have an effect on the final ranks and hence conclusions? Does it matter which method is chosen?

In a recent biometry training session in Papua New Guinea, 13 researchers were asked to indicate the relative importance of five problems which were thought to be constraining agricultural production in the country. They were asked to do this using two different methods, so that the two methods could be compared:

- *System 1 (ranking):* they ranked the problems in order of importance using a simple 1, 2, 3, 4, 5 ranking system (as described above).

- *System 2 (scoring)*: they were asked to score the same five problems by allocating a total score of 25 'votes' between each of the five problems. Thus if a respondent considered all problems to be equally important, a score of 5 would be given to each; if only one problem is considered important this could be given 25 votes and the remaining problems no votes (see Table 2).

The total scores obtained from the researchers using Systems 1 and 2 are given in Table 2. It can be seen that although there is general agreement between the two ranking methods, different total rank orders are obtained. For example, using System 1, it seems that 'lack of suitable varieties' was regarded as the most important constraint. However, looking at the results from System 2, it appears that 'weed problems' are the greatest constraint. Thus it seems that the two methods have produced different results from the same group of people, to the same questions under the same conditions!

Statistical analysis was carried out to see if these rank orders were equivalent and, if not, which one was correct. The analysis revealed that there were statistically significant differences between the five totals. Using the same statistical technique for the corresponding 13 datasets in System 2, significant differences were also revealed between the five System 2 totals.

Thus, the statistical analysis indicates the differences in the scores are large enough for us to say that some problems really are more important than others. However, when multiple comparisons are calculated between the total scores within each system separately, the results indicate that there are different significant groupings in the two different total rank orderings and thus the order of the lists is not equivalent. If there were similar significantly different groupings in the two datasets, the two systems could be assumed to have generated equivalent results. The same would apply if no significant groupings were found in both datasets. Thus the statistical analysis indicates that the two methods have produced a different set of preferences, either or none of which may be correct.

Clearly we would not like our conclusions to be dependent upon the method used to collect the preferences. In the above example, the two methods produced different results from the same group of people, to the same questions under the same conditions. This is a worrying observation and it would require triangulation of the results (checking the results using data from other sources, if they exist) to verify which ordering is correct.

Table 2. Results from two preference methods

Problem	System 1 Only ranks 1 to 5 allowed	System 2 Scoring with a total of 25 'votes'
Weed problems	43	86
Pests and disease problem	43	73
Lack of suitable varieties	47	64
Poor market price	30	53
Lack of markets	32	49

Two preference methods discussed

System 1 is rigid and does not allow respondents to express their views in as much detail as System 2. An exercise which allows 15 votes to be split at will between 5 items and so allows for ties, is a system 2-type method. However, use of only 15 votes reduces the range of preferences which can be expressed**.** For this reason, the second ranking method may be preferred. However, the second method can be more confusing for participants; some of the researchers in this experiment failed to allocate exactly 25 votes. If they had been given exactly 25 counters and told to apportion all of them between the five problems, this would have overcome the problem. As a result of these demonstrated differences, it is reasonable to ask what properties a 'good' ranking/scoring method should have? We feel that an appropriate method should:

- allow each item to be equally preferred (i.e. given the same score);
- allow extreme preferences; and,
- be simple to administer/explain.

Ranking (System 1) always fails the first two properties listed above, whilst scoring (System 2) always allows equal ranking of the items. Scoring is also easy to administer/explain if the number of items being compared is small (six or less). If the number of items to be scored is exceptionally large (more than six), we suggest that for practical reasons five times as many counters as items are allocated between the items. So if there are eight items, 40 (5 times 8) counters are given to each person. Scoring also allows the possibility of greater differentiation between items which then makes it easier for the researcher to focus on items with similar scores and those which are extremely different to the next ranked item.

Because, as we have seen, the method chosen can influence the results, the interpretation of the scores must be made with caution. Simply applying statistical tests is not a solution because different tests can give different results (Riley & Fielding, 1998). Probably the best solution is for researchers to collect data from as many respondents as possible and look at groupings within the total rankings and investigate how each grouping differs from the next grouping down the ranked list. Then triangulation of results with other sources should be done whenever possible, in order to assess the validity of the study.

- **William J. Fielding,** P. O. Box CB 13789, Cable Beach, Nassau, The Bahamas. Email: fielding@batelnet.bs **and Janet Riley**, IACR-Rothamsted, Harpenden, AL5 2JQ. Email: janet.riley@bbsrc.ac.uk

ACKNOWLEDGEMENTS

The authors are grateful to participants of the Biometrics Training Course conducted at the Oil Palm Research Association, Papua New Guinea, for providing the data in this article. The course was funded by the European Union.

IACR-Rothamsted receives grant-aided support from the BBSRC of the UK.

Feedback… Feedback… Feedback…

Preference ranking: a response from Robert Chambers

Fielding and Riley have illustrated the point made by Maxwell and Bart in *PLA Notes 22* (February 1995) that adding up numbers in ranking is statistically illegitimate and can mislead. One reason is that ranking does not show the size of interval. In this context, as they say, scoring methods are better. It is useful that they have also given a clear example of different results from adding up.

Two sets of comments seem pertinent. The first set is within the numbers paradigm. The second set is beyond it.
Within the numbers paradigm, the weaknesses of ranking compared with scoring have been explained in the article.

It is useful to go further and recognise that there are many different methods of scoring and estimating with matrices and that these have varied strengths and weaknesses. They can also be conducted with individuals or with groups. Those with statistical backgrounds may tend to prefer individuals to generate commensurable numbers. Some of the more commonly used methods (following and qualifying Maxwell and Bart 1995) include the following.

- *Scoring each box out of the same total.* This is the most common method. Each box is scored out of, say, 5 or 10 (16 has been used in Bangladesh for rice varieties). Boxes can be left blank.
- *Restricted scoring by column or row*. This allows the same fixed number of units to be distributed in each column or row.
- *Free scoring*. Allowing any number of counters to be placed in any box.
- *Voting systems*. Each person has a fixed number of counters and places these as votes. This is often preferred by people in the North, but is weak because late voters have more influence, having seen how earlier votes have gone.
- *Restricted overall scoring*. A fixed number of counters, often 100, are distributed between the boxes as scores.
- *Restricted overall estimating*. A fixed number of counters, often 100, are distributed between the boxes as estimates.

This is not the place for a detailed analysis but, in my judgement, the first, third and sixth points are the strongest.

- Scoring each box out of the same total is easily understood by participants, who usually quickly take over the process and argue freely. It is easy to inspect, interpret and discuss. It allows fine judgements and compromises.
- Free scoring overcomes problems of weighting. It allows an individual item to be heavily scored for a key characteristic in a way which is not as easy with any other method.
- Matrix estimating is a powerful method for estimating and allows adding up. If a total of 100 units are used, rows and columns can be added up to give percentages, for example of the relative importance of different crops in different agro-ecological zones. Beyond the numbers paradigm, there are several key points. Some of these are illustrated in Drinkwater's fascinating and revealing account of matrix scoring in Zambia (see *RRA Notes* 17, March 1993 pp. 24-28) 'Sorting fact from opinion: the use of a direct matrix to evaluate finger millet varieties':
- *Learning process.* Much of the value of matrix scoring is the process of analysis. Hearing the arguments, watching the interactions, sensing the power relations between individuals, and especially seeing how scores are debated and changed, can reveal more about the complexities, subtleties and variance in the realities and judgements than the scores themselves.

Feedback… Feedback… Feedback…

- *Whose analysis? Whose learning?* Farmer and other analysts themselves learn much from the process. As Drinkwater shows, they learn from one another. At the end of a good process they have a different understanding from the beginning.

- *Whose reality is expressed?* In any group certain individuals may dominate. Observing and interpreting interactions, and understanding the interests and views of dominant and subordinate individuals within a group deserves attention.

- *Empowerment*. Well facilitated, matrix scoring, like many other PRA-type methods, empowers the 'lowers' who carry them out. In some of my recent field experience in Eritrea, it was after matrix estimating that the village leader changed from *"whatever the government says is best"* to *"now you can see why your land policy will not work"*.

A valuable part of the process is often the discussion which follows, using the matrix as a visible agenda from which items can be picked out. This can be a wonderful experience of learning, often about topics one did not know to ask about.

Thus it is often the non-numerical aspects which matter. As practitioners, we need to exercise care to keep the reductionism of numbers and statistics in their place as useful tools, applied where appropriate to illuminate and enlighten, but not allowed to divert us from process and learning. For the realities which are analysed and shared are often complex, diverse and dynamic and are mediated through judgements and interactions. It is through facilitating good judgement and interactions that we can find ways of doing better.

- **Robert Chambers,** c/o Institute of Development Studies (IDS), University of Sussex, Brighton, BN1 9RE, UK. Email: r.chambers@ids.ac.uk

Authors' reply

We are grateful to Dr. Chambers for his comments and we just like to add a few of our own. As Dr. Chambers points out, the discussion process involved in any scoring/ranking method may be more informative than the numbers which are collected. Therefore, the exercise is useful in the *understanding* aspect of participatory studies, which is one of its key features. This is something not necessarily reflected in the data.

However, it is increasingly difficult to escape the 'numbers paradigm' and, as decision makers look for justification for their choices, numbers are an obvious support to them. Therefore, the researcher must, however reluctantly, be able to assess the limitations of the conclusions associated with a data set, so that a reliable interpretation is put on the data. In this context, statistical methods can be useful.

As Maxwell & Frankenberger (1995) write: 'It is important to recognise that both quantitative and qualitative techniques are tools that play a useful and complementary role in improving our understanding' and that 'sampling considerations also apply to qualitative information'. What this real example shows is that the interpretation on a set of preferences depends upon the method of data collection and the method of analysis. Therefore, any interpretation of this type of data probably should not be made solely via statistical analysis. Thus we conclude that there is a need for checking results from other studies.

REFERENCE

Maxwell, S and Frankenberger, T R (1995) Household Food Security: Concepts, Indicators, Measurements. UNICEF/IFAD.