9 Conclusions

The objective of this thesis is to answer the following three questions:

1. Can the existing methods of paired comparisons be used to reasonably rank a set of objects in an experiment that is based on an incomplete and unbalanced (non) design? That is, can they be used when there is no control over the design of the experiment or when the data to be analysed are taken from an observational study?

2. Is it possible to develop a new method of paired comparisons (or further develop an existing method) that has a higher level of validity when applied to an incomplete and unbalanced (non) design?

3. Can the method of paired comparisons be used to help understand what variables (from a set of variables) are important in making a particular preference decision?

Presented in this thesis is a reasonably thorough background to the method of paired comparisons and an overview of the existing methods for collating paired comparison data into a final ranking. This by itself is not particularly new research as there are a number of books and articles (for example, David, 1988) that do the same.

However, where this research differs from existing texts and journal articles is in the following areas:

- It considers a set of pairwise preferences as a link-node network and introduces the concept of ‘filtering’ through the network of preferences to utilise the maximum amount of information available;

- It has, as a result of the extensive review of existing collation methods, progressed logically to the new collation method;

- It has extensively tested the performance of the new method and those of David (1987) and Allen (1992) by way of a simulation exercise and highlights the performance of the collation methods under different scenarios in terms of experiment size, experiment completeness and judgement consistency, as well as by considering the number of direct
comparisons and the strength of competition – something that has not been done before (or at least not reported in the literature);

- It has applied the new collation method, as well as the existing collation method of Allen (1992) to a set of real world data (the school case study in Chapter 7) and compared the outcomes of the two methods; and

- It has considered the usefulness of paired comparisons in understanding the way judges use information to construct their own criteria when instructed to make preference decisions at a broad level and has reported on a real world application of this approach performed by this author.

Considering a set of paired comparisons as a link-node network and introducing the concept of ‘filtering’ through the network of preferences is not entirely original but develops on previous work of other authors. David (1987) considers the notion of the wins of the wins and losses of the losses in paired comparisons, so in a sense this represents the first step in the filtering process. But David (1987) does not extend this analysis or present the analysis in a way that assists the reader in understanding the importance of the linked nature of the network of preferences and the amount of information indirectly available. This is probably the case because David’s methodology does not require this additional, indirect information when collating a final ranking.

Further, the typical size of the experiments reported in the texts and journals referenced in this thesis have been restricted to small numbers of objects, which may have further influenced the approach taken by David and others in terms of using all the available information. Certainly, if the number of objects in the experiment is less than 10 or so then the new method does not offer much improvement in performance over the existing methods. Part of the motivation of this current research is the applicability of paired comparisons to large data sets (with 100 or so objects). It is this motivation that influenced the consideration of using all available information (that is, indirect comparisons as well as direct comparisons) when collating the pairwise preferences into a final ranking. The new collation methodology has taken and explored existing methods and has extended these in a logical way to enable the collation of a ranking from a large experiment when there is little or no control over the experimental design.
The new method aims to provide a final rank for each object in the experiment that is not affected by the number and strength of the objects with which the object is compared directly. It does this by powering the $P$ matrix and its transpose, until convergence in the ranking is reached, unlike the methods of David (1987), Andrews and David (1990) and Allen (1992), which only power the matrices once. Powering the $P$ and $P'$ matrices until convergence is reached allows each object to be compared, albeit indirectly at times, with every other object within the experiment. Allowing each object to be compared with every other object removes the bias of an object being directly compared with only a subset of objects.

The new method takes a balanced approach in determining the final ranking of all objects by considering the difference in the wins and losses rather than the ratio of Ramanujacharyulu (1964), which is more concerned with identifying the ‘winner’ of the experiment.

The research then focuses on testing the new collation method and comparing its performance to the existing methods of David and Allen. This is done by way of a Monte Carlo simulation in which differences in judgement consistency as well as experiment completeness and size are considered. The simulation exercise makes the following findings:

- The new methodology generally provides an improved performance when there are more than 10 objects to be ranked;
- Compared with non-replication, replication of each pairwise judgement certainly improves the accuracy of the overall ranking, regardless of the level of judgement inconsistency. The exception is when there is perfect consistency – in which case the matrix of pairwise preference data is identical under both approaches;
- In the case of non-replication, the accuracy of the result greatly improves as judgement consistency improves. In other words, if it is not possible to replicate individual pairwise judgements then high judgement consistency is important for a reasonable result;
- In the case of replication, the accuracy of the returned ranking improves with judgement consistency only in the case of the new method and only when the overall judgement consistency increases above 85%; For the existing methods, the accuracy actually decreases marginally with the
improvement of judgement consistency, particularly if there is a low level of experiment completeness;

- In terms of experiment completeness, for non-replicated experiments, there is an increase in the accuracy of the returned ranking as the proportion of possible pairwise preferences completed increases, but not to the same extent as occurs with an increase in judgement consistency. That is, judgement consistency is actually more important than experiment completeness. This suggests that control over the design of the experiment (the extent of completeness and which pairwise preferences are completed) is less important than judgement consistency and replication – certainly a finding not found reported in the literature;

- The new method outperforms the existing methods when there is perfect consistency or when overall judgement consistency is at least 85%.

The collation method of Allen (1992) and the new method are applied to a real set of student assessment data to show the potential impact on the ranking of individual objects (in this case, students). It is shown that the change to the ranking of individual students is large enough to be of some real world significance.

In particular, the analysis of the school data set shows the new collation method, by way of filtering through the network of preferences, takes into account the underlying nature of the student curriculum pattern. In general terms, this is analogous to taking into account the indirect linkages of objects in an incomplete non-designed experiment. This example also shows that the method of Allen (1992) is not capable of doing this.

It appears that the method of paired comparisons can be used to assist in understanding the way judges make preference decisions. The midwifery workforce case study (Chapter 8) demonstrates that paired comparisons, together with secondary analysis of the results, can identify the variables that are important to a judge when making a preference decision and improve the understanding of the way in which the variables interact and are used.

In summary, the objectives of this thesis have been satisfied

1. Existing methods for collating paired comparisons can be used to reasonably rank a set of objects in an experiment that is based on an
incomplete and unbalanced (non) design providing there is a high level of judgement consistency (greater than 85%) and the level of completeness is greater than 70%. If overall judgement consistency is less than 85% then it is necessary to involve replication of judgements.

2. The existing collation methods of Kendall-Wei (Kendall, 1955), David (1987) and Allen (1992) are extended to produce the new collation method, which, in general, has in most cases a higher level of validity in the final ranking when applied to an incomplete and unbalanced (non) design.

3. It has been demonstrated by example that the method of paired comparisons can help understand what variables (from a set of variables) are important in making a particular preference decision.
References

1. Adler, H (1965), see Fechner, GT (1965).


37. Miller, GA (1956), The magical number seven, plus or minus two: some limits on our capacity for processing information, *The Psychological Review*, 63, 81 – 97.


deviations and equal correlations are assumed, *Psychometrika*, 16, 203 – 206.


## Attachments

Provided in the DVD are the following attachments.

<table>
<thead>
<tr>
<th>File name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thesis examples and tables.xls</td>
<td>A Microsoft Excel spreadsheet containing each of the worked examples in the thesis from Chapters 3 to 5. Each example is labelled in the spreadsheet with the same table number used in the thesis.</td>
</tr>
<tr>
<td>Simulation exercise – Stage One Analysis.xls</td>
<td>A Microsoft Excel spreadsheet containing the summary analyses of stage one of the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – Stage One Raw Data.xls</td>
<td>A Microsoft Access Database containing the raw results for stage one of the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – Stage Two Analysis.xls</td>
<td>A Microsoft Excel spreadsheet containing the summary analyses of stage two of the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – Stage Two Raw Data.xls</td>
<td>A Microsoft Access Database containing the raw results for stage two of the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – Stage Three Analysis.xls</td>
<td>A Microsoft Excel spreadsheet containing the summary analyses of stage three of the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – Stage Three Raw Data.xls</td>
<td>A Microsoft Access Database containing the raw results for stage three of the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – Stylised probability of consistent response.xls</td>
<td>A Microsoft Excel spreadsheet containing the probability function described in equation (6.1) in Chapter 6.</td>
</tr>
<tr>
<td>Simulation exercise – VBA Code.xls</td>
<td>A Microsoft Excel spreadsheet containing the VBA code written to undertake the simulation exercise presented in Chapter 6.</td>
</tr>
<tr>
<td>Case Study Two – Midwifery Analysis.xls</td>
<td>A Microsoft Excel spreadsheet containing the summary analyses of the midwifery case study presented in Chapter 8.</td>
</tr>
</tbody>
</table>