DOCUMENT RESUME

ED 412 228		TM 027 454
AUTHOR	Garner, Mary; Engelhard, George, Jr.	
TITLE	Rasch Measurement Theory, the Method of Paired Comparisons, and Graph Theory.	
PUB DATE	1997-03-00	
NOTE	72p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).	
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC03 Plus Postage.	
DESCRIPTORS	*Comparative Analysis; *Estimation (Mathematics); *Graphs; *Item Response Theory; Mathematical Models; Statistical Distributions	
IDENTIFIERS	*Paired Comparisons; *Rasch Model	

ABSTRACT

This paper considers the following questions: (1) what is the relationship between the method of paired comparisons and Rasch measurement theory? (2) what is the relationship between the method of paired comparisons and graph theory? and (3) what can graph theory contribute to the understanding of Rasch measurement theory? It is specifically shown how the method of paired comparisons can lead to the Rasch model, just as consideration of the Rasch model can lead to a pairwise algorithm for estimating the parameters of the Rasch model. Furthermore, both graph theory and previously unexplored aspects of the method of paired comparisons are used to increase understanding and utility of a pairwise algorithm for estimating parameters of the Rasch model as presented by B. Choppin (1985). Bringing together these three lines of inquiry enhances understanding of the Rasch model and provides more effective means of analysis. Appendixes present the conversion of the cumulative distribution function and Statistical Analysis System routines for the analysis. (Contains 8 figures, 7 tables, and 66 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 228

Running head: RMT, MPC, AND GT

Rasch Measurement Theory,

The Method of Paired Comparisons,

and Graph Theory

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Mary Garner
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Mary Garner
&
George Engelhard, Jr., Ph.D.

Emory University
Division of Educational Studies

A paper presented at the annual meeting of the American Educational Research Association, Chicago, March, 1997.

TM027454

BEST COPY AVAILABLE

Abstract

The purpose of this paper is to answer the following questions: (a) What is the relationship between the method of paired comparisons and Rasch measurement theory? (b) What is the relationship between the method of paired comparisons and graph theory? (c) What can graph theory contribute to our understanding of Rasch measurement theory? Specifically, it is shown how the method of paired comparisons can lead to the Rasch model, just as consideration of the Rasch model can lead to a pairwise algorithm for estimating the parameters of the Rasch model. Furthermore, both graph theory and previously unexplored aspects of the method of paired comparisons are used to increase understanding and utility of a pairwise algorithm for estimating parameters of the Rasch model as presented by Choppin (1985). Bringing together these three lines of inquiry enhances our understanding of the Rasch model, as well as provides more effective means of analyzing the Rasch model.

ERIC

Rasch Measurement Theory, The Method of Paired Comparisons, and Graph Theory

The traditional approach to measurement in education is to administer a test and count the number of items a person gets correct. This is surely the simplest approach, but one that is lacking in at least two regards: (a) scores on different tests cannot be compared meaningfully, and (b) the score that supposedly reflects a person's ability can change from one test to the other or even on different administrations of the same test. A norm referencing system could be used to compare scores on different tests, but in that situation, scores can be interpreted only within a certain population.

In the 1950's and 1960's, the Danish mathematician Georg Rasch proposed a new approach to educational measurement (Rasch, 1980). Rasch measurement theory provides a simple method for measuring person ability that does not depend on a particular set of items or a particular reference population, and that includes the possibility of variation in performance. One of the results of a Rasch analysis is the creation of a linear scale along which items are located according to difficulty. This scale is a ruler that can be used to measure person abilities.

Rasch (1961, 1966, 1977, 1980) repeatedly pointed out that a key characteristic of his measurement model is that the relative difficulty of any two items does not depend on the characteristics of a particular population. In other words, a comparison between any two items is independent of a particular population; likewise, a comparison between any two persons is independent of a particular set of items. Rasch placed great importance on such comparisons; indeed, he developed a theory regarding the generality and validity of scientific statements based on the idea that "comparisons form an essential part of our recognition of our surroundings . . . both in everyday life and in scientific studies." (Rasch, 1977, p. 68-69). He stated that any good measurement model, like any good scientific model, is based on objective comparisons.

> It is my opinion that only through systematic comparisons -- experimental or observational -- is it possible to formulate empirical laws of sufficient generability to be -- speaking frankly -- of real value, whether for furthering theoretical knowledge or for practical purposes.
>
> I see systematic comparisons as a central tool in our investigation of the outer world. (Rasch, 1977, p. 74).

The method of paired comparisons and graph theory are both based on comparisons between pairs of objects. Therefore, it seems very appropriate to bring these two areas together with Rasch measurement theory.

The method of paired comparisons is a widely used technique for describing preference behavior, based on the principles described by Thurstone (1927a,1927b,1927c) in his Law of Comparative Judgment. The method reduces preference behavior to its most basic and most easily grasped element: a person's choice between two objects. The result is a linear scale along which the objects are ordered. Building from Rasch's work (1960), Choppin (1985) described several methods of estimating the item difficulties of the Rasch model by comparing performance on pairs of items. Andrich (1978) linked Thurstone's Law of Comparative Judgment to Rasch Measurement Theory, and Engelhard (1984) described the parallels between Thurstone's and Rasch's approaches to measurement. Through the method of paired comparisons, Rasch measurement theory can also be linked with other scaling methods and with graph theory.

Graph theory is a branch of mathematics that provides a language, a set of procedures, and a way of visualizing a system that is built on the relationships between pairs of objects. It has been useful for this reason in assessing the outcome of paired comparisons experiments, and it holds promise as an analytic framework for examining aspects of Rasch measurement theory.

Purpose

The purposes of this paper are to extract from the literature the parallels between the method of paired comparisons and Rasch measurement theory, to describe their intersection in pairwise (PW) algorithms for estimating the parameters of the Rasch model, and to bring forward the graph theoretical concepts that can be used in analyzing links between items which would enhance the use of the PW algorithms as well as other methods of parameter estimation. Specifically, the questions addressed are the following: (a) What is the relationship between the method of paired comparisons and Rasch measurement theory? (b) What is the relationship between the method of paired comparisons and graph theory? (c) What can graph theory contribute to our understanding of Rasch measurement theory?

The paper is divided into five sections. In the first section, the connections between Rasch measurement theory and the method of paired comparisons are presented, along with the pairwise algorithms for estimating parameters of the Rasch model as described by Choppin (1985). The second section includes an introduction to the language of graph theory and a description of what role graph theory has played in applications of the method of paired comparisons. In the third section, initial steps are taken to explore the relationship between Rasch measurement theory and graph theory. In the fourth section, the techniques presented in the previous three sections are applied to a small data set. The last section consists of summary, conclusions, and suggestions for additional research.

Rasch Measurement Theory and The Method of Paired Comparisons

This section provides an introduction to the Rasch model and to the method of paired comparisons. These two areas are then brought together in pairwise algorithms for estimating parameters of the Rasch model as presented by Choppin (1985).

Rasch Measurement Theory

Rasch measurement theory is based on a mathematical model that describes the probability of a student achieving a certain score on a test as a function of the difference between the student's ability and the difficulty of the items on the test. Specifically, the probability that a person v will score correctly on particular item i ($a_{vi} = 1$) is expressed in terms of the person's ability b_v and the difficulty of the item d_i as follows:

$$\Pr(a_{vi} = 1) = \frac{e^{(b_v - d_i)}}{1 + e^{(b_v - d_i)}} \qquad (1)$$

This model is remarkable for at least two reasons. First of all, it is a stochastic rather than deterministic model; in other words, a student of a certain ability is not predicted to obtain a certain score but may obtain a range of scores with varying probabilities. A second characteristic of the model is what Rasch termed specific objectivity; that is, the mathematical structure of the model allows one to eliminate person abilities and be left with a model describing the relationship among item difficulties regardless of the persons involved; conversely, item difficulties can be eliminated to leave a model describing the relationship among person abilities regardless of the items used. Rasch (1966) described specific objectivity as follows:

> ... the comparison of any two subjects can be carried out in such a way that no other parameters are involved than those of the two subjects - neither the parameter of any other subject nor any of the stimulus parameters. Similarly, any two stimuli can be compared independently of all other parameters than those of the two stimuli, the parameters of all other stimuli as well as the parameters of the subjects having been replaced with observable numbers. It is suggested that comparisons carried out under such circumstances be designated as "specifically objective." (p. 104-105)

It is interesting that Rasch chose to define specific objectivity in terms of paired comparisons.

How does one obtain the item difficulties and person abilities? The most frequently used methods for estimating these parameters are maximum likelihood methods, particularly Conditional Maximum Likelihood (CML), Joint Maximum Likelihood (JML), and Marginal Maximum Likelihood (MML) estimation algorithms. These methods involve setting up equations that describe the likelihood of the observed scores in terms of the unknown item difficulties and/or person abilities. Values for the item difficulties and person abilities are then sought that maximize the likelihood of the observed scores.

Three important properties of estimators are consistency, sufficiency, and unbiasedness (Neter, Kutner, Nachtsheim, and Wasserman, 1996). An estimator is consistent if the estimate approaches the true value of the parameter as sample size increases. It is unbiased if its average value over repeated trials is the true value of the parameter. A statistic is sufficient if it contains all the information needed about the parameter being estimated; that is, parameter estimation cannot be improved by considering any aspect of the data other than our estimator. Under any of the maximum likelihood procedures, it can be shown that total score for any particular item is a sufficient statistic for item difficulty; and likewise, the raw score for any person is a sufficient statistic for person ability. The JML procedure is perhaps the simplest computationally, but has been shown to lack consistency (Wright & Masters, 1982; Baker, 1992). The MML technique requires assumptions regarding the distribution of abilities in the population (Baker, 1992). The CML technique is the only one of the maximum likelihood procedures that capitalizes on the specific objectivity of the model, and proceeds by first eliminating person abilities from the model, then estimating item difficulties, and finally estimating person abilities. However, although the CML algorithm is consistent and efficient, it involves computation of complicated functions that are sensitive to round-off errors. Baker (1992) points out that the computational difficulties have been lessened with the creation of more efficient algorithms, but the programs incorporating these algorithms are not readily available in the United States. Adams and Wilson (1996) point out that the complexity of CML estimation remains a disadvantage.

In all parameter estimation algorithms, a persistent problem is how to handle missing data. Baker (1992) suggests that missing values might be filled in at random. An algorithm specifically designed to deal with incomplete data, the EM algorithm, has been successfully paired with the MML algorithm for estimating the paramters of the Rasch model (Adams & Wilson, 1996; Baker, 1992). On the other hand, Linacre (1989) points out that JML estimation can also tolerate some missing values. However, it is unclear how the structure of the missing data and the extent of the missing data affect parameter estimates.

The Method of Paired Comparisons

The method of paired comparisons was first suggested by Fechner in 1860. In 1927, Thurstone (1927a, 1927b, 1927c, 1959) popularized the method by providing a rigorous formulation of the method through his Law of Comparative Judgment. Since that time, it has been applied in a variety of fields including dentistry, economics, epidemiology, optics, preference and choice behavior, sensory testing, ecology, acoustics, food science, psychology, medicine, and sociology (David, 1988). In all cases, the method of paired comparisons is used to construct a scale for the measurement of the relative magnitude of some perceived stimulus or non-physical trait, and assign scale values to the observed phenomena. In 1927, for example, Thurstone (1927b) constructed a scale for the measurement of the perceived seriousness of criminal offenses; the scale value for rape was the highest of all offenses at 3.275, while the value for vagrancy was 0.

In a paired comparisons experiment, a subject is presented with pairs of objects and is asked to indicate a preference for one of the objects according to some characteristic. A balanced paired comparison experiment is one in which every judge compares every possible pair of objects. In an unbalanced experiment, there are unequal numbers of comparisons between pairs. In the simplest case, ties are not allowed; however, the method of paired comparisons has been extended to include ties.

Based on the preferences between pairs of objects, a scale is constructed. Noether (1960) presented a simple and very general approach for describing how to obtain scale values from the paired comparison experiment. Noether (1960) considered the problem of estimating the true values V_i (i = 1, 2, ... t) of some set of objects, ordered along a linear scale, when judged pairwise on some characteristic. One restriction must be placed on the set to permit unique estimation of the V_i and the usual restriction is that the V_i sum to zero. The probability of preferring i to j, P_{ij}, is given by

$$P_{ij} = H(V_i - V_j) \quad (i,j = 1,2, \ldots t; j \neq i), \tag{2}$$

where H is the cumulative distribution function (cdf) of the differences, according to the model chosen. David (1988) derived the formula for the V_i regardless of the form of the cdf and assuming that the sum of the values is zero as

$$V_i = \frac{1}{t}\sum_j (V_i - V_j) \tag{3}$$

To see that this is true, note the following:

$$\frac{1}{t}\sum_j (V_i - V_j) = \frac{1}{t}(V_i - V_1 + V_i - V_2 + V_i - V_3 + \ldots + V_i - V_t)$$

$$= \frac{1}{t}(tV_i) - \frac{1}{t}\sum_j V_j$$

$$= V_i \text{ (because the sum of the } V_j \text{ is zero).}$$

To estimate the V_i one can estimate the P_{ij} with p_{ij}, the observed proportion of preferences for object i over object j, find approximations $d_{ij} = d_i - d_j$ to $V_i - V_j$ by

$$d_{ij} = H^{-1}(p_{ij}) \tag{4}$$

and then find the estimates d_i by using the relationship already described in (3) as

$$d_i = \frac{1}{t}\sum_j d_{ij} \tag{5}$$

David also showed that the d_i obtained in this way are the (unweighted) least squares estimates of the V_i regardless of the cdf used. More specifically, the solution minimizes the expression

$$\sum_j (d_{ij} - V_i + V_j)^2$$

What is H^{-1}? According to Case V of Thurstone's Law of Comparative Judgment, the cdf is the normal curve and the d_{ij} is the unit normal deviate. According to the Bradley-Terry-Luce model, the cdf takes the form (Bradley, 1953; David, 1988)

$$p_{ij} = H(d_{ij}) = \frac{1}{2}\left[1 + \tanh(\frac{1}{2}d_{ij})\right] \tag{6}$$

which is shown in Appendix A to lead to the following relationship between d_{ij} and the probabilities p_{ij} and p_{ji}.

$$d_{ij} = H^{-1}(p_{ij}) = \ln(p_{ij} / p_{ji}). \tag{7}$$

As in Case V of the Law of Comparative Judgment, there are two assumptions implicit in this approach: (a) each distribution of the d_i's has the same standard deviation, called discriminal dispersion by Thurstone, with mean V_i, and (b) the d_i's are equally correlated. Mosteller (1951b) explored the consequences when the assumption of equal discriminal dispersions is violated, and Davison, McGuire, Chen, and Anderson (1995) described a means of testing for equality of discriminal dispersions.

Noether's approach provides a least squares estimate of the scale values. By using equation (2), however, the scale values can also be obtained by maximizing the probability of the observed differences. David (1988) called the least squares approach used with the normal cdf, the Thurstone-Mosteller approach, and he called the maximum likelihood approach used with the logistic cdf, the BTL approach; however, it is clear from Noether's treatment that either estimation technique is appropriate for either cdf. Both approaches are also associated with goodness of fit measures.

The results of a paired comparisons experiment are summarized in a preference matrix such as the one shown below:

Objects:	A	B	C	D	
A	0	12	5	2	(8)
B	2	0	1	2	
C	1	8	0	6	
D	8	7	4	0	

The four rows and four columns correspond to the four objects in the experiment. The entry in the ith row and jth column corresponds to the number of times i was preferred to j. Each off-diagonal entry in the matrix is converted to a proportion p_{ij} as follows:

Objects:	A	B	C	D	
A	0	12/14	5/6	2/10	(9)
B	2/14	0	1/9	2/9	
C	1/6	8/9	0	6/10	
D	8/10	7/9	4/10	0	

Since this is not a balanced experiment, each off-diagonal entry is divided by the total number of comparisons for that pair of objects. Each entry p_{ij} is then divided by p_{ji}:

Objects:	A	B	C	D
A	0	12/2	5/1	2/8
B	2/12	0	1/8	2/7
C	1/5	8/1	0	6/4
D	8/2	7/2	4/6	0

(10)

Note that we could have gone directly from (8) to (10). Applying the BTL model, we can then apply equation (5) to the d_{ij} estimates described by equation (7). The scale values of the objects are then the row means of the natural logarithm of the matrix in (10).

Objects:	A	B	C	D
A	0	1.79	1.61	-1.39
B	-1.79	0	-2.08	-1.25
C	-1.61	2.08	0	.41
D	1.39	1.25	- .41	0

(11)

Scale Values:	A	B	C	D
	.50	-1.28	.22	.56

A different approach, but one that is still based on least squares estimation, is described by Bock and Jones (1968), Beaver (1977), and McGuire and Davison (1991). Their approach is based on the set of equations defined in (4). For example, a system of paired comparisons involving three objects would take on the following form:

$$\begin{bmatrix} H^{-1}(p_{12}) \\ H^{-1}(p_{13}) \\ H^{-1}(p_{23}) \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \quad (12)$$

After applying a constraint such as requiring the d_i to sum to zero, standard regression software could be applied to obtain regression coefficients d_i and associated statistics. McGuire and Davison (1991) used this approach to test group differences.

A preference matrix such as (8) is also called a tournament matrix. This tournament matrix might reflect the outcome of varying numbers of games (no ties allowed) between every pair of players. Ranking of players in such a tournament is traditionally accomplished by summing the rows of the original matrix (8), rather than the matrix (11) as in the least squares approach described above. Kendall (1955), among others (Cowden, 1974; Daniels, 1960; David, 1987), described a simple way of accommodating ties and compensating for missing data. Rather than summing the rows of the original tournament matrix, each player can also be given the score of every player that he has beaten. For example, the row sums of the above matrix are

$$
\begin{array}{l}
0 + 12 + 5 + 2 = 19 \\
2 + 0 + 1 + 2 = 5 \\
1 + 8 + 0 + 6 = 15 \\
8 + 7 + 4 + 0 = 19
\end{array}
$$

If we assign to the winner of each game all the wins of his opponent, the scores would change as follows:

$$
\begin{array}{l}
0 + 8(5) + 9(15) + 2(19) = 213 \\
2(19) + 0 + 2(15) + 3(19) = 125 \\
1(19) + 8(5) + 0 + 6(19) = 173 \\
8(19) + 7(5) + 4(15) + 0 = 247
\end{array}
$$

Thus, player 4 and player 1 are no longer tied. Kendall (1955) showed that such reallocation of wins was equivalent to summing the rows of the square of the original preference matrix. He also demonstrated that such a reallocation could take place a second or third time, corresponding to the third or fourth powers of the matrix. Kendall (1955) observed that if this process continues, as larger and larger powers of the matrix are taken, the vector of scores settles down to the eigenvector associated with the largest eigenvalue of the preference matrix.

Cowden (1974) and Andrews and David (1991) later recommended that Kendall's method should be modified to accommodate unbalanced paired comparison experiments, i.e. those experiments in which each pair played a different number of games, and experiments in which comparisons are missing, by using the proportions of games won rather than the count of games won. With this adjustment to Kendall's method, it is possible to see the relationship between Kendall's row-sum approach and Noether's scheme. The key is in the choice of the cdf in equation (4). The cdf in Kendall's method is simply the identity function, so that $d_{ij} = p_{ij}$ and therefore $d_i = \Sigma p_{ij}$.

Connections Between Rasch Measurement Theory and The Method of Paired Comparisons:

The method of paired comparisons and the Rasch measurement model have the same goal: to construct a scale for the measurement of some latent trait, a scale that is independent of the particular items used or the particular group being measured (Rasch, 1980). Rasch (1966, 1980) suggested a pairwise algorithm for obtaining parameters of the Rasch model. A pairwise procedure would take advantage of the specific objectivity that is unique to the Rasch model; indeed, as already noted, Rasch (1966) described specific objectivity in terms of paired comparisons.

Choppin (1985) developed Rasch's suggestion into two techniques for using paired comparisons to estimate item difficulties: a maximum likelihood approach and a least squares approach. In the maximum likelihood approach, the model parameters are chosen so that the probability of the observed test scores is maximized, whereas in the least squares approach, the model parameters are chosen such that the sum of the squared differences between the observed values and the estimated parameters is minimized. The maximum likelihood approach has received much attention in the Rasch literature (Andrich, 1988; Fischer & Tanzer, 1994; Linacre, 1989; van der Linden & Eggen, 1986; Zwinderman, 1995), perhaps because of the original emphasis on maximum likelihood estimation of parameters of the Bradley-Terry-Luce (BTL) model for paired comparisons, a model that is strongly related to the Rasch model (Andrich, 1978). On the other hand, the least squares approach is appealingly simple, has been explored extensively outside the Rasch literature, and can be linked to graph theoretical analysis of tournaments (Cowden, 1974, Kendall, 1955), and to Saaty and Vargas' (1991) method involving eigenvectors of preference matrices.

Least Squares Pairwise Algorithm for Estimating Parameters of the Rasch Model. Assuming that performance on each item is independent of the performance on any other items, a standard assumption in Rasch measurement, Choppin showed that the person ability parameter can be eliminated entirely from equation (1). This can be done by using equation (1) for item *i* and another for item *j*, to derive the conditional probability of a person giving a correct response to item *i*, given that the sum of the scores on item i and item j is 1 ($a_{vi} + a_{vj} = 1$). The result is that

$$\Pr(a_{vi} = 1 | a_{vi} + a_{vj} = 1) = \frac{e^{d_j}}{e^{d_i} + e^{d_j}} \tag{13}$$

This probability can be empirically estimated by observing the number of people who respond correctly to item i and incorrectly to item j, b_{ij}, among those that respond correctly either to item i or item j. Then we can write:

$$\Pr(a_{vi} = 1 | a_{vi} + a_{vj} = 1) = \frac{b_{ij}}{b_{ij} + b_{ji}} \tag{14}$$

Thus,

$$\frac{b_{ij}}{b_{ij} + b_{ji}} = \frac{e^{d_j}}{e^{d_i} + e^{d_j}} \tag{15}$$

This relationship can be rewritten as

$$\frac{b_{ij} / b_{ji}}{b_{ij} / b_{ji} + 1} = \frac{e^{(d_j - d_i)}}{e^{(d_j - d_i)} + 1} \tag{16}$$

and thus,

$$b_{ij}/b_{ji} \text{ estimates } e^{(d_j - d_i)}$$

or

$$b_{ij} / b_{ji} = e^{d_j} / e^{d_i}$$

which is equivalent to

$$\ln (b_{ij}/ b_{ji}) = d_j - d_i \tag{17}$$

So the difference in item difficulties can be estimated by $\ln (b_{ij}/ b_{ji})$, which involves observed values. If we add the constraint that the item difficulties must sum to zero, equation (17) defines a system of equations with a unique solution.

Thus, Rasch's goal of achieving measurement that does not depend on the abilities of the people measured is demonstrated mathematically. Furthermore, this method of pairwise comparisons for obtaining item difficulties arises naturally from a consideration of the properties of the model.

In order to solve the system of equations described in (17), Choppin (1985) recommended setting up a matrix B, with entries b_{ij} representing the number of people who got item *i* right and item *j* wrong. The matrix is shown in Figure 1 and is contrasted with the usual approach to measurement that begins with a persons by items matrix. The result is an asymmetric matrix of entries, with zeros on the diagonal. The matrix B is then converted to a matrix D with entries d_{ij} equal to b_{ji}/b_{ij}. D is then converted to ln D with entries $\ln(b_{ji}/b_{ij})$. These entries in ln D represent the log odds of getting item i correct given that either item i or item j is correct but not both. Choppin (1985) then showed that the item difficulties can be calculated from the matrix ln D using the following formula:

$$d_i = \frac{1}{t}\sum_j \ln(b_{ji}/b_{ij}). \qquad (18)$$

where t is the number of items. Equation (18) amounts to obtaining the means of the rows of the natural logarithm of the matrix D. Once the item difficulties have been calculated, the original model (1) can be used to set up another set of equations to solve for the ability parameters.

The approach described above is exactly the same as the approach described by Noether for obtaining scale values from paired comparisons experiments using the BTL model. Equation (17) is the same as equation (7) and equation (18) is the same as

ERIC

equation (5), except for a factor of -1, which can be attributed to the fact that the scales are reversed; that is, choosing an item more often means it is easier, and therefore lower, on the difficulty scale, whereas the usual case in a paired comparison experiment is that an item that gets chosen more often would have a higher value on the scale. Thus, Choppin's method is equivalent to a least squares estimate of item difficulties using the BTL model for an unbalanced paired comparisons experiment. Furthermore, the matrix B is a tournament matrix or a preference matrix from a paired comparisons experiment like the one shown in (8), matrix (10) is the D matrix described in Choppin's method, and matrix (11) is ln D.

The only difficulty in using this approach for estimating Rasch item difficulties arises when any of the B matrix entries are zero, which must be expected when the same person does not take two items or when both items are always right or both are always wrong. Noether suggested that 0 be replaced by 1/(2N) where N is the number of items. Choppin, on the other hand, showed algebraically that the entries of B^2 rather than B may be used in equation (18). This technique is equivalent to Kendall's (1955) approach of reallocating wins in a tournament. Choppin (1985) implied that this technique essentially replaces the results of the direct comparisons between i and j with the sum of the indirect comparisons of i and j through an intermediate k. If the items are adequately linked, all off-diagonal entries of the squared matrix will be non-zero. Rasch provided support for this approach in the following "rule of transitivity":

> The rule of transitivity seems to generalize one of the most fundamental properties of measurement. If, for instance, we wish to measure the distance between two points A and C on a straight line we may do it directly or we may interpose

> a third point B, measure the distance AB, and on top of that measure the distance BC to obtain the total AC. (Rasch, 1961, p. 332).

Extensions of Choppin's Least Squares Algorithm and Connection To the Analytic Hierarchy Method. Choppin's use of the square of the B matrix is equivalent to Kendall's (1955) technique of reallocating wins in a tournament. As Kendall pointed out, this reallocation could be repeated by using higher powers of the matrix. As higher powers of the matrix are used, the solution converges to the eigenvector associated with the largest eigenvalue; in fact, this approach is equivalent to using the power method for obtaining the dominant eigenvalue and associated eigenvector (Watkins, 1991). Cowden (1974) points out that this convergence will result if, in every possible partition of the players into two non-empty sets, some player in each set has won at least once from some player in the other set. In a later section, this requirement will be seen to be equivalent to a requirement for the convergence of the maximum likelihood algorithm for paired comparisons, and will also be shown to be equivalent to a simple graph theoretical concept.

There is thus a connection between the item difficulties of the Rasch model and the eigenvectors of the paired comparisons matrix B. This approach is further justified by consideration of Saaty and Vargas' (1991) analytic hierarchy process, which also makes use of eigenvectors. In Saaty and Vargas' analytic hierarchy process, subjects are asked to indicate not just a preference for two objects, but they are asked to estimate the strength of the preference in terms of pairwise ratios. The resulting comparisons matrix, called a reciprocal matrix, looks like the one shown below:

ERIC

$$\begin{array}{ccccc} w_1/w_1 & w_1/w_2 & w_1/w_3 & \cdots & w_1/w_N \\ w_2/w_1 & w_2/w_2 & w_2/w_3 & \cdots & w_2/w_N \\ w_3/w_1 & w_3/w_2 & w_3/w_3 & \cdots & w_3/w_N \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ w_N/w_1 & w_N/w_2 & w_N/w_3 & & w_N/w_N \end{array}$$

where w_i is the scale value of the ith object. The following equation must be true:

$$\begin{bmatrix} w_1/w_1 & w_1/w_2 & w_1/w_3 & \cdots & w_1/w_N \\ w_2/w_1 & w_2/w_2 & w_2/w_3 & \cdots & w_2/w_N \\ w_3/w_1 & w_3/w_2 & w_3/w_3 & \cdots & w_3/w_N \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ w_N/w_1 & w_N/w_2 & w_N/w_3 & & w_N/w_N \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_N \end{bmatrix} = N \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_N \end{bmatrix}$$

By definition of eigenvectors and eigenvalues, the solution vector of w_i's is the eigenvector associated with the eigenvalue N.

To connect the above system with the pairwise algorithm, recall that each entry in the D matrix, b_{ji}/b_{ij}, as shown in equation (13), estimates e^{d_j}/e^{d_i}. Thus, D is a reciprocal matrix as described by Saaty and Vargas and the item difficulties we seek are the natural logarithm of the eigenvector association with the largest eigenvalue of the D matrix as shown below.

$$\begin{bmatrix} e^{d_1}/e^{d_1} & e^{d_1}/e^{d_2} & e^{d_1}/e^{d_3} & \cdots & e^{d_1}/e^{d_N} \\ e^{d_2}/e^{d_1} & e^{d_2}/e^{d_2} & e^{d_2}/e^{d_3} & \cdots & e^{d_2}/e^{d_N} \\ e^{d_2}/e^{d_1} & e^{d_2}/e^{d_2} & e^{d_2}/e^{d_3} & \cdots & e^{d_2}/e^{d_N} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ e^{d_{1N}}/e^{d_1} & e^{d_N}/e^{d_2} & e^{d_N}/e^{d_3} & & e^{d_N}/e^{d_N} \end{bmatrix} \begin{bmatrix} e^{d_1} \\ e^{d_2} \\ e^{d_3} \\ \cdot \\ \cdot \\ \cdot \\ e^{d_N} \end{bmatrix} = N \begin{bmatrix} e^{d_1} \\ e^{d_2} \\ e^{d_3} \\ \cdot \\ \cdot \\ \cdot \\ e^{d_N} \end{bmatrix}$$

Saaty and Vargas also show that a necessary and sufficient condition for matrix W to be consistent is that the maximum eigenvalue λ_{max} be equal to N. As a measure of deviation from consistency, the authors use consistency index:

$$C.I. = (\lambda_{max} - N)/(N - 1) \tag{19}$$

This application of Saaty and Vargas' analytic hierarchy process to the scaling of choice preferences can only be accomplished through the BTL or Rasch models, because only those models transform a difference in scale values to a ratio.

Connection Between Choppin's Least Squares Algorithm and Multidimensional Scaling. It has already been observed by Chen and Davison (1996) that item difficulties may be obtained through nonmetric MDS; this method also provides an opportunity to verify the unidimensionality of the scale. Nonmetric MDS (Krusdal & Wish, 1976) as applied by Chen and Davison to a paired comparisons matrix seeks to minimize the squared deviations of the differences between estimates and the empirical observations. Both MDS and Noether's technique are based on a least squares approach. It is beyond the scope of this paper to make the relationship between the two approaches more explicit.

Maximum Likelihood Pairwise Algorithm for Estimating Parameters of the Rasch Model. Assuming that pairs of items are independent, the likelihood of the paired comparison matrix B can be expressed as

$$Pr[B|N] = \prod^{i,j,i\neq j}\left[\frac{(b_{ij} + b_{ji})!}{b_{ij}!b_{ji}!}\right]\left[\frac{e^{(b_{ij}d_j + b_{ji}d_i)}}{(e^{d_i} + e^{d_j})^{(b_{ij} + b_{ji})}}\right] \tag{20}$$

Zwinderman (1995) described the same function in terms of the traditional persons by items matrix. The derivative of the log of this likelihood function is

$$\frac{dL}{dd_i} = \sum^{j,i\neq j} b_{ji} - \sum^{j,i\neq j}\frac{e^{d_i}(b_{ij} + b_{ji})}{(e^{d_i} + e^{d_j})} \tag{21}$$

Setting this derivative to 0, and adding the constraint that the sum of the item difficulties must be zero, we have a set of N equations in N unknowns which Choppin recommended solving iteratively in two stages. He recommended that an initial approximation to the solution be obtained by using the iteration:

$${}^{(n+1)}d_i = \ln\left[\sum_{j,i\neq j} b_{ji}\right] - \ln\left[\sum_{j,i\neq j}\frac{(b_{ij} + b_{ji})}{(e^{{}^{n}d_i} + e^{{}^{n}d_j})}\right] \tag{22}$$

which is a fixed-point iteration method (Conte and de Boor, 1980) for solving the equations in (21). After setting the initial value of the item difficulties to 0, equation (22) is used three or four times to provide the initial item difficulties for the following iteration:

$${}^{(n+1)}d_i = {}^{n}d_i - \frac{\displaystyle\sum_{j,i\neq j} {}^{n}d_{ij} - \sum_{j,i\neq j}\frac{e^{d_i}(b_{ij} + b_{ji})}{(e^{d_i} + e^{d_j})}}{\displaystyle -\sum\frac{e^{(d_i + d_j)}(b_{ij} + b_{ji})}{(e^{d_i} + e^{d_j})^2}} \tag{23}$$

which constitutes the Newton-Raphson method (Conte and De Boor, 1980). The iterations defined by equation (22) are recommended because the Newton-Raphson technique might not converge if the initial estimates are not close enough to the solution.

Equations (22) and (23) differ slightly from the equations given in Choppin's article in that several typographical errors were corrected.

This approach has been used by Andrich (1978, 1988), Fischer and Tanzer (1994), van der Linden and Eggen (1986), Wright and Masters (1982) and Zwinderman (1995) to estimate the item parameters of the Rasch model. David (1988) pointed out that through this maximum likelihood approach, in the context of the BTL model for the method of paired comparisons, the row sums of the paired comparisons matrix, which can be considered item raw scores, are sufficient statistics for the item difficulties. In fact, Buhlmann and Huber (1963) showed that these scores are sufficient statistics only under the BTL model. Zwinderman (1995) showed that the method provides a consistent estimate of item difficulties that is comparable to conditional and marginal maximum likelihood methods. Fischer and Tanzer(1994; David, 1988) cited the Zermelo-Ford condition for uniqueness of the maximum likelihood solution: the solution is unique if and only if, for any partition of the items into two subsets, at least one item in the first set has been preferred to at least one item in the second set.

The Issue of Dependencies Between Pairs. The use of maximum likelihood estimation as described above hinges on the assumption that the pairs of items are independent. Zwinderman (1995) called the pairwise maximum likelihood algorithm a pseudo-likelihood method for this reason. Van der Linden and Eggen (1986) suggested the possibility of removing those dependencies.

Another perspective on this issue is provided by Wasserman and Faust (1994) in their comprehensive text on social network analysis. Starting with the description of social interactions as random directed graphs (defined in a later section), the authors present a family of models "which use the (natural) log of probabilities as their basic modeling unit. The models posit a structural form for the (natural) logarithm of the probability that actor i chooses actor j at one strength while actor j chooses actor i at a possibly different strength" (p. 606). One model that they present assigns scale values associated with friendliness to each actor. The model does accommodate ties unlike Choppin's model, but otherwise the BTL and Rasch approaches are hiding here in a different form. Maximum likelihood estimation is used, and the authors discuss the issue of dependencies between pairs. They review studies of a maximum pseudo-likelihood (MPL) estimation procedure that does not assume independence between pairs, and conclude that the effect of dependencies is inconsequential. MPL and ML parameter estimates were the same "even under conditions where the assumption of dyadic independence is known to be violated." They conclude that the simpler ML methods are justified.

In deriving the least squares algorithm, the requirement of independence between pairs does not arise. The only assumption is local independence, which is the standard assumption in Rasch measurement theory; in other words, performance on any one item is independent of the performance on any other item. Choppin (1985) suggested that a comparison of the item difficulties obtained using the B matrix with the item difficulties obtained using the square of the B matrix, assuming that both matrices have no zero entries, would show violations of the assumption of local independence. He suggested that the comparison could be used to test local independence even when other maximum likelihood parameter estimation is used.

The Method of Paired Comparisons and Graph Theory

Graph Theory

Graph theory is a branch of mathematics that traces its origins to a paper by Leonard Euler written in 1736 (Gould, 1988). In the paper, Euler analyzed the oldest known problem in graph theory, the bridges of Konigsberg problem. Konigsberg was situated on the river Pregel. There were two islands in the middle of the river, connected to the banks of Konigsberg and to each other by a system of bridges. The inhabitants of Konigsberg amused themselves by trying to determine a path that would start and end at the same point in the system and that would cross each bridge only once.

The tools of graph theory are graphs, which are composed of a finite nonempty set of elements called vertices, and a set of edges connecting those vertices. A graph with five vertices and seven edges is shown in Figure 2. The vertices may represent cities on an airline route, or phones in a telephone network, or tasks in a production line, or items in an item bank. Given a set of vertices and edges, graph theory provides answers to questions such as: Are every pair of vertices connected through some sequence of edges? What's the shortest route between two vertices? Could the graph be disconnected by eliminating just one edge? Graph theorists have built a set of computer algorithms that may be used to answer such questions. The user has to only supply substantive meaning to the vertices and the connections between them.

ERIC

> Graphs and multigraphs often appear under other names: sociograms (psychology), simplexes (topology), electrical networks, organizational charts, communication networks, family trees, etc. It is often surprising to learn that these diverse disciplines use the same theorems. The primary purpose of graph theory was to provide a mathematical tool that can be used in all these disciplines. (Berge, 1985, p. 3)

One way of representing a graph is through an adjacency matrix. An adjacency matrix is a square matrix A with n rows and n columns, where n corresponds to the number of vertices in the graph. For each entry $a_{ij} = 1$, there is an edge from vertex i to vertex j. If $a_{ij} = 0$ or if $i=j$, there is no edge. The adjacency matrix associated with the graph in Figure 2 is as follows:

Vertex:	1	2	3	4	5
1	0	1	1	1	1
2	1	0	0	1	0
3	1	0	0	0	1
4	1	1	0	0	1
5	1	0	1	1	0

Another way of representing a graph is through an incidence matrix, in which the rows represent vertices and the columns represent edges. The incidence matrix for the graph in Figure 2 would appear as follows.

	Edges:	1	2	3	4	5	6	7
Vertex:	1	1	0	1	1	0	0	1
	2	0	0	0	0	0	1	1
	3	0	0	0	1	1	0	0
	4	0	1	1	0	0	1	0
	5	1	1	0	0	1	0	0

There are several useful variations on the basic graph described above. A multigraph is a graph that may have more than one edge between vertices. A digraph is a graph whose edges have a specific direction; in this case the edges are called arcs (Berge, 1973). Such digraphs lead to nonsymmetric adjacency matrices and incidence matrices in which the entry is 1 if the arc initiates at the vertex and -1 if the arc terminates at the vertex. A digraph is shown in Figure 3. Its associated adjacency matrix is:

$$
\begin{array}{cccccccccc}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{array}
$$

The associated incidence matrix is:

$$
\begin{array}{cccccccccccccc}
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 0 & -1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
$$

A well established property of adjacency matrices is that the entries a_{ij}^{m} of the powers of the original matrix A^m provide the number of distinct walks of length m between the vertices i and j. A walk is an alternating sequence of vertices and edges that begins with the vertex i and ends with the vertex j and "in which each edge in the sequence joins the vertices that precedes it in the sequence to the vertex that follows it in the sequence" (Gould, 1988, p. 8). The walk may repeat edges and/or vertices. A path between vertices is a walk in which edges and vertices are visited only once. A cycle in a graph or circuit in a directed graph is a path that begins and ends with the same vertex.

To see that the powers of the adjacency matrix, A^m, provide the number of distinct walks of length m between the vertices i and j, consider the entries a_{ij}^2 of A^2 for example. Each entry a_{ij}^2 of A^2 is formed by summing the products $a_{ik} \times a_{kj}$ over all k; but this product is 0 if either term is 0 (that is, if there is no arc from vertex i to vertex k or no arc from k to j), and the product is 1 if both terms are 1 (that is, if there is an arc from i to k and one from k to j). Thus, a_{ij}^2 represents the number of times there is an arc from i to k and one from k to j, where k is any other vertex besides i or j. In other words, a_{ij}^2 is the number of walks of length 2 from vertex i to vertex j.

Each edge in a graph may have a number associated with it. This number may represent a weight, cost, or distance associated with crossing that edge, or some allowed flow through the edge. The entries in the adjacency matrix or the incidence matrix could then contain the numbers associated with each edge. A random digraph is a graph in which the edge weight represents the probability associated with the existence of that edge.

The following characteristics of graphs are important in answering questions regarding the connectivity of a set of vertices:

1. A graph is connected if there is a path between every pair of vertices. In a digraph, if you can find a path between any two vertices by following the direction of the

arcs, then the digraph is strongly connected. If you can find a path only by disregarding the direction of the arcs, then the graph is weakly connected. The graph in Figure 3 is weakly but not strongly connected. The graph in Figure 4 is disconnected; in this case, the associated adjacency matrix clearly has a block structure that shows the disconnection:

$$\begin{matrix} 0 & 1 & 1 & 0 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 1 & 0 & 1 & 0 & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 1 & 1 & 0 & 1 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 1 & 0 & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 1 & 0 & 1 & 0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 1 & 1 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & 0 & 1 & 0 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & 1 & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & 0 & 1 & 0 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 & 0 & 0 & 1 & 0 \end{matrix}$$

2. A component of a graph is a maximal connected subgraph; that is, the subgraph is maximal in the sense that it is as large as possible without being disconnected. A digraph may have strongly connected components or weakly connected components.

3. A connected graph is *k*-connected if a minimum of *k* vertices must be deleted to disconnect the graph. If a graph is k-connected, then any two vertices can be joined by k independent paths (Bollobas, 1979).

4. A connected graph is k-edge-connected if a minimum of k edges is required to disconnect the graph.

5. A cut vertex (or articulation vertex) is a vertex whose deletion disconnects the graph, while a bridge (or isthmus) is an edge whose deletion disconnects the graph.

6. The degree sequence of a graph is a listing of the degree of each of its vertices, where degree of a vertex refers to the number of edges that are incident to the vertex. In a

digraph, the <u>indegree</u> of a vertex is the number of arcs that terminate in a vertex, and the <u>outdegree</u> is the number of arcs that begin at the vertex. Note that the column sums of the adjacency matrix are the indegrees of the vertices and the row sums are the outdegrees.

7. A <u>clique</u> is a subset of the vertices in which an edge exists between every pair of vertices in the subset. A <u>maximum clique</u> is the largest possible clique.

8. A graph may be divided into <u>independent</u> sets of vertices; these are sets of vertices that are not directly connected to each other; that is, these are sets in which no edges exist betwen any pair of vertices.

It should be noted that most graph theorists (Gould, 1988) define a graph as described above, and then define a directed graph or digraph as a variation on the original definition of graph. However, Berge (1985) and Carre (1979) take a different approach. They both define graphs as a collection of vertices and arcs; their definition of graph is our definition of digraph. They describe undirected graphs as graphs whose edges have no specific direction; in other words, undirected graphs can be considered a variation of directed graphs in which each undirected edge actually consists of two oppositely directed arcs. According to Berge (1985):

> It would be convenient to say that there are two theories and two kinds of graphs: directed and undirected. This is not true. All graphs are directed, but sometimes the direction need not be specified. (p. 3)

Despite Berge's assertion, results are usually described in terms of either directed or undirected graphs. This issue in graph theory is brought up to emphasize that the connectivity issues that are usually defined in terms of undirected graphs in general, can be applied to digraphs with some modification.

Connections Between the Method of Paired Comparisons and Graph Theory

Kendall (Kendall, 1955; Kendall & Smith, 1940) used digraphs to visualize the results of a paired comparisons experiment. Only one arc existed between any two edges, and that arc pointed to the loser in the comparison. The paired comparisons matrix is the adjacency matrix for this graph, with edges weighted according to how many times the vertex at the initial end of the arc won against the vertex at the terminal end. The vertices can be ordered according to indegree or outdegreee, and this ordering goes from least able player (highest indegree, lowest outdegree, most losses) to most able player (lowest indegree, highest outdegree, fewest losses). The reallocation of wins that takes place by squaring the adjacency matrix can be visualized as utilizing all the walks of length two between every pair of vertices in the digraph.

In a balanced paired comparison experiment with a single judge, Kendall showed how graph theory could be used to analyze the consistency of the judgments. An inconsistency in the set of preferences would reveal itself as a circuit in the digraph, a situation in which i is preferred to j, j is preferred to k, and k is preferred to i. Kendall and Smith (1940) used the number of such triads as a measure of the inconsistency of a preference system. For example, Riechard (1990; 1991) used the number of circular triads to examine inconsistencies in paired comparisons experiments related to age, gender, and socioeconomic setting.

ERIC

Masuda (1988) presented a method of analyzing all cycles, not just triads, in a digraph that arises from a paired comparison experiment. In this method, the fundamental cycles of the graph are identified and relations among these cycles are made clear. Any cycles indicate inconsistences in the preference structure, and Masuda's technique would be useful for a system with a large number of inconsistencies.

Graph theoretic interpretations have also played a role in several other approaches to scaling in paired comparisons experiments. Shamus (1994) provided a graph theoretic interpretation of Chebotarev's (1994) generalized row sum method, a method in which direct comparisons between items carry the most weight, whereas indirect comparisons through other items decrease in weight with increasing distance from the item in question. Lattin (1990) used a network flow algorithm to obtain scale values from a paired comparisons experiment by minimizing absolute residuals. This method appeared to be more stable in the presence of aberrant proportions. The algorithm involved a digraph created from a linear programming problem that involved minimizing absolute residuals, and used a software program designed to analyze flow in networks.

Rasch Measurement Theory and Graph Theory

Connections Between Rasch Measurement Theory and Graph Theory

Table 1 presents generally how graph theoretical principles can be linked to the description and analysis of measurement principles. Vertices represent test items; edges represent comparisons between those items. Vertices could also represent raters, with edges between raters representing a basis for some comparison between raters. Vertices that are not connected directly by an edge may be connected through a path of intermediate vertices and edges; in measurement terms, items that are not directly comparable because no one has taken both items, may be compared through a series of other items. Modeling an assessment network in this manner makes explicit the ways in which two items or raters might be compared.

A connection between Rasch measurement theory and graph theory has been made on two occasions, through a discussion of the PW algorithm for estimating parameters of the Rasch model. Fischer and Tanzer (1994) and van der Linden and Eggen (1986) used digraphs to provide an interpretation of the Zermelo-Ford condition for uniqueness of the maximum likelihood solution. The digraph is defined by the original paired comparison matrix B, with a directed edge from item i to item j if there is a nonzero entry in the matrix for b_{ij}. The B matrix can thus be considered an adjacency matrix for a digraph. If the digraph is strongly connected, the maximum likelihood estimates are unique. The digraph must also be strongly connected for the matrix powers to converge to the eigenvector associated with the largest eigenvalue, as indicated by Cowden (1974). This is equivalent to requiring that the operation of raising the B matrix to successive powers eventually supplies a matrix with no zero entries.

If the digraph is not strongly connected, there is at least one item (or set of items) that has incident arcs in only one direction; in other words, there is at least one item that is always the correct one out of every pair or always the incorrect one out of every pair. Such a situation has always been recognized as unacceptable in Rasch measurement. Items on which all persons have succeeded or on which all persons have failed should be eliminated from consideration since their position on the item difficulty scale is indeterminable except to say that these items are beyond the item difficulty range measured by the other items. It would seem that any item that always matches performance on some other paired item adds nothing to the scale.

If a digraph associated with a paired comparison matrix is not strongly connected, the strongly connected components of the graph may be easily identified. There is, however, more that graph theory can provide, especially in the case of data sets with missing data. The connectivity of the digraph can be determined and used to indicate how well connected the system is. For example, if a digraph is 2-connected or biconnected, then it would imply that the graph remains strongly connected even when any one item is removed from the system; this is also equivalent to the condition that there are two unique paths comparing any pair of items. For a graph that is 1-connected, identification of the cut vertices, the vertices that could break the graph into a weakly connected system with strongly connected components, would allow examination of the quality of those items that are crucial for the connectivity of the whole system. A parallel analysis could be built from the determination of edge connectivity and identification of bridges. Furthermore, two items might be compared via different paths to assess the consistency of the system, or what Rasch might have described as adherence to the rule of transitivity.

Choppin (1968), Wright & Stone (1979), Engelhard and Osberg (1983), Masters (1984), Wright and Bell (1984), and Engelhard (1997) resorted quite naturally to graphical illustrations of the principles of item banks and test networks. An item bank is a large collection of items that have been calibrated according to difficulty and can be used to measure person ability as a ruler is used to measure height. Similarly, test networks are groups of tests whose relative difficulties are known. Both tests networks and item banks can be described as graphs with vertices being tests or items, and edges between vertices representing some result of a comparison between two tests or two items. Figure 5 shows some of the graphs that have appeared in the publications cited above. The links between items or tests could be identified and characterized precisely through graph theory.

Recognizing that the paired comparisons matrix B is an adjacency matrix for a digraph is not the only way to link Rasch measurement theory and graph theory. The system of equations described by Engelhard and Osberg (1983) for determining the linking constants for networks of tests is the same approach described in equation (12) and used by Bock and Jones (1968), Beaver (1977) and McGuire and Davison (1991) for obtaining least squares estimates of scale values from paired comparisons. The matrix shown in (12) is the transpose of a typical incidence matrix of a three-vertex, three-edge digraph. Such matrices are commonly used to represent electrical networks.

Data Analysis

Data from a study by Monsaas and Engelhard (1996) are used to illustrate the techniques described in this paper.

Instrument

An eleven-item subtest of the Home Observation for Measurement of Environment (HOME) instrument was used. The subtest is designed to describe the type of learning stimulation available in a child's home. Each item is scored dichotomously. Two-thirds of the items were scored by a teacher who was trained in the use of the test and who visited the child's family and observed the environment. About one-third of the items were scored on the basis of parental reports.

Participants

The data shown in Table 2 reflect the results of the HOME subtest for forty preschool children who had been defined as being at risk for school failure, as described in Monsaas and Engelhard (1996). There were 23 males and 17 females. Twenty-seven were African-American, and 13 were white.

Procedures

SAS routines shown in Appendix B were designed to estimate item difficulties according to the following methods: (a) Choppin's PW algorithm using maximum likelihood (PW Maximum Likelihood) described in equations (20) through (23); (b) Choppin's PW least squares algorithm using the B matrix of paired comparisons, with zero entries replaced with 1/(2N) (PW Least Squares - B) described by equation (18); (c) Choppin's least squares algorithm using the nth power of the B matrix (PW Least Squares - B^n). The FACETS computer program was used to obtain estimates of the item difficulties and standard errors using JML estimation.

Using the B matrix obtained from the HOME data as an adjacency matrix, the connectivity of the system of items was explored using Mathematica (Wolfram, 1993). Specifically, the following were obtained: (a) strongly connected components, (b) biconnected components, (c) cut vertices, (d) bridges, (e) vertex connectivity, and (f) edge connectivity. This analysis was also performed on an incomplete version of the HOME data set, in order to illustrate the results of the very simple PW algorithm on incomplete data and to illustrate the application of graph theory to analyzing the connectivity of the system of items. Table 3 shows the incomplete data set. Only the first 10 students were rated on items 1 through 5; the next 10 were rated on items 7 through 11; the next 10, on items 1, 2, 3, 10, and 11. The last ten students were rated on all items, but one item was deleted randomly from each student.

Results

Table 4 shows the B matrix for the HOME data. Table 5 shows the item difficulty estimates obtained through JML estimation, PW Maximum Likelihood, PW Least Squares using the matrix B, and PW Least Squares using the matrix B^2. The estimates using PW maximum likelihood and those using the least squares algorithm on B^2 are usually well within one standard error of the estimates obtained using JML. The estimates using the least squares algorithm on the B matrix, however, are often more than one standard error from the JML estimates. It appears that the method of handling the missing data in matrix B is inadequate. Table 6 shows the item difficulty estimates obtained through applying the least squares method to successive powers of the B matrix. As Kendall observed, the item difficulties appear to settle down with successive powers of the B matrix. The only dramatic difference in values occurred in using B^2 rather than B; perhaps the dramatic change can be attributed to the fact that B^2 had no zero entries. Consistent with Saaty and Vargas' analytic hierarchy method, the solution converges to the natural logarithm of the eigenvector associated with the maximum eigenvalue of the D matrix derived from B^2.

Figure 6 shows the digraph associated with the B matrix. This digraph was strongly connected and was characterized by 4-vertex-connectivity and 5-edge-connectivity, indicating that it would take the deletion of at least 4 items or 5 comparisons to disconnect the digraph. In other words, comparisons between items can be made through at least four independent paths through the digraph. For example, there is a direct comparison between items 1 and 9, but these items can also be compared through item 2, through item 11, or through items 3 and 8.

The paired comparisons matrix of the incomplete HOME data set corresponds to the digraph shown in Figure 8 and the item difficulties obtained using the least squares algorithm are shown in Table 7. The system was so poorly connected that the fourth power of the B matrix was the first matrix to contain no zero entries. It appears that not until this fourth power did the estimates of the item difficulties settle down. The system illustrated by the digraph in Figure 8 is strongly connected, but only 1-connected. There are two cut vertices, items 1 and 3; in other words, deletion of either of these items would change the strongly connected graph to a weakly connected graph and prevent proper parameter estimation. Items 1 and 3 would have to be examined to determine whether the connectivity of the system should rest with either of these items. If item 1 is deleted, for example, the system breaks into two strongly connected components: one including items 2, 3, 5, 7, 8, 9, and 10, and the other including items 4, 6, and 11. Figure 9 shows subgraphs of the graph shown in Figure 8, illustrating the opportunity to examine connections among subsets of the items. Clearly, the component involving items 4, 6, and 11 is minimally connected. Interestingly, item 6 is the only item that differs by more than one standard error from the JML estimates. All comparisons between items 4, 6, and 11 and other items, must be mediated by item 1 because of the connectivity.

Summary and Conclusions

This study was motivated by three questions. The first question was: What is the relationship between the method of paired comparisons and Rasch measurement theory? The method of paired comparisons and Rasch measurement theory have the same goal: to construct a linear scale along which a set of objects or items can be located. RMT has the additional goal of placing persons on that scale after the calibration of objects or items. Through Choppin's work it was shown that item difficulties in the Rasch model could be estimated by methods that are equivalent to least squares or maximum likelihood estimation of item difficulties using the BTL model for an unbalanced paired comparisons experiment. Applying the least squares algorithm to powers of the paired comparison matrix appeared to be more effective than arbitrarily filling in values for missing data in the comparison matrix as shown in Table 5. This power method was tied to Saaty and Vargas' analytic hierarchy method in which the scale values are components of the eigenvector associated with the maximum eigenvalue of the appropriate matrix. The item difficulties obtained are similar to JML estimates. The connectivity required in the system of paired comparisons is parallel to the situation in RMT in which items that are always correct or always incorrect cannot be properly placed on the scale with the other items in the system.

The second question was: What is the relationship between the method of paired comparisons and graph theory? It was shown that the paired comparisons matrix is an adjacency matrix for a digraph with edges weighted according to how many times the vertex at the initial end of the arc won against the vertex at the terminal end. Using graph theory, the connectivity of a system built from pairs of objects or items may be made explicit.

An alternate least squares method used by Bock and Jones (1968) and shown in equation (12) involves a system of equations that can be described in part by an incidence matrix for a digraph with unweighted edges. This latter method can be further explored and compared to results obtained through Choppin's PW algorithm. Standard regression software can be used, providing a great deal of valuable information regarding the fit of the data to the model.

The third question was: What can graph theory contribute to our understanding of Rasch measurement theory? Table 1 summarizes how some of the language and methods of graph theory might be used in measurement. It was shown that graph theory is essential in analyzing the connectivity of the system produced by the paired comparisons algorithm. Graph theory provides a well-established language and framework for discussing any systems based on pairwise comparisons. The influence of different degrees of connectivity must be explored. Network flow algorithms may provide new graph theoretical means of analyzing and estimating parameters of the Rasch model. Other applications of graph theory might be in determining goodness of fit measures for the item difficulties produced by the PW algorithm. The well-developed application of graph theory in social network analysis might suggest other ways to use graph theory in measurement.

The PW methods of estimating item difficulties are important in that they provide a way of utilizing the specific objectivity of the Rasch model without the computational disadvantage associated with CML estimation. By separating estimation of item difficulties from person abilities, it becomes possible to establish a measuring instrument that can be used consistently across different populations. Choppin (1968) was particularly interested in PW estimation for this reason. It is an ideal procedure for setting up item banks.

> The idea (of an item bank) is that a large collection of test items, the characteristics of which are known, be made available at some central place so that individuals who wish to construct achievement tests, but who lack the resources to carry out detailed standardization and validation procedures, can select items from the bank to form a test of known characteristics. (p. 870)

The usefulness of a PW algorithm was expressed as follows:

> The advantages of these procedures over classical item analysis techniques are several. First, because the model allows the separation of person and item parameters, we can make the estimation for any pair of items, without much regard for which set of individuals provides the data. People who score one on the item pair contribute to the estimation. People who score two or zero contribute nothing but do not spoil it. (p. 872)

Choppin also pointed out how simple the least squares algorithm is. In Appendix B is shown the few lines of code that are necessary to generate item difficulties, even in the presence of missing data. Choppin lamented that the technique was so easy that it allows one to produce item difficulties even when the data are not sufficiently interconnected. However, by extending his technique to powers of the comparison matrix, exploiting the link to Saaty and Vargas' technique using eigenvectors, and applying graph theoretical analysis of the paired comparison matrix, the simple technique might be successfully

ERIC

applied and thoroughly analyzed. Standard errors might be generated using a bootstrap technique or a technique recommended by Thurstone (1927b), and standard MDS techniques might be applied to test the unidimensionality of the data.

This paper sets the foundation necessary for a comprehensive treatment of a very simple procedure for calibrating achievement items according to the Rasch model. The statistical properties of the method must be explored further and the method must be extended to situations in which items are not dichotomously scored, but graded on a scale, and to situations involving raters. In addition, the applications of graph theory must be explored further and can certainly be extended to other graph theoretical constructs.

References

Adams, R. J., & Wilson, M. (1996). Formulating the Rasch model as a mixed coefficients multinomial logit. In G. Engelhard & M. Wilson (Eds.), Objective measurement III: Theory into practice (pp. 143-166). Norwood, NJ: Ablex.

Andrews, D. M., & David, H. A. (1990). Nonparametric analysis of unbalanced paired-comparison or ranked data. Journal of the American Statistical Association, 85(412), 1140-1146.

Andrich, D. (1978). Relationships between the Thurstone and Rasch approaches to item scaling. Applied Psychological Measurement, 2(3), 449-460.

Andrich, D. (1988). Rasch models for measurement. Newbury Park, CA: Sage Publications, Inc.

Baker, F. B. (1992). Item response theory. Parameter estimation techniques. New York: Marcel Dekker, Inc.

Berge, C. (1985). Graphs. New York: North Holland Publishing Company.

Biggs, N. (1993). Algebraic Graph Theory. Cambridge: Cambridge University Press.

Bock, R.D., & Jones, L.V. (1968). The measurement and prediction of judgment and choice. San Francisco, CA: Holden-Day.

Bollobas, B. (1979). Graph Theory. New York, NY: Springer-Verlag.

Bradley, R. A. (1953). Some statistical methods in taste testing and quality evaluation. Biometrics, 9, 22-38.

Bradley, R. A., & Terry, M.E. (1952). The rank analysis of incomplete block designs. The method of paired comparisons. Biometrika, 39, 324-345.

Buhlmann, H., & Huber, P. (1963). Pairwise comparison and ranking in tournaments. Annals of Mathematical Statistics, 34, 501-510.

Carre, B. (1979). Graphs and networks. Oxford: Clarendon Press.

Chebotarev, P. Y. (1994). Aggregation of preferences by the generalized row sum method. Mathematical Social Sciences, 27, 293-320.

Chen, T. H., & Davison, M. L. (1996). A multidimensional scaling, paired comparisons approach to assessing unidimensionality in the Rasch model. In G. Engelhard & M. Wilson (Eds.), Objective measurement III: Theory into practice (pp. 309-333). Norwood, NJ: Ablex.

Choppin, B. (1968). Item banking using sample-free calibration. Nature, 219(5156), 870-872.

Choppin, B. (1985). A fully conditional estimation procedure for Rasch Model parameters. Evaluation in Education, 9, 29-42.

Conte, S. D., & de Boor, C. (1980). Elementary numerical analysis: An algorithmic approach. New York: McGraw-Hill.

Cowden, D. J. (1974). A method of evaluating contestants. The American Statistician, 29(2), 82-84.

Critchlow, D. E., & Fligner, M. A. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. Psychometrika, 56(3), 517-533.

David, H. A. (1987). Ranking from unbalanced paired-comparison data. Biometrika, 74(2), 432-436.

David, H. A. (1988). The method of paired comparisons. New York, NY: Oxford University Press.

Davison, M. L., McGuire, D. P., Chen, T., & Anderson, R. O. (1995). Testing the equality of scale values and discriminal dispersions in paired comparisons. Applied Psychological Measurement, 19(3), 257-267.

Eggen, T. J. H. M., & van der Linden, W. J. (1987). The use of models for paired comparisons with ties. In W. E. Saris & I. N. Gallhofer (Eds.). Sociometric research. Volume I: Data collection and scaling (pp. 109-122). London: Macmillan.

Engelhard, G. (1984). Thorndike, Thurstone and Rasch: A comparison of their methods of scaling psychological tests. Applied Psychological Measurement, 8, 21-38.

Engelhard, G. & Osberg, D. W. (1983). Constructing a test network with a Rasch measurement model. Applied Psychological Measurement, 7(3), 283-294.

Fischer, G. H., & Tanzer, N. (1994). Some LBTL and LLTM relationships. In G. H. Fischer & D. Laming (Eds.), Contributions to mathematical psychology, psychometrics, and methodology (pp. 277-303). New York, NY: Springer Verlag.

Gould, R. (1988). Graph theory. Menlo Park, CA: The Benjamin Cummings Publishing Company.

Harary, F., Norman, R. Z., & Cartwright, D. (1965). Structural models: An introduction to the theory of directed graphs. New York: John Wiley & Sons, Inc.

Kendall, M. G., & Smith, B. (1940). On the method of paired comparisons. Biometrika, 31, 324-345.

Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. Biometrics, 11, 43-62.

Lattin, J. M. (1990). A minimum-cost network-flow solution to the Case V Thurstone scaling problem. Psychometrika, 55(2), 353-370.

Linacre, J. M. (1989). Many-facet Rasch measurement. Chicago: MESA Press.

Linacre, J. M. (1995). Paired comparisons with ties: Bradley-Terry and Rasch. Rasch Measurement Transactions, 9(2), 425.

Luce, R. D. (1961). A choice theory analysis of similarity judgments. Psychometrika, 26, 151-163.

Masters, G. N. (1984). Constructing an item bank using partial credit scoring. Journal of Educational Measurement, 21(1), 19-32.

Masuda, T. (1988). An analysis method for strongly connected structures obtained by binary paired comparisons. Memoirs of the Osaka Institute, 33(1), 1-14.

McGuire, D. P., & Davison, M. L. (1991). Testing group differences in paired comparisons data. Psychological Bulletin, 110(1), 171-182.

Monsaas, J., & Engelhard, G. (1996). Examining changes in the home environment with the Rasch Measurement Model. In G. Engelhard & M. Wilson (Eds.), Objective measurement III: Theory into practice (pp. 127-140). Norwood, NJ: Ablex.

Mosteller, F. (1951a). Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations. Psychometrika, 16(1), 3-9.

Mosteller, F. (1951b). Remarks on the method of paired comparisons: II. The effect of an aberrant standard deviation when equal standard deviations and equal correlations are assumed. Psychometrika, 16(2), 203-206.

Mosteller, F. (1951c). Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed. Psychometrika, 16(2), 207-218.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models. New York, NY: Richard D. Irwin, Inc.

Noether, G. E. (1960). Remarks about a paired comparison model. Psychometrika, 25(4), 357-367.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Psychology, 4, 321-333.

Rasch, G. (1966). An individualistic approach to item analysis. In P.F. Lazarfeld & N. W. Henry (Eds.), Readings in mathematical social science (pp. 89-107). Chicago, IL: Science Research Associates, Inc.

Rasch, G. (1977). On specific objectivity: An attempt at formalizing the request for generality and validity of scientific statements. Danish Yearbook of Philosophy, 14, 58-94.

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. Chicago: The University of Chicago Press. (Original work published 1960.)

Riechard, D. E. (1990). Paired-comparisons intransitivity related to age of subjects.

Educational and Psychological Measurement, 50(1), 105-110.

Riechard, D. E. (1991). Intransitivity of paired comparisons related to gender and community socioeconomic setting. Journal of Experimental Education, 59(2), 197-203.

Saaty, T. L., & Vargas, L. G. (1991). Prediction, projection, and forecasting. Boston, MA: Kluwer Academic Publishers.

Shamus, E. (1994). Graph-theoretic interpretation of the generalized row sum method. Mathematical Social Sciences, 27, 321-333.

Thurstone, L. L. (1927a). A law of comparative judgment. Psychological Review, 34, 278-286.

Thurstone, L. L. (1927b). The method of paired comparisons for social values. Journal of Abnormal and Social Psychology, 21, 384-400.

Thurstone, L. L. (1927c). Psychophysical analysis. American Journal of Psychology, 38, 368-389.

Thurstone, L. L. (1959). The measurement of values. Chicago, IL: University of Chicago Press.

Torgerson, W. S. (1958). Theory and methods of scaling. New York, NY: John Wiley & Sons, Inc.

van der Linden, W. J., & Eggen, J. H. M. (1986). An empirical Bayesian approach to item banking. Applied Psychological Measurement, 10(4), 345-354.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.

Watkins, D. S. (1991). Fundamentals of matrix computations. New York, NY: John

Wiley & Sons.

Wilson, M., & Hoskens, M. (1996). Multidimensional item responses: Multimethod-multitrait perspectives. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Wolfram, S. (1993). Mathematica (Version 2.2) (Computer programming language). Champaigne, IL: Wolfram Research Inc.

Wright, B. D., & Bell, S. (1984). Item banks: What, why, how. Journal of Educational Measurement, 21(4), 331-345.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: MESA press.

Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago: MESA press.

Zwinderman, A. H. (1995). Pairwise parameter estimation in Rasch models. Applied Psychological Measurement, 19(4), 369-375.

Appendix A

Shown here is the conversion of the cumulative distribution function

$$p_{ij} = H(d_{ij}) = \frac{1}{2}\left[1 + \tanh(\frac{1}{2}d_{ij})\right] \tag{6}$$

(where $d_{ij} = d_i - d_j$)

to the form

$$d_{ij} = H^{1}(p_{ij}) = \ln(p_{ij}/p_{ji}) \tag{7}$$

By definition of tanh:

$$p_{ij} = \frac{1}{2}\left[1 + \tanh(\frac{1}{2}(d_i - d_j)\right] = \frac{1}{2}\left[1 + \frac{\sinh\frac{1}{2}(d_i - d_j)}{\cosh\frac{1}{2}(d_i - d_j)}\right]$$

which is equal by definition of sinh and cosh to:

$$= \frac{1}{2}\left[1 + \frac{e^{\frac{1}{2}(d_i - d_j)} - e^{\frac{-1}{2}(d_i - d_j)}}{e^{\frac{1}{2}(d_i - d_j)} + e^{\frac{-1}{2}(d_i - d_j)}}\right]$$

$$= \frac{1}{2}\left[1 + \frac{1 - e^{-(d_i - d_j)}}{1 + e^{-(d_i - d_j)}}\right]$$

$$= \frac{1}{2}\left[\frac{2}{1 + e^{-(d_i - d_j)}}\right]$$

So, $p_{ij} = \dfrac{1}{1 + e^{-(d_i - d_j)}} = \dfrac{e^{(d_i - d_j)}}{1 + e^{(d_i - d_j)}}$

To obtain H^{1} note that $p_{ij} = \dfrac{e^{(d_i - d_j)}}{1 + e^{(d_i - d_j)}} \Rightarrow e^{(d_i - d_j)} = \dfrac{p_{ij}}{1 - p_{ij}}$

ERIC

Thus, $d_i - d_j = \ln\left(\frac{p_{ij}}{1 - p_{ij}}\right)$

which implies that $d_i - d_j = \ln\left(\frac{p_{ij}}{p_{ji}}\right)$ if we assume that $p_{ji} = 1 - p_{ij}$ which would be true as long as ties are not allowed.

Appendix B: SAS Routines

Routine #1: Input: X matrix shown in Figure 1.
Output: B matrix shown in Figure 1.

```
NITEM=NCOL(X);        * NITEM IS THE NUMBER OF ITEMS;
B=J(NITEM,NITEM,0.0); * INITIALIZE THE COMPARISON MATRIX;

* CREATE THE B MATRIX OF PAIRED COMPARISONS ;
DO K=1 TO N;
 DO I=1 TO NITEM;
  DO J=1 TO NITEM;
   IF X[K,I] ^= 9 & X[K,J] ^= 9 THEN DO;          *9 indicates missing value;
    IF X[K,I] > X[K,J] THEN B[I,J] = B[I,J] + 1.0 ;
END; END; END; END;
```

Routine #2: Input: B matrix or power of B matrix with no zero entries.
Output: Item difficulties according to least squares routine.

```
D = B` / B;              *See Figure 2 for description of D matrix.
LOGIT = LOG(D);
G = LOGIT[,:];
```


ERIC

Routine #3:	Input: B matrix. Output: Item difficulties according to maximum likelihood routine.

```
* Initial approximation for input to the Newton-Raphson procedure;
 G = J(NITEM,1,0.0);   * Item difficulties initialized to zero;
 DO S = 1 TO 4;
 DO I=1 TO NITEM;
  SUMONE = 0.0;
  SUMTWO = 0.0;
   DO J = 1 TO NITEM;
    IF J ^= I THEN DO;
     SUMONE = SUMONE + C1[J,I];
     TEMP = (C1[I,J]+C1[J,I])/(EXP(G[I,1])+EXP(G[J,1]));
     SUMTWO = SUMTWO + TEMP;
    END;
   END;
  G[I,1] = LOG(SUMONE) - LOG(SUMTWO);
 END;
 END;

 * The Newton-Raphson procedure ;
  DO I = 1 TO NITEM;
   NEWG=G[I,1];
   OLDG=50.0;
   COUNT = 0;
   DO WHILE(ABS(OLDG-NEWG) > .001 & count < 100);
     COUNT = COUNT + 1;
     SUMONE = 0.0;
     SUMTWO = 0.0;
     SUMTHREE = 0.0;
     DO J = 1 TO NITEM;
      IF I ^= J THEN DO;
       SUMONE = SUMONE + C1[J,I];
       TEMP = EXP(G[I,1])+EXP(G[J,1]);
       SUMTWO = SUMTWO + (C1[I,J]+C1[J,I])*EXP(G[I,1])/TEMP;
       SUMTHREE = SUMTHREE+(C1[I,J]+C1[J,I])*EXP(G[I,1]+G[J,1])/TEMP**2;
      END;
    END;

 OLDG = G[I,1];
 NEWG = OLDG +  (SUMONE - SUMTWO)/SUMTHREE;
```

```
G[I,1] = NEWG;
END;
END;

ONETEMP = 0.0;
DO I = 1 TO NITEM;
  ONETEMP = ONETEMP + G[I,1];
END;
MN = ONETEMP/NITEM;
DO I = 1 TO NITEM;
  G[I,1] = G[I,1] - MN;
END;
```

Table 1.

Measurement Interpretations of Graph Theoretical Constructs.

Graph Theory	Measurement
Basic Structure of Graphs/Basic Structure of Assessment Networks	
A graph	An assessment network
Vertices of a graph	Items (or raters or tests) in an assessment network
Edges of a graph	Direct comparisons between items; for example the same examinee takes two different items
Directed edges or arcs	A representation of the result of a direct comparison between two tasks; specifically, showing that task i is easier than task j
A directed graph or digraph	An assessment network in which the results of all direct comparisons are shown
Properties of Vertices & Edges in a Graph/Comparisons in an Assessment Network	
Degree of a vertex	The number of other items with which a certain item has been directly compared
Indegree of a vertex in a digraph	The number of items that have be found to be easier than a certain item after direct comparison
Outdegree of a vertex in a digraph	The number of items that have been found to be harder than a certain item after direct comparison
Length (or weight or value) of an edge	Number of times i was preferred to j or found easier than j
Associated Matrices	
Adjacency matrix	A paired comparison or preference matrix in which rows are items and columns are items
Incidence matrix	The transpose of the incidence matrix is equivalent to the matrix used in equation (12)
Powers of the adjacency matrix	The number of unique ways of indirectly comparing two items

Graph Theory	Measurement
Paths in a Graph/Direct and Indirect Comparisons in an Assessment Network	
Path between two vertices	An indirect comparison between two items; for example, item i is compared with k, k is compared with l, and l is compared with j, so there is a basis for comparison between i and j
Path in a digraph	Item i is easier than k, k is easier than l, and l is easier than j, so i must be easier than j
Connectivity	
Connected graph	An assessment system in which there is a basis for either direct or indirect comparison between any two items
Disconnected graph	An assessment system in which there are two or more sets of items that have no basis for comparison between them
A k-connected graph	There are k independent ways of comparing any two items in the assessment network (either directly or indirectly). In other words, a minimum of k items must be deleted to disconnect the system.
A cut vertex	An item that will disconnect the system if that item is eliminated
A k-edge connected graph	A minimum of k comparisons between pairs of items must be deleted to disconnect the assessment network
A bridge	A link between two items upon which the connectivity of the whole system depends
Connectivity of Subsets	
Component	A subset of items in which there is a basis for comparison between any two items
Biconnected component	A subset of items in which there are at least two means of comparison between any two items
A clique	A subset of items in which there is a direct comparison between any two items
A maximum clique	The largest subset of items for which there is a direct comparison between any two items
Independent sets of vertices	Subsets of items in which no two items have been compared directly

Table 2.

HOME Data

	Items										
Persons	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	0	0	1	0	1	1	1	1	0	1
3	1	1	0	1	1	1	1	1	0	0	1
4	1	0	1	1	1	1	1	1	0	1	1
5	1	0	1	1	1	0	1	0	0	1	1
6	1	1	0	1	1	1	1	1	1	1	1
7	1	1	0	1	0	0	0	0	0	0	0
8	1	1	1	1	1	1	1	0	1	1	1
9	1	1	1	1	1	1	1	1	1	1	1
10	1	0	0	1	1	1	0	1	1	0	1
11	1	1	0	1	0	1	1	1	0	0	1
12	1	0	0	1	1	1	1	1	0	0	1
13	1	0	1	1	0	1	0	1	0	0	1
14	1	0	1	1	0	1	0	1	0	0	1
15	1	1	1	1	1	1	1	1	0	0	1
16	1	0	0	1	0	1	0	0	1	0	1
17	1	0	0	1	1	1	1	0	0	0	1
18	1	1	0	1	0	1	1	0	1	0	1
19	0	0	0	0	0	1	0	0	0	0	1
20	1	0	0	1	0	1	0	1	0	0	1
21	1	0	0	1	1	0	0	0	0	0	1
22	0	0	1	1	0	1	1	1	1	0	1
23	1	0	1	1	1	0	1	1	0	1	1
24	0	0	1	1	1	0	0	1	1	0	1
25	1	1	0	1	1	1	1	1	0	1	1
26	0	0	0	0	0	0	1	0	1	0	1
27	1	1	1	1	1	1	1	1	0	1	1
28	1	1	1	1	1	1	1	1	1	1	1
29	1	1	1	1	1	1	1	1	0	0	1
30	1	1	0	1	0	1	1	0	1	0	1
31	1	1	0	1	1	1	1	1	1	0	1
32	1	1	0	1	1	1	1	0	0	0	1
33	0	0	0	1	0	0	0	0	0	0	1
34	1	1	0	1	1	1	1	1	1	1	1
35	1	0	0	1	0	1	0	1	0	1	1
36	1	0	0	0	0	1	0	0	0	0	0
37	1	1	0	1	1	1	0	0	0	0	1
38	1	0	0	1	1	1	1	0	0	1	1
39	1	1	1	1	1	1	1	1	1	0	1
40	1	1	0	1	0	1	1	1	1	0	1

Table 3.

HOME Data Set With Missing Values

Persons	Items										
	1	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1						
2	1	0	0	1	0						
3	1	1	0	1	1						
4	1	0	1	1	1						
5	1	0	1	1	1						
6	1	1	0	1	1						
7	1	1	0	1	0						
8	1	1	1	1	1						
9	1	1	1	1	1						
10	1	0	0	1	1						
11							1	1	0	0	1
12							1	1	0	0	1
13							0	1	0	0	1
14							0	1	0	0	1
15							1	1	0	0	1
16							0	0	1	0	1
17							1	0	0	0	1
18							1	0	1	0	1
19							0	0	0	0	1
20							0	1	0	0	1
21	1	0	0							0	1
22	0	0	1							0	1
23	1	0	1							1	1
24	0	0	1							0	1
25	1	1	0							1	1
26	0	0	0							0	1
27	1	1	1							1	1
28	1	1	1							1	1
29	1	1	1							0	1
30	1	1	0							0	1
31	1	1	0	1	1	1	1	1		0	1
32	1	1	0	1	1	1	1	0		0	1
33	0		0	1	0	0	0	0	0	0	1
34	1	1	0	1		1	1	1	1	1	1
35	1	0	0		0	1	0	1	0	1	1
36	1	0	0	0	0	1		0	0	0	0
37	1	1		1	1	1	0	0	0	0	1
38	1	0	0	1	1	1	1	0	0	1	
39	1		1	1	1	1	1	1	1	0	1
40	1	1	0	1	0	1	1		1	0	1

Table 4.

Comparison Matrix B for HOME Data.

	Items										
Items	1	2	3	4	5	6	7	8	9	10	11
1	0	15	22	1	12	4	10	12	21	22	2
2	0	0	12	0	5	1	2	6	9	12	1
3	2	7	0	0	3	3	3	2	8	7	0
4	3	17	22	0	13	6	11	12	21	24	1
5	1	9	12	0	0	4	4	7	14	12	0
6	2	14	21	2	13	0	9	10	18	22	1
7	2	9	15	1	7	3	0	8	13	15	0
8	2	11	12	0	8	2	6	0	13	15	0
9	3	6	10	1	7	2	3	5	0	11	0
10	0	5	5	0	1	2	1	3	7	0	0
11	5	19	23	2	14	6	11	13	21	25	0

Table 5.

Item Difficulty Estimates Based on Choppin's Pairwise Algorithm

Compared to Joint Maximum Likelihood Estimates for HOME Data

		Pairwise Estimates		
Item Number	JML	Maximum Likelihood	Least Squares-B	Least Squares - B^2
10	2.34(.43)	2.18	3.05	2.20
3	1.99(.41)	1.75	2.23	1.70
9	1.66(.40)	1.36	1.59	1.47
2	1.35(.39)	1.25	1.89	1.49
5	.58(.40)	.62	1.20	.60
8	.42(.40)	.45	1.13	.52
7	.09(.42)	.06	.34	.08
6	-1.15(.51)	-.78	-.86	-.65
1	-1.73(.57)	-1.37	-2.13	-1.25
4	-2.51(.68)	-2.49	-3.91	-2.74
11	-3.05(.80)	-3.02	-4.51	-3.41
Mean	0	0	0	0
SD	1.86	1.72	2.53	1.84

Table 6.

Item Difficulty Estimates for the HOME Data Based on Choppin's

Pairwise Least Squares Algorithm Using Powers of the B Matrix

Item No.	B^2	B^3	B^4	B^5	B^{10}
10	2.20	2.25	2.23	2.24	2.23
3	1.70	1.76	1.73	1.74	1.74
9	1.47	1.42	1.43	1.43	1.43
2	1.49	1.28	1.30	1.29	1.30
5	.60	.63	.62	.62	.62
8	.52	.53	.53	.53	.53
7	.08	.08	.08	.08	.08
6	-.65	-.69	-.68	-.68	-.68
1	-1.25	-1.29	-1.28	-1.28	-1.28
4	-2.74	-2.74	-2.73	-2.73	-2.73
11	-3.41	-3.23	-3.23	-3.23	-3.23
Mean	0	0	0	0	0
SD	1.84	1.80	1.80	1.80	1.80

Note: The item difficulties constitute the natural logarithm of the eigenvector associated with the maximum eigenvalue of the D matrix for the HOME data.

ERIC

Table 7.

Item Difficulty Estimates for the HOME Data with Missing Values,

Based on Choppin's PW Least Squares Algorithm Using Powers of the B Matrix

Item	JML	B	B^2	B^3	B^4	B^5	B^6	B^{10}	B^{11}
10	3.33(.57)	3.04	3.65	3.22	2.89	2.90	2.90	2.90	2.90
3	2.89(.53)	4.10	3.01	3.69	3.21	2.64	2.65	2.64	2.64
9	2.23(.62)	2.36	3.01	2.68	2.31	2.29	2.29	2.28	2.28
2	1.68(.51)	1.69	2.86	2.28	2.01	1.99	1.98	1.98	1.98
5	.74(.64)	1.15	1.75	1.38	.98	1.00	.99	.99	.99
8	.82(.56)	1.51	2.12	1.78	1.31	1.27	1.26	1.26	1.26
7	.74(.54)	1.08	1.81	1.33	.93	.90	.89	.89	.89
6	-3.32(1.34)	-3.64	-5.67	-7.85	-5.78	-5.25	-5.17	-5.18	-5.18
1	-1.47(.77)	-3.31	-.59	-.34	-.71	-.71	-.71	-.71	-.71
4	-3.44(1.28)	-3.73	-5.65	-3.81	-3.28	-3.19	-3.21	-3.20	-3.20
11	-4.20(1.12)	-4.25	-6.29	-4.36	-3.89	-3.85	-3.86	-3.86	-3.86
M	0	0	0	0	0	0	0	0	0
SD	2.68	3.09	3.93	3.72	3.02	2.85	2.84	2.83	2.83

Note: The items are ordered according to the difficulty of the items determined using the complete HOME data set.

Matrix X -- Raw Data

$$
\begin{array}{ll|llll|}
 & \text{Items} & 1 & 2 & 3 \; \ldots & N \\
\hline
\text{Persons} & 1 & x_{11} & x_{12} & x_{13} & \\
 & 2 & x_{21} & & & \\
\hline
\end{array}
$$

Data in the form $x_{ij} = 0$ if person i got item j incorrect, and $x_{ij} = 1$ if person i got item j correct.

Joint Maximum Likelihood

$$
\begin{array}{ll|llll|}
 & \text{Items} & 1 & 2 & 3 \; \ldots & N \\
\hline
\text{Raw Score} & 0 & 0 & 0 & 0 & \\
 & 1 & a_{11} & a_{12} & a_{13} & \\
 & 2 & a_{21} & & & \\
\hline
\end{array}
$$

Data in the form a_{ij} = number of people in score group i who responded correctly to item j

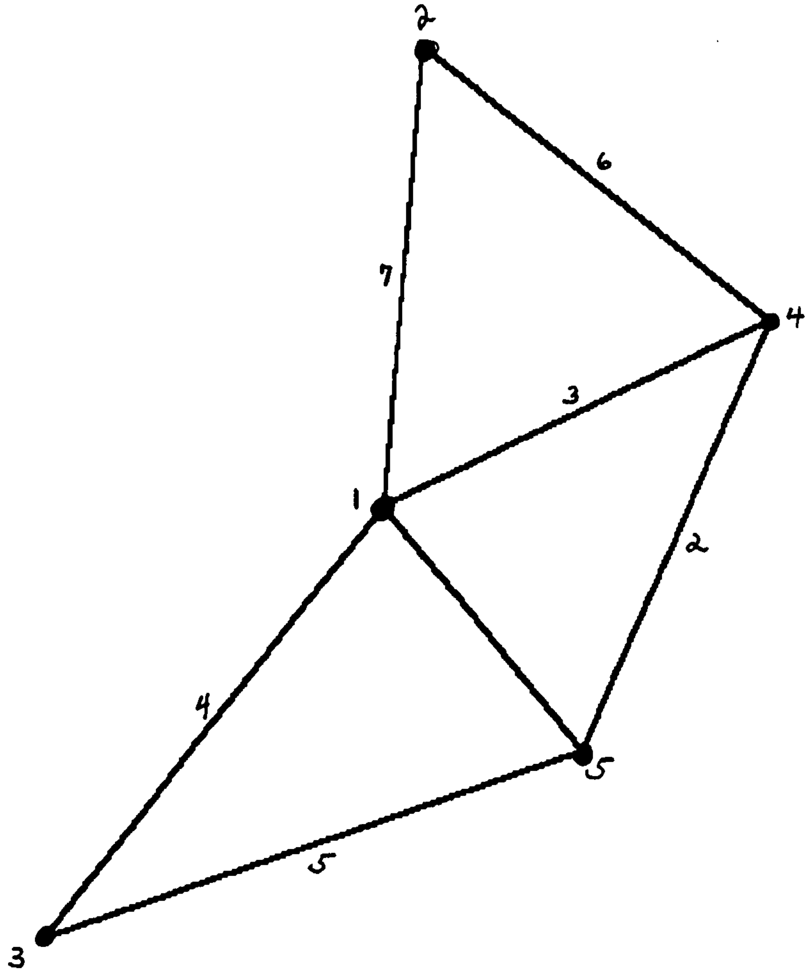
Pairwise Algorithms

B Matrix of Paired Comparisons

$$
\begin{array}{ll|lllll|}
 & \text{Items} & 1 & 2 & 3 \; \ldots & & N \\
\hline
\text{Items} & 1 & 0 & b_{12} & b_{13} & \ldots & \\
 & 2 & b_{21} & 0 & & & \\
 & N & & & & & \\
\hline
\end{array}
$$

Data in the form b_{ij} = number of people who responded correctly to item i and incorrectly to item j

Figure 1. Data summary in matrix form required for JML estimation contrasted with data summary required for the pairwise algorithm. (Adapted from Choppin, 1985).



<u>Figure 2.</u> A graph with 5 vertices and 7 edges.

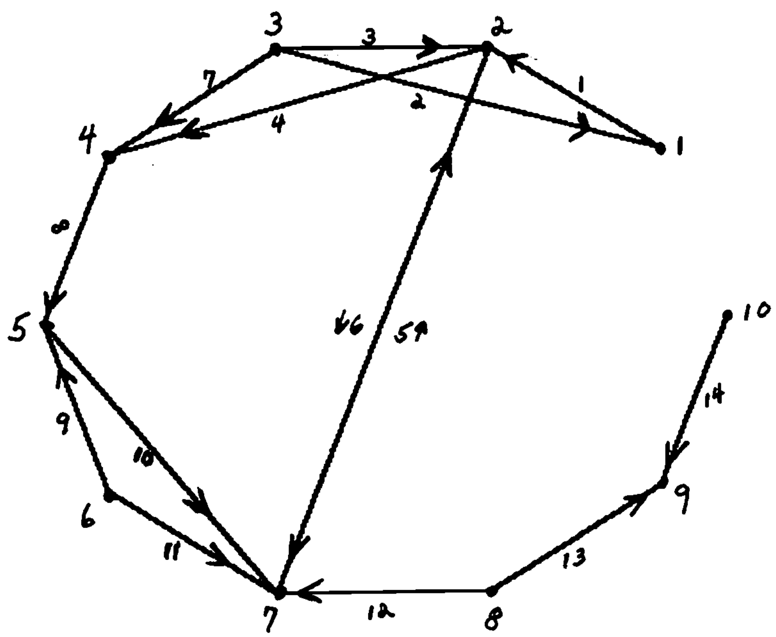


Figure 3. A digraph.

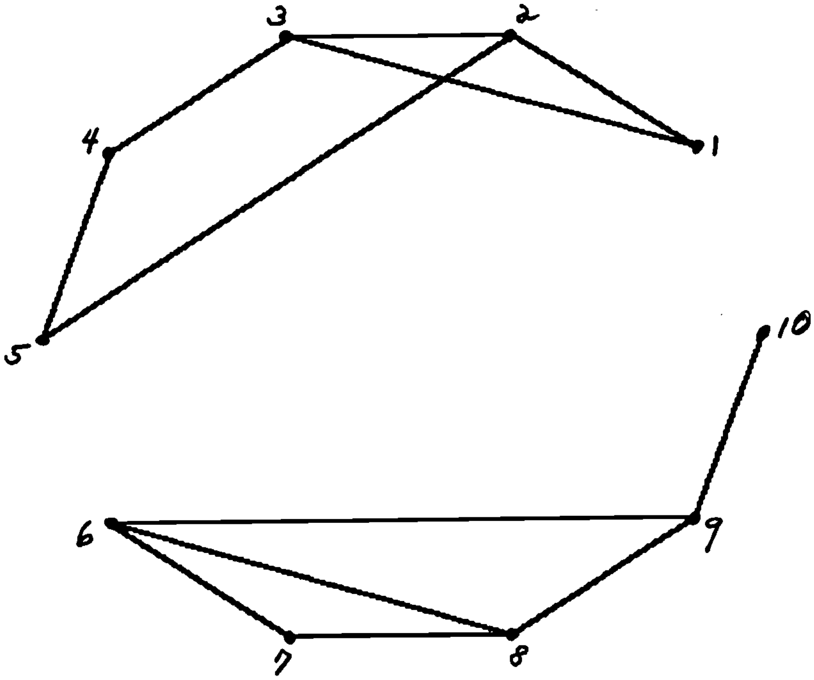
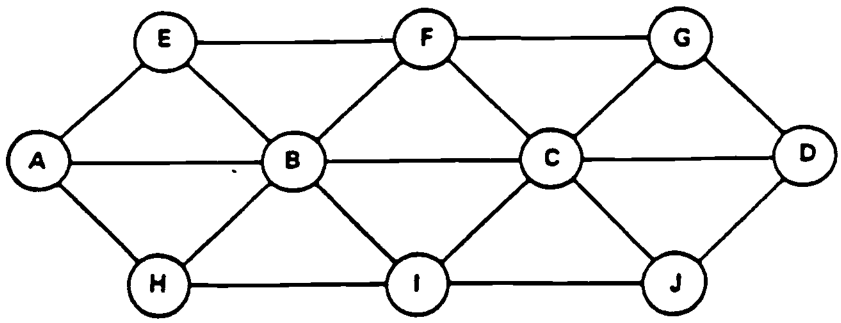
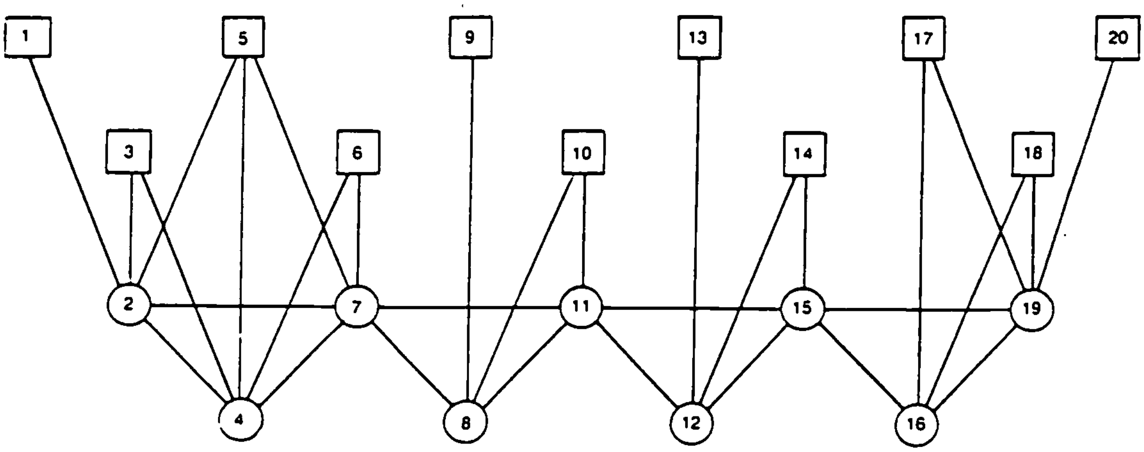


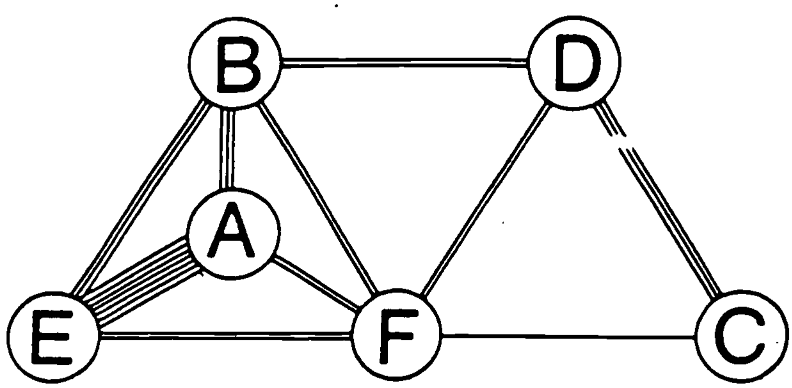
Figure 4. A disconnected graph with two components.



Graph from Wright & Stone (1979, p.101)



Graph from Engelhard (1983, p.287)



Multigraph from Masters (1984, p. 20)

Figure 5. Graphs from published articles on item banks and test networks.

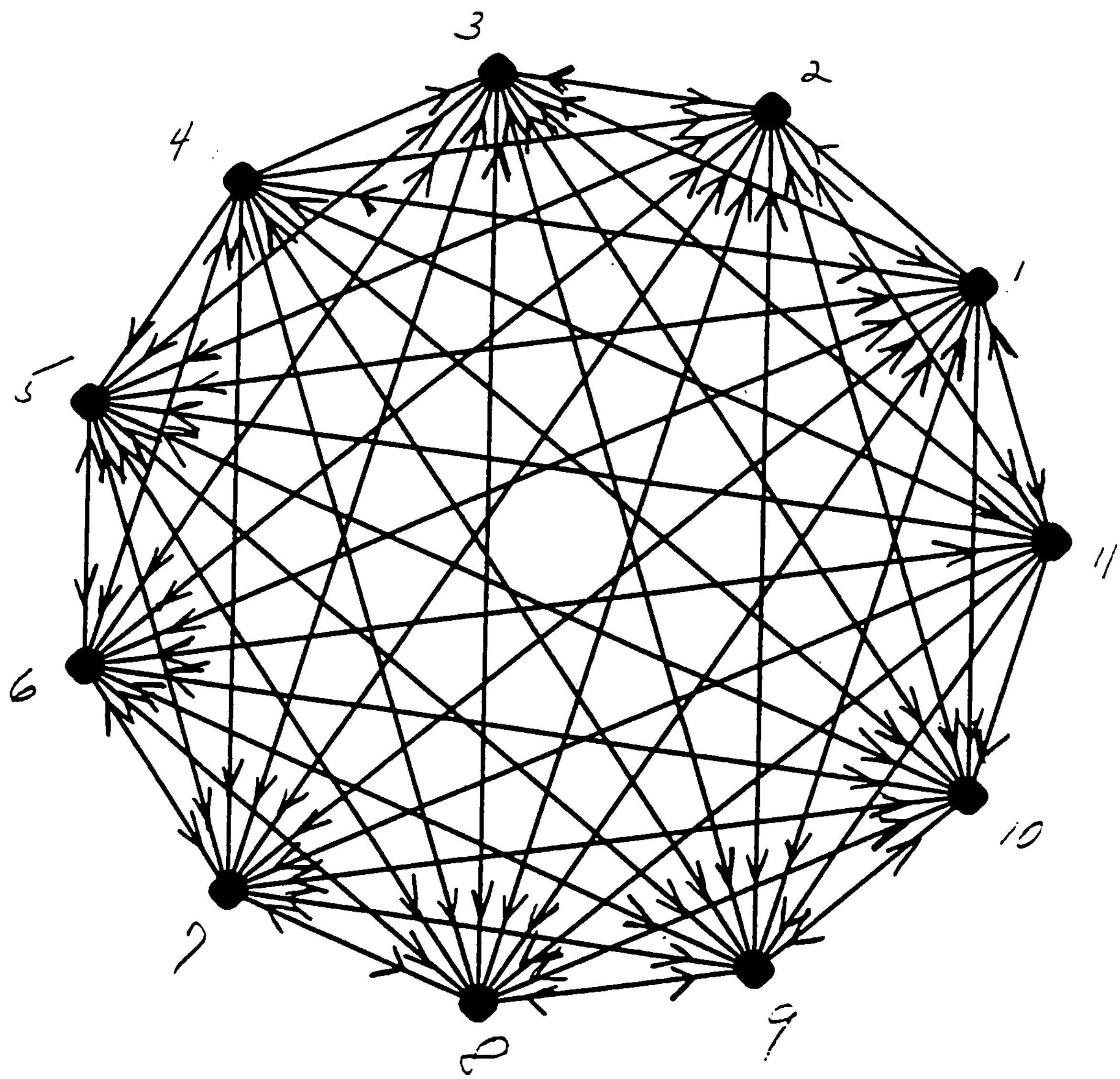


Figure 6. The digraph associated with the adjacency matrix shown in Table 4 and the HOME data shown in Table 2.

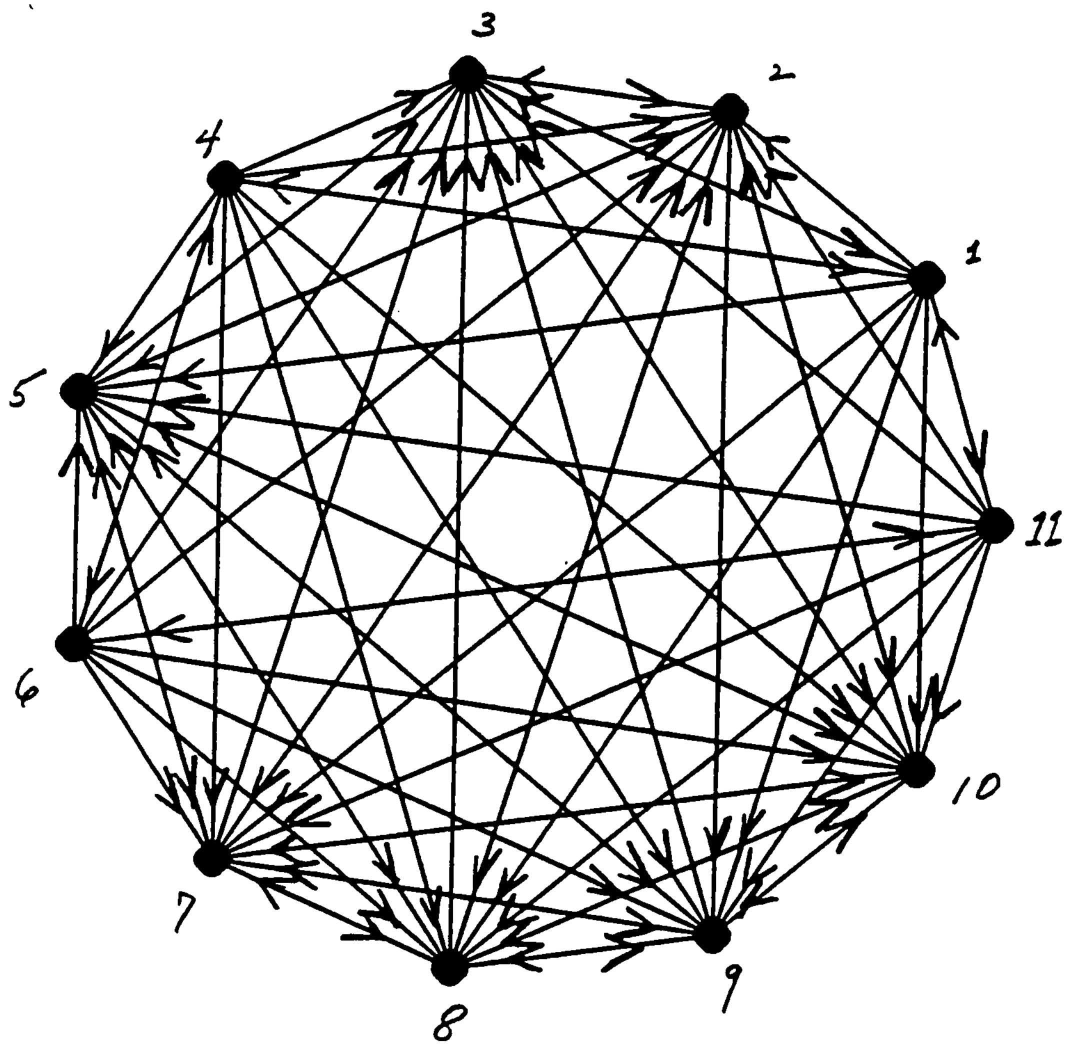
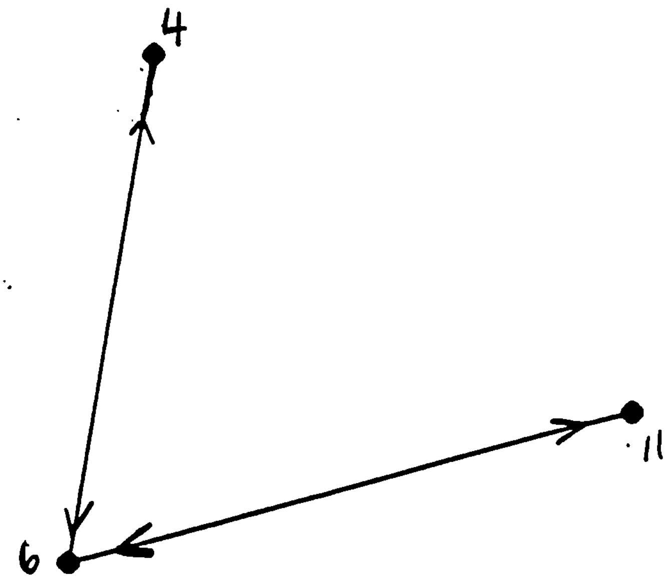


Figure 7. The digraph associated with the paired comparison matrix for the data shown in Table 3.



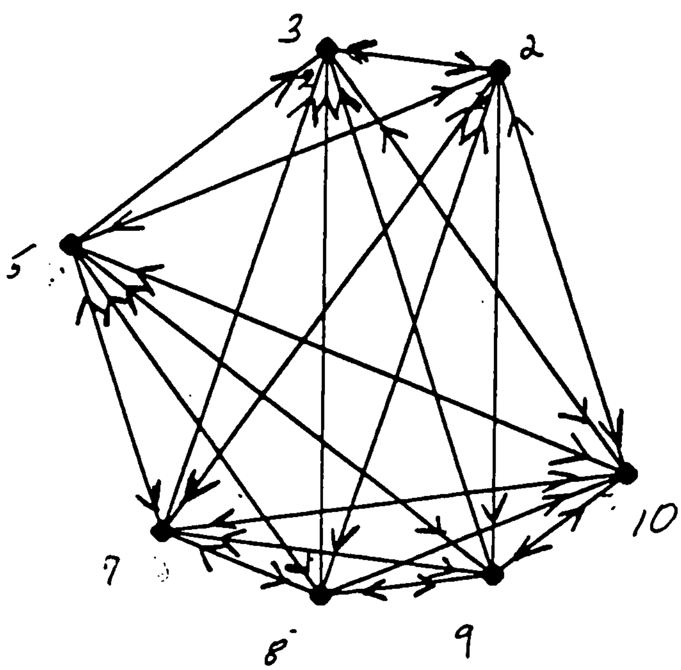


Figure 8. The strongly connected subgraphs of the digraph shown in Figure 7 with item #1 deleted.

AREA 1997

TM027454

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

ERIC®

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Rasch Measurement Theory, The Method of Paired Comparisons, and Graph Theory	
Author(s): MARY Garner & George Engelhard	
Corporate Source:	Publication Date: March, 1997

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

☑ **Sample sticker to be affixed to document**

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document ☐

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

or here
Permitting reproduction in other than paper copy.

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Mary Garner	Position: Graduate Student
Printed Name: MARY GARNER	Organization: Emory University
Address: 3231 ROCKINGHAM DR Atlanta GA 30327	Telephone Number: (404) 327 6468
	Date: 5/24/97

OVER

CUA

THE CATHOLIC UNIVERSITY OF AMERICA
Department of Education, O'Boyle Hall
Washington, DC 20064
202 319-5120

February 21, 1997

Dear AERA Presenter,

Congratulations on being a presenter at AERA[1]. The ERIC Clearinghouse on Assessment and Evaluation invites you to contribute to the ERIC database by providing us with a printed copy of your presentation.

Abstracts of papers accepted by ERIC appear in *Resources in Education (RIE)* and are announced to over 5,000 organizations. The inclusion of your work makes it readily available to other researchers, provides a permanent archive, and enhances the quality of *RIE*. Abstracts of your contribution will be accessible through the printed and electronic versions of *RIE*. The paper will be available through the microfiche collections that are housed at libraries around the world and through the ERIC Document Reproduction Service.

We are gathering all the papers from the AERA Conference. We will route your paper to the appropriate clearinghouse. You will be notified if your paper meets ERIC's criteria for inclusion in *RIE*: contribution to education, timeliness, relevance, methodology, effectiveness of presentation, and reproduction quality. You can track our processing of your paper at http://ericae2.educ.cua.edu.

Please sign the Reproduction Release Form on the back of this letter and include it with **two** copies of your paper. The Release Form gives ERIC permission to make and distribute copies of your paper. It does not preclude you from publishing your work. You can drop off the copies of your paper and Reproduction Release Form at the **ERIC booth (523)** or mail to our attention at the address below. Please feel free to copy the form for future or additional submissions.

Mail to:

AERA 1997/ERIC Acquisitions
The Catholic University of America
O'Boyle Hall, Room 210
Washington, DC 20064

This year ERIC/AE is making a **Searchable Conference Program** available on the AERA web page (http://aera.net). Check it out!

Sincerely,

Lawrence M. Rudner, Ph.D.
Director, ERIC/AE

[1]If you are an AERA chair or discussant, please save this form for future use.

ERIC Full Text Provided by ERIC