DOCUMENT RESUME

ED 307 338 TM 013 526

AUTHOR Linacre, John M.
TITLE Fundamental Measurement of Rank-Ordered Objects.
PUB DATE Mar 89
NOTE 13p.; Paper presented at the International Objective Measurement Workshop (5th, Berkeley, CA, March 26, 1989).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Analysis; Equations (Mathematics); Error of Measurement; *Goodness of Fit; *Latent Trait Theory; *Mathematical Models; Research Methodology
IDENTIFIERS Parametric Analysis; *Rank Order; *Rasch Model

ABSTRACT

A Rasch measurement model can be constructed to meet the requirements of rank ordered data. If multiple rankings of the same objects are available, then the parameters of the objects can be estimated, along with their standard errors and also with statistics summarizing the fit of the data to the measurement model. This paper summarizes the relevant theoretical principles associated with rank ordering and presents an example of this sort of analysis. The example includes H. Polskin's (1988) rankings of seven play-by-play baseball announcers on six specific items of performance. The application of the principles of fundamental measurement to rank ordered data provides the means to convert entirely local rankings into generalizable measures of the latent abilities. Moreover, fit statistics for each object and for each ordering enable a determination of the success of the ranking process as a measurement operation. (TJH)

Fundamental Measurement of

Rank-Ordered Objects

John M. Linacre

M E S A

University of Chicago

5835 S. Kimbark Avenue

Chicago Il 60637

Paper presented at the
Fifth International Objective Measurement Workshop
University of California, Berkeley, CA, March 26, 1989.

Abstract:

A Rasch measurement model can be constructed to meet the requirements of rank ordered data. If multiple rankings of the same objects are available, then the parameters of the objects can be estimated, along with their standard errors and also statistics summarizing the fit of the data to the measurement model. An example analysis is provided.

Key-words: Rank order, Rasch measurement.

I. Introduction.

"Examiners who are asked to place answer books in rank order, or order of merit, are asked to do a task which is far simpler for human judgement than is the assigning of absolute marks" (Harper 1976 p. 255). The use of judge-created rankings of the performance of examinees on each test item, instead of judge-awarded scores, removes from the test analysis the severity of the judges, the difficulty of the test items and the arbitrary nature and idiosyncratic implementation of rating scales.

Ranking would also appear to remove the foundational component identified by Rasch for measurement models in psychology: "The possible behavior of a pupil is described by means of a probability that he solves the task" (Rasch 1980 p. 11). A ranking of examinees contains no indication of what level of success the examinees attained on the particular item on which they were judged. It does, however, contain information about their relative levels of success.

It has been observed that judges differ considerably in the rankings they assign: "The [examinee's performance with] the highest degree of agreement still covered nearly one-third of the range of ranks, while the average [range of ranking a performance] included nearly two-thirds of the available ranks" (Harper 1976 p. 14). It is this variation in the rankings across judges which provides the stochastic element necessary for Rasch measurement.

II. The fundamental measurement model for paired objects.

A comparison of the performance of two objects (e.g. Om and On) across numerous replications of a given agent (e.g. a test item) yields counts of the three possible outcomes:

1) Fmn, the frequency with which Om out-performs On.
2) Fnm the frequency with which On out-performs Om.
3) the frequency with which they perform at the same level.

For the purposes of this discussion, the discrimination of performance is assumed to be so fine that identical performance levels never occur. Thus, a comparison of the performance levels of these objects is Fmn/Fnm, which becomes, in the limit, Pmn/Pnm, where Pmn is the probability that Om out-performs On and Pnm is similarly defined.

We can also define object O0, whose performance level is at the local origin of the measurement scale. We can similarly compare the performance of Om with O0 yielding Pm0/P0m, and also On with O0 yielding Pn0/P0n.

Following Rasch, "if a relationship between two or more variables is to be considered really important, as more than an ad hoc description of a very limited set of data - if a more or less general interdependence may be considered in force - the relationship should be found in several sets of data which differ materially in some relevant respects" (Rasch 1980 p. 9). In our case, this implies that the results of a direct comparison of Om and On should lead to the same conclusion as a comparison of Om with On via O0. This requirement for generalizability thus leads to

$$\frac{Pmn}{Pnm} = \frac{Pm0}{P0m} \Big/ \frac{Pn0}{P0n} \qquad (1)$$

but Pm0/P0m is the performance of Om relative to a measure defined to be at the origin of the scale, and so is a constant, Am. Similarly Pn0/P0n is a constant, An. Then taking logarithms,

$$\log(Pmn/Pnm) = \log(Am) - \log(An) \qquad (2)$$

or, reparameterizing, this becomes the measurement model for paired objects,

$$\log(Pmn/Pnm) = Bm - Bn \qquad (3)$$

where

Pmn is the probability that object Om out-performs object On
Pnm is the probability that object On out-performs object Om
Bm is the measure of object Om
Bn is the measure of object On.

III. Extending the measurement model from pairs to rankings.

If a ranking is of only two objects, then the measurement model for paired objects applies directly. Thus the probability, Rab of observing object Oa ranked higher than object Ob is given by

$$Rab = Pab = \frac{Pab}{Pab + Pba} \qquad (4)$$

where the denominator contains 2! = 2 terms, representing all the possible valid numerators for ordering two objects.

The ranking of three objects, Oa, Ob, Oc, can be regarded as a set of three paired rankings, but with the constraint that if Oa is ranked higher than Ob, and Ob is ranked higher than Oc, then Oa must be ranked higher than Oc. The probabilities of their eight theoretically possible paired relationships are shown in Table 1.

Probability of independent pairing	Representation as rank order
Pab*Pac*Pbc	R(Oa,Ob,Oc)
Pab*Pac*Pcb	R(Oa,Oc,Ob)
Pab*Pca*Pcb	R(Oc,Oa,Ob)
Pba*Pac*Pbc	R(Ob,Oa,Oc)
Pba*Pca*Pbc	R(Ob,Oc,Oa)
Pba*Pca*Pcb	R(Oc,Ob,Oa)
Pab*Pca*Pbc	inconsistent
Pba*Pac*Pcb	inconsistent

Table 1. Probabilities of all possible paired comparisons of of three objects. The contents of R( ) represent the ordering of the objects.

The effect of the constraint on the pairings imposed by ranking is the determination that two of the possible paired combinations of objects are inconsistent and can never be observed. Apart from this constraint, the probability of observing any particular rank ordering is assumed to depend only on the paired comparison of the objects and not to involve any other characteristics of the sample of objects. This is equivalent to the "local independence" axiom of other Rasch models. Thus the comparison of the objects manifested in the ranking is as "sample-free" as possible. If, for any particular data set of rankings, this is not the case, then the data set can not be expected to fit the measurement model presented here. Fit statistics can diagnose this eventuality.

Considering the possible rankings, if Rab is the probability that Oa is ranked higher than Ob in the rank ordered data, then

$$\frac{Rab}{Rba} = \frac{\text{Probability of observing Oa higher than Ob}}{\text{Probability of observing Ob higher than Oa}} \qquad (5)$$

$$= \frac{Pab*Pac*Pbc + Pab*Pac*Pcb + Pab*Pca*Pcb}{Pba*Pac*Pbc + Pba*Pca*Pbc + Pba*Pca*Pcb} \qquad (6)$$

$$= \frac{Pab * (1 - Pca*Pbc)}{Pba * (1 - Pac*Pcb)} \qquad (7)$$

since Rab = 1 - Rba, then

$$Rab = \frac{Pab * (1 - Pca*Pbc)}{Pab * (1 - Pca*Pbc) + Pba * (1 - Pac*Pcb)} \qquad (8)$$

$$Rab = \frac{Pab * (1 - Pca*Pbc)}{1 - Pab*Pca*Pbc - Pba*Pac*Pcb} \qquad (9)$$

$$Rab = \frac{\text{Probability of observing Oa ranked higher than Ob}}{\text{Probability of all possible rankings}} \qquad (10)$$

Rbc and Rac are similarly obtained. These probabilities are not independent as the following identity makes clear:

$$Rabc \equiv Rab \cap Rbc \equiv Rab \cap Rbc \cap Rac \qquad (11)$$

where Rabc is the probability of observing the ranking R(Oa,Ob,Oc)
$\cap$ is the intersection of the sample spaces.

In particular, in general,

$$Rabc < Rab * Rbc \qquad (12)$$

with the precise result that

$$Rabc = \frac{\text{Probability of } R(Oa,Ob,Oc)}{\text{Probability of all possible rankings}} \qquad (13)$$

$$= \frac{Pab * Pac * Pbc}{1 - Pab*Pca*Pbc - Pba*Pac*Pcb} \qquad (14)$$

Numbering the objects arbitrarily, O1, O2, O3, then the probability of the observed rank ordering, whatever it is, is given by

$$R(\{3\}) = \frac{\prod_{j=1}^{3} \prod_{k=j+1}^{3} (Xjk * Pjk + Xkj * Pkj)}{\Sigma R(\{3\})} \qquad (15)$$

where

R({3}) is the probability of a particular ranking of 3 objects
Xjk = 1 if Oj is ranked higher than Ok,
= 0 otherwise
Xkj = 1 - Xjk
ΣR({3}) is the sum of all possible numerators and contains one term for every permutation of 3 objects, i.e. 3! = 6 terms.

IV. Rank ordering of n objects.

For convenience of generalization, let us arbitrarily number the objects O1, O2,..,On with corresponding parameters B1, B2,..,Bn. For some rank ordering of the objects, R({n}),

$$R(\{n\}) = \frac{\prod_{j=1}^{n} \prod_{k=j+1}^{n} (Xjk * Pjk + Xkj * Pkj)}{\Sigma R(\{n\})} \qquad (16)$$

with the same conventions as before. In particular, $\Sigma R(\{n\})$ is a sum including one term for each of the possible numerators, identical to that numerator. The number of possible numerators is the number of ways of permuting n objects, that is n!.

The measurement model defining the relationship between objects Oj and Ok is, rewriting (3) with $Pkj = 1 - Pjk$,

$$Pjk = \exp(Bj) \,/\, (\mathrm{Exp}(Bj)+\mathrm{Exp}(Bk)) \qquad (17)$$

and

$$Pkj = \exp(Bk) \,/\, (\mathrm{Exp}(Bj)+\mathrm{Exp}(Bk)) \qquad (18)$$

so that the probability of a rank ordering in terms of the underlying parameters is

$$R(\{n\}) = \frac{\prod_{j=1}^{n} \prod_{k=j+1}^{n} \dfrac{Xjk*\exp(Bj) + Xkj*\exp(Bk)}{\exp(Bj)+\exp(Bk)}}{\Sigma R(\{n\})} \qquad (19)$$

V. Independent rank orderings of n objects.

If independent rank orderings of the same n objects have been compiled by T judges, then the likelihood of the data set becomes

$$\Lambda\{n\} = \prod_{r=1}^{T} \frac{\prod_{j=1}^{n} \prod_{k=j+1}^{n} \dfrac{Xrjk*\exp(Bj) + Xrkj*\exp(Bk)}{\exp(Bj)+\exp(Bk)}}{\Sigma R(\{n\})} \qquad (20)$$

For estimability of all parameters in one frame of reference, it is required that the orderings of the objects overlap in such a way that every object can be compared to every other object, either directly or indirectly, in terms of both relative successes and relative failures. If, for instance, one object is always ranked highest, then its parameter is inestimable. A more subtle example of inestimability is a set of orderings in which the objects form two groups, the high group and the low group, and no object in the high group is ever ranked below any object in the low group.

If all objects do not participate in every rank ordering, the overall likelihood becomes the product of the likelihood of homogeneous subgroups in which the same set of objects has been ranked by one or more judges. Thus if n objects have been ranked by T judges, and m objects (including some of the n objects) have been ranked by S judges, then

$$\Lambda\{m \cup n\} = \left(\prod_{r=1}^{T} \frac{\displaystyle\prod_{j=1}^{n} \prod_{k=j+1}^{n} \frac{Xrjk*\exp(Bj) + Xrkj*\exp(Bk)}{\exp(Bj)+\exp(Bk)}}{\Sigma R(\{n\})} \right) * \left(\prod_{r=1}^{S} \frac{\displaystyle\prod_{j=1}^{m} \prod_{k=j+1}^{m} \frac{Xrjk*\exp(Bj) + Xrkj*\exp(Bk)}{\exp(Bj)+\exp(Bk)}}{\Sigma R(\{m\})} \right) \quad (21)$$

where $\Lambda\{m \cup n\}$ is the likelihood of the entire data set. The following derivation can then be adapted to this formulation of the data, but, for clarity, we return to the consideration of a homogeneous data set.

The factor

$$\prod_{j=1}^{n} \prod_{k=j+1}^{n} \frac{1}{\exp(Bj)+\exp(Bk)}$$

is common to every term in the numerator and denominator of (20), and so can be cancelled out. Thus (20) becomes

$$\Lambda\{n\} = \prod_{r=1}^{T} \frac{\displaystyle\prod_{j=1}^{n} \prod_{k=j+1}^{n} (Xrjk*\exp(Bj) + Xrkj*\exp(Bk))}{\displaystyle\sum_{s=1}^{n!} \prod_{j=1}^{n} \prod_{k=j+1}^{n} (Xsjk*\exp(Bj) + Xskj*\exp(Bk))} \quad (22)$$

The denominator includes all the possible numerators corresponding to all valid rankings and so consists of n! terms corresponding to the n! ways of ordering n objects.

Taking logarithms, the log-likelihood of a set of T rank orderings of n objects is

$$\log(\Lambda\{n\}) = \Psi = \sum_{r=1}^{T} \sum_{j=1}^{n} \sum_{k=j+1}^{n} (\log (Xrjk*\exp(Bj) + Xrkj*\exp(Bk)))$$

$$- T*\log(\sum_{s=1}^{n!} \prod_{j=1}^{n} \prod_{k=j+1}^{n} (Xsjk*\exp(Bj) + Xskj*\exp(Bk)))) \quad (23)$$

VI. Estimation equations for rank ordered objects.

The Newton-Raphson estimation equations for the parameters can be obtained using first and second derivatives of the log-likelihood function (23).

To estimate B_m, partially differentiate with respect to B_m,

$$\frac{\delta \Psi}{\delta B_m} = \sum_{r=1}^{T} \sum_{j=1,\neq m}^{n} X_{rmj}$$

$$- T * \sum_{r=1}^{n!} \left(\sum_{l=1,\neq m}^{n} X_{rml}\right) \frac{\prod_{j=1}^{n} \prod_{k=j+1}^{n} (X_{rjk}*\exp(B_j) + X_{rkj}*\exp(B_k))}{\sum_{s=1}^{n!} \prod_{j=1}^{n} \prod_{k=j+1}^{n} (X_{sjk}*\exp(B_j) + X_{skj}*\exp(B_k))} \quad (24)$$

The first term represents the observed score and is a count of the number of objects higher than which O_m is ranked in all the observed rank orderings. The second term represents the expected score is the sum, across all possible rank orderings, of the number of objects higher than which O_m is ranked in each rank ordering, multiplied by the probability of that rank ordering, all multiplied by the number of rank orderings in the observed data.

Differentiating the log-likelihood again with respect to B_m,

$$\frac{\delta^2 \Psi}{\delta B_m^2} = T*\left(\frac{\sum_{r=1}^{n!} \left(\sum_{l=1,\neq m}^{n} X_{rml}\right) \prod_{j=1}^{n} \prod_{k=j+1}^{n} (X_{rjk}*\exp(B_j) + X_{rkj}*\exp(B_k))}{\sum_{s=1}^{n!} \prod_{j=1}^{n} \prod_{k=j+1}^{n} (X_{sjk}*\exp(B_j) + X_{skj}*\exp(B_k))} \right)^2$$

$$- T*\left(\frac{\sum_{r=1}^{n!} \left(\sum_{l=1,\neq n}^{n} X_{rml}\right)^2 \prod_{j=1}^{n} \prod_{k=j+1}^{n} (X_{rjk}*\exp(B_j) + X_{rkj}*\exp(B_k))}{\sum_{s=1}^{n!} \prod_{j=1}^{n} \prod_{k=j+1}^{n} (X_{sjk}*\exp(B_j) + X_{skj}*\exp(B_k))} \right) \quad (25)$$

This provides the specific form of the terms for the general form of the Newton-Raphson estimation equation for B'_m, the improved estimate of B_m, which is the measure corresponding to object O_m,

$$B'_m = B_m - \frac{\delta \Psi}{\delta B_m} \Big/ \frac{\delta^2 \Psi}{\delta B_m^2} \quad (26)$$

When the iterative process has converged, the asymptotic standard error of the estimate, B_m, is given by

$$S.E.(B_m) = 1/\sqrt{\left(- \frac{\delta^2 \Psi}{\delta B_m^2}\right)} \qquad (27)$$

Rasch model fit statistics, both information-weighted and outlier-sensitive, can also be calculated (Wright and Masters 1982 p. 100).

VII. Tied rankings.

In some judging situations, two or more objects may be given the same ranking. If two objects Oj and Ok are given the same ranking, then this is equivalent to the statement that orderings (Oj, Ok) and (Ok, Oj) are equally probable as representations of the ordering of the objects on the latent variable. Consequently, if orderings (Oj, Ok) and (Ok, Oj) are each given a weighting of one-half, then the sum is equivalent to the tied ordering. Thus, if Oj and Ok are tied in the ordering, then $X_{jk} = 0.5$ and $X_{kj} = 0.5$ for the purposes of determining empirical scores. Considered in this way, the admissability of tied rankings does not add any more orderings into the scheme of all possible rank orderings described above.

VIII. An application of the Rasch model for rank ordered objects.

In Polskin (1988), and reproduced in Table 2, are rankings of seven play-by-play baseball announcers on six specific items of performance. For the purposes of this analysis, the six rankings are considered to be independent manifestations of the same latent abilities.

Calling the game	Broadcasting ability	Quality of anecdotes
1. Vin Scully	1. Vin Scully	1. Vin Scully
2. Bob Costas	2. Al Michaels	2. Bob Costas
3. Al Michaels	3. Bob Costas	3. Al Michaels
4. Skip Caray	4. Skip Caray	4. Skip Caray
5. Harry Caray	5. Harry Caray	5. Ralph Kiner
6. Steve Zabriskie	6. Steve Zabriskie	6. Harry Caray
7. Ralph Kiner	7. Ralph Kiner	7. Steve Zabriskie

Working with analyst	Knowledge of baseball	Enthusiasm level
1. Bob Costas	1. Vin Scully	1. Harry Caray
2. Al Michaels	2. Ralph Kiner	2. Al Michaels
3. Vin Scully	3. Bob Costas	3. Bob Costas
4. Skip Caray	4. Al Michaels	4. Vin Scully
5. Steve Zabriskie	5. Harry Caray	5. Steve Zabriskie
6. Ralph Kiner	6. Skip Caray	6. Skip Caray
7. Harry Caray	7. Steve Zabriskie	7. Ralph Kiner

Table 2. Rankings of Play-by-Play Announcers.

The Rasch rank-order measurement model can be used to answer such questions as "How much better is one announcer than another ?". "Which announcers have the most consistent quality level ?" and "Do the items cooperate in defining one "Quality of Announcing" variable ?"

In answer to the question, "How much better is one announcer than another ?", Table 3 lists the estimates of the measures obtained for this data set. The relationship between the sum of each announcer's ranks and his measure is close to linear, as can be seen by inspection of Figure 1. The most consistently ranked announcer, with a mean-square fit statistic of 0.34, is Al Michaels, and the one most inconsistently ranked is Ralph Kiner with a fit of 2.12. The distinction between "information-weighted" fit statistics and "outlier-sensitive" fit statistics does not exist here because the variance term for each estimate is uniform across rank orderings. How difficult misfit is to determine, by eye, from lists of rank orderings is indicated by the different conclusion reached by Polskin. According to the analysis given in the text of his article, Polskin had the impression that Harry Caray was the least consistently ranked announcer, due to his first place on "Enthusiasm".

A basic question to the success of the measurement operation is the uni-dimensionality of the "Quality of Announcing" variable. Are the six orderings independent manifestations of the same latent parameters ? Table 4 summarizes the degree of fit within each ordering. Since ordering provides no information on, say, how difficult it is to "call the game", no difficulty calibrations are shown.

Ability Order	Sum of Rankings	Measure (Logits)	S.E.	Mean-Square Fit Statistic	Announcer
1	11	0.98	0.41	1.51	Vin Scully
2	14	0.67	0.35	0.40	Bob Costas
3	16	0.50	0.32	0.34	Al Michaels
4	28	-0.26	0.28	0.43	Skip Caray
5	29	-0.33	0.29	1.73	Harry Caray
6	34	-0.69	0.33	2.12	Ralph Kiner
7	36	-0.87	0.37	0.54	Steve Zabriskie
	Mean:	0.00		1.01	

Table 3. Ability of Baseball Announcers

Information-weighted Mean-Square fit	Outlier-sensitive Mean-Square fit	Name of Ordering Item
0.29	0.32	Calling the game
0.35	0.39	Broadcasting ability
0.38	0.41	Quality of anecdotes
0.91	0.91	Working with analyst
1.77	1.80	Knowledge of baseball
2.29	2.22	Enthusiasm level

Table 4. Fit statistics for items as manifested in the rank orderings.

The items are generally acting in a coherent manner in defining the variable. It may well be that "Calling the game" and "Broadcasting ability" are somewhat synonymous and not independent items, leading to a redundancy in the data. "Enthusiasm" displays the most misfit, and may be multi-dimensional. This is because, according to Polskin, it is easier to announce when you are doing it for the "home-team fans", as Harry Caray does.

Table 5 shows those rankings which were the least expected. This Table is an aid to the diagnosis of aberrations in the measuring process, which Polskin's analysis apparently lacked, since he failed to comment on the most unexpected ranking, that of Ralph Kiner on "Knowledge".

Ordering	Announcer	Rank	Expected	Difference	S.E.	Z-Score
Knowledge	Ralph Kiner	2	5.7	3.67	1.23	2.97
Enthusiasm	Harry Caray	1	4.8	3.83	1.42	2.71
Working	Bob Costas	1	2.3	1.33	1.18	1.13
:						
Working	Vin Scully	3	1.8	-1.17	0.99	-1.18
Working	Harry Caray	7	4.8	-2.17	1.42	-1.53
Enthusiasm	Vin Scully	4	1.8	-2.17	0.99	-2.19
Mean for all ranks:				0.0		0.00
Variance for all ranks:						1.00

Table 5. Most unexpected rankings of announcers arranged in Z-score order.

IX. Conclusion.

The application of the principles of fundamental measurement to rank ordered data has provided the means to convert entirely local rankings into generalizable measures of the latent abilities. Moreover, fit statistics for each object and for each ordering enable a determination of the success of the ranking process as a measurement operation.

Bibliography:

Harper, A. E. Jr & Misra, V. S. Research on Examinations in India. National Council of Educational Research and Training, New Delhi, India. 1976.

Polskin, H. The Best Sportscasters in Baseball. TV Guide, July 30, 1988, 36:31.

Rasch, G. Probabilistic Models for Some Intelligence and Attainment Tests. Copenhagen 1960 and Chicago: University of Chicago press. 1980.

Wright, B.D. & Masters, G.N. Rating Scale Analysis: Rasch Measurement. Chicago: MESA Press. 1982.

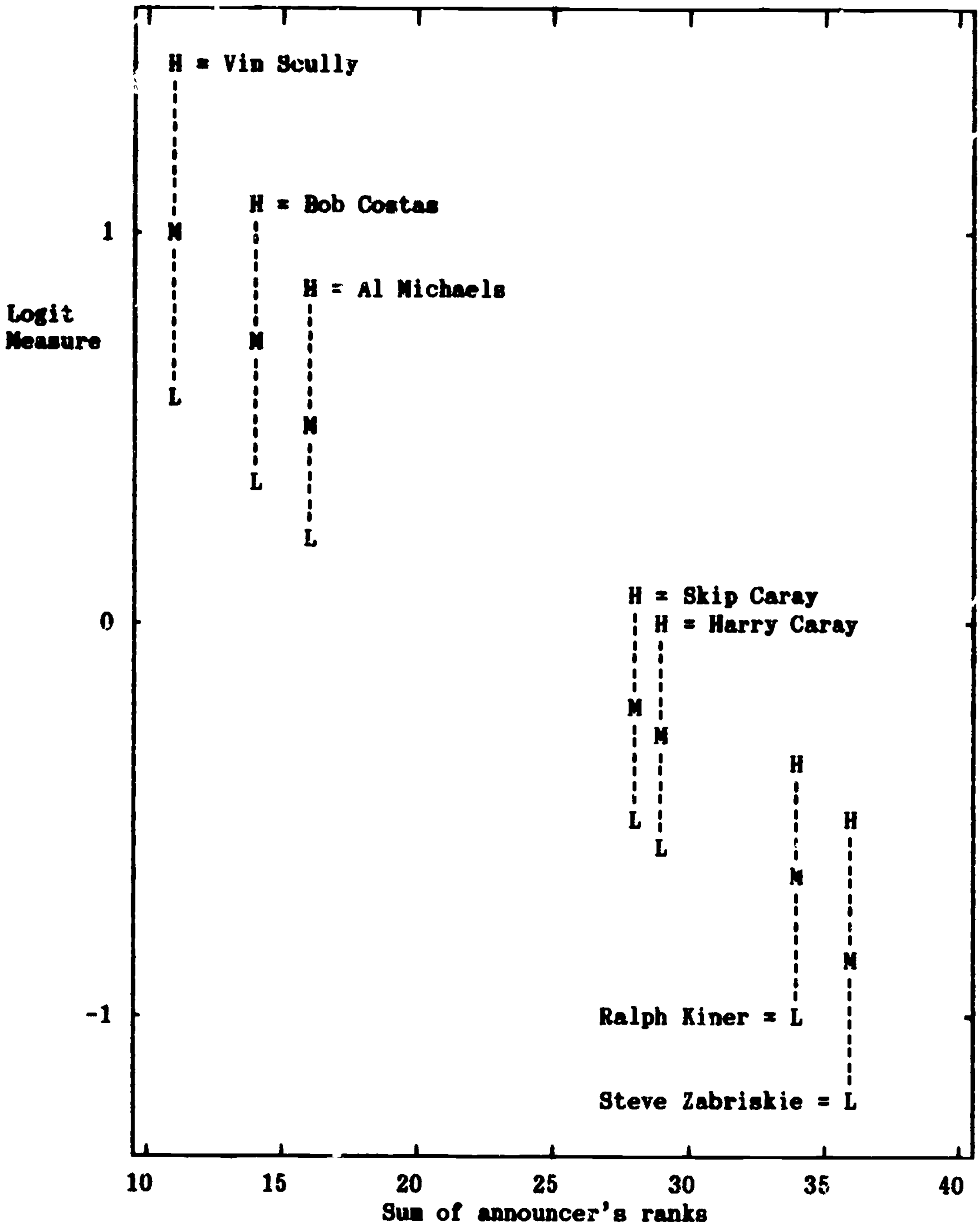

Figure 1. Announcers' measures plotted against sum of rankings.
M = estimated measure of the announcer.
H = M + standard error, L = M - standard error.