DOCUMENT RESUME

ED 236 164 TM 830 614

AUTHOR Wilson, Mark; Wright, Benjamin D.
TITLE Finite Measures from Perfect Scores.
PUB DATE Mar 83
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Factor Analysis; *Item Analysis; *Latent Trait Theory; Mathematical Models; *Scores; *Scoring; Testing Problems
IDENTIFIERS Partial Credit Model; *Perfect Scores

ABSTRACT

A common problem in practical educational research is that of perfect scores which result when latent trait models are used. A simple procedure for managing the perfect and zero response problem encountered in converting test scores into measures is presented. It allows the test user to chose among two or three reasonable finite representations of these boundary scores. A convenient approximation is given which is accurate in all but highly assymetric situations when the procedure is applied to the Partial Credit model. (Author/HFG)

Finite Measures from Perfect Scores

Mark Wilson and Benjamin D. Wright
MESA Psychometric Laboratory
University of Chicago
March, 1983

Paper presented at the
American Educational Research Association Meeting
Montreal, Canada, April 1983.

Paper printed in U.S.A.

Abstract

A simple procedure for managing the perfect and zero response problem encountered in converting test scores into measures is presented. It allows the test user to choose among two or three reasonable finite representations of these boundary scores to suit the particular situation at hand. A convenient approximation is given which is accurate in all but highly asymmetric situations.

The problem.

Those who actually use latent trait models in their research and in the solution of practical educational problems are inevitably confronted with the conundrum of perfect scores. This is particularly so in situations involving Mastery Learning and Mastery Testing where the goal of, say, an 80 per cent pass rate guarantees that many students will achieve the maximum possible test score. For the sake of completeness, and as an essential to the reporting of every individual's performance, a finite measure is needed to represent these perfect scorers.

That no such finite measure can be estimated from the interaction of a perfect scorer with a test was noted by Wright and Stone,

> "All we can do in those situations is to observe that the person who scored all incorrect or all correct is substantially below or above the operating level of the test they have taken. If we wish to estimate a finite measure for such a person, then we will have to find a test for them which is more appropriate to their level of ability."
>
> (Wright & Stone, 1979, p.6)

Whilst the application of of a more appropriate test is obviously preferred in such a situation, it is often inconvenient.

A solution

One alternative is to introduce the assumption that perfect scores can be represented by some distribution of measures over the latent trait and to rationalise a scaled score from this distribution. This, however, undoes one of the major advantages of latent trait measurement - the "sample-free" calibration of items and "test-free" measurement of persons (Wright & Masters, 1981, p.5).

Is there any other way to establish a reasonable estimate of how far a perfect scorer might be above the operating level of the test? We can answer this in a probabalistic sense. Consider an L item test with a maximum score of m_i for each item, and let

$$M = \sum_{i=1}^{L} m_i$$

be the perfect score. Then, assuming local independence, the probability of a perfect score, Q_M, is

$$Q_M = \Pr(r=M) = \prod_{i=1}^{L} \Pr(x_i = m_i),$$

where x_i indicates the score on each item, and r is the score on the test. Thus, given any measure b one can always calculate the probability that a person with that measure b will score M. An example of the probability curve generated by this equation is shown in Figure 1. The assumption being made is that the perfect scorers would fit on our latent trait if they were given more difficult items (ie. that it was reasonable to give them this sort of test in the first place).

Insert Figure 1 about here

This approach does not tell us where the measures that might go with perfect scores actually are on the latent trait, but it does provide us with information which we can use to represent the occurence of perfect scores in a reasonable fashion. To this end several points on the probability curve are worth noting:

b_x ...the ability at which the score characteristc curves for a perfect score M and for a one less than perfect score M-1 intersect

$b_{.5}$ ...the minimum ability at which a person is more likely to score M than anything else

$b_{.9}$ ...the ability at which a person is likely to make a score other than M only 10% of the time

$b_{.95}$ ...the ability at which a person is likely to make a score other than M only 5% of the time.

These values can help us to address the perfect score problem. There would seem to be no circumstances under which a perfect scorer would be assigned a value below b_x, for below that value a score of M-1 is more likely. Of course the highest possible ability that could be given to a perfect scorer is infinity, but a more realistic maximum value might be indicated by the point at which a less than perfect score would be observed with probability 0.10 or 0.05: thus $b_{.9}$ or $b_{.95}$ might be reasonable maximum values. A median estimate is provided by $b_{.5}$: this is the lowest point at which a person is more likely to score M than to make any other score. In the absence of further information about the requirements of the specific testing situation, it would seem that this median value, $b_{.5}$, might be the best choice.

Application to the Partial Credit Model

The Partial Credit model (Wright & Masters, 1982, p.43) states that for an m_i+1 category item, the probability of a person who scored r, with scaled score b_r , getting x on item i, with step values d_{i1} , d_{i2} , ... , d_{im_i} is

$$P_{rix} = \frac{\exp \sum_{j=0}^{x} (b_r - d_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (b_r - d_{ij})} \qquad x = 0,1,2,\ldots,m_i$$

$$\text{where } \sum_{j=0}^{0} (b_r - d_{ij}) = 0 .$$

For a test with L items, the maximum score M is $\sum_{i=1}^{L} m$.

Now, for each interesting probability level Q, we want to find the ability level b_M for which

$$Q = \prod_{i=1}^{L} Pr(x_i = m_i) = \prod_{i=1}^{L} P_{Mim_i} ,$$

that is

$$Q = \prod_{i=1}^{L} \frac{\exp \sum_{j=0}^{m_i} (b_M - d_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (b_M - d_{ij})}$$

Since the solution of this equation is not explicit, we solve it by iteration: first take logarithms

$$\sum_{i=1}^{L}\sum_{j=0}^{m_i}(b_M-d_{ij}) - \sum_{i=1}^{L}\log\sum_{k=0}^{m_i}\exp\sum_{j=0}^{k}(b_M-d_{ij}) - \log Q = 0 ,$$

which simplifies to

$$Mb_M - D - \sum_{i=1}^{L}\log\sum_{k=0}^{m_i}\exp\sum_{j=0}^{k}(b_M-d_{ij}) - \log Q = 0 ,$$

where $D = \sum_{i=1}^{L}\sum_{j=1}^{m_i} d_{ij}$.

The derivative of this equation with respect to b_M is

$$M - \sum_{i=1}^{L}\sum_{k=0}^{m_i} kP_{Mik}$$ (Wright & Masters, 1982, p.87) .

Thus if b_M^n is an initial approximation to b_M, we may use Newton's method to find a better approximation, b_M^{n+1}

$$b_M^{n+1} = b_M^n - \frac{Mb_M^n - D - \sum_{i=1}^{L}\log\sum_{k=0}^{m_i}\exp\sum_{j=0}^{k}(b_M^n-d_{ij}) - \log Q}{M - \sum_{i=1}^{L}\sum_{k=0}^{m_i} kP_{Mik}^n}$$

where P_{Mik}^n is P_{Mik} evaluated at b_M^n .

A very good approximation to b_M is given by

$$b_M^0 = b_{M-1} - \log\log(1/Q) + 1/(M-1).$$

To see why this approximation is reasonable, consider the case of an L item dichotomous test (ie. M=L) with all items of difficulty 0.

The probability of a person with ability b_L getting each item right is $(1 + \exp(-b))^{-1}$ so

$$Q = (1 + \exp(-b_L))^{-L}$$

which from $\exp(x) \simeq 1 + x$

is approximately

$$(\exp(\exp(-b)))^{-L}$$

so that

$$L(\exp(-b_L)) \simeq \log(1/Q)$$

or $b_L \simeq \log L - \log\log(1/Q)$.

But $b_{L-1} = \log(L-1)$

so $\log L = b_{L-1} + \log(L/(L-1))$

$$\simeq b_{L-1} + 1/(L-1).$$

Substituting this in the equation above gives the approximation

$$b_L \simeq b_{L-1} - \log\log(1/Q) + 1/(L-1).$$

The introduction of b_{L-1} relaxes the requirement that all items be of difficulty 0. Since the two b's cancel the difficulty level of the test, all that is required is that the items be of similar difficulty.

The crossover ability b_x is the ability at which

$$Pr(r=M) = Pr(r=M-1).$$

Now $\quad Pr(r=M) = \prod_{i=1}^{L} (P_{b|m_i})$

$$= \frac{\prod_{i=1}^{L} \exp \sum_{j=0}^{m_i} (b-d_{ij})}{\prod_{i=1}^{L} \sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (b-d_{ij})}$$

and $Pr(r=M-1) = \sum_{i=1}^{L} (P_{b|(m_i-1)} \prod_{\substack{j=1 \\ j\neq i}}^{L} (P_{b|m_j}))$

$$= \frac{\prod_{i=1}^{L} \exp \sum_{j=0}^{m_i-1} (b-d_{ij}) (\sum_{i=1}^{L} \prod_{\substack{j=1 \\ j\neq i}}^{L} \exp(b-d_{jm_j}))}{\prod_{i=1}^{L} \sum_{k=0}^{m_i} \exp(\sum_{j=0}^{k} (b-d_{ij}))} .$$

Equating these two

$$\prod_{i=1}^{L} \exp(b_x-d_{im_i}) = \sum_{i=1}^{L} \sum_{\substack{j=1 \\ j\neq i}}^{L} \exp(b_x-d_{jm_j}) .$$

We divide to get

$$1 = \sum_{i=1}^{L} (\exp(b_x-d_{im_i}))^{-1}$$

$$= \exp(-b_x) \sum_{i=1}^{L} (\exp(-d_{im_i}))^{-1}$$

so

$$b_x = \log(\sum_{i=1}^{L} \exp(d_{im_i})).$$

In particular, when all items are dichotomous and of zero difficulty

$$b_x = \log L$$

Some examples

(i) For the L=10 example of Figure 1 :

$b_9 = 2.20$

$b_x = 2.30$

$b_{.5} = 2.63 \quad b^0_{.5} = 2.68$

$b_{.9} = 4.55 \quad b^0_{.9} = 4.56$

$b_{.95} = 5.27 \quad b^0_{.95} = 5.28$

This shows that the approximation works for the circumstances from which it was derived.

(ii) For a group of items from a blind persons' activity scale (Schulz et al., 1982), with step difficulties:

Items	1	2	3	4	5	6	7	8	9	10
Step1	-0.37	-2.91	-0.85	-2.47	-0.47	3.35	-4.05	-1.36	1.11	-0.20
Step2	-	-0.01	2.35	1.58	-	-	-0.47	1.99	-	2.80

The maximium score for these 16 steps is 16. These step difficulty values give the following boundary abilities

$b_{15} = 4.06$

$b_x = 4.30$

$b_{.5} = 4.59 \quad b^0_{.5} = 4.49$

$b_{.9} = 6.54 \quad b^0_{.9} = 6.38$

$b_{.95} = 7.26 \quad b^0_{.95} = 7.10$

(iii) An example of items with differing difficulty spreads

	20 items uniformly spread over -3.9 to 3.7	20 items uniformly spread over -1.9 to 1.9
b_{19}	4.60	3.48
b_x	4.81	3.59
$b_{.5}$	5.11	3.92
$b^0_{.5}$	5.02	3.90

The approximation suffers a little with greater spread.

(iv) An example that is highly asymmetric

19 items spread uniformly from -3.9 to -0.3 and one item at 3.9

$b_{19} = 2.77$

$b_x = 3.98$

$b_{.5} = 4.04$

$b^0_{.5} = 3.19$

In this asymmetric case the approximation is poor.

Zero scores

The analysis above can be applied to zero scores with a slight adjustment:

Newton's method becomes

$$b_0^{n+1} = b_0^n - \frac{\sum_{i=1}^{L} \log \sum_{k=0}^{m_i} \exp \sum_{j=0}^{k} (b_0^n - d_{ij}) - \log Q}{\sum_{i=1}^{L} \sum_{k=0}^{m_i} k P_{0ik}^n},$$

the approximation b_0^0 is

$$b_0^0 = b_1 + \log\log(1/Q) - 1/(M-1),$$

the cross-over point is

$$b_x = -\log\left(\sum_{i=1}^{L}\exp(-d_{i1})\right)$$

which, for the case where all items are dichotomous and have the same difficulty becomes

$$b_x = -\log L .$$

The case of zero and perfect items

The same method can be applied to manage the perfect and zero items which occasionally arise during item calibration. The resultant difficulty values would be interpreted in the same way as the ability values above.

Conclusion

Although there can be no substitute for the aquisition of further information regarding perfect scorers, it is often impractical to do so. The procedure described above uses the already-known features of the test (ie. item difficulties) to provide a readily interpreted measure for a perfect scorer. When this is applied to the Partial Credit model an iterative solution is required, however an approximation is available which gives excellent results. Generalization to zero scores and perfect and zero items is direct.

References

Schulz, E.M., Wright, B.D., Lambert, R.W., & Becker, S. A Measure of Activity Capacity in Blind Rehabilitation. Paper presented at the Conference on Rehabilitation Engineering, Houston, 1982.

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

Figure 1. Score Characteristic Curves for a test of 10 dichotomous items following the simple logistic model where all items have zero difficulty.