A closed form approximation is given for the variance of examinee proficiency estimates in the Rasch model for dichotomous items, under the condition that only estimates, rather than true values, of item difficulty parameters are available. The term that must be added to the usual response-sampling variance is inversely proportional to both the number of examinees in the item calibration sample and the length of the test. Illustrative numerical values suggest that the impact of uncertainty about Rasch item parameters on subsequent estimates of examinee proficiencies is less than has been observed in the two- and three-parameter logistic item response theory models. Two tables present approximate error variances, and an eight-item list of references is included. (Author/SLD)
THE VARIANCE OF RASCH ABILITY ESTIMATES
FROM PARTIALLY-KNOWN ITEM PARAMETERS

Robert J. Mislevy
Educational Testing Service

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-88-K-0304
R&T 4421552

Robert J. Mislevy, Principal Investigator
Educational Testing Service
Princeton, New Jersey

Reproduction in whole or in part is permitted
for any purpose of the United States Government.

Approved for public release; distribution unlimited.

BEST COPY AVAILABLE
The Variance of Rasch Ability Estimates from Partially-Known Item Parameters (Unclassified).

A closed-form approximation is given for the variance of examinee proficiency estimates in the Rasch model for dichotomous items, under the condition that only estimates, rather than true values, of item difficulty parameters are available. The term that must be added to the usual response-sampling variance is inversely proportional to both the number of examinees in the item calibration sample and the length of the test. Illustrative numerical values suggest that the impact of uncertainty about Rasch item parameters on subsequent estimates of examinee proficiencies is less than has been observed in the 2- and 3-parameter logistic IRT models.

Key words: Cohen's approximation, Rasch model, variance components.
The Variance of Rasch Ability Estimates from Partially-Known Item Parameters

Robert J. Mislevy

Educational Testing Service

January, 1992

This work was supported by Contract No. N00014-88-K-0304, R&T 4421552, from the Cognitive Sciences Program, Cognitive and Neural Sciences Division, Office of Naval Research. I am grateful to Nancy Allen, Kathleen Sheehan, and Benjamin Wright for comments on earlier drafts.
The Variance of Rasch Ability Estimates from Partially-Known Item Parameters

Abstract

A closed-form approximation is given for the variance of examinee proficiency estimates in the Rasch model for dichotomous items, under the condition that only estimates, rather than true values, of item difficulty parameters are available. The term that must be added to the usual response-sampling variance is inversely proportional to both the number of examinees in the item calibration sample and the length of the test. Illustrative numerical values suggest that the impact of uncertainty about Rasch item parameters on subsequent estimates of examinee proficiencies is less than has been observed in the 2- and 3-parameter logistic IRT models.

Key words: Cohen's approximation, Rasch model, variance components
Introduction

It is common practice in applications of item response theory (IRT) to estimate item parameters from a calibration sample of examinees, then use these estimates as if they were known true parameters when estimating the proficiency parameters of these and subsequent examinees. Tsutakawa and his colleagues have pointed out that ignoring the uncertainty about item parameters leads to inferential errors, and offered a Taylor series approximation to incorporate this uncertainty into proficiency estimates and statements about their precision. Results for the two-parameter logistic IRT model appear in Tsutakawa and Soltys (1988) and for the three-parameter logistic model appear in Tsutakawa and Johnson (1990). (Also see Lewis, 1985, for an alternative approach to the problem, and Mislevy, Sheehan, & Wingersky, in press, for the application of these ideas to test equating.) Wright and Panchapekesan (1969) present without derivation a closed-form approximation for the variance of examinee proficiency estimates under the Rasch model for dichotomous items (Rasch, 1960/1980). Utilizing a series of computing approximations and simplifying assumptions, the present paper gives a somewhat simpler approximation. Neither numerical accuracy nor practical application is the point of this presentation—the aforementioned alternatives are preferable in this regard. Rather, by giving an approximation with fewer “moving parts,” the objective is to aid intuition into the sources, the structures, and the consequences of uncertainty.

The Rasch Model

The Rasch model for dichotomous items gives the probability of a correct response to Item j from an examinee with proficiency \( \theta \) as

\[
P_j(\theta) = \frac{\exp(\theta - b_j)}{1 + \exp(\theta - b_j)}
\]

(1)
where $b_j$ is the difficulty parameter of Item $j$. Letting $x_j$ denote the response to Item $j$, 1 for correct and 0 for incorrect, the assumption of local independence gives the conditional probability of the response vector $x=(x_1,\ldots,x_L)$ on an $L$-item test as

$$P(x|\theta,b) = \prod_{j=1}^{L} P_j(\theta)^{x_j}[1-P_j(\theta)]^{1-x_j},$$

(2)

where $b=(b_1,\ldots,b_L)$.

Given item difficulty parameters and a realization of $x$, (2) can be interpreted as a likelihood function for $\theta$. Unless all the responses are correct or all are incorrect, it attains its maximum at a unique point, the maximum likelihood estimate (MLE) $\hat{\theta}$. The negative reciprocal of its second derivative at this point approximates the sampling variance of $\hat{\theta}$, or $\text{Var}(\hat{\theta}|\theta,b)$. Given response vectors from a calibration sample of $N$ examinees, the product of expressions like (2) serves similarly as a basis for the estimation of $b$ and the sampling variance of the resulting estimate $\hat{b}$.

**Cohen's Approximations**

Cohen (1979) gives closed form approximations for $\hat{\theta}$, $\text{Var}(\hat{\theta}|\theta,b)$, $\hat{b}$, and $\text{Var}(\hat{b}|b)$ for long tests when both $\theta$ and $b$ are normally distributed, in the samples of examinees and items respectively. Under these circumstances, item difficulty parameters can be approximated by rescaling the logits, or log-odds, of item percents-correct in the calibration sample, and examinees' proficiencies can be approximated by rescaling logits of their total scores by functions of item parameters. Adapting Wright's (1977) presentation, we begin with a calibration data set from which zero and perfect scoring examinees and items with all or no correct responses have been excised, then calculate the following quantities:
Partially Known Item Parameters

The proportion of the calibration sample that answered Item $j$ correctly, for $j=1, \ldots, L$,

$$\bar{P}_j$$

the number of examinees with a total score of $r$, for $r=1, \ldots, L-1$,

$$n_r$$

the log ratio of wrong to right answers on Item $j$,

$$t_j = \ln \left( \frac{i_{-j}}{P_j} \right)$$

the mean of $t_j$ over $L$ items,

$$t_r = \sum_{j=1}^{L} \frac{t_j}{L}$$

the variance of $t_j$ over $L$ items,

$$T = \sum_{j=1}^{L} \frac{(t_j-t.)^2}{(L-1)}$$

the log ratio of right to wrong answers for score $r$,

$$y_r = \ln \left( \frac{r}{L-r} \right)$$

the mean of $y_r$ over $N$ examinees,

$$y_r = \sum_{r=1}^{L-1} \frac{n_r y_r}{N}$$

the variance of $y_r$ over $N$ examinees, and

$$Y = \sum_{r=1}^{L-1} \frac{n_r (y_r-y)^2}{(N-1)}$$

$$W = \sqrt{\frac{1+Y/2.89}{1-TY/8.35}}$$

an expansion factor, in which 1.7 is used to match the logistic and normal distributions, and $2.89=1.7^2$ and $8.35=1.7^4$.

Then, from Cohen (1979), item parameters can be approximated as follows:

$$\hat{b}_j = W(t_j-t.)$$

and

(3)
Partially Known Item Parameters

\[ \text{Var}(\hat{b}_j | b_j) = \frac{1}{N} \left( \frac{W_j^2}{\hat{P}_j(1-\hat{P}_j)} \right) = \frac{C_i}{N}, \]

where \( C_i = \frac{W^2}{\hat{P}_j(1-\hat{P}_j)} \). The sampling covariances among the parameters of different items tend toward zero as \( L \) increases, as long as the items are not too dissimilar (Mislevy, 1981, shows that the error correlations among equivalent items are \(-1/(L-1)\)). We shall therefore approximate the \( L \)-by-\( L \) covariance matrix of sampling errors among item parameters as simply

\[ \text{Var}(\hat{b}_b) = \text{diag}(\frac{C_1}{N}, \ldots, \frac{C_L}{N}). \] (4)

If item parameters are known, examinee proficiencies can be similarly approximated by rescaling logits of their total scores. Define three additional quantities:

\[ b_\cdot = \frac{1}{L} \sum_{j=1}^{L} b_j \]

the mean of \( b_j \) over \( L \) items,

\[ U = \frac{1}{L} \sum_{j=1}^{L} (b_j - b_\cdot)^2 \]

the variance of \( b_j \) over \( L \) items, and

\[ X = (1 + U/2.89)^{1/2} \]

an expansion factor.

Then, for a total score of \( r \),

\[ \hat{\theta} = b_\cdot + X \ln \left( \frac{L-r}{L} \right) = b_\cdot + X \ln \left( \frac{p_r}{1-p_r} \right), \] (5)

where \( p_r = r/L \), and

\[ \text{Var}(\hat{\theta} | b, b) = X^2 \left( \frac{L-r}{r(L-r)} \right) \approx \frac{1}{L} \left( \frac{X^2}{p_r(1-p_r)} \right). \] (6)
Estimating Ability with Item Parameter Estimates

When \( b \) is known, (5) gives a proficiency estimate from \( x \), and (6), evaluated with the true value of \( \theta \), gives its sampling variance. What is the sampling variance of \( \hat{\theta} \) when (5) is evaluated with \( \hat{b} \) rather than \( b \)? In this presentation, this amounts to using item parameter estimates rather than true parameter values to calculate \( b \) and \( X \). The setup corresponds to a two-stage experiment:

Stage 1: A sample of \( N \) examinees is drawn from a normal population. They respond to the \( L \) items in the test. Their responses are used to calculate \( \hat{b} \) via (3).

Stage 2: A new examinee is administered the test and the test score \( r \) is observed. A proficiency estimate is calculated using (5) with \( \hat{b} \). We can write this estimate as \( \theta(r, \hat{b}) \) to emphasize that it is a function of two statistics—which, it will be noted, are independent, given \( b \) and \( \theta \).

The variance of \( \hat{\theta} \) in this setting can be decomposed into two terms:

\[
\text{Var}(\hat{\theta} | \theta) = \mathbb{E}_b \left[ \text{Var}(\hat{\theta} | \theta, \hat{b}) \right] + \mathbb{E}_b \left[ \text{Var}(\hat{\theta} | \theta, \hat{b}) \right].
\] (7)

The first term on the right of (7) is the expectation of (6) as evaluated with \( \hat{b} \), over the distribution of \( \hat{b} \) given \( b \). It depends on \( b \) only through \( X^2 \), or \( 1+U/2.89 \), where \( U \) is the variance of the true item parameters. The expected variance of the item parameter estimates is the variance of the item parameters themselves plus the average estimation error variance, \( C/N=(\sum C_j/L)/N \). Thus,

\[
\mathbb{E}_b \left[ \text{Var}(\hat{\theta} | \theta, \hat{b}) \right] = \mathbb{E}_b \left[ \frac{1}{L Pr(1 - Pr)} \right]
\]

\[
= \frac{1}{L} \left[ 1 + \frac{U + C/N}{2.89} \right] Pr(1 - Pr)
\]
Partially Known Item Parameters

\[
\frac{1}{L} \left( \frac{X^2}{p_r (1-p_r)} \right) + \frac{C}{LN} \left( \frac{1}{2.89 \frac{p_r}{1-p_r}} \right).
\]

This is the sum of the error variance for \( \hat{\theta} \) with known item parameters and a term depending on their error variance that is inversely proportional to both \( L \) and \( N \).

The second term in (7) concerns the variation in \( \hat{\theta} \) associated with the expected test score \( r \) for the true \( \theta \), as induced by variation in \( \hat{b} \). This term can be approximated by the delta method as follows:

\[
\text{Var}_{\hat{b}}[E(\hat{\theta}|\theta, \hat{b})] = \frac{\partial \hat{\theta}}{\partial \hat{b}} \times \text{Var}(\hat{b} | b) \times \frac{\partial \hat{\theta}}{\partial b}.
\]

\[
= \sum_{j=1}^{L} \left[ \left( \frac{\partial \hat{\theta}}{\partial b_j} \right)^2 \text{Var}(\hat{b}_j | b_j) \right] \quad \text{(by (4))}
\]

\[
= \sum_{j=1}^{L} \left[ \left( \frac{\partial \hat{\theta}}{\partial b_j} \left[ \hat{b}_j + \hat{X} \ln \left( \frac{p_r}{1-p_r} \right) \right] \right)^2 \text{Var}(\hat{b}_j | b_j) \right]
\]

\[
= \sum_{j=1}^{L} \left[ \frac{1}{L} + \frac{(b_j - b) \ln \left( \frac{p_r}{1-p_r} \right)}{2.89 LN} \right] \frac{C_j}{N}.
\]

Approximating all \( C_j \)s by their average, say \( C \), leads to the following simplification:

\[
\text{Var}_{\hat{b}}[E(\hat{\theta}|\theta, \hat{b})] = \frac{C}{LN} \left[ \frac{1}{1 + \frac{U \left( \ln \left( \frac{p_r}{1-p_r} \right) \right)^2}{8.35 \ X^2}} \right].
\]

Substituting (8) and (9) back into (7) yields the final approximation...

\[
\text{Var}(\hat{\theta}) = \frac{1}{L} \left( \frac{X^2}{p_r (1-p_r)} \right) + \frac{C}{LN} \left( 1 + \frac{1}{2.89 \frac{p_r}{1-p_r}} + \frac{U \left( \ln \left( \frac{p_r}{1-p_r} \right) \right)^2}{8.35 \ X^2} \right).
\]
Thus the sampling variance of $\hat{\theta}$ with item parameter estimates can be approximated as the sum of its variance with true item parameters and a correction term. The following observations can be made about the correction term:

1. It is always positive.
2. It is inversely proportional to test length ($L$).
3. It is inversely proportional to the size of the examinee calibration sample ($N$).
4. It is directly proportional to $C_j$. Since

$$C_j = \frac{\left(1-Y/2.89\right)}{\left(1-TY/8.35\right)P_j(1-P_j)},$$

where $Y$ is the variance of the logits of item percents-correct and $P_j$ is the percent-correct for Item $j$ in the calibration sample, the impact of the uncertainty about item parameters is greater when...

a. the items are more dispersed, so that $Y$ departs from zero, or
b. items are farther from 50-percent correct, so that $P_j(1-P_j)$ terms fall farther from the maximum value of .25.

5. Values of $\theta$ for which expected proportions of correct response are near 50-percent are less affected by uncertainty about item parameters than more extreme values of $\theta$. This follows from the second and third terms inside the brackets in (10): As $p_r$ approaches .5, the second term approaches its minimum value of $4/2.89$, and, because the logit of .5 is zero, the third term approaches zero.
Illustrative Numerical Values

To provide a feel for the relative contributions of sources of uncertainty this section evaluates the preceding formulae with some illustrative values. Standard normal distributions are assumed for $\theta$ and $b$ in the item calibration sample, implying the following values for true parameter values:

- $t. = y. = b. = 0$
- $T = Y = .743$
- $X = W = 1.160$
- $U = 1$

Equation (10) is used to approximate $\theta$ sampling variances, for $\theta = 0$ and 2, for test lengths of 10, 20, 40, 80, and 160 items, after item calibration with examinee samples of 50, 100, 250, 500, 1000, 2500, and 10000. For $C_r$, we employ a representative value, namely that for an item one standard deviation from the mean. For such an item, $\tilde{P}$ is about .7, in which case

$$C_r = W^2/(.7 \times .3) = 6.40$$

The values of $p_r$ that correspond to $\theta = 0$ and 2 are .50 and .85.

Table 1 gives results for $\theta = 0$, in terms of approximate sampling variances and proportional increases in sampling variance over those based on known item parameters. The proportional increases in standard errors are the square roots of the values in the rightmost column. Table 2 gives similar results for $\theta = 2$.

The most striking feature of these results is how small the corrections are. Even with a calibration sample of only 50 examinees estimation error variances for subsequent $\theta$ estimates increase by only about 5-percent. This contrasts with the increases of up to 30-percent Tsutakawa and Soltys (1988) observed in posterior variances for $\theta$ with a
calibration sample of 100 under the 2-parameter logistic model. Tsutakawa and Johnson's (1990) results for the 3-parameter logistic were even more extreme, with increases in posterior variances more than doubling at higher levels of proficiency with a calibration sample of 400.

Conclusion

Cohen’s (1979) closed-form approximations for the parameters in the Rasch model support a closed-form approximation for the variance of these examinee proficiency estimates when they are calculated with item parameter estimates rather than true item parameter values. The approximation is the sum of the sampling variance with known item parameters and a correction term. The correction term is inversely proportional to both test length and the size of the examinee item-calibration sample. Illustrative numerical values suggests the additional variance is quite small, yielding increases in standard errors of less than 5-percent even with calibration samples of only 50 examinees.
References


### Table 1
Approximate Error Variances at $\theta=0$

| N   | 10  | 20  | 40  | 80  | 160 | Var($\hat{\theta}|\theta, b)$ | Var($\hat{\theta}|\theta, \hat{b}$) |
|-----|-----|-----|-----|-----|-----|-------------------------------|---------------------------------|
| 50  | 0.5689 | 0.2845 | 0.1422 | 0.0711 | 0.0356 | 1.0567                      |
| 100 | 0.5537 | 0.2768 | 0.1384 | 0.0692 | 0.0346 | 1.0283                      |
| 250 | 0.5445 | 0.2723 | 0.1361 | 0.0681 | 0.0340 | 1.0113                      |
| 500 | 0.5415 | 0.2707 | 0.1354 | 0.0677 | 0.0338 | 1.0057                      |
| 1000| 0.5399 | 0.2700 | 0.1350 | 0.0675 | 0.0337 | 1.0028                      |
| 2500| 0.5390 | 0.2695 | 0.1348 | 0.0674 | 0.0337 | 1.0011                      |
| 10000| 0.5386 | 0.2693 | 0.1346 | 0.0673 | 0.0337 | 1.0003                      |
| $\infty$ | 0.5384 | 0.2692 | 0.1346 | 0.0673 | 0.0337 | 1.0000                      |
### Partially Known Item Parameters

#### Table 2

Approximate Error Variances at $\theta=2$

| N  | 10     | 20    | 40    | 80    | 160    | $\frac{\text{Var}(\theta|\theta, \mathbf{b})}{\text{Var}(\theta|\theta, \hat{\mathbf{b}})}$ |
|----|--------|-------|-------|-------|--------|----------------------------------|
| 50 | 1.1066 | 0.5533| 0.2767| 0.1383| 0.0692 | 1.0483                           |
| 100| 1.0812 | 0.5406| 0.2703| 0.1351| 0.0676 | 1.0241                           |
| 250| 1.0659 | 0.5329| 0.2665| 0.1332| 0.0666 | 1.0097                           |
| 500| 1.0608 | 0.5304| 0.2652| 0.1326| 0.0663 | 1.0048                           |
| 1000|1.0582 | 0.5291| 0.2646| 0.1323| 0.0661 | 1.0024                           |
| 2500|1.0567 | 0.5284| 0.2642| 0.1321| 0.0660 | 1.0010                           |
| 10000|1.0559| 0.5280| 0.2640| 0.1320| 0.0660 | 1.0002                           |
| $\infty$| 1.0557| 0.5278| 0.2639| 0.1320| 0.0660 | 1.0000                           |