ABILITY ESTIMATION FOR CONVENTIONAL TESTS

JWA K. KIM
MIDDLE TENNESSEE STATE UNIVERSITY

W. ALAN NICEWANDER
UNIVERSITY OF OKLAHOMA

Five different ability estimators—maximum likelihood [MLE (θ)], weighted likelihood [WLE (θ)], Bayesian modal [BME (θ)], expected a posteriori [EAP (θ)] and the standardized number-right score [Z (θ)]—were used as scores for conventional, multiple-choice tests. The bias, standard error and reliability of the five ability estimators were evaluated using Monte Carlo estimates of the unknown conditional means and variances of the estimators. The results indicated that ability estimates based on BME (θ), EAP (θ) or WLE (θ) were reasonably unbiased for the range of abilities corresponding to the difficulty of a test, and that their standard errors were relatively small. Also, they were as reliable as the old standby—the number-right score.

Key words: item response theory, ability estimation, reliability.

Estimates of the latent trait (either ability or proficiency) are necessary as test scores in adaptive test procedures based on item response theory (IRT). For conventional tests, ability estimates offer the possibility of an invariant test score (i.e., a score that does not change according to the test difficulty). Unlike classical test theory, in which true scores for persons vary with the scaling of the items, IRT provides a score for a person that does not depend on item difficulty. The focus of the present inquiry is the feasibility of using ability estimates as scores for conventional tests. Since ability estimates do provide the possibility of the desirable characteristics discussed above, and since these estimates are being provided by some major testing companies, the accuracy of ability estimates needs to be carefully scrutinized. In our investigation, we compared five ability estimates in terms of their reliability, standard error and degree of bias—the methods compared were the maximum likelihood estimate [MLE (θ)], the weighted likelihood estimate [WLE (θ)], the Bayes modal estimate [BME (θ)], the expected a posteriori estimate [EAP (θ)] and the standardized number-right score [Z (θ)] (the standardization was done so that "number-right" had the same mean and variance as true θ).

Methods for Estimating θ

The theoretical basis for the comparison of ability estimators is provided by the three-parameter logistic (3-PL) model, in which $P_j(θ)$, the probability of answering the $j$-th item correctly is

$$P_j(θ) = c_j + \frac{(1 - c_j)}{1 + \exp(-1.7a_j(θ - b_j))}, \quad (1)$$

Requests for reprints should be sent to Jwa K. Kim, Department of Psychology, Middle Tennessee State University, Murfreesboro, Tennessee 37132.
where for the $j$-th item, $a_j$ is the slope parameter, $b_j$ is the location or difficulty parameter, $c_j$ is the lower asymptote or guessing parameter and $\theta$ is ability parameter for a specific examinee.

Several estimates of $\theta$ have been proposed over the past two decades. The MLE ($\theta$) (Birnbaum, 1968), BME ($\theta$) (Samejima, 1969), and EAP ($\theta$) (Bock & Aitkin, 1981) constitute the major estimation methods. More recently, Warm (1989) proposed the WLE ($\theta$) method which unlike the other methods, yields unbiased estimates of ability. Each of these methods is detailed below.

**Maximum Likelihood Estimator.** For the three-parameter logistic model, if item parameters are assumed known, the ability of each examinee can be estimated from the likelihood function,

$$L = L(u|\theta) = \prod_{j=1}^{n} P_j^{u_j} Q_j^{1-u_j},$$

where $u$ is the vector of observed item responses; $u_j = 1$ if item $j$ is correctly answered, and 0 if item $j$ is incorrectly answered. Let $P_j = P_j(\theta)$, and $Q_j = 1 - P_j$. If $\ln (L)$ is the log likelihood, then

$$\ln (L)' = \sum \frac{P_j'(u_j - P_j)}{P_j Q_j},$$

where $P_j$ and $\ln (L)'$ are the partial derivatives of $P_j$ and $\ln (L)$ with respect to $\theta$. The maximum likelihood estimate of $\theta$, MLE ($\theta$), can be obtained by setting (3) equal to zero and solving for $\theta$ using the Newton-Raphson method or some other suitable numerical strategy.

**Bayesian Modal Estimator.** If the prior distribution of ability is represented as $\phi(\theta)$, then the marginal distribution of the item responses is given by

$$L(u) = \int_{-\infty}^{+\infty} L(u|\theta) \phi(\theta) \, d\theta,$$

where $L(u|\theta)$ is the conditional likelihood given in (2). The posterior distribution of $\theta$, given $u$, can be derived as

$$p(\theta|u) = \frac{L(u|\theta) \phi(\theta)}{\int_{-\infty}^{+\infty} L(u|\theta) \phi(\theta) \, d\theta},$$

using Bayes' Theorem. The mode of this distribution is the Bayesian modal estimation of ability, BME ($\theta$), and was first presented by Samejima (1969).

**Expected A Posteriori (EAP) Estimator.** Bock and Aitkin (1981) introduced the EAP ($\theta$) estimator which is the mean of the posterior distribution of $\theta$. The mean of the posterior distribution may be expressed as

$$\text{EAP}(\theta) = \frac{1}{\int_{-\infty}^{+\infty} L(u|\theta) \phi(\theta) \, d\theta} \int_{-\infty}^{+\infty} L(u|\theta) \phi(\theta) \theta \, d\theta.$$