Supplementary Figure 1. Path Diagram for the Multilevel Rasch Model

Path diagram for multilevel Rasch model: Item response data can be considered as level 0, nested within respondents with covariate information such as demographic or clinical data considered in level 1 that can be further nested in groups considered in level 2. Higher levels may be needed to model complex data structure which is common in survey research. The levels in the rectangular box illustrate the nesting of item observations in individuals and individuals in groups. Levels 1 and 2 constitute the multilevel model for $\theta_{ij}$. There are uncertainties involved at each level, i.e. at the level of observations, at the individual level and at the group level. Explanatory information $x_{ij}$ and $w_j$ at levels 1 and 2 explain variability in the latent abilities between individuals within groups and across groups respectively. The dotted inverse L-box describes the Rasch model where item parameters, $\eta_k, \gamma_y$ are not influenced by the nested data structure.
Model Codes

```r
model {
  for (i in 1:N) {
    for (k in 1:K) {
      Y[i,k] ~ dcat(p[i,k,1:C])
    }
  }
  for (c in 1:C) {
    q[i,k,c] <- exp((c-1)*(theta[i]-eta[k])-sum(gamma[1:c]))
    p[i,k,c] <- q[i,k,c]/sum(q[i,k,1:C])
  }

  ## Difficulty parameter
  for (k in 1:(K-1)) { eta[k] ~ dnorm(0,priore) }
  eta[K] <- (-sum(eta[1:(K-1)]))

  ## Threshold
  gamma[1] <- 0
  for (c in 2:(C-1)) { gamma[c] ~ dnorm(0,priorg) }
  gamma[C] <- (-sum(gamma[1:(C-1)]))

  for (i in 1:N) {
    mutheta[i] <- inprod(X[i,], beta[])
    theta[i] ~ dnorm(mutheta[i], priort)
  }

  for (i in 1:Q) { beta[i] ~ dnorm(0,priorb) }
}
```

For readers with limited knowledge of Bayes, WinBUGS software is recommended since it provides GUI for model building and MCMC sample post-processing. An elementary introduction to WinBUGS can be found in: Fryback D.G., Stout, N.K., Rosenberg, M.A. An elementary introduction to Bayesian computing using WinBUGS. International Journal of Technology Assessment in Health Care. 2001 Winter; 17(1):98-113.

For readers with advanced knowledge of Bayes and programing, C/C++ would be preferred by using Metropolis–Hastings within Gibbs algorithm.
Simulation Study

We simulated our data as follows: two independent covariates ($X_{11}, X_{12}$), a continuous variable data such as standardized age was drawn from standard normal distribution and a binary variable such as gender drawn from binomial distribution with equal probability of being male or female gender (i.e. probability 0.5). The association effects ($\beta_1, \beta_2$) of these two covariates with the latent visual functioning ability parameter were fixed for a range from -1 to 1 by steps of 0.5 (i.e. $\beta_k = -1, -0.5, 0, 0.5, 1$) and hence, these were considered as the “true” association effects for our simulated datasets. The calibration of nine item difficulty parameters, $\eta_k$ was fixed according to Table 3 of a study conducted by Ecosse L. Lamoureux et. al.,\textsuperscript{36} that performed a systematic evaluation of the reliability and validity of the visual functioning questionnaire (VF-11) using Rasch analysis that was later modified to nine items (VF-9) to tailor fit to the Asian population. The threshold parameters were generated from a normal distribution of mean 0 and standard deviation 1 once, then ordered and fixed. Hence, multinomial response data for each of the nine items with five response categories were then generated for a sample size of 300 with probability of response determined by the Andrich rating scale model with the model parameters specifications described above, to form our pseudo-visual functioning questionnaire (modified VF-9) data (i.e. $n = 300, k= 9$ and $y$ in the range of integers 1 to 5 for five response categories for each item resulted in 2,700 response data generated).

Analysis of the generated pseudo-visual functioning sample data was performed with our one-stage HB approach and the frequently used two-stage procedure. Based on 100 replicates for each pair of our specified “true” association effects (25 pairs of “$\beta_1, \beta_2$”), average association estimates and their standard errors from the one and two-stage approach were computed to assess their performance in comparison to our pre-specified “true” effects. Moreover, we performed another 200 simulations for continuous and categorical variable separately to investigate and compare the empirical power between one and two-stage approach. Finally, to demonstrate the effect of sample size on bias, another 200 simulations were performed for continuous variables only since continuous and categorical variables show similar bias pattern in the previous analysis (Figure 1).
Supplementary Table 1. Impact of Sample Size on Level of Bias: Comparison of Proposed One-Stage HB and Observed Two-Stage Analysis Framework from Simulation Results

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Beta=0.2</th>
<th></th>
<th>Beta=0.5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One-stage</td>
<td>Two-stage</td>
<td>One-stage</td>
<td>Two-stage</td>
</tr>
<tr>
<td>50</td>
<td>0.011</td>
<td>-0.023</td>
<td>-0.010</td>
<td>-0.098</td>
</tr>
<tr>
<td>100</td>
<td>0.001</td>
<td>-0.032</td>
<td>-0.006</td>
<td>-0.097</td>
</tr>
<tr>
<td>300</td>
<td>-0.001</td>
<td>-0.030</td>
<td>-0.002</td>
<td>-0.092</td>
</tr>
<tr>
<td>600</td>
<td>-0.001</td>
<td>-0.032</td>
<td>-0.003</td>
<td>-0.091</td>
</tr>
</tbody>
</table>

Bias estimated based on 200 simulations for each sample size
## Supplementary Table 2. Power Analysis: Comparison of Proposed One-Stage HB and Observed Two-Stage Analysis Framework from Simulation Results

<table>
<thead>
<tr>
<th>alpha level</th>
<th>N=100</th>
<th>N=300</th>
<th>Continuous Variable</th>
<th>Categorical Variable</th>
<th>Effect Size</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One-stage</td>
<td>Two-stage</td>
<td>One-stage</td>
<td>Two-stage</td>
<td>Beta=0.2</td>
<td>Beta=0.5</td>
</tr>
<tr>
<td>1%</td>
<td>0.245</td>
<td>0.250</td>
<td>0.970</td>
<td>0.970</td>
<td>One-stage</td>
<td>0.045</td>
</tr>
<tr>
<td>5%</td>
<td>0.490</td>
<td>0.505</td>
<td>0.995</td>
<td>0.995</td>
<td>Two-stage</td>
<td>0.150</td>
</tr>
<tr>
<td>10%</td>
<td>0.610</td>
<td>0.605</td>
<td>1</td>
<td>1</td>
<td>One-stage</td>
<td>0.225</td>
</tr>
</tbody>
</table>

Based on 200 simulations using K=9; C=5
Data represented as empirical power of Two-stage and One-stage estimates of Beta, where the true model is given by Beta*Continuous or Beta*Categorical