

## ON THE CONSISTENCY OF Q-MATRIX ESTIMATION: A REJOINDER

JIMMY DE LA TORRE AND CHIA-YI CHIU

RUTGERS, THE STATE UNIVERSITY OF NEW JERSEY

This rejoinder responds to the commentary by Liu (Psychometrika, 2015) entitled “On the consistency of Q-matrix estimation: A commentary” on the paper “A general method of empirical Q-matrix validation” by de la Torre and Chiu (Psychometrika, 2015). It discusses and addresses three concerns raised in the commentary, namely the estimation accuracy when a provisional Q-matrix is used, the consistency of the Q-matrix estimator, and the computational efficiency of the proposed method.

Key words: cognitive diagnosis, G-DINA, Q-matrix, validation, MMLE.

### 1. Overview

Liu (2015), in his commentary on the Q-matrix validation method proposed by de la Torre and Chiu (2015), raised several concerns that include (a) the accuracy of the posterior distribution  $w$  and the item response probabilities  $p(\alpha) = P(Y = 1|\alpha)$ , as they depend on the provisional Q-matrix  $Q_0$ , which may or may not be correctly specified, (b) the consistency of the estimated  $\hat{Q}$ , and (c) the computational efficiency of the algorithm. We appreciate the opportunity to further explain in detail the proposed method. This rejoinder responds to these concerns by providing some insights and discussion of the three issues.

### 2. The Accuracy of the Posterior Distribution $w$ and the Item Response Probabilities $p(\alpha)$

It can be shown that, when the number of misspecifications in the Q-matrix is small to moderate, as assumed in de la Torre and Chiu’s (2015) study and as is the case with most real assessments, the accuracy of the estimated  $w$  would not be significantly affected by using the provisional  $Q_0$ . In the maximum likelihood estimation procedure, the likelihood is used in a relative, not absolute sense; thus,  $w$  can be well recovered by the item parameter estimates of other, correctly specified items. In other words, it might be the case that the estimates of  $p(\alpha)$  for some misspecified items deviate from the true probabilities, but when the number of misspecified q-entries is small, the whole likelihood is sufficiently robust to produce a posterior  $w$  that is very similar to, if not the same as, that arrived at using the correct Q-matrix  $Q^*$ . Of course, this does not guarantee that some unexpected peculiar combinations of the misspecifications would not result in a larger-than-expected distortion of  $w$ .

### 3. The Consistency of $\hat{Q}$

Given the above notion (i.e., that the estimation of  $w$  will not be significantly affected by a slightly misspecified Q-matrix compared with the correct Q-matrix), it is questionable whether the

Correspondence should be made to Jimmy de la Torre, Department of Educational Psychology, Rutgers, The State University of New Jersey, 10 Seminary Place, New Brunswick, NJ 08901, USA. Email: j.delatorre@rutgers.edu

hypothesis  $P(Q_\infty = Q^*) \rightarrow 1$  is always true, because from some point on along the iterations,  $\hat{Q}$  may result in an estimated  $w$ , and therefore  $\hat{Q}$ , that can no longer be updated despite the fact that  $\hat{Q} \neq Q^*$ . In the commentary, the author introduced the concept of “equilibrium point” to rule out the inconsistent cases discussed above. It was mentioned that it is not difficult to construct examples where  $Q_0$  converges to the equilibrium points. It would have been helpful if such examples had been included so that readers could learn more about the asymptotic behavior of  $\hat{Q}$ , and the relation between  $\hat{Q}$  and the equilibrium points in the context of cognitive diagnosis. Given the possibility of converging to different stationary points, not surprisingly, additional conditions are needed to make the consistency hold. However, it is unclear from the commentary what these additional conditions might be, and whether it is even possible to establish such conditions. The conjecture that “one conceivable key condition would be the closeness of  $Q_0$  to  $Q^*$ ” in the commentary was established without additional elaboration. It seemed to point to the direction of the convergence of  $Q_0$  to  $Q^*$  rather than to the fact that the equilibrium points can be established if  $Q_0$  is within a small neighborhood of  $Q^*$ , which may or may not be true. At present, there exists no systematic research indicating that, when the true q-vector is (11100), recovering the true specification from (11110) is easier compared with (10000), although (11110) is “closer” to (11100) than (10000) in terms of distance. This issue is further exacerbated when there exist other misspecifications in the Q-matrix.

#### 4. Computational Efficiency of the Algorithm

In the commentary, it was stated that, when the effects of attributes  $1 : (K' - 1)$  are small or  $\alpha_{1:(K'-1)}$  is well predicted by  $\alpha_{K':K''}$ ,  $\varsigma_{K':K''}^2 > \epsilon \varsigma_{1:K}^2$  can be true for a fixed  $\epsilon$ . This raises the question of whether the procedure that de la Torre and Chiu (2015) proposed can only produce one type of errors—underestimations of the true q-vectors. Also, if in fact the effects of attributes  $1 : (K' - 1)$  are small or a more compact subset of  $\alpha$  can be found, should one deem  $\alpha_{1:(K'-1)}$  necessary? It was suggested in the commentary that  $\epsilon = 1 - (\log n)^{-1}$  should be chosen instead of a prefixed  $\epsilon$  to overcome possible underestimation of the q-vector. Although we agree that the use of a prefixed  $\epsilon$  may not necessarily be ideal across various sample sizes, it is not clear that setting  $\epsilon = 1 - (\log n)^{-1}$  would be the solution. De la Torre and Chiu (2015) found the current cutoff of  $\epsilon = 0.95$  to be optimal across the conditions examined in their study. To match this cutoff using  $\epsilon = 1 - (\log n)^{-1}$ , the sample size needs to be in the vicinity of  $n = 5 \times 10^8$ . However, in practical testing situations, it would be more realistic to find  $n$  between 200 and 2000, which translates to  $\epsilon$  between 0.81 and 0.87. In the context of the de la Torre and Chiu (2015) study, it is a concern that these cutoffs may be suboptimal.

Finally, although not related to the three main issues above, we would like to address the issue of whether or not the estimation of  $p(\alpha)$  is independent of the fitted cognitive diagnosis model (CDM). The claim that “[the]  $p(\alpha)$  for a particular item is not based on the estimated G-DINA model” may need some elaboration. This is because the posterior distribution  $p(\alpha|Y_i)$  would definitely change depending on which CDM is used, except when a more general model is used in place of a true reduced model. Given that  $p(\alpha)$  is a function of  $p(\alpha|Y_i)$ , using different CDMs, some of which may be incorrect, can be expected to have an impact on the estimation of  $p(\alpha)$ .

#### References

- de la Torre, J., & Chiu, C.-Y. (2015). A general method of empirical Q-matrix validation. *Psychometrika*. doi:10.1007/s11336-015-9467-8.
- Liu, J. (2015). On the consistency of Q-matrix estimation: a commentary. *Psychometrika*. doi:10.1007/s11336-015-9487-4.

*Manuscript Received:* 29 JAN 2016  
*Published Online Date:* 23 JUN 2016