Published in final edited form as:

Qual Life Res. 2020 October ; 29(10): 2609–2610. doi:10.1007/s11136-020-02526-1.

Letter to Editor Regarding “RespOnse Shift ALgorithm in Item response theory (ROSALI) for response shift detection with missing data in longitudinal patient-reported outcome studies”

Heather J. Gunn

Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, USA

Dear Editor,

I read with interest the article by Guilleux and colleagues [1] published in your journal. The authors proposed a user-implemented algorithm, ROSALI, for detecting response shift across time using item response theory (IRT) models. However, I noticed four issues with the study, which calls into question the validity of the results and algorithm.

First, it is unclear why Step 0 is a valid or necessary step. In this step, the authors estimate two models according to Figure 1: a PCM at time 1 to obtain the difficulty parameters and a longitudinal PCM to obtain the latent variable variances. The authors do not provide justification for this step nor reference any research that has used this step to estimate an IRT model. The authors admit that the difficulty parameter estimates are just that, estimates. Fixing the difficulty parameters (δ) in Model 1 and Model 2 to equal the difficulty parameters in Step 0 does not allow for estimation of the standard errors of these parameter estimates. Thus, the test is not if parameters are equal across time, but if a parameter at time 2 is equal to a particular value, which is a different test. I am concerned that this will lead to erroneous conclusions of difficulty parameter differences across time (i.e., recalibration) and erroneous conclusions of the true change in the latent variable since the mean of the latent variable is affected by the difficulty parameters. As for the second model of Step 0, this model is overly strict. The discrimination parameters at both time points are equal to 1 and the latent variable variances across time are constrained to be equal. If the latent variable variance and/or any of the discrimination parameters are truly different across time, then this will bias the estimate of the latent variable variance.

Second, ROSALI specifies users to calculate a likelihood ratio test (LRT) to compare Model 1 and Model 2. LRTs require the models being compared to be nested. Model 2 is nested in Model 1 if Model 2 can be derived from Model 1 by constraining parameters in Model 1

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Correspondence concerning this article should be addressed to Heather J. Gunn, UCLA Global Center for Children and Families, 10920 Wilshire Boulevard, STE 350, Los Angeles, CA, 90024. HGunn@mednet.ucla.edu.

Publisher's Disclaimer: This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

without freeing any parameters. At first glance, it appears that the two models are not nested. When moving from Model 1 to Model 2, the parameters that change are:

1. latent variable mean (μ) at time 2 freed,
2. latent variable variance (σ) at time 2 freed,
3. change in item difficulties (η) at time 2 constrained, and
4. discrimination parameters (α) at time 2 constrained.

The LRT for this specific comparison works because Model 1 produces equivalent fit to a model that is clearly nested in Model 2 (i.e., a model that identifies the scale of the latent variable at time 2 by using an anchor item). To clarify, the latent variable variance at time 2 is fixed to equal the latent variable variance at time 1 at Step 1 only to identify the model, but the variances are not equal for any other tested models. To avoid confusion, Model 1 should be identified in the same way as Model 2 (e.g., use the same anchor item to link the time points in both Model 1 and Model 2). An added benefit to doing so is that the parameter estimates can be compared across models.

Third, the ROSALI results in the illustrative example are inaccurate. The authors present the results of the final model based on their algorithm in Table 3. As presented, this model is not identified; thus, the results and conclusions based off ROSALI for the applied example are not trustworthy. Every latent variable in an IRT model needs to be identified by either constraining a discrimination parameter or the variance of the latent variable to a value (usually 1). The results presented in Table 3 show that none of the discrimination parameters nor the variance of the latent variable are constrained to a value at time 2. Further, there are no equality constraints across time on any of those parameters. Thus, this model is not identified because the scale of the latent variable at time 2 is not established. As part of the algorithm, the authors state that the discrimination parameters at time 2 can be released from equality with time 1 “until there are no more items displaying reprioritization remaining.” However, this is not true because at least one item is needed as an anchor item to identify the model. Further, users of the algorithm should be made aware of the limits of partial measurement invariance. There is not a clear rule as the purpose of the measure drives how much partial invariance should be tolerated [2], but it is hard to argue that latent variables can be validly compared if a majority of items are non-invariant.

Fourth, the treatment and discussion of missing data handling is confusing and perhaps erroneous. The authors discuss that Rasch-based IRT models have the property of specific objectivity, but the algorithm they propose allows the discrimination parameters to vary over time. If the discrimination parameters are not all 1 (e.g., there is partial measurement invariance), does the specific objectivity property still hold? If not, then the ROSALI algorithm does not produce unbiased estimates if data are MNAR and this property cannot be considered an advantage. Further, the authors handle missing data differently depending on whether they are using the IRT or the SEM procedure. Because they use a robust maximum likelihood estimator for the SEM procedure, it is unclear why they also use mean imputation for missing item scores. No imputation is needed when using maximum

likelihood, which uses all of the observed information to provide unbiased estimates when the data are MCAR or MAR.

Given these concerns, I argue that the results of the applied example are incorrect for both procedures and ROSALI should not be used as described. Instead, readers should refer to Liu and colleagues [3] for an SEM-based procedure to test longitudinal invariance/response shift for categorical variables.

Acknowledgments

The first author's work on this paper was supported by a grant from the National Institute of Mental Health (T32MH109205).

References

1. Guilleux A, Blanchin M, Vanier A, Guillemin F, Falissard B, Schwartz CE, ... Sébille V (2015). RespOnse Shift ALgorithm in Item response theory (ROSALI) for response shift detection with missing data in longitudinal patient-reported outcome studies. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care, and Rehabilitation*, 24(3), 553–564.
2. Millsap RE, & Kwok O-M (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, 9, 93–115. doi:10.1037/1082-989X.9.1.93 [PubMed: 15053721]
3. Liu Y, Millsap RE, West SG, Yun-Tein J, Tanaka R, & Grimm KJ (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. doi:10.1037/met0000075 [PubMed: 27213981]