

Introduction to special section on response shift at the item level

Carolyn E. Schwartz^{1,2}

Accepted: 15 April 2016 / Published online: 28 April 2016
© Springer International Publishing Switzerland 2016

Methodological developments for detecting response shift have enabled increasingly sophisticated analyses over the past decade. While most response shift research has focused on effects on patient-reported outcome (PRO) subscales, the emergence of item response theory applications to PRO research led to the following questions: Does response shift at the item level have particular importance or significance, and is it worthwhile to address such effects in addition to subscale-level analysis? The International Society of Quality of Life Research (ISOQOL) Response Shift Special Interest Group (SIG) undertook an international collaboration to address these questions, and the resulting special issue is presented in this issue of *Quality of Life Research*.

In this special section, we present five scientific papers that address item-level response shift using three broad methodological approaches. These five papers have been selected through a competitive peer-review process. We first issued a call for papers from the ISOQOL Response Shift SIG and then accepted the top papers. Each paper underwent a peer-review process that required multiple iterations. Additionally, we implemented a survey of SIG members to elicit their input on what item-level response shift means and why it is important. We present the results of this survey briefly below, followed by a short introduction to the five papers included in this special section. This special section can be divided into method-based

papers that utilize one of three response shift detection methods: (1) the retrospective pretest (i.e., then-test) [1–4], (2) the Oort Structural Equation Modeling (SEM) approach [5], and (3) the RespOnse Shift ALgorithm in Item response theory (ROSALI) [6].

Survey results

Our SIG survey was implemented online in March 2015, and we received 20 responses with complete data from SIG members. Respondents came from 10 countries: from Europe (the Netherlands, Belgium, UK, France, Switzerland, and Germany), Africa (Nigeria, Uganda), and North America (Canada, USA). About half of the respondents identified themselves as response shift researchers, and the other half indicated that they were interested in response shift but had not yet done such research.

In response to our query about what response shift at the item level means, respondents stated that it concerns the relationship between item-level responses and the latent construct of interest, just as (sub)scale-level response shift concerns the relationship between (sub)scale scores and the latent construct of interest. Still, respondents felt that item-level response shift may provide more and complementary insight into the understanding of response shift.

In discussing the interpretation of item-level response shift, respondents stated that it is distinct from subscale-level response shift, is important, and is meaningful. The primary concern raised was that response shift effects at the item level may cancel each other out, making response shift effects seem negligible when they actually add noise to the data, thereby reducing statistical power. While some respondents expressed a concern that studying response shift at the item level would lead to a higher Type I error

✉ Carolyn E. Schwartz
carolyn.schwartz@deltaquest.org

¹ DeltaQuest Foundation, 31 Mitchell Road, Concord, MA 01742, USA

² Departments of Medicine and Orthopaedic Surgery, Tufts University School of Medicine, Boston, MA, USA

rate (e.g., false positives at the item level, false negatives at the subscale level), others noted that such item-level effects might bias relationships between covariates and scores, although the approach might be useful because it could identify particularly vulnerable items. Almost all respondents deemed both item- and subscale-level response shift effects relevant and distinct.

In interpreting item- versus subscale-level response shift effects, respondents suggested that recalibration response shift effects at the item- versus subscale-level may not have the same meaning from a methodological or conceptual perspective, and were concerned about effects cancelling each other out at the item level. In interpreting reprioritization response shifts at the item versus subscale level, respondents also felt that the effects may not have the same meaning from a methodological or conceptual perspective, and noted that item-level response shift effects provide less information about the latent variable being measured. In interpreting reconceptualization response shift effects, they noted a concern about items operating independently of one another and of items cancelling each other out. For both reprioritization and reconceptualization response shift effects, they noted a difficulty interpreting either outside of the context of other items.

In identifying measurement areas of particular relevance to item-level response shift effects, respondents suggested that any area where domains are covered by single items would be particularly relevant. This would include, for example, global health measures or symptom scales in randomized trial research. Such item-level effects were also deemed particularly relevant in intervention studies where outcomes tap aspects or activities that were part of the (intervention) training, in utility assessment which is often based on single items, and in psychometric studies focused on estimating item responsiveness of changing values of minimally important differences.

Item-level study using the then-test

For many years, the retrospective pretest was the primary method for detecting (recalibration) response shift effects. In this method, respondents are asked at the posttest to reevaluate their pretest level of functioning on selected items or subscales, with their current frame of reference. Difference scores (then-minus-pre, then-minus-post) were then used to quantify response shift and true change effects, respectively. Originating from work in educational and management sciences research, this approach gained popularity because it was easy to implement and easy to analyze. However, a number of researchers have documented that the then-test is confounded with recall bias [7, 8] and reflects a number of cognitive processes in addition to

recalibration response shift effects [9, 10]. Tamineau-Bloem et al. [11] conducted qualitative studies and used cognitive interviewing to examine whether two key assumptions of the then-test approach are valid. Their findings further undermine the credibility of the then-test.

Item-level studies using the Oort SEM approach

One of the most frequently applied methods for detecting response shift over the past decade, the Oort SEM approach [5, 12], has the advantage of being codified and interpretable using available software. In this special section, three research teams apply and demonstrate the application of the Oort SEM at both the item and subscale levels in distinct patient populations. Nolte et al. [13] apply the Oort SEM method to examine item-level response shift effects in psychosomatic inpatients during their hospital stay. Gandhi et al. [14] apply the method to a pediatric sample with asthma and examine the impact of response shift effects on measurement bias [15]. Verdam et al. [16] apply the Oort SEM to data from cancer patients and demonstrate how to apply the method to discrete item responses, using SF-36 data as an example.

Item-level study using the ROSALI approach

While the Oort SEM method is useful for examining response shift effects at both the subscale level (with linear relationships) and the item level (with linearized relationships), the ROSALI approach attempts to investigate item-level response shifts using item response models with logistic relationships. In their contribution to this special section, Blanchin et al. [17] develop a statistical method for identifying response shift effects at the individual level using Guttman errors to identify discrepancies in respondent's answers to items compared to an expected response pattern. They then created two patient groups (showing or not showing discrepancies) and applied the ROSALI algorithm to investigate recalibration and reprioritization response shift effects in both groups.

As a group, this set of five articles demonstrates novel and useful developments in response shift methods that can be applied in the field of PRO research. Both the results of our survey input and the papers that are presented in this special section imply that the issue of item-level response shift is worthy of further research. When studying response shift effects, it might be worthwhile to check both item- and subscale-level response shift effects in PRO data. Further methodological development might enable simultaneous study of subscale- and item-level response shift, by combining the linear, linearized, and nonlinear

relationships of the Oort SEM and ROSALI approaches. New research may also use mixed methods, for example enriching quantitative findings with qualitative data [10, 18] or using data mining approaches as an exploratory first step, prior to using quantitative methods to test formal hypotheses (e.g., [19]). It is our hope that bringing together this eclectic set of papers will stimulate further creative and rigorous work in the emerging field of item-level response shift research.

Acknowledgments We gratefully acknowledge Dr. Frans Oort for helpful comments on an earlier draft of this manuscript.

References

- Howard, G. S., & Dailey, P. R. (1979). Response shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, *64*(2), 144–150.
- Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. W., & Gerber, S. K. (1979). Internal invalidity in pretest–posttest self-report evaluations and a re-evaluation of retrospective pretests. *Applied Psychology Measurement*, *3*(1), 1–23.
- Sprangers, M., & Hoogstraten, J. (1989). Pretesting effects in retrospective pretest–posttest designs. *Journal of Applied Psychology*, *74*(2), 265–272.
- Sprangers, M. (1988). *Response shift and the retrospective pretest: On the usefulness of retrospective pretest-posttest designs in detecting training related response shifts*. Gravenhagen: het Instituut voor Onderzoek van het Onderwijs S.V.O.
- Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Quality of Life Research*, *14*, 587–598.
- Guilleux, A., Blanchin, M., Vanier, A., Guillemin, F., Falissard, B., Schwartz, C. E., et al. (2015). Response shift algorithm in item response theory (ROSALI) for response shift detection with missing data in patient-reported outcomes in longitudinal clinical trials. *Quality of Life Research*, *24*(3), 553–564.
- Schwartz, C. E., Sprangers, M. A. G., Carey, A., & Reed, G. (2004). Exploring response shift in longitudinal data. *Psychology and Health*, *19*(1), 51–69.
- Ahmed, S., Mayo, N. E., Corbiere, M., Wood-Dauphinee, S., Hanley, J., & Cohen, R. (2005). Change in quality of life in people with stroke over time: True change or response shift? *Quality of Life Research*, *14*, 611–627.
- Schwartz, C. E., & Sprangers, M. A. G. (2010). Guidelines for improving the stringency of response shift research using the then-test. *Quality of Life Research*, *19*, 455–464.
- Schwartz, C. E., & Rapkin, B. D. (2012). Understanding appraisal processes underlying the then-test: A mixed methods investigation. *Quality of Life Research*, *21*(3), 381–388.
- Taminiau-Bloem, E. F., Schwartz, C. E., van Zuuren, F. J., Koeneman, M. A., Visser, M. R. M., Tishelman, C., et al. (2016). Using a retrospective pretest instead of a conventional pretest is replacing biases: A qualitative study of cognitive processes underlying responses to then-test items. *Quality of Life Research*. doi:10.1007/s11136-015-1175-4.
- Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2005). An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Quality of Life Research*, *14*, 599–609.
- Nolte, S., Mierke, A., Fischer, H. F., & Rose, M. (2016). On the validity of measuring change over time in routine clinical assessment: A close examination of item-level response shifts in psychosomatic inpatients. *Quality of Life Research*. doi:10.1007/s11136-015-1123-3.
- Gandhi, P. K., Schwartz, C. E., Reeve, B. B., DeWalt, D. A., Gross, H. E., & Huang, I.-C. (2016). An item-level response shift study on the change of health state with the rating of asthma-specific quality of life: A report from the PROMIS[®] Pediatric Asthma Study. *Quality of Life Research*. doi:10.1007/s11136-016-1290-x.
- Oort, F. J., Visser, M. R. M., & Sprangers, M. A. G. (2009). Measurement and conceptual perspectives on response shift: Formal definitions of measurement bias, explanation bias, and response shift. *Journal of Clinical Epidemiology*, *62*, 1126–1137.
- Verdam, M. G. E., Oort, F. J., & Sprangers, M. A. G. (2016). Using structural equation modeling to detect response shifts and true change in discrete variables: An application to the items of the SF-36. *Quality of Life Research*. doi:10.1007/s11136-015-1195-0.
- Blanchin, M., Sébille, V., Guilleux, A., & Hardouin, J.-B. (2016). The Guttman errors as a tool for response shift detection at subgroup and item levels. *Quality of Life Research*. doi:10.1007/s11136-016-1268-8.
- Tsai, T. I., Morisky, D. E., Kagawa-Singer, M., & Ashing-Giwa, K. T. (2011). Acculturation in the adaptation of Chinese-American women to breast cancer: A mixed-method approach. *Journal of Clinical Nursing*, *20*(23–24), 3383–3393. [Research Support, Non-U.S. Gov't].
- Li, Y., & Schwartz, C. E. (2011). Using classification and regression tree to examine evidence of quality of life response shift in patients with multiple sclerosis. *Quality of Life Research*, *20*(10), 1543–1553.