Authors' Response to: Alternatives to principal components analysis to derive asset-based indices to measure socio-economic position in low- and middle-income countries: the case for multiple correspondence analysis

From DAVID GORDON, LAURA D HOWE,* BRUNA GALOBARDES, ALICIA MATIJASEVICH, DEBORAH JOHNSTON, OBINNA ONWUJEKWE, RITA PATEL, ELIZABETH A WEBB, DEBBIE A LAWLOR and JAMES R HARGREAVES

*Corresponding author: MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK. E-mail: Laura.Howe@bristol.ac.uk

Downloaded from https://academic.oup.com/ije/article/41/4/1209/690658 by guest on 16 August 2022

We welcome the comments of Traissac and Martin-Prével and share their concerns about the use of principal components analysis (PCA) to derive weights for asset-based household welfare indices. We are in complete agreement that the use of PCA for this purpose is problematic and that several other potentially better methods are available, such as multiple correspondence analysis, non-linear canonical correlation analysis and latent trait analysis. We did not discuss in detail these alternatives and the related issue of item response theory, because the article aimed to provide a broader overview of the topic of measuring socio-economic position in low- and middle-income country study populations. We welcome the opportunity to discuss these issues further here.

The 'problem' of weights

It is a self-evident truth that some assets and/or deprivations are more important than others, for example, it is much worse not to be able to afford to feed or clothe your children adequately than not to be able to buy a bicycle. Thus, many researchers wish to calculate differential weights for asset indices to reflect the differences in standard of living implied by possession (or lack of possession) of each of the items in the index. It seems intuitively obvious that this is a desirable thing to do—hence, the use of PCA and other methods to derive differential weights for asset index components.

Unfortunately, statistical theory and intuitive truths do not always coincide. Classical test theory is the philosophical model that underlies the development of tests and measurement indices across the natural, medical and social sciences. Classical test theory assumes that there are an infinite (or large) number of asset/deprivation measures. If we could have answers to this infinite number of questions about assets, then we would have 'perfect knowledge' (we would know everything) about each person's assets. No set of weights could add any additional information, as we would already know everything, that is, the infinite asset index is self-weighting.

The square root of the Cronbach's alpha statistic can be considered to be the correlation between the asset index and the 'perfect' index made from the answers to the infinite set of asset questions.[1] The standard advice in statistical textbooks is that a 'good' test/index should aim to achieve a Cronbach's alpha of 0.8, (or at least 0.7).[2] Thus, the correlations with the perfect infinite set of asset indicators and a 'good' asset index would be 0.9 (or higher), so there is little additional information that any differential weights could add. Even if perfect error-free differential weights could be developed, the results from a 'good' unweighted asset index and a weighted asset index would be essentially identical.

Thus, we agree with Kline[3] who argues that:

> While much effort goes into discussing and determining differential item weights, Ghiselli, Campbell, and Zedek[4] (1981) are persuasive in arguing that differential item weighting has virtually no effect on the reliability and validity of the overall total scores. Specifically, they say that "empirical evidence indicates that reliability and validity are usually not increased when nominal differential weights are used" (p. 438). The reason for this is that differential weighting has its greatest impact when there (a) is a wide variation in the weighting values, (b) is little intercorrelation between the items, and (c) are only a few items. All three are usually the opposite of what is likely to occur in test development. That is, if the test is developed to assess a single construct, then if the developer has done the job properly, items will be intercorrelated. As a result, the weights assigned to one item over another are likely to be relatively small. In addition, tests are often 15 or more items in length, thus rendering the effects of differential weighting to be minimized. Finally, the correlation between weighted and unit-weighted test scores is almost 1.0. **Thus, the take-home message is pretty simple—don't bother to differentially weight items. It is not worth the effort**.[3] (Emphasis in the original.)

Item response theory

It is possible to obtain useful additional information on the properties of each individual item in an asset-based household welfare index using item response theory (IRT) models. Classical test theory provides information on the reliability of a scale/asset index as a whole, whereas IRT provides additional information on the reliability of each individual item in the scale/index. IRT, also known as latent trait analysis, is a set of statistical models that describes the relationship between a person's response to questionnaire items and an unobserved latent trait, such as knowledge of biology, level of happiness or amount of wealth. This is generally shown with item characteristic curves, which are modelled by the main parameters (ability and discrimination). It is often used for the selection of questions in assessment and for psychological testing. It has also been used for the measurement of poverty.[5–9]

The IRT model assumes that 'wealth' is an unobservable latent trait that cannot be measured directly, like say height or weight, as it is a concept rather than a physical entity. However, it is assumed that this concept of 'wealth' can be measured indirectly using social survey questions about the respondent's ability to afford certain consumer durables and household items. The amount of wealth measured by an asset item in an IRT model is defined by the likelihood that the person/household will possess/lack that item; thus, it is desirable that an asset indicator should include items with a range of different 'wealth' scores, that is, some low-wealth items, some medium-wealth items and some high-wealth items. IRT models produce what are termed 'ability' scores, which could be used as differential item weights in an asset-based household wealth index—if such weights are required (see discussion above). In two parameter IRT models, a 'discrimination' score is calculated for each index item. The discrimination of an asset item measures how well this item differentiates between the wealthy and the poor; thus, high 'discrimination' scores are desirable. The purpose of an asset indicator is to measure how much of the latent trait 'wealth' a person/household has achieved.

If differential weights are required for an index, then IRT models can provide a framework for calculating robust theory-driven weights, as long as the assumptions of the IRT model are not violated. It is assumed that the items included in the asset index measure only one latent trait (unidimensionality). There is no consensus on how unidimensionality may be established, but multiple correspondence analysis, cluster or factor analysis may give researchers an idea of the data structure. It should be noted that unidimensional IRT models are robust to moderate degrees of multidimensionality as defined by factor analyses, particularly where the dimensions are highly correlated and/or where the test/index length is >20 items and/or the sample size is >250.[10] Local independence is also an important assumption, that is, responses to a test item do not depend on other test item responses once trait is taken into account. This is an assumption shared by most statistical models and is partly a reflection of the unidimensionality assumption.

Conclusion

We would like to thank Traissac and Martin-Prével for highlighting the important issue of the problematic use of PCA for constructing and weighting asset-based household welfare indices. In the 1970s, Dutch researchers at the Social and Cultural Planning Office developing the Leefsituatie (life situation) index used non-linear canonical correlation analyses to derive weights after identifying the problematic nature of PCA derived weights.[11,12] Advances in statistical methods and computing power have now made a range of robust weighting methods widely available, such as multiple correspondence analysis and latent trait analysis.

However, we would caution researchers developing a test or index to consider carefully whether differential item weights are likely to improve the accuracy and precision of their measure before weighting an index. Regardless of the choice of weighting method, a key issue is likely to be the selection of a broad and context-appropriate set of assets at the data collection stage.

References

1. Nunally JC. *Psychometric Theory*. 2nd edn. New York: McGraw-Hill, 1978.
2. Nunally JC, Bernstein IH. *Psychometric Theory*. 3rd edn. New York: McGraw-Hill, 1994.
3. Kline TJB. *Psychological Testing: A Practical Approach to Design and Evaluation*. London: Sage, 2005, p. 150.
4. Ghiselli EE, Campbell JP, Zedek S. *Measurement Theory for the Behavioral Sciences*. San Francisco: W.H. Freeman, 1981.
5. Cappellari L, Jenkins SP. Summarizing multiple deprivation indicators. In: Jenkins SP, Micklewright J (eds). *Inequality and Poverty: Re-examined*. Oxford: Oxford University Press, 2007, pp. 166–84.
6. Fusco A, Dickes P. The Rasch model and multidimensional poverty measurement. In: Kakwani N, Silber J (eds). *Quantitative Approaches to Multidimensional Poverty Measurement*. Chapter 3. Houndmills: Palgrave Macmillan, 2008, pp.49–62.
7. Guio AC, Gordon D, Marlier E. *Measuring Material Deprivation in the EU: Indicators for the Whole Population and Child-specific Indicators*. Luxembourg: EUROSTAT, 2012.
8. Martini MC, Vanin C. *A Measure of Poverty Based on the Rasch Model*. http://homes.stat.unipd.it/mgri/SIS2010/Program/contributedpaper/792-1494-1-DR.pdf.
9. Szeles MR, Fusco A. Item response theory and the measurement of deprivation: evidence from Luxembourg data. *Qual Quant* 2011;doi: 10.1007/s11135-011-9607-x [Epub 4 October 2011].
10. Kirisci L, Hsu TC, Yu L. Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Appl Psychol Meas* 2001;**25:**146–62.
11. Boelhouwer J, Stoop I. Measuring well-being in the Netherlands. The SCP-index from 1974 to 1997. *Soc Indic Res* 1999;**48:**51–75.
12. Boelhouwer J. *Wellbeing in the Netherlands. The SCP Life Situation Index Since 1974*. The Hague: The Netherlands Institute for Social Research, 2010.

doi:10.1093/ije/dys120
© The Author 2012; all rights reserved.

Downloaded from https://academic.oup.com/ije/article/41/4/1209/690658 by guest on 16 August 2022