

A conversation with Louis Tay, November 30, 2015

Participants

- Louis Tay – Assistant Professor of Industrial-Organizational Psychology, Purdue University
- Luke Muehlhauser – Research Analyst, Open Philanthropy Project

Note: These notes were compiled by the Open Philanthropy Project and give an overview of the major points made by Professor Tay.

Summary

The Open Philanthropy Project spoke with Professor Louis Tay as part of an investigation into measures of subjective well-being (SWB). Conversation topics included the relative scarcity of systematic reviews of the psychometric properties of subjective well-being scales, and potential ways to improve the measurement of SWB.

Relative scarcity of systematic reviews of the psychometric properties of SWB scales

The following are potential reasons why systematic reviews of the psychometric properties of particular SWB scales are rarely conducted:

1. It is relatively expensive to conduct studies that use some measures, such as Ecological Momentary Assessment (EMA) or Day Reconstruction Method (DRM) measures, as participants in such studies usually receive compensation to report on their SWB.
2. For some scales, reliability and validity have been firmly established, so there is little motivation to conduct a systematic review. For example, the reliability of the Satisfaction with Life Scale (SWLS) is well-established. A systematic review of its validity, however, could be valuable.
3. Few scales are used so commonly that a systematic review of their reliability and validity would be warranted. It is common for researchers to develop new scales or adapt older scales for their own purposes, which is a barrier to scale standardization.

Professor Tay is not aware of any tools (*a la* the Cochrane Risk of Bias tool) that provided a standardized way for evaluating the reliability and validity evidence for a particular measure of SWB although there are principles set forward by *The Standards for Educational and Psychological Testing*.

Efforts to standardize SWB measures

Despite the challenges involved in standardizing SWB measures, the Academy of Management Research Methods Division (<http://rmdiv.org/>) is attempting to

develop a “Measure Chest” in an attempt to encourage a standardized set of the best measures for use in the field.

One way to promote standardization is for a particular set of measures to be used widely by the largest polling organizations (e.g. Gallup) and by governments that maintain national accounts of SWB. Researchers would likely be inspired to adopt the same measures for comparability purposes, while also continuing to develop new or adapted scales in some cases.

Potential ways to improve the measurement of SWB

App for measuring SWB in daily life

Methods for measuring SWB in daily life could be improved. A smartphone app could be an effective way to deploy a daily-life SWB measure at a large enough scale that its psychometric properties can be evaluated and improved. Moreover, such an approach could lead to the development of a bank of SWB measurement items that could be correlated with other data, for example users’ health or activity metrics, or their use of social media.

Item Response Theory (IRT) would be the most effective model to use in the app’s development. It would employ the same technology that the College Board uses for standardized tests, such as the Scholastic Assessment Test (SAT) and the Graduate Record Examination (GRE).

Advantages of using IRT

IRT offers advantages over Classical Test Theory (CTT) for some purposes. For example:

1. IRT allows one to construct items that measure different levels of a construct. For example, one or more items on the scale could be particularly sensitive to the difference between medium-high levels of SWB and very high levels of SWB.
2. IRT allows scores to be equated over time and over societies, which results in more accurate comparisons. For example:
 - a. Testing companies might equate test scores over time to prevent them from fluctuating according to the difficulty of the test or how good the test takers were in a particular year.
 - b. Since IRT allows researchers to compute item parameters (e.g. item difficulties) that are independent of sample mean levels (e.g. country mean levels). This can be especially useful for cross-national comparisons of SWB. It might reveal, for example, that a particularly low mean score in one country was the result of item translation rather than a result of genuinely low sample mean levels of SWB.

An IRT model can be used both to create a new scale and to make inferences from raw data produced using a CTT-developed scale, but it is most powerful when it is used during scale development.

Adding IRT-developed items to the Gallup World Poll

One way to improve our understanding of cross-national comparisons of SWB would be to add relevant IRT-developed items to the Gallup World Poll. Gallup is willing to accept new items, but these efforts would likely encounter some challenges:

1. Most questions in the Gallup poll are single-item, whereas IRT measures require at least 3, and ideally many more than 3, items.
2. Adding even a single item question to the Gallup World Poll can be fairly expensive, because the Gallup World Poll is delivered to so many people worldwide.

Other people to talk to

- Edward Diener, Alumni Distinguished Professor of Psychology (Emeritus), University of Illinois at Urbana-Champaign
- Ulrich Schimmack, Associate Professor of Psychology, University of Toronto Mississauga
- Michael Eid, Professor of Psychology at the Free University of Berlin
- Johannes Eichstaedt, Founding Research Scientist of World Well-Being Project, University of Pennsylvania

All Open Philanthropy Project conversations are available at <http://www.givewell.org/conversations>