Evaluation applications of instrument calibration research findings in psychology for very small samples

W P Fisher Jr¹, P Petry²

¹Research Associate, BEAR Center, Graduate School of Education, University of California, Berkeley, CA USA
²Principal, PaulaPetry.com
Miami, FL USA
E-mail: wfisher@berkeley.edu

Abstract. Many published research studies document item calibration invariance across samples using Rasch’s probabilistic models for measurement. A new approach to outcomes evaluation for very small samples was employed for two workshop series focused on stress reduction and joyful living conducted for health system employees and caregivers since 2012. Rasch-calibrated self-report instruments measuring depression, anxiety and stress, and the joyful living effects of mindfulness behaviors were identified in peer-reviewed journal articles. Items from one instrument were modified for use with a US population, other items were simplified, and some new items were written. Participants provided ratings of their depression, anxiety and stress, and the effects of their mindfulness behaviors before and after each workshop series. The numbers of participants providing both pre- and post-workshop data were low (16 and 14). Analysis of these small data sets produce results showing that, with some exceptions, the item hierarchies defining the constructs retained the same invariant profiles they had exhibited in the published research (correlations (not disattenuated) range from 0.85 to 0.96). In addition, comparisons of the pre- and post-workshop measures for the three constructs showed substantively and statistically significant changes. Implications for program evaluation comparisons, quality improvement efforts, and the organization of communications concerning outcomes in clinical fields are explored.

1. Introduction
An extensive body of published research documents the methods, models and processes by which item calibration invariance across samples can be established [1-3]. Application of these kinds of calibration research results in workshop evaluation contexts may provide previously unavailable advantages for demonstrating value obtained. Workshop providers and clinical practitioners may find new degrees of scientific credibility in being able to demonstrate repeatable and reproducible effect sizes expressed in comparable units, and as a result of establishing baselines for their own internal benchmarking and improvement activities. Workshop sponsors and funding organizations will be better able to separate successful and unsuccessful programs, enabling them to more clearly define returns on investments. Workshop participants will have the means to assess their status on any given construct at any time, to better maintain momentum on the therapeutic trajectory, and may also become informed enough about

¹ To whom any correspondence should be addressed.
² To whom any correspondence should be addressed.
measurement to demand comparable outcome measures from competing workshop providers. Future research will focus on developing practical scaffolding from the therapeutic progressions defined by each scale’s invariant item hierarchy.

Participants participated in workshop programs designed to build awareness of the impact of misgivings, victim/bully/rescue cycles, and conscious engagement on their lives. The programs included leadership coaching sessions presented in a positive, uplifting context, to open participants’ perspectives to harmony and flow. Participants kept journals recording their observations of how misgivings and rescue dynamics impact their lives, and about the approaches they used to step into conscious engagement with aspects of life they had previously taken for granted.

The data analyzed here were originally collected with the intention of evaluating the effectiveness of a stress-reduction workshop, which was not funded by federal research money. The original evaluation was completed to the satisfaction of the sponsoring organizations. In the course of examining the data for the purposes of the providing the evaluation, some unexpected results emerged. These pre-existing data gathered for the evaluation were stripped of all identifying information, such that there was no responses could be linked back to the persons originally providing them through a key to a coding system or any other means. There being no interaction at all between the research analyst and any individual workshop participants, and as no identifiable private information was involved, this project did not require review by any committee or office for the protection of human subjects. (For more information, see, for instance, http://cphs.berkeley.edu/secondarydata.pdf.)

2. Method

Typical workshop evaluations ask participants to rate the quality and organization of the presentations, the presenters’ skills and knowledge, the efficiency of the registration process, and the comfort and cleanliness of the facility. Results are usually compiled by calculating percentages of responses indicating various levels of agreement with each item. Measures of the actual constructs intended to be impacted are not commonly obtained.

The Joyful Living Workshop evaluations, in contrast, were designed and conducted with the goals of establishing measures exhibiting consistent interpretability and meaningfulness, building on past research, and measuring change with enough sensitivity and power to obtain meaningful and statistically significant differences in individual comparisons of the pre- and post-workshop measures.

3. Samples

The Joyful Living Workshop program is designed to expand consciousness as to a productive sense of personal identity and to help shift away from misgivings, victim/bully/rescue cycles, and toward conscious engagement in everyday life. Several one-to-two day, and two-and-a-half day workshops were held over three-month periods in 2012 and 2014. Sixteen participants provided pre- and post-workshop measures on all three constructs in 2012, and 14 did in 2014.

In 2012, there were 19 total workshop participants, and 12 of the 15 who provided demographic information were employees of a Miami (FL) community resource center for parents of children with disabilities. Ages ranged from 20s to 60s, with the mode and median in the 40s. All 15 respondents identified themselves as Hispanic, 13 of the 15 were female, and education levels ranged from high school to doctorate, with the mode and median at completed college. The 2014 participants were of similar demographic composition.

4. Instruments

Rasch-calibrated instruments measuring the intended outcomes of the workshop (reduced depression, anxiety and stress, and enhanced joy for living as an effect of mindfulness practice) were identified in the research literature. These instruments were the 21-item Depression, Anxiety and Stress Scales (DASS) [4-6] and the Solloway Mindfulness Scale (SMS) [7,8].
5. Results
Correlations (not disattenuated) of the two sets of small-sample workshop item calibrations with the published values for the three scales range from 0.85 to 0.99. In addition, pre- and post-workshop item calibrations showed the same patterns of invariance (correlations (not disattenuated) range from 0.86 to 0.92). Finally, comparisons of the pre- and post-workshop measures for the three constructs showed substantively and statistically significant changes.

6. Discussion
There were five primary outcomes. Each of these is a significant innovation in its own right, and will be found in other workshop evaluations only rarely, if at all.

- First, three well defined scales were constructed by adapting previously researched instruments.
  
  Long experience with instruments of these kinds guided identification of tools likely to be both useful and well-based in evidence and strong theory. That experience also indicated what kinds of changes might prove workable.

  The vast majority of workshop evaluations are not based in previously researched instruments.

- Second, each of the adapted scales maintained the scientific qualities expected of it.
  
  The new small-sample calibrations retained their expected values when compared to their original published values (see the figure below).

  The pre-workshop scale values were invariant when compared with the post-workshop scales.

- Third, linear measures with known uncertainty values and internal consistency (model fit) indices were constructed.

- Fourth, measurement precision was sufficient to distinguish pre- and post-workshop measures to statistically significant degrees. (See Table 1.1.)

- Fifth, the substantive significance of the differences between the pre- and post-workshop measures is evident in the content of the survey items positioned on the scales between the two measures. (See Table 1.1.)

Generalized application of these results should anchor the items in each scale at established scale values, or could be converted to more intuitively accessible scale ranges, if connections to published values were maintained.

7. Conclusion
Application of Rasch instrument calibration research results in workshop evaluation contexts may provide previously unavailable advantages for demonstrating the value of results. Most workshop
evaluations do not maintain instrument comparability with published calibrations, nor do they experimentally establish a basis for comparing pre- and post-workshop measures. Workshop providers and clinical practitioners may find new degrees of scientific credibility in being able to demonstrate repeatable and reproducible effects, and as a result of establishing baselines for their own internal benchmarking and improvement activities. Workshop sponsors and funding organizations will be better able to separate successful and unsuccessful programs, enabling them to more clearly define returns on investments. Workshop participants will have the means to assess their status on any given construct at any time, to better maintain momentum on the therapeutic trajectory, and may also become informed enough about measurement to demand comparable outcome measures from competing workshop providers. Future research will focus on developing practical scaffolding from the therapeutic progressions defined by each scale’s invariant item hierarchy.

8. References

[1] Fisher W P Jr 1997 Physical Med & Rehabil State Art Revs 11 357-373
[2] Fisher W P Jr, Harvey R F, & Kilgore K M 1995 NeuroRehabil 5 3-25
[3] Fisher W P Jr 1999 J Louisiana St Med Soc 151 566-578
[4] Antony M M, Bieling P J, Cox B J, Enns M W, Swinson R P 1998 Psychol Assessment 10 176-181
[5] Lovibond S H, Lovibond P F 1995 Manual for the Depression Anxiety Stress Scales. 2nd edition Sydney Psychology Foundation of Australia
[6] Shea T L, Tennant A, & Pallant J F 2009 BMC Psychiatry 9 [doi:10.1186/1471-244X-9-21]
[7] Solloway S, & Fisher W P Jr 2007 International J Transpersonal Studies 26 58-81
[8] Solloway S, & Fisher W P Jr 2007 J Appl Meas 8 359-372
### MAP OF PERSON AND ITEM

**MEASURE**
- **person**
  - less
  - more

**ITEM**
- I see ordinary things as extraordinary.
- I have felt peaceful.
- I feel free to choose what my life will be like.
- I have self-control.
- There's been a lot to look forward to.
- I enjoyed being lighthearted.
- I've affirmed my value to others.
- I am aware that other people appreciate who I am.
- I experience life in new ways.
- Unexpected things brought me joy.
- I enjoy discovering amazing things in nature.
- I breathed freely.
- Life seems full of meaning.
- I find joy in ordinary things.
- I see how things change from moment to moment.

**EXPECTED RATINGS ON ITEMS AT THIS LEVEL**
- **preworkshop:** 20% S or VS agree
- **postworkshop:** 45% S or VS agree

**AVERAGE EXPECTED RATING HERE IS 2.4**

**AVERAGE POSTWORKSHOP MEASURE**

**INPUT:** 52 person 73 item reported 35 person 34 item 4 cat 8 winsteps 3 72 0