Introduction

Background

Questions in psychology are commonly set up to ask respondents to provide a response on a specified rating scale (Fur, 2011). Depending on the questionnaire item, ratings can be provided on an agreement scale, a frequency scale, or through meaningful numeric response options on a frequency scale for a specified time frame. The Center for Epidemiologic Studies Depression Scale (CES-D, Radloff, 1977) queries respondents regarding the frequency of depressive symptoms experienced in the previous week, providing both verbal as well as numeric labels for response options (See Table 1). When the questionnaire is translated into different languages, the consistency in the translation of the verbal response labels bears consideration. Collection of response process data may be informative with regards to respondents’ use of verbal and numeric labels, as well as equivalency of translation of verbal options.

This study examines English and Mandarin (Zhang et al., 2010) versions of the CES-D, and usage of numeric and verbal response labels among respondents from diverse language backgrounds (LB) using two post-questionnaire completion interview protocols.

Research Questions

RQ1. How do respondents utilize numeric, verbal, or both response labels? Does usage of labels vary for language background and version of the instrument? What reasons do respondents provide for use of numeric or verbal labels?

RQ2. Are the translations of the response labels on the English and Mandarin forms equivalent? If different, where are the differences? And, do different response process probing protocols yield different patterns?

Methods

Participants

Participants were undergraduate students (N = 106) with a mean age of 20 (SD=2.25); 58% were female. Due to intended recruitment strategy, most participants were non-native English speakers (79%) with 61% of Mandarin language background (LB).

Procedure and Measure

Participants completed the CES-D, followed by a post-questionnaire completion interview yielding information pertaining to their response label usage, reasons for their decision, and demographic information including language background. 60 participants completed the English form. An additional forty participants completed both the English and Mandarin forms.

Among the 40 participants who completed both forms, half were probed regarding the translations of the response labels verbally with the questionnaires in front of them (Protocol 1) and half were probed with the additional usage of a graphic (Protocol 2) (see Fig 1).

CES-D: The CES-D is a 20 item measure of symptoms of depression over the past week. Response options are on a 4-point scale.

Focal Post-Questionnaire Completion Interview Transcript Coding

Coders examined transcripts individually to identify possible themes for focal questions, and worked collaboratively on the thematic coding.

1. Which RL Used/Helpful (Numeric, Verbal, or Both)
2. Reasons for Choice (e.g., “Days More Specific”)
3. Equivalence/Comparability of Translation of RL Options (e.g., “MR < ER”, “MS = ES”)

Definitions of codes and themes with exemplar quotes are provided.

Analysis

Analysis of coded interviews were conducted. Frequency, chi-square analysis, and tests on means were conducted as appropriate. Analyses were considered at both 0.05, as well as reduced alpha levels.

Results

RQ1 – Use/Helpfulness of Numeric and/or Verbal RL. When completing the English form, 67% of Native English (NE) speakers indicated the numeric RL (NRL) as more helpful; in contrast, Non-Native English (NNE) speakers indicated greater use of the verbal RL (VRL) on both English (78%) as well as on Mandarin (78%) forms (See Fig 2.). Respondents’ interviews indicate different reasons (e.g., “Days More Specific”) for their use of numeric or verbal labels (see Fig 3); exemplar quotes are provided.

RQ2 – Comparability of Translations of Verbal RL option pairs (e.g., ER vs MR, ES vs MS). Table 2 illustrates the number of option pairs that were perceived as comparable overall and by protocol, with more pairs being perceived as comparable under the protocol without the graphic aid. For the first option pair (ER, MR), if perceived as different, 92% respondents indicated MR < ER; for the second pair (ES, MS), if perceived as different, 92% respondents indicated ES < MS. With the other two pairs, respondents were split on whether EO and EM were greater or less than MO and MM respectively.

Use of the graphic aid (Protocol 2) further permits assessing the magnitude of the differences between RL options and the difference in patterns of comparability of translations across the VRL pairs (e.g. ER vs MR, ES vs MS). The boxplot in Figure 4 illustrates how respondents rated each option.

Conclusion

Post-questionnaire completion interviews are a valuable tool to inform response label use when completing questionnaires. Questionnaires using both numeric and verbal response labels allow respondents to choose the type of label (numeric or verbal) to which they prefer to use. Some respondents completely ignore one type of response label; others indicate a strong preference of one label over the other, voicing in some instances tremendous antipathy regarding their less preferred label type. Interestingly, in the current sample, respondents of different language backgrounds showed different preference profiles, with NES preferring numeric and NNEs preferring verbal labels. With many respondents showing a reliance on the verbal labels, translation comparability of labels when moving from one language form to another is important. The current study’s findings of differences in “meaning” on some translated label pairs highlights the importance of translation of response labels as part of the creation of forms of questionnaires across languages. As such, the use of post-questionnaire completion interview protocols for informing understanding of the response process aspect in validity considerations (AERA, APA, NCME, 2014).

Exemplar Quotes

The following provides exemplar quotes from post-questionnaire completion interviews from 11 participants grouped by use/helpfulness of (a) Numeric RL: P1., P4., (b) Verbal RL: P5., P11. Sample themes (e.g., “Ignore Days”) are grouped together and indicated using corresponding colours from Fig 2.

(a) Numeric RL

P1: “I prefer the quantitative because it’s more specific.”

P2: “The word descriptiveness, they’re very vague. I wasn’t really sure what “rarely,” exactly referred to, so I used all the quantitative ones.”

P3: “Days were more accurate form because you can actually count out how many days you’ve been angry or you’ve been happy.”

P4: “I will remember how long for the days, but we’re not really for the description; I know the times [days?], not the description.”

(b) Verbal RL

P5: “I actually didn’t even pay attention to the days at all.”

P6: “The words stuck out to me the most.”

P7: “I don’t think like my feeling can be quantified. I will prefer with descriptive, like I can describe how I feel that.”

P8: “The time is more important than the days because you count the days that you’re depressed, you just think about it.”

P9: “Because like for the number of days, it’s more detailed; like 1-2 days is too detailed; you will not even count when you feel sad.”

P10: “I felt hard to count in this case; kind of hard to count the number of days.”

P11: “I wasn’t able to actually remember which were the days or how many days exactly. I was just evaluating my previous week generally.”

Figures and Tables

Figures and Tables

Table 1. Proportion and Conditional Comparability of Translations of Verbal Response Labels (VRLs)

Table 2. Proportion of Respondents & # of VRL Pairs Parassembled

Table 3. Proportion-Saving, # for Same/Different

Notes: 1This research was funded by an SFU Psychology Department Research Grant to the PI, R. T. Fouladi … 3 The research team acknowledges SFU is situated on unceded Coast Salish Territory in Burnaby, British Columbia, Canada.