

Running head: DETECTING TRANSLATION ERRORS IN THE JAPANESE VERSION

Detecting Translation Errors in the Japanese Version of the MAACL-R

Tomoyuki Yasuda

The Pennsylvania State University

DETECTING TRANSLATION ERRORS IN THE JAPANESE VERSION OF THE MAACL-R

One of the widely used instruments that measures a broad range of human affect is the Multiple Affect Adjective Check List – Revised (MAACL-R: Lubin & Zuckerman, 1999). Using a total of 66 scored adjectives composing five subscales, the MAACL-R measures affects that are of clinical relevance (i.e., Anxiety, Depression, Hostility, Positive Affect, and Sensation Seeking). Due to its brevity, the range of affects that it gauges, and its adequate reliability and validity, the MAACL-R, together with the original version of the Multiple Affect Adjective Checklist (MAACL: Zuckerman & Lubin, 1965), have been utilized frequently in both research and practice (Lubin, Swearingin, & Zuckerman, 1997). The MAACL-R was also translated into different languages (e.g., Chinese, French, Hebrew, Norwegian, & Spanish).

The Japanese translated version of the MAACL-R was developed in order to facilitate the use of “imported instruments” (e.g., Sue & Chang, 2003) for screening psychopathology and stress-related mental-health problems that are also reported in the Japanese society (Yasuda, Lubin, Kim, & Whitlock, 2003). Appropriate use of psychological tests for individuals with limited English proficiencies, such as recent Japanese immigrants, was another impetus for constructing the Japanese version (e.g., Okazaki & Sue, 1995). Although the use of back-translation procedure (e.g., Brislin, 1970) revealed the linguistic equivalence between the original English and Japanese versions of the MAACL-R (Yasuda et al., 2003), adequate cross-cultural instrument adaptation also requires that psychometric equivalence be present (e.g., Okazaki & Sue, 1995). Furthermore, the comparable reliabilities estimates and the factor and other correlational analyses results provided only limited information about the psychometric equivalence of the items between the two versions (Yasuda et al., 2003).

Despite the growing interests in the cross-cultural adaptation of mental health measures (Geisinger, 1994; Okazaki & Sue, 1995; Sue & Chang, 2003), little attention has been paid to examine translation equivalence from psychometric point of view. In particular, items on the Japanese version of the MAACL-R have yet to be examined for the psychometric equivalence with the original English version. In order to ensure the psychometrically equivalence of items, the conditional statistical procedure known as differential item functioning (DIF) should be conducted. Thus, the present study reported the DIF analyses of the two versions of the MAACL-R. Special attention has been paid to identify items containing translation errors, as indicated by the presence of uniform DIF, in addition to identify items signifying cultural impact as indicated by non-uniform DIF (e.g., van de Vijver & Poortinga, 1997). Since DIF is normally not expected before administration of the instruments (e.g., Allalouf, Hambleton, & Sireci, 1999), no specific hypotheses are generated.

Method

Participants and Procedures

The American participants consisted of 287 undergraduate students from a large state university in a mid-Atlantic state (Mean age = 19.89; $SD = 3.38$. There were 89 males and 198 females. The Japanese sample consisted of 299 undergraduate students (130 females and 169 males) recruited from two universities in the Tokyo and Kumamoto areas (Mean age = 19.87; $SD = 2.06$).

Measures

Trait MAACL-R: Trait MAACL-R is a 66-item adjective checklist for which respondents are asked to check items that describe “How they generally feel” (Lubin & Zuckerman, 1999). The items checked are scored as one, whereas those not checked are scored

as zero. The numbers of items for each scale were as follows: Anxiety (10), Depression (12), Hostility (15), Sensation Seeking (8), and Positive Affect (21).

The Japanese version of the MAACL-R: The Japanese version of the MAACL-R (Yasuda et al., 2003) was developed by translating the original English version of the Trait MAACL-R.

Data Analyses

A total of 66 logistic regressions (i.e., one for each item) were conducted, given that multiple traits, as well as the interaction term, can readily be incorporated in its analytic model. The MAACL-R subscales were used as matching variables. Detection of DIF was based on the statistical criteria, with the adjusted alpha levels of .002 for Positive Affect (.05/21), .006 for Sensation Seeking (.05/8), .005 for Anxiety (.05/10), .004 for Depression (.05/12), and .003 for Hostility (.05/15) in order to control Type I errors due to multiple analyses.

Results

A total of 12 items were found to show DIF (see Table 1). More specifically, nine items (13.6%) exhibited uniform DIF, indicating translation errors. The remaining three items were shown to have non-uniform DIF or cultural impact. Out of the nine items with the uniform DIF, six adjectives were positively related to group membership, with the odds ratios being 6.97 (*free*), 24.72 (*mean*), 10.39 (*peaceful*), 11.76 (*secure*), 8.63 (*shaky*), and 15.72 (*tormented*), whereas three adjectives (i.e., *enthusiastic*, *forlorn*, and *impatient*) had negative association, with the odds ratios being .15 (*enthusiastic*), .16 (*forlorn*), and .11 (*impatient*). Thus, probabilities¹ of the Japanese sample endorsing *free* (.87), *mean* (.96), *peaceful* (.91), *secure* (.92), *shaky* (.89), and *tormented* (.94) were higher than those of the American sample, given that all possible trait

¹ Probabilities were calculated by the formula: probability = odds ratio/(1+ odds ratio).

levels, as measured by the MAACL-R subscales, were held constant. On the other hand, probabilities of the Japanese sample endorsing the adjectives *enthusiastic* (.13), *forlorn* (.14), and *impatient* (.10) were lower than those of the American sample conditioning on all the trait levels.

On the other hand, three adjectives – *alone*, *annoyed*, and *whole* – revealed non-uniform DIF. First, *alone* and *whole* contained positive coefficients (i.e., β_1 and β_2), suggesting that the Japanese sample had higher probability of endorsement than the American sample. Probability became even higher as Depression increased. Second, negative (β_1) and positive (β_2) coefficients in the adjective *annoyed* suggested that the Japanese individuals with lower levels of Hostility had the lower probabilities of endorsement than those of the American individuals. Nevertheless, the Japanese individuals with higher levels of Hostility had the higher probabilities as compared to the American individuals.

Conclusions

This study examined DIF between the original English and Japanese versions of the MAACL-R. Particular attention has been paid to the identification of uniform DIF that is indicative of translation errors. A total of nine items of the Japanese version were shown to include translation errors. Thus, use of these items for research or clinical purpose (e.g., screening, assessment, and diagnosis), especially when involving cross-cultural assessment, may not be appropriate. In addition, the three items displaying non-uniform DIF may not be utilized appropriately, given that these items include different degrees of relevance for the two groups of individuals. For example, the items should be submitted to sensitivity reviews in search of plausible explanations for possible cultural impact before they are adapted for further use. The remaining 54 items, on the other hand, did not include any DIF and thus seemed to be psychometrically equivalent between the American and Japanese samples.

References

- Allalouf, A., Hambleton, R.K., & Sireci, S.G. (1999). Identifying the sources of differential item functioning in translated verbal items. Journal of Educational Measurement, 36, 185-198.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. Journal of Cross Cultural Psychology, 1, 185-216.
- Lubin, B., & Zuckerman, M. (1999). Manual for the Multiple Affect Adjective Check List – Revised (3rd ed.). San Diego, CA: Educational and Industrial Testing Service.
- Lubin, B., Swearingin, S. E., & Zuckerman, M. (1997). Research with the Multiple Affect Adjective Check List (MAACL and MAACL-R): 1960 – 1996. San Diego, CA: Educational and Industrial Testing Service.
- Okazaki, S., & Sue, S. (1995). Methodological issues in assessment research with ethnic minorities. Psychological Assessment, 7, 367-375.
- Sue, S., & Chang, J. (2003). The state of psychological assessment in Asia. Psychological Assessment, 15, 306-310.
- van de Vijver, F. J. R., & Poortinga, Y. (1997). Towards an integrated analysis of bias in cross-cultural assessment. European Journal of Psychological Assessment, 13, 29-37.
- Yasuda, T., Lubin, B., Kim, J., Whitlock, R. V. (2003). The Japanese version of the Multiple Affect Adjective Check List – Revised: Development and validation. Journal of Clinical Psychology, 59, 93-109.
- Zuckerman, M., & Lubin, B. (1965). Manual for the Multiple Affect Adjective Check List. San Diego, CA: Educational and Industrial Testing Service.

γ_3

Table 1. Parameter Estimates of the Items Flagged as Uniform and Non-uniform DIF.

Item	Scale	-2Loglikelihood	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7
<u>Uniform DIF</u>										
Enthusiastic	SS	371.34	-3.14 (.52)	<u>-1.89</u> (.65)	.06 (.17)	.81 (.13)	.07 (.02)	.027 (.08)	-.03 (.07)	.07 (.06)
Forlorn	D	270.79	-5.28 (.90)	<u>2.93</u> (.83)	-.10 (.15)	.69 (.10)	.02 (.03)	-.21 (.10)	.24 (.09)	-.19 (.07)
Free	PA	595.92	-3.04 (.49)	<u>1.94</u> (.53)	-.04 (.05)	.25 (.04)	.19 (.06)	-.10 (.06)	-.01 (.06)	.01 (.05)
Impatient	A	501.12	-1.25 (.38)	<u>-2.16</u> (.44)	-.19 (.12)	.62 (.11)	-.04 (.02)	.07 (.06)	-.28 (.07)	.35 (.06)
Mean	H	217.93	-7.60 (1.03)	<u>3.20</u> (.97)	.46 (.17)	.06 (.04)	.06 (.04)	.11 (.09)	-.02 (.10)	-.04 (.08)
Peaceful	PA	544.54	-3.60 (.56)	<u>2.34</u> (.61)	.00 (.05)	.37 (.04)	-.15 (.06)	-.03 (.06)	.10 (.06)	-.25 (.05)
Secure	PA	572.66	-4.50 (.64)	<u>2.46</u> (.69)	.09 (.05)	.28 (.03)	-.14 (.06)	-.18 (.06)	-.06 (.06)	.05 (.05)
Shaky	A	364.84	-4.09 (.70)	<u>2.16</u> (.72)	.09 (.61)	.78 (.11)	.14 (.03)	.14 (.08)	.08 (.06)	-.08 (.06)
Tormented	D	227.34	-6.70 (.94)	<u>2.91</u> (.87)	.17 (.14)	.47 (.08)	.02 (.04)	.17 (.10)	.08 (.10)	.01 (.07)
<u>Non-uniform DIF</u>										
Alone	D	406.17	-3.38 (.51)	1.46 (.47)	<u>.87</u> (.21)	.64 (.09)	-.03 (.03)	.09 (.07)	-.05 (.08)	-.04 (.06)
Annoyed	H	408.13	-1.92 (.46)	-.92 (.48)	<u>.66</u> (.15)	.57 (.08)	-.02 (.03)	-.09 (.07)	-.03 (.07)	-.12 (.07)
Whole	PA	391.21	-7.41 (.89)	2.78 (1.02)	<u>.22</u> (.06)	.19 (.04)	.09 (.07)	-.05 (.07)	-.05 (.08)	.16 (.06)

Note. Standard errors are given in parentheses. PA: Positive Affect; SS: Sensation Seeking; A: Anxiety; D: Depression; H: Hostility. Parameters are estimated by the following logistic regressions: 1) $z = \beta_0 + \beta_1 G + \beta_2 (G \times PA) + \beta_3 PA + \beta_4 SS + \beta_5 A + \beta_6 D + \beta_7 H$ for the Positive Affect items; 2) $z = \beta_0 + \beta_1 G + \beta_2 (G \times SS) + \beta_3 SS + \beta_4 PA + \beta_5 A + \beta_6 D + \beta_7 H$ for the Sensation Seeking items; 3) $z = \beta_0 + \beta_1 G + \beta_2 (G \times A) + \beta_3 A + \beta_4 PA + \beta_5 SS + \beta_6 D + \beta_7 H$ for the Anxiety items; 4) $z = \beta_0 + \beta_1 G + \beta_2 (G \times D) + \beta_3 D + \beta_4 PA + \beta_5 SS + \beta_6 A + \beta_7 H$ for the Depression items; 5) $z = \beta_0 + \beta_1 G + \beta_2 (G \times H) + \beta_3 H + \beta_4 PA + \beta_5 SS + \beta_6 A + \beta_7 D$ for the Hostility items, where $\text{prob.}(u=1) = e^z / (1 + e^z)$. The modeled group = Japanese (i.e., $G = 0$ for Americans; $G = 1$ for Japanese). Underlined values are statistically significant in the respective alpha levels (i.e., alpha = .002 for Positive Affect; .006 for Sensation Seeking; .005 for Anxiety; .004 for Depression; and .003 for Hostility).