

## A Method for Social Scientists to Adapt Instruments From One Culture to Another: The Case of the Job Descriptive Index

Ramzi Nasser

Notre Dame University, PO Box 72 Zouk Mosbeh, Lebanon

---

**Abstract:** This study outlined an adaptation procedure for the Job Descriptive Index Subscale of “type of work”. The cross-translation or committee translation procedure asks two or more translators to translate a text from source to target language, and then an expert assesses the validity of these translations. Empirically, this method has three or more translators translate the instrument from English to Arabic and then an expert assesses the translations made by the three translators. A selection of 180 bilinguals attempt the source language and later attempt the target language instrument or the translated instrument supported by this method. The two versions are then compared through the ANOVA, correlation analyses and factor analyses. The results indicated a high reliability for the Arabic and English versions. The committee translation approach provides a valid method for translation, the results however, showed that the instrument in both languages do not show item-to-item similarity or equivalence.

**Key words:** Cross-translation, adaptation of instruments, job descriptive index

---

### INTRODUCTION

Social scientists working in cross-cultural contexts use instruments (questionnaire, attitude scale, interview schedules, special techniques, observations, instructions or tests) to measure or evaluate constructs in different settings from the source culture. These constructs are often adapted by translators or colleagues that provide a translation to a target language, although these methods have been marginal and understated at large. A standardized cross-cultural procedure have gained recognition in the Standards for Educational and Psychological Testing (American Educational Research Association), American Psychological Association and National Council of Measurement in Education.

Methodological procedures in the adaptation of instruments for the use in different cultures can be derived from three methods, these are: the cross-translation or committee translation method<sup>[1]</sup>, the back-translation<sup>[2]</sup> and decentering method<sup>[3]</sup>. The committee translation has a panel of experts who translate from a source to a target language. If all the translation are the same then the translation can be considered valid in the target language. The back-translation procedure involves the translation of a source to a target language and the back-translation of the target to the source. If the back-translated version is similar to the source, the translations are adequate to the target; however, if the back-translations are not similar to the source, further validation or decentering takes place i.e., change the source language of the instrument to satisfy the target. The decentering method, has translators attempt to

translate the target to the source language and modify the source to satisfy the meaning of the target language. Although, the specificity and application of this method have been used on narratives and essay texts, its use in the translation of psychometric or single item questionnaires has been limited in the social sciences.

The importance of reliability and validity measures for both language versions has had little attention in the research literature. Equivalence can essentially be achieved through specific measures as validity or reliability where language and cultural difference are substantial to draw attention to the problem of reliability agreement between translators and consistency between versions, such that to insure that the meaning of the instruments is the same across languages<sup>[4]</sup>. Most studies published on cross-cultural adaptation of instruments in the social sciences have emanated from Western countries as source language constructs which implicates the Western knowledge structure as having an over-arching and supreme framework for understanding concepts and the use of English language as the medium for promoting claims of a dominant culture<sup>[5]</sup>. It is rare for instance, to find research work that emanates from non-western cultures by researchers residing in developing/transition countries and uses tools developed in dominant cultures to validate in their own culture.

Recent studies<sup>[6-8]</sup> have provided validity and reliability of target versions of instruments irrespective of how responses relate to the source language. Furthermore, research studies have not been systemized and corroborated in their approaches to establishing a

translation method. Typically they have lay translators or professionals translate the instruments with out concern for the method it was translated. In addition translators may not be familiar with context, purpose or operational definitions of the instrument and hence could alter the translation according to their understanding of the genre or nuances of the language. In cases where single words (e.g., adverbs or stimuli for word association) are translated without a contextual frame; it is extremely problematic to grasp the meaning of the source language and provide its equivalence in the target language.

To provide an amenable solution for the adaptation of instruments, a popular and most likely used method known as the back-translation method. The disadvantage in the translation is that it might produce inaccurate translations from source to target consequently that results in wrong translations from the target language back to the source<sup>[9]</sup>. In many instances cultures often have vernacular, classical and written languages that are used interchangeably, which result in a difficult instrument to adapt to a specified group<sup>[10]</sup>. In some instances words in English language have no equivalence in a language like Arabic or Urdu and overall translations do not satisfy the meaning of the construct without the modification of the linguistic construct in the source language. When the source language does not have an equivalent term in the target language the translation will result in partiality and does not fulfill the construct domain of items. As a result the psychometric properties or constructs could be lost in the translation to the target. The decentering translation requires the modification of the source as a result of the language translation, which often requires the researchers to go through construct validity and reliability procedure so that the original language is adjusted to the target.

Few studies have explored the use of the more basic committee translation approach as being one of many sound approaches for the translation or adaptation of instruments with single words or adjectives. In this process a group of bilinguals translate from a source to a target language and other members of a committee or an independent body of professional translators consensually assess the translations.

One of the disadvantages of the cross-translation approach in connection to the use of materials, or instructions, is the low number of translators who translate from one language to another, i.e., from source to target language. These types of translations rely on a single person often with a large number of inadequacies, and inappropriate translations. In some cases phrasing questions that are supposed to be equivalent in both languages i.e., source and target could elicit a different type of stimuli from different respondents, with the possibility of large differences

between translators in supplanting appropriate closure on the translation.

In this study a committee approach is proposed for the Job Descriptive Index Subscale of "work." This method suggests a sound method for the translation of single words or items. In addition this study obtains validity and reliability of the translation instrument in its target form to give some soundness to the translation process. Particularly, when the back-translation is not feasible as a method, the cross-translation method should overcome much of the tedium involved in the back-translation procedure. This study does not compare the three methods of translation; instead, it specifically determines the reliability and validity of an application of the cross-translation method, and provides an accumulation of information about a theory and method of translation equivalence as applied to empirical data.

**The job descriptive index (JDI):** One of the widely used instruments for studying job satisfaction is known as the JDI<sup>[11]</sup>. It has been translated into several languages: Hebrew<sup>[12]</sup>, Tagalog<sup>[13]</sup> and French<sup>[1]</sup>. This instrument has been administered to staff at all organizational levels all over the United States.

The Job Descriptive Index measures job satisfaction in the five areas of pay, promotion, supervision, type of work and co-workers. The JDI consists of 72 items that are allocated among five areas as follows: work, supervision, co-workers, pay and promotion, the first three have 18 items each and the last two have 9 items each. The instrument is reliable and valid in its five areas. The split half internal consistency coefficient is reported at above  $r=+.80$  for each of the five scales<sup>[14]</sup>.

The scoring scheme of the JDI asks the respondents to write "yes" if they agree, "no" for disagreements and "?" for the undecided. Agreement responses receive a score of 3; "yes" to positive items and "no" to negative items, disagreement responses receive a score 1 for "no" in response to positive items and "yes" for negative items. Undecided responses "?" receive a score of 2.

In this study this scoring scheme was changed from a three point scale to a 9 point scale, from strongly agree to strongly disagree. The internal consistency coefficients were high for the Likert format with an alpha level of  $.87$ <sup>[15]</sup>. The alternative scoring scheme, indicates little difference in providing an overall estimate of convergent validity. However, a slight advantage of the non-significant skewed responses on the Likert scale or a 5 point scale over the three point scale<sup>[14]</sup>. In this study the sample consisted of university students and one area of the Job Descriptive Index by using the "type of work" reconceptualized into "school-work" satisfaction.

**METHODS**

This research is concerned with three important procedures and analyses in the cross-translation process.

- \* A construct validation of the translation.
- \* Administration of source version of the instrument, followed by a post administration on the translated target.
- \* Demonstration of comparable ratings on the Arabic and English versions of the instrument by bilinguals that provide further evidence for construct validity.

The Committee method was used to translate the instrument from source to target language. Three expert translators translated the instrument, which included instruction, i.e., the items as well as the information gathering questions. In addition these translators were asked to translate the responses. Translators examined each item phrase carefully and tailored the translation to the target culture.

The translators professional background of education, degrees and previous experience in the profession are reported on Table 1.

Table 1: Professional Translators Background

Translator	Degree	Specialization	Years in Profession
A	Ph.D.	Linguistics	7
B	M.A.	Philosophy	6
C	M.A.	English as a Second Language	5
Expert Translator (Judge)	Ph.D.	Linguistics	9

A coding scheme was devised to compare the translations made by the committee of translators. Then a comparison was made among the three translators on syntactical, vocabulary and structure equivalence. A blind expert rater was asked to rate the translations based on a degree of convergence of the translated items by the three main translators. Those items that were found to have a correct translation as judged by the expert rater were scored as "3," a translation that slightly diverged from the true meaning was scored as "2" (using a verb instead of an adjective) and a translation that completely diverged from the meaning was scored as "1." The general paradigm of this scoring procedure is based on Kerlinger's<sup>[16]</sup> method of congruence, where a panel is to reflect, define or translate a number of items based on specifications and operationalization of a construct or a translation. A second step is to have an expert whether the items are logically or adequately reflecting the objective specification<sup>[17]</sup> for the items to be translated. Agreement among raters based on the judge rating and whether it supports the correct translation. The

specification or equivalence which is evidence for validity of the instrument across languages. Assessing the degree of agreement between raters and criterion is the more appropriate analysis than interjudge agreement<sup>[17]</sup>.

Once the translation procedure is complete a target sample of 180 Bilingual students were selected from an American University in Beirut, Lebanon. These students were given the English version of the instrument. In an eight week period these same students took the Arabic version. All questionnaires with incomplete answers were not considered in this study.

**RESULTS**

Two research questions were investigated in this study. The first asked if the translation conducted were in agreement. A high agreement would be a reflection of an equivalence of items. The second research question investigated whether the instruments achieved equivalence through item responses on the Arabic and English instruments. Similar responses indicated high equivalence among item meaning and translation using the judged cross-translation method. In connection to these questions a repeated measure ANOVA, factor analysis, and correlational analysis was performed. The first question in this study addressed "the inter-rater agreement when all the judges on all activities are analyzed as a group. This was done based on the judged ratings of the translations. Table 2, presents the frequencies and percentages of inter-rater agreements type across all translations. The square root of agreement percentages approximates the inter-rater correlation (agreement) coefficients<sup>[16]</sup>. The inter-rater agreement was at  $r=0.57$  which is a modest correlation value. The first analysis attempted to find the reliability of the subscale for the "school-work" Job Descriptive Index of the students. The Cronbach alpha reliability was found at 0.67 (N=144) for the 18 items. The reliability of the 18 translated (Arabic Versions) items was at 0.81 (N=47). The strong alphas for the subscales demonstrate that the subscale attempts to assess i.e., it does so equally reliably among students who perform the Arabic and English version of the instrument. However, the alpha on the English version was shown to be slightly lower than the Arabic version. An explanation to these results suggests that students' primary language, which is Arabic, may be a source to the low consistency. It is also found that the number of students who were first year and second year students accounted for almost all the respondents who had come from an Arab monolingual home.

The procedure devised had three professionals, translate the items with only one expert judge who evaluated the convergence of the translations. As all translators could all have the "wrong" translations and have agreed with each other, a deviation score between each rating and the correct response was computed so

Table 2: Percentages of Inter-Rater Agreements by Type of Agreement Across All Translations

Type of Agreement	Frequency	Percent	Cumulative percent	R
All three translators agreed	6	33.3	33.3	0.57
Two of three raters agreed and one partially agreed	4	22.2	55.5	0.47
Two of three raters agreed; one disagreed	2	11.11	66.6	0.57
Two of three raters or only partially agreed	6	33.33	100	0.56

Table 3: Rotated Factor Structure of the "School-Work" Area of the English Version of the Job Descriptive Index (N=144).

	Factor1	Factor 2	Factor 3	Factor 4	Factor 5	h <sup>2</sup>
1. Fascinating	.62*	.17	-.08	-.14	-.15	.46
2. Routine	.03	.02	.35	.68*	.02	.58
3. Satisfying	.05	.74*	-.06	.12	.13	.59
4. Boring	-.20	.00	.11	.81*	.13	.73
5. Good	-.19	.80*	-.09	-.07	.03	.69
6. Creative	.06	.80*	-.14	.00	-.07	.60
7. Respected	.38	.30	.25	-.35	.15	.44
8. Hot	.80*	-.21	.03	.12	-.12	.70
9. Pleasant	.65*	.39	-.22	-.04	.05	.62
10. Useful	.07	.44	.45	-.52*	.08	.67
11. Tiresome	-.03	-.06	.71*	.00	-.32	.61
12. Healthful	.71*	.02	.21	-.24	.23	.67
13. Challenging	.27	.47	.31	-.18	-.30	.51
14. On Your Feet	.72	-.07	.14	-.04	.08	.55
15. Frustrating	.05	-.01	.68*	.31	-.03	.57
16. Simple	.04	.05	-.07	.06	.90*	.78
17. Endless	.06	-.09	.71*	.02	.11	.53
18. Gives Sense of Accomplish	.16	.53*	.21	-.37	.01	.48
Total						h <sup>2</sup> =60%

\* Factor Loading

Table 4: Rotated Factor Structure of the "School-Work Area" of the Arabic Version of the Job Descriptive Index (N=47)

	Factor1	Factor 2	Factor 3	Factor 4	Factor 5	h <sup>2</sup>
1. Fascinating	.77*	.16	-.21	.13	-.36	.81
2. Routine	-.01	-.40	.43	.50*	.34	.70
3. Satisfying	.43	.66*	.17	.11	.01	.66
4. Boring	-.01	-.59*	.60*	.10	.16	.74
5. Good	.56*	.12	.28	.61*	-.05	.78
6. Creative	.68*	.15	.28	.02	.31	.65
7. Respected	.58*	.19	.38	-.41	.35	.80
8. Hot	.51*	.29	.03	-.65*	.03	.77
9. Pleasant	.64*	.60*	-.17	-.09	.01	.81
10. Useful	.25	.64*	.02	-.01	.19	.50
11. Tiresome	-.02	.15	.87*	.12	-.03	.79
12. Healthful	.81*	.20	.00	-.07	.16	.72
13. Challenging	.19	.18	.82*	-.09	-.03	.75
14. On Your Feet	.05	.67*	.57*	-.14	-.12	.82
15. Frustrating	-.04	-.09	.72*	.23	.12	.60
16. Simple	.04	.19	.10	.80*	-.12	.70
17. Endless	.12	.12	-.01	-.09	.90*	.84
18. Gives Sense of of Accomplish.	.16	.85*	.10	.10	.08	.77
Total						h <sup>2</sup> =70%

\* Factor Loading

Table 5: Correlations Between Factor Loadings for the Two Versions of the Instrument

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Factor 1	0.505*				
Factor 2		0.362			
Factor 3			0.469*		
Factor 4				0.290	
Factor 5					0.049

\* Correlation is significant at the 0.05 level (2-tailed)

that this relativity was removed, making the results easily interpretable<sup>[17]</sup>. Each rater translated the 18 items, the ratings or translations were correlated. Therefore to take account of the correlations, a repeated measure ANOVA was measured, to assess the degree of intra and inter-agreement of judges.

A one-way repeated measures ANOVA on the 18 translations done by 3 translators. As can be seen from Table 3 no significant difference were found between Judges or translators ( $F=.47;df=2;p>.001$  and  $F=1.86, df=17, p>.001$ ) respectively. These three raters agreed with each other and translated the 18 adjective phrases according to the criterion of the expert rater.

The data were factor analyzed using principle component analysis with unities in the diagonals, an eigen cut-off value of 1.0, and a varimax rotation. Table 3 presents the results. First, the 18 items of the English version of the instrument were factor analyzed, followed by the Arabic version. The two factor structures were then compared. On the English version, factor analysis reduced the 18 variables to 5 factors. The first factor accounted for 20.6% of the variance and all the five factors accounted for 60% of the variance. On the Arabic version of the instrument, factor analysis reduced the 18 variables to 5 factors. The first factor accounted for 29.5% of the variance. The 5 factors accounted for 73.3% of the variance. Table 4 presents the rotated factor analysis results of the Arabic version of the JDI, using principle component with unities in the diagonals, an eigen cut of value of 1.0 and a varimax rotation of the Arabic version of the JDI.

The main purpose of the factor analysis is to determine if the structures of the source and target instruments were similar. The English version of the JDI showed items 1, 8, 9, 12 and 14 had relatively high loading on the items, were the Arabic version had shown a greater number of items load on the first factor including items 1, 5, 6, 7, 8, 9 and 12. Item 14 which loaded on the first factor of the English version of the instrument, did not load on the second factor of the Arabic version of the instrument. Common item loading on the second factor for both version of the instruments were items 3, 10 and 18. Item loadings on the third factor were items 11 and 15, on the fourth factor was item 2. The fifth factor had the 13th item load in the English & Arabic version. In total 9 items loaded on the factors; although, the instruments have shown some similarities, the factor loadings comparisons do not contribute to a similar factor structure. However, if one examines the communalities on the two previous analysis on Tables 3 and 4, one finds that communalities were relatively high between both versions of the instruments, suggesting a common feature for the schoolwork subscale of the JDI. The translation and expert ratings on the instrument showed convergence on the translation specification, which provides evidence for construct validity of the Arabic version of the instrument. In the final analyses, the correlation between factors for each of the versions the source and target language. Each factor from the

varimax rotation was correlated with the same factor preconceptualized in the source target language version. Factor analysis shows some similarities that do not warrant a significant equivalence of both tests.

## DISCUSSION

The most practical procedure for the translation of instruments is known as the committee approach method. In the committee procedure a bilingual translates the instrument from a source language to a target language. This is particularly true when a domain content is represented as a short set of items or an adjective checklist<sup>[3]</sup>. Nevertheless, the translator might not be familiar with the context of the specified research. A cross validation is necessary in a cross-translation of an instrument. This is accomplished by using the convergent validity paradigm<sup>[19]</sup> where one or more expert, rates the translations based on their adequacy and appropriateness.

This study attempted to present a methodology for the translation of the Job Descriptive Index "school-work" subscale. To account for the equivalence in the translation procedure, several phases of the study were established. Three translators translated the instrument into the target language. An expert judge conducted a criterion based rating of the translation, by comparing the number of errors based on the equivalence criteria. Within this process, an expert rater examines the source with the target versions and scores the target translation for its clarity, adequacy and appropriateness. Consequently, a final version was developed and a group of students were given the English version of the instrument.

Equivalence of source language instruments to target language should establish adaptability universality of measures. Adequacy of instrument-translation founded upon similar validity, reliability, and factor structures across languages that insure consensus and construct substantiation. The established validity of the JDI in the English version has been reported by Smith, *et. al*<sup>[11]</sup>. In this study, the construct validity of the Arabic version of 18 items "school-work" subscale was assessed by having three professional translators, translate the items and an expert rater i.e., judge establish criteria for the "correct translations." One should note that inter-rater agreement is not the same as reliability of observational measures, meaning inter-rater agreements could concur with one another for which translators may have the wrong translations, and all translations agree with each other, hence reliability is not the same of inter-rater agreement. It is necessary to have an expert rater rate the translations. No significant differences were found among the ratings of the translations by the judge; hence, translations were construct validated based on expert criterion.

The final aspect of adaptability is to measure the consistency of factor scores; the correlation analysis procedure was conducted between the factors for both

languages in source and target form. This should give some sense to the social scientist whether a large group of bilinguals understand the social constituents in the same way across languages. The low correlations for the bilinguals on the two instruments of the source (English) and target (Arabic) versions of the factors, do not grant equivalence between them. These low correlations might have occurred due to the administration time span between the two instrument versions. Although the reliability scores for both the English and Arabic versions of the JDI was high, the factor analysis do not lend support for a similar factor structure. However, some similarities were found in the first factor, which accounted for 20.6% of the variance for the English version and 29.5% for the Arabic version. The latter results do not provide a strong reliable support as the respondent to item ratio was well below the minimum five-to-one ratio. These results are not comparable to provide equivalent form-validity between the English and Arabic versions of the instrument, however, moderate construct validity was established as translators showed agreement based on an expert criteria. In conclusion, the results provide evidence that the Arabic and English version were not similar in structure.

One should keep in mind that exact translation is impossible in principle and more important the committee approach does not satisfy equivalent translation or hold the original language for revision. A condition for equivalence of translation is to have bilinguals respond to items in both languages. Evidence from these results suggest that test-retest on a group reflect some linguistic and cultural factor differences; students who responded on the English version responded differently on the Arabic version as reflected on the results. The selected sample of first year and second year students had come from an Arab monolingual home where the saliency of these items in the native language reflected a different attitudinal responses in the time gap between the pre-administration (English version) and post-administration (Arabic version) of the instrument.

A social science researcher who follows a method of translation in cultures or linguistic backgrounds different from his native one might face substantial problems with conditions that respondents might not corroborate translations. Instruments can not be ideally translated into equivalent forms, however, some specific methods could be applied to provide reliable translations. The results of this study suggest that cross-translation may have not been a viable method but could be used with the back-translation or the decentering method to provide a more valid and reliable results.

#### REFERENCES

1. Candell, G. and C. Hulin, 1986. Cross-language and cross-cultural comparisons in scale translations: Independent sources of information about item nonequivalence. *J. Cross-Cultural Psychol.*, 17: 417-440.

2. Chapman, D. and J. Carter, 1979. Translation procedures for the Cross-cultural use of measurement instruments. *Educational Evaluation and Policy Analysis*, 1: 71-76.
3. Werner, O. and D. Campbell, 1970. Translating, working through interpreters and the problem of decentering, In R. Naroll and R. Cohen (Eds.). *A handbook of cultural anthropology*. New York, American Museum of Natural History.
4. Ervin, S. and R. Bower, 1953. Translation problems in international surveys. *The Public Opinion Quart.*, 16: 595-604.
5. Temple, B. and R. Edwards, 2002. Interpreters/translators and cross-language research: Reflexivity and border crossings. *Intl. J. Qualit. Methods*, 1. [www.ualberta.ca/~ijqm/backissues/1-2Final/html/temple.htm](http://www.ualberta.ca/~ijqm/backissues/1-2Final/html/temple.htm)
6. Nuhr, M., R. Crevenna, M. Quittan, A. Auterith, G. Wiesinger, T. Brockow, P. Nuhr, V. Fialka-Moser, A. Haig and G. Ebenbichler, 2004. *J. Rehabilitation Med.*, 36: 267-272.
7. Luo, N., L. Tan, S. Li, L. Soh and J. Thumboo, 2005. Validity and reliability of the Chinese (Singapore) version of the Parkinson's Disease Questionnaire (PDQ-39). *Qual. of Life Res.*, 14: 273-280.
8. Montazeri, A., A. Goshtasebi, M. Vahdaninia and B. Gandek, 2005. The short form health survey (SF-36): Translation and validation study of the Iranian version. *Qual. of Life Res.*, 14: 875-882.
9. Brislin, R., 1970. *Back-translation for Cross-Cultural Research Methods*. New York, John Wiley & Sons.
10. Sechrest, L., T. Fay and S. Zaidi, 1972. Problems of translation in cross-cultural research. *J. Cross-Cultural Psychol.*, 3: 41-56.
11. Smith, P., L. Kendall. and C. Hulin, 1969. *The measurement of satisfaction in work and retirement*. Chicago, II. Rand McNally & Company.
12. Hulin, L. and L. Mayer, 1986. Psychometric equivalence of a translation of the JDI into Hebrew. *J. Appl. Psychol.*, 62: 385-391.
13. deVera, M., 1985. *Establishing cultural relevance and measurement equivalence using emic and etic items*. Unpublished Ph.D. Thesis, University of Illinois at Urbana-Champaign.
14. Hulin, C., 1966. Job satisfaction and turnover in female clerical population. *J. Appl. Psychol.*, 50: 280-285.
15. Johnson, S., P. Smith and S. Tucker, 1982. Response format of the job descriptive index: Assessment of reliability and validity by the multitrait-multimethod matrix. *J. Appl. Psychol.*, 67: 500-505.
16. Kerlinger, F., 1986. *Foundations of Behavioral Research*, New York: Holt, Rinehart and Winston.
17. Dagostino, L. and J. Carifio, 1993. Establishing the logical validity of instructional activities for teaching reading evaluatively. *North Eastern Res. Assoc. Ann. Conf.*
18. Frick, T. and M. Semmel, 1978. Observer agreement and reliabilities of classroom observational measures. *Rev. Educat. Res.*, 48: 157-184.
19. Campbell, D. and D. Fiske, 1964. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psycholog. Bull.*, 56: 81-105.