

Cross-Cultural Questionnaires and the Necessity of Using Native Translators: A Croatian-Swedish Case

Molander, Bo; Holmström, Stefan; Takšić, Vladimir

Source / Izvornik: **Psihologijske teme, 2020, 29, 191 - 198**

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.31820/pt.29.1.11>

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:186:060734>

Rights / Prava: [In copyright](#)

Download date / Datum preuzimanja: **2021-07-27**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



Cross-Cultural Questionnaires and the Necessity of Using Native Translators: A Croatian-Swedish Case

Bo Molander, Stefan Holmström,

Umeå University, Department of Psychology, Umeå, Sweden

Vladimir Takšić

University of Rijeka, Faculty of Humanities and Social Sciences,
Department of Psychology, Rijeka, Croatia

Abstract

In this paper, we discuss problems of comparing two European cultures in a study of emotional intelligence by relying on traditional back translation of the questionnaire and the scales used in the study (Holmström, Molander, & Takšić, 2008; Molander, Holmström, & Takšić, 2009, 2011). We compared Croatian and Swedish university students in using *The Emotional Skills and Competence Questionnaire*, which is an original Croatian questionnaire (Takšić, 1998; Takšić, Mohorić, & Duran, 2009) based on the emotional intelligence theory of Mayer and Salovey (1997). Initially, we found small differences in responding to emotional items between the two countries as revealed by traditional statistical methods. Here we illustrate a large increase of the initial differences by applying differential item functioning (DIF) procedures (Zumbo, 1999), and then reducing again differences by taken several important steps in analyzing the translated items. Most important in these latter procedures was a new translation to Swedish by a native Croatian-speaking translator.

Keywords: cross-cultural measures, back translation, native translators, latent variables, DIF-analyses

✉ Bo Molander, Department of Psychology, Umeå University, 901 87 Umeå, Sweden. E-mail: Bo.Molander@umu.se

Author Note:

We appreciate very much the collaboration and helpful suggestions of translator Vesna Bušić in early parts of our Croatian-Swedish research.

The Department of Psychology and the Faculty of Behavioral and Social Sciences, Umeå University, kindly provided facilities for the cooperation of Bo Molander.

This paper contains part of information which was presented in the *7th Psychometric Symposium in honor of Professor Alija Kulenović* (2015). The symposium was conducted at the 22nd Ramiro and Zoran Bujas' Days, Zagreb, Croatia.

Introduction

An inspection of a recent volume of the *Journal of Cross-Cultural Psychology* (Volume 50, Issues 1-8, 2019) revealed that among the 52 papers available on comparisons between cultures, using questionnaires or other written or spoken materials, included no text in the abstract or in keywords about the translation procedures. Perhaps authors thought that such information would suit better in the method sections. However, neither in those sections did we find much about translation. Among the 52 papers, 22 of them did not mention clearly, or not at all, the linguistic origin of one or more of the questionnaires or scales used. In the other 30 papers, the origin was British English or American English. Furthermore, the procedure for translation to another language was not always described. We suggest that authors provide, at a minimum, information about a) the linguistic merit of the translator for both languages involved, b) if there are several translators, results of comparing them, and c) translator knowledge of the meaning of important latent variables (Harkness, 2008; Zumbo, Gelin, & Huble, 2002). Often the inspected papers referred to translators only as „bilingual” without giving information about native language and level of experience in both languages. Back translation was mentioned explicitly only in eight of the 52 papers, and in varying degrees of clearness.

In addition to translation procedures, it is also of high importance in cross-cultural studies to demonstrate measurement equivalence. In only 11 of the papers, authors used factor analysis, alone, or together with structural equation calculations, and in some cases only discussed. Specific methods for item bias, e.g. DIF-methods, were seldom used.

First ESCQ Results in Croatian-Swedish Cooperation

Although it is depressing to read in the JCCP's 2019's volume papers which are lacking important methodology, for valid reasoning about obtained results we should remember that only 20 years ago procedures for translation and measurement equivalence were discussed even less. The Croatian-Swedish cooperation in the area of emotional intelligence started around 2005, and description of a project collecting data based on the ESCQ- questionnaire from six different countries was first published in 2006 in a Portuguese journal (Faria et al., 2006), with Croatian data from high school students and Swedish data from bus drivers and nurses. However, in Tables 1 and 2 below, we present a more fair comparison between the two countries, with data from a later publication, based on Croatian and Swedish university students (Molander et al., 2011).

Table 1

Typical Example of Number and Proportions of Participating Men and Women, and Estimated Mean Age in the Croatian and Swedish Samples (Molander et al., 2011)

	Men N (proportion)	Women N (proportion)	Total Sample	Mean Age (years)
Croatia	236 (.31)	522 (.69)	758	≈22
Sweden	79 (.37)	133 (.63)	212	≈25

Table 2

Means, Standard Deviations, and Levels of Significance for ESCQ Total Scores and Subscale Scores for Men and Women (Molander et al., 2011)

		Croatia M	SD	Sweden M	SD	p
ESCQ	All	160.4	16.29	161.0	12.60	.632
	Men	158.0	16.59	159.1	13.25	.579
	Women	161.4	15.94	162.1	12.15	.679
PU	All	53.4	6.96	52.2	5.81	.022
	Men	51.9	7.11	51.1	5.75	.345
	Women	54.0	6.81	52.8	5.79	.063
EL	All	48.6	7.22	51.0	6.28	.000
	Men	47.9	6.54	50.5	6.26	.002
	Women	48.8	7.45	51.2	6.33	.001
MR	All	58.5	5.99	57.8	4.71	.150
	Men	58.1	6.20	57.5	5.09	.416
	Women	58.6	5.87	58.0	4.50	.285

Note. ESCQ = Emotional Skills and Competence Questionnaire (45 items); PU = Perceive and Understand (15 items); EL = Express and Label (14 items); MR = Manage and Regulate (16 items).

In addition, alpha values were high and very similar for Croatia and Sweden: for total scores .85 - .88; for PU .84 - .84, for EL .80 - .81; and for MR .59 - .69. The lower values for MR may be due to a higher complexity of variables in that scale, and thus perhaps more difficult to translate to other cultures. Overall, Table 2 shows that Croatian and Swedish university students seem to respond very similar to the ESCQ instrument. We found no significant statistical differences for Total results and the PU and MR scale results, whereas there was a significant difference between Croatia and Sweden for the EL scale. Thus, and overall, the cultural difference between Croatian and Swedish university students, as shown by the ESCQ questionnaire, seems to be very modest. The original American theory behind ESCQ (Mayer & Salovey, 1997) does not make any cross-cultural predictions, usable for evaluation of these results, but studies like World Value Study (Inglehart & Welzel,

2011) show that cultural differences between these countries in sociological and psychological areas could be substantial.

Measurements of Item Bias

During the 90s and the beginning of 2000, several papers were published addressing *the requirement of equivalence* for making possible comparisons between cultures and languages. We were of course quite interested in this development. Methods for making safer conclusions about cultural differences, especially *construct equivalence* and *item bias* were needed. In particular we started to investigate a method called Differential Item Functioning (DIF), and where *DIF is said to exist if an item is more difficult, discriminating or easily guessed for one group than for another* (Zumbo, 1999). Zumbo described this method as logistic regression modeling, and as a framework for binary and Likert-type item scores. Chi-squared tests and effect sizes are part of the procedure. In our paper (Molander et al., 2011), we made an investigation of possible DIF effects. Table 3 shows the results with calculations on total scores and scale scores based on criteria for two different effect sizes. It should be noted that DIF effects are calculated separately for each of the three scales.

Table 3

Number of DIF Items and Proportions for Total Scores (45 items) and Subscales for Croatian (758) and Swedish (212) University Students according to Zumbo & Thomas (1997) and Jodoin & Gierl (2001) Effect Size Criteria (Molander et al., 2011)

Criteria	Total	PU	EL	MR
<i>Jodoin & Gierl (2001)</i>	14 .31	4 .27	2 .14	5 .31
<i>Zumbo & Thomas (1997)</i>	<i>1</i> .02	<i>0</i> .00	<i>0</i> .00	<i>1</i> .06

Note. Jodoin & Gierl (bold): effect size > .035; Zumbo & Thomas (italics): effect size > .13; PU = Perceive and Understand (15 items); EL = Express and Label (14 items); MR = Manage and Regulate (16 items).

We first used the Zumbo and Thomas (1997) criterion for deciding DIF items because their logistic regression method was the first we came across. In addition to the effect size criterion, there was also a demand that Chi² calculations should be significant on .01 level. According to the criteria by Zumbo and Thomas (1997) the results looked good and in line with the obtained results shown in Table 2. Almost no difference at all between Croatia and Sweden. However, this was a period of great statistical activity in DIF calculations and soon the requirements for acceptable items increased quite a lot. The Jodoin and Gierl (2001) criterion is an example of this development. The arguments for their criterion seemed to be reasonable, and we have used this criterion since the day we read their paper. Finding 14 DIF items in a

questionnaire of 45 items seemed anyway to be too much, and we were concerned about the effect of these items on the scores. However, it should also be remembered that getting DIF items in your instruments is not only bad. DIF measurements may also reveal cultural differences or other differences, which were not thought of earlier.

A New Look on ESCQ Items

Although procedures for checking item bias were welcome around the beginning millennium, there was also a need for improvements in translation procedures. One book that arrived in 2011 and sometime after we performed our first DIF-analyses made a strong impression. The title of the book was „Cross-Cultural Research Methods in Psychology” edited by David Matsumoto and Fons J.R. van de Vijver (2011). Several chapters were very valuable for our research at that time, and still are.

In the beginning of the Croatian-Swedish cooperation in the emotional intelligence field, we first used the commonly suggested back-translation procedure. The ESCQ questionnaire was Croatian original and was translated to English in a version that was adapted to Swedish. The present Swedish authors made the translation to Swedish. We then consulted a teacher in English at the Linguistic Department in our university for correctness of the translation. Some smaller changes were suggested, and the whole translation was then discussed with Vladimir Takšić, present author, and the creator of the ESCQ instrument. After this talk, we considered the translation to be acceptable. However, after making DIF measurements, and after having read the Matsumoto-van de Vijver (2011) book, we understood that more had to be changed in the Swedish version.

We went through the English version of the questionnaire following the advice found in a chapter by Hambleton and Zenisky (2011), which was part of the Matsumoto and van de Vijver book. These authors listed 25 critical aspects of the text in five different categories, i.e., *General, Item format, Grammar and phrasing, Passages, Culture*. We found that about 40-50% of the items in our Swedish instrument were affected in one or several categories. We also found that DIF-items were more often affected than non-Dif items. It seemed like a good idea to improve the original translation.

Discovering a Native Croatian Translator

Somewhat later and by sheer luck, we found the name Vesna Bušić on a door in the Department of Linguistics, situated in another building at Umeå University but very close to the Department of Psychology. We never heard of her before, but by knocking at the door and talking with her it turned out that she was a native Croatian speaker, and had spent the last 20 years in Sweden. She worked at the University as

a teacher in Swedish as well as in Swedish as a Second language. Moreover, she was very good at English, as she had to talk English with some of her students. After talking to Vesna Bušić about our Croatian-Swedish research and the need to get a native translator for the questionnaire she agreed to be that translator. We handed the original Croatian questionnaire and the Swedish translation to her, and she immediately recognized some obvious faults in the Swedish translation. She then made a very new translation with cooperation also from the three of us, mainly on questions of the expected intent of an item. Finally, it became time for us to collect new data on Swedish university students based on the new translation. We collected this new sample in very much the same courses as the old sample.

Comparing again Croatian-Swedish Results

Table 4 shows the result of DIF-analyses based on a new sample of 272 university students (Molander, Holmström, & Takšić, 2015) and compared with the old Croatian sample (Molander et al., 2011). DIF effects are calculated separately for each of the three scales.

Table 4

Number of DIF Items and Proportions for ESCQ Scores (45 items) and Subscales for Croatian Sample (758) and the New Sample (272) of Swedish University Students (Molander et al., 2015)

Criteria	Total	PU	EL	MR
<i>Jodoin & Gierl (2001)</i>	4 .09	3 .20	1 .07	3 .19
<i>Zumbo & Thomas (1997)</i>	<i>1</i> .02	<i>0</i> .00	<i>0</i> .00	<i>1</i> .06

Note. Jodoin & Gierl (bold): effect size > .035; Zumbo & Thomas (italics): effect size > .13; PU = Perceive and Understand (15 items); EL = Express and Label (14 items); MR = Manage and Regulate (16 items).

Results according to the Jodoin and Gierl (2001) criterion show a large reduction of number of DIF-items on total scores from 14 to 4, and a substantial reduction for the three subscales from a total of 11 to 7. In addition, the new translation also increased the total scores and subscale scores, as well as the Cronbach's alpha values. It should be noted also, that if DIF-items are found, there are several actions that can be taken before there is a conclusion that the translation has to be improved. We will not go through these actions here, and there were strong indications anyway that the native translation improved the Swedish ESCQ questionnaire.

Conclusion

We have shown that recent papers in the cross-cultural field are still lacking in acceptable methods for handling questionnaires, scales or other instruments for measuring cross-cultural effects. In particular, we have argued for more of native translators and for specific analyses of items. We are fully aware of many other fields in psychology, which have similar problems, and we are of course aware that we are not the first to point out these problems. In the past, several authors have suggested routines for how translation of questionnaires should proceed. We have mentioned the Hambleton and Zenisky (2011) chapter, but there are earlier and more extensive papers on that issue, as for example van de Vijver and Hambleton (1996), Harkness, Pennell, and Schoua-Glusberg (2004), and Harkness (2008). A recent interesting publication on translation of information on informed consent is Brelsford, Ruiz, and Beskow (2018). Psychology uses many measuring tools. We need these tools to be sharp, not blunt and non-efficient. Good advice in the cross-cultural field, and still needed, is to make native translators an early and more important part of the development of questionnaires and other measuring tools.

References

- Brelsford, K. M., Ruiz, E., & Beskow, L. (2018). Developing informed consent materials for non-English-speaking participants: An analysis of four professional firm translations from English to Spanish. *Clinical Trials*, *15*, 557-566. doi:10.1177/1740774518801591
- Faria, L., Lima Santos, N., Taksic, V., Rätty, H., Molander, B., Holmström, S., Jansson, J., Avsec, A., Extremera, N., Fernandez-Berrocal, P., & Toyota, H. (2006). Cross-cultural validation of the Emotional Skills and Competence Questionnaire (ESCQ). *Psicologia*, *20*, 95-127.
- Hambleton, R. K., & Zenisky, A. L. (2011). Translating and adapting tests for cross-cultural assessments. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 46-74). Cambridge: Cambridge University Press.
- Harkness, J. A. (2008). *Round 4 ESS translation strategies and procedures*. European Social Survey, ESS document.
- Harkness, J. A., Pennell, B. E., & Schoua-Glusberg, A. (2004). Questionnaire translation and assessment. In S. Presser, J. Rothgeb, M. Couper, J. Lessler, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 453-473). New Jersey: John Wiley and Sons.
- Holmström, S., Molander, B., & Takšić, V. (2008, July). Analysis of item bias in the Emotional Skills and Competence Questionnaire: A cross-cultural comparison. In V. Taksic (Chair), *Cross cultural validation of various measures of emotional intelligence*. Symposium conducted at the XXIX International Congress of Psychology, Berlin. *International Journal of Psychology*, *43*, 354.

- Inglehart, R., & Welzel, C. (2011). *The WVS cultural map of the world*. Retrieved from http://www.worldvaluessurvey.org/wvs/articles/folder_published/article_base_54
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect-size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education, 14*, 329-349. doi:10.1207/S15324818AME1404_2
- Matsumoto, D., & van de Vijver, F. J. R. (Eds.) (2011). *Cross-cultural research methods in psychology*. Cambridge: Cambridge University Press.
- Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Implications for educators* (pp. 3-31). New York: Basic Books.
- Molander, B., Holmström, S., & Takšić, V. (2009). Cross-cultural and sex differences in the Emotional Skills and Competence Questionnaire scales: Challenges of differential item functioning analyses. *Horizons of Psychology, 18*, 35-53.
- Molander, B., Holmström, S., & Takšić, V. (2011). South and North: DIF analyses of university-student responses to the Emotional Skills and Competence Questionnaire. *Psychological Topics, 20*, 425-447.
- Molander, B., Holmström, S., & Takšić, V. (2015, April). Differential item functioning in the Croatian-Swedish study of the Emotional Skills and Competence Questionnaire. In V. Buško (Chair), *7th Psychometric Symposium in honor of Professor Alija Kulenović*. Symposium conducted at the 22nd Ramiro and Zoran Bujas' Days, Zagreb, Croatia.
- Takšić, V. (1998). *Validacija konstrukta emocionalne inteligencije. [Validation of the Emotional Intelligence Construct]*. (Unpublished doctoral dissertation). Faculty of Humanities and Social Sciences, University of Zagreb, Zagreb.
- Takšić, V., Mohorić, T., & Duran, M. (2009). Emotional skills and competence questionnaire (ESCQ) as a self-report measure of emotional intelligence. *Horizons of Psychology, 18*, 7-21.
- van de Vijver, F. J. R., & Hambleton, P. M. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89-99.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for Binary and Likert-type (Ordinal) item scores*. Ottawa, ON: Directorate of Human resources Research and Evaluation, Department of National Defense of Canada.
- Zumbo, B. D., Gelin, M. N., & Huble, A. M. (2002). The construction and use of psychological tests and measures. In the Psychology theme of the *Encyclopedia of Life Support Systems (EOLSS)*. Oxford, UK: Eolss Publishers.
- Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Working paper of the edge worth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia, Prince George, BC, Canada.

Received: December 1, 2019