Psihološka obzorja / Horizons of Psychology, 18, 3, 35-53 (2009)
© Društvo psihologov Slovenije 2009, ISSN 1318-187
Znanstveni empiričnoraziskovalni prispevek

Cross-cultural and sex differences in the Emotional Skills and Competence Questionnaire scales: Challenges of differential item functioning analyses[#]

Bo Molander[1], Stefan Holmström[1] and Vladimir Takšić[2]*
[1]Department of Psychology, Umeå University, Sweden
[2]Department of Psychology, University of Rijeka, Croatia

Abstract: University students in Croatia, Slovenia, and Sweden ($N = 1129$) were examined by means of the Emotional Skills and Competence Questionnaire (Takšić, 1998). Results showed a significant effect for the sex factor only on the total-score scale, women scoring higher than men, but significant effects were obtained for country, as well as for sex, on the *Express and Label* (EL) and *Perceive and Understand* (PU) subscales. Sweden showed higher scores than Croatia and Slovenia on the EL scale, and Slovenia showed higher scores than Croatia and Sweden on the PU scale. In subsequent analyses of differential item functioning (DIF), comparisons were carried out for pairs of countries. The analyses revealed that a large proportion of the items in the total-score scale were potentially biased, most so for the Croatian-Swedish comparison, less for the Slovenian-Swedish comparison, and least for the Croatian-Slovenian comparison. These findings give doubts about the validity of mean score differences in comparisons of countries. However, DIF analyses of sex differences within each country show very few DIF items, indicating that the ESCQ instrument works well within each cultural/linguistic setting. Possible explanations of the findings are discussed, and improvements for future studies are suggested.

Key words: Emotional Skills and Competence Questionnaire (ESCQ), cross-cultural differences, sex differences, DIF analyses

[#] Acknowledgements: Slovenian data were provided by the courtesy of Professor Dr Andreja Avsec. We are indebted to Andreas Stenling for assistance in the DIF-analyses work, and to John Jansson for help in collecting data. We would also like to thank two anonymous reviewers who provided useful improvements. Part of the Croatian-Swedish analysis was presented at the XXIX International Congress of Psychology, Berlin 2008. This research was supported by grants from IF Metall, Sweden, and the Ministry of Science and Sport, Croatia.

[*] Naslov / Address: Dr. Bo Molander, Department of Psychology, Umeå University, 901 87 Umeå, Sweden, e-mail: bo.molander@psy.umu.se

Medkulturne razlike in razlike med spoloma na vprašalniku emocionalne inteligentnosti ESCQ: Izzivi analize diferencialnega funkcioniranja postavk

Bo Molander[1], Stefan Holmström[1] in Vladimir Takšić[2]*
[1]Oddelek za psihologijo, Univerza v Umeå, Švedska
[2]Oddelek za psihologijo, Univerza v Reki, Hrvaška

Povzetek: Hrvaški, Slovenski in Švedski študenti ($N = 1129$) so izpolnili samoocenjevalni vprašalnik emocionalne inteligentnosti ESCQ (Takšić, 1998). Rezultati so pokazali pomembne razlike med spoloma v skupnem rezultatu, pri čemer imajo ženske v povprečju višje rezultate, in pomembne razlike med državami in med spoloma na lestvicah sposobnosti izražanja in poimenovanja emocij ter sposobnosti zaznavanja in razumevanja emocij. Švedski študenti poročajo o višje izraženi sposobnosti izražanja in poimenovanja emocij, slovenski pa višje rezultate na lestvici sposobnosti zaznavanja in razumevanja emocij. Pri analizi diferencialnega funkcioniranja postavk (DIF) smo primerjali med seboj po dva vzorca. Analiza je pokazala, da je velika večina postavk potencialno pristranskih, najbolj v primerjavi hrvaškega in švedskega vzorca in najmanj v primerjavi hrvaškega in slovenskega vzorca. Ti rezultati kažejo na problematičnost direktnih primerjav povprečnih vrednosti med različnimi državami. Po drugi strani pa DIF analiza kaže, da je znotraj posamezne države zelo malo potencialno pristranskih postavk glede na spol. Zaključimo lahko, da vprašalnik ESCQ dobro funkcionira znotraj posameznega kulturnega/jezikovnega okolja. Predstavljene so možne razlage rezultatov kot tudi predlogi za izboljšanje nadaljnjih raziskav.

Ključne besede: Vprašalnik emocionalne inteligentnosti ESCQ, medkulturne razlike, razlike med spoloma, DIF analiza

CC = 2220, 3120

The concept of emotional intelligence has developed fast from being judged by researchers as quite suspicious in the early nineties to the present state of relative decency (e. g., Mayer, Salovey, & Caruso, 2008; Roberts, Schulze, Zeidner, & Matthews, 2005; Zeidner, Roberts, & Matthews, 2008; see also the paper by Takšić in the present issue). During the same period research and applications in the field have increased greatly, as has the interest in emotional intelligence by laymen. As a result of this activity various versions of tests of emotional intelligence have been assessed and applied in a number of settings, countries and cultures. In this article we will focus on cross-cultural comparisons, and, in particular, the possible problem of item bias.

Not much has been published as yet on the present theme. As revealed by sources of publications, such as PsychNet, Web of Science, and Science Direct, there are very few papers published where outcomes of assessment of tests or scales are discussed in terms of cultural differences. Most papers are cross-cultural only in the meaning that assessments of tests of emotional intelligence or correlations of

such tests with other tests are performed in countries other than USA. Furthermore, such studies do not always report proper national assessments (e. g., control of factor structure) for comparison with the original test, which often is MSCEIT (Mayer, Salovey, & Caruso, 2002) or EQ-i (Bar-On, 1997). Thus, many studies in the field are likely to contribute to mystification rather than clarification. Also, recent summaries of research on emotional intelligence are surprisingly quiet about cross-cultural comparisons and the methodological and conceptual problems which are involved (e. g., Mayer, Roberts, & Barsade, 2008; Mayer, Salovey, & Caruso, 2008; Murphy, 2006; Roberts et al., 2005; Zeidner et al., 2008), although Mayer et al. (2008) acknowledge the need of greater attention to the factors of culture and sex, and their impact on theories and measurement of emotional intelligence.

As indicated by these recent sources there is no clear picture of the extent to which present measures of emotional intelligence can be used for comparisons among different cultural settings. This state of affairs is understandable, of course, considering the massive work that has to be done for proper assessment, this work including such aspects as repeated factor analyses, item analyses, replication of original procedures, and collection of normative data in each culture. In addition, the lack of theory in the field about cross-cultural variations contributes to the slow progress. Contributing is probably also the fact that the two most influential tests (MSCEIT and EQ-i) are only commercially available. Even if assessments of these tests have been performed in other countries by the distributing company or associated researchers, results may not be easily accessible to other researchers. Anyway, international publications of assessments of MSCEIT and EQ-i seem to be directed to psychometric aspects rather than cultural aspects. The comments by J. E. Helm (1992), where she argued for studies of cultural equivalence in standardized cognitive ability testing are valid also for the field of emotional intelligence.

Although research on emotional intelligence from a cross-cultural point of view is quite limited there is a vast literature on various aspects of emotion and culture. Some of the fields of research are of high relevance also for the study of emotional intelligence. One important field is the study of emotion expressions, much of which has been directed to facial expressions (e. g., Ekman, Friesen, & Ellsworth, 1972; Matsumoto, 2001; Russell, 1994). A related field is emotion perception, where recognition of emotion and various parameters of emotion, such as intensity, are of interest (e. g., Biehl et al., 1997; Izard, 1971; Matsumoto, Wallbott, & Scherer, 1987). Another field of interest is how people in different cultures experience emotions, that is, the question is if people have the same type of subjective and bodily reactions for a specific emotion (e. g., Markus & Kitayama, 1991; Scherer, Summerfield, & Wallbott, 1983). The study of emotion appraisal is an area of research dealing with "…the process by which people evaluate the events, situations, or occurrences that lead to their having emotions" (Matsumoto & Juang, 2004, p. 250), thus an area of seemingly high importance for the study of emotional intelligence (see e. g., Mauro, Sato, & Tucker, 1992; Scherer, 1997). A last field of importance is concerned with

the understanding of the concept of emotion and other related concepts and how this understanding can differ among cultures (e. g., Brandt & Boucher, 1986; Levy, 1973; Russell, 1991). An extreme example is the observation by Levy that Tahitians do not have a word for emotion.

The message from the research findings in these different fields is that there are commonalities among cultures, and sometimes, as in the study of facial expressions, universal commonality. However, research in all of the listed fields also tells us that there are differences among cultures. These differences are sometimes small and sometimes large. Differences are found also among seemingly similar cultures. Thus, the cross-cultural study of emotional intelligence should not be content with relying on findings from just a few countries. Neither should there be reliance on a translation of an instrument for assessing emotional intelligence from one language to another without performing several types of controls, people's understanding of the concept of emotion being one important control.

In this article we proceed from studies by Faria et al. (2006), Toyota, Morita, and Takšić (2007), and Takšić et al. (2009), where the instrument Emotional Skills and Competence Questionnaire (ESCQ) was assessed in seven countries in Europe and Asia (i. e., Croatia, Finland, Japan, Portugal, Slovenia, Spain, and Sweden). The ESCQ was developed in Croatia by Vladimir Takšić (1998) and based on the Mayer and Salovey (1997) model of emotional intelligence. The instrument comprises three subscales: *Perceive and Understand emotion* (PU), *Express and Label emotion* (EL), and *Manage and Regulate emotion* (MR), with a total of 45 items to be answered by the participants through ratings on a 5-point scale. This instrument is thus considered to measure subjective or self-perceived emotional intelligence. The subscales are supposed to correspond to three of the four dimensions of emotional intelligence postulated in the Mayer and Salovey model. The fourth dimension, that is, *Emotional Facilitation*, has, so far, been very difficult to find evidence for (e. g., Zeidner et al., 2008).

In their cross-cultural assessment, based mostly on high-school and university students, Faria et al. (2006) found good agreement among most of the seven countries with respect to reliability and various forms of validity. Throughout, the factor structures of the national instruments are rather similar, as is the level of internal consistency (Takšić et al., 2009). Interestingly, the internal consistency (i. e., Cronbach's alfa) of the MR subscale was lower in all countries as compared to the other two subscales. Lower reliability of the MR dimension has been observed also in studies using the Mayer and colleagues' (2002) MSCEIT performance-based battery (e. g., Föllesdal & Hagtvet, 2009; Kafetsios, 2004; Lopes, Salovey, & Straus, 2003). Thus, some caution is recommended in interpreting MR scores.

Judging from the analyses performed in Faria et al. (2006) and Takšić et al. (2009) it looks as if the ESCQ measure of emotional intelligence can be used for generalizing results among countries or cultures and that the concept of emotional intelligence is pretty much the same in the studied countries. Unexplained variance exists in the study, however, and part of that variance might very well be due to cul-

tural variation. Actually, there were indications that some items in some countries were difficult to interpret as related to emotion. Hence, it is of importance to make analysis on item level, so items that do not function equivalently over groups can be identified. This is also a necessary step in evaluating the psychometric properties of the ESCQ scales, because presence of item bias affects the validity of the scales. If there are differences in response patterns among subgroups, this can be a sign of item bias known as differential item functioning (Smith, 2002; Swaminathan & Rogers, 1990; Zumbo, 1999).

Differential item functioning (DIF) exists if an item is more difficult, discriminating, or easily guessed for one group than for another. For example, all persons at a given level on one of the ESCQ scales should have the same probability of endorsing an item in the same way regardless of subgroup (e. g., sex, age, country). DIF methods focus on the trait continuum at the item level, rather than testing for differences across groups of items, which is the typical approach (Asçi, Fletcher, & Çağlar, 2009; Holmström, 2008). Variance in item means may simply reflect divergence among the groups on the construct being measured rather than differences in how the items are functioning (Bann, Iannacchione, & Sekscenski, 2005). When translating items into other languages DIF analysis is especially valuable for evaluating the agreement among items translated into different languages (Bann et al.). Items that display language-related DIF do not measure the same concepts, and therefore the results across languages are inappropriate to combine or compare. Thus, the existence of DIF could indicate a problem with the translation or be a sign of possible cultural difference.

There are several methods existing for evaluation and identification of DIF (c.f. Kristjansson, Aylesworth, McDowell, & Zumbo, 2005). In the present study ordinal logistic regression is used to evaluate DIF in ESCQ items. One of the reasons for selecting ordinal logistic regression was that this procedure is considered to be more general and flexible than the other DIF procedures (Swaminathan & Rogers, 1990). Furthermore, ordinal logistic regression is a suitable method for detecting DIF in ordinal items (Kristjansson et al.).

By using data from the Faria et al. (2006) and Takšić et al. (2009) studies it is possible to further illuminate possible cross-cultural differences by examining item bias. This will be done here by comparing the data from Croatia, Slovenia, and Sweden. The choice of these three countries is based on the fact that the first two countries share the cultural and linguistic environments to a great extent although there are still some cultural and linguistic differences among them, and that Sweden differs from both countries both culturally and linguistically, and due to socio-economic factors, possibly a little less from Slovenia than from Croatia. Also, previous studies suggest that the difference among Croatia, Slovenia, and Sweden is big enough culturally to expect some differences in ratings of emotionality (Arar & Molander, 1996; Molander & Arar, 1998, 2000; Schwartz & Rubel, 2005). To our knowledge there are no previous psychological studies examining the relationship among Croatia, Slovenia,

and Sweden in the area of emotional skill and competence except what is reported in the studies by Faria et al. (2006) and Takšić et al. (2009). However, all European countries investigated by these authors are also part of the European Social Survey (ESS, see www.europeansocialsurvey.org), a survey sponsored by European Union and European Research Council, and presently comprising 34 countries. Mostly, sociologically based questions are asked in the survey, but occasionally questions of relevance for emotional skills are examined. Unfortunately, Croatia has not been part of the first three rounds, but in future rounds it will be possible to make finer predictions than what is possible in the present study about cultural differences among Croatia, Slovenia, Sweden and the other European countries, where the ESCQ instrument has been assessed.

To summarize, we will present the ESCQ data for Croatia, Slovenia, and Sweden, incorporating in the Method section a description of the samples and procedures used for collecting these data. After running analyses of variance on the scores we will perform DIF analyses in order to study possible biases among the three countries. In these analyses we will also examine possible sex differences, among countries, as well as within countries.

Method

Participants

The samples in this study include a total of 1129 university students from Croatia, Slovenia, and Sweden. See Table 1 for sample characteristics.

Table 1. *Number (and percentage in parentheses) of participating men and women, mean age, and standard deviation of age in the Croatian, Slovenian, and Swedish samples*

	Men	Women	Total	*M*	*SD*
Croatia	201 (43%)	263 (57%)	464	20.6	2.16
Slovenia	93 (31%)	207 (69%)	300	21.5	3.08
Sweden	164 (45%)	201 (55%)	365	25.0	7.00

In Table 1 the number of participants in the Swedish sample is larger and different from what is presented in the Faria et al. (2006) study. The reason for this deviation is that the size of the Swedish sample in that study (i. e., $n = 190$) was considered being too small for a reliable DIF-analysis. Also, the present sample comprises university students instead of bus drivers and nurses, who were examined in the Faria et al. study.

Instrument

The ESCQ instrument comprises 45 statements to be answered by self-ratings on a five-point scale (i. e., Never-Seldom-Occasionally-Usually-Always). In addition to total scores the instrument also provides scores for three subscales: *Perceive and Understand* (15 items), *Express and Label* (14 items), and *Manage and Regulate* (16 items). For a detailed description of the instrument, see the paper by Takšić in the present issue. The Slovenian version of the ESCQ instrument was translated directly from the original Croatian version, whereas the Swedish version was translated from an English version of the Croatian original version (Takšić, Tkalčić, & Brajković, 2001). For both the Slovenian and Swedish versions the technique of back translation was applied (e. g., van de Vjer & Hambleton, 1996).

Procedure

The ESCQ was administrated to students in classes during regular school hours. Participation was voluntarily and no monetary reward was given. Before the start of the testing, the participants were introduced in general terms to the purpose of the study, and informed consent was obtained. Instructions were given about how to use the scale of the instrument. The questionnaire took about 30 minutes to finish, and the whole session lasted approximately 45 minutes.

Statistical methods

Internal consistencies were determined by means of Cronbach's alpha, using .70 as an acceptable level (Nunnally & Bernstein, 1994). Differences in mean values on ESCQ-subscales for country and gender were evaluated with multivariate analysis of variance (MANOVA), and analyses of variance (ANOVAs) were conducted as post-hoc tests. For the post-hoc tests Bonferroni adjustments were used. All analyses were performed in SPSS version 15.0.

In the present study ordinal logistic regression was used to evaluate DIF in ESCQ items. This DIF technique (e. g., Zumbo, 1999) is based on the ordinal logistic regression equation

$$y = b_0 + b_1 TOTAL + b_2 GRP + b_3 TOTAL \times GRP_i + \varepsilon_i,$$

where TOTAL stands for total scores and GRP stands for group. In this equation ε_i is distributed with mean zero and variance $\pi^2/3$. Scripts for calculating DIF in SPSS are provided by Zumbo.

Results

Means and standard deviations of the ESCQ total scores and scores of the subscales are presented in Table 2 for country as well as for gender. Reliability analyses of the internal consistencies for country were performed on both the total scale and the subscales. Cronbach´s alpha values varied between .88 and .90 for the total scale. The internal consistency for the subscales ranged from .81–.89 and .82–.87 for PU and EL, respectively, to .67–.73 for the MR scale. In the later subscale alpha levels are lower than .70 for both the Swedish and Croatian samples.

Table 2. *Means and standard deviations of ESCQ scores for Croatian, Slovenian, and Swedish samples*

	Croatia				Slovenia				Sweden			
	Men		Women		Men		Women		Men		Women	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
ESCQ total	157.3	16.0	161.1	17.2	158.6	16.0	162.1	17.2	158.6	16.0	162.0	13.3
PU	51.8	6.8	53.6	7.1	53.1	7.4	55.3	7.2	50.7	6.1	52.5	5.9
EL	47.7	6.5	49.0	7.5	47.0	7.3	49.3	7.4	50.5	7.1	51.7	6.9
MR	57.8	5.9	58.5	5.9	58.6	6.3	57.5	6.3	57.4	5.2	57.8	4.8

Note. PU = Perceive and Understand (15 items); EL = Express and Label (14 items); MR = Manage and Regulate (16 items).

Due to multicollinearity between the ESCQ-total scale and the subscales, a univariate analysis of variance was made for the ESCQ-total scale. The only significant difference in this analysis was obtained for Sex, $F(1, 1123) = 12.43$, $p < .000$, $\eta^2 = .01$, women showing higher score than men. A 3 (Country) × 2 (Sex) multivariate analysis of variance (MANOVA) was conducted to determine the extent to which Countries and Sex differed in the subscales. The assumption of homogeneity was broken, and therefore the more robust Pillai´s trace was used when reporting significant differences. The MANOVA revealed multivariate effects for both Country and Sex, for Country $F(6, 2244) = 20.22$, $p < .000$, Pillai´s trace = .10, $\eta^2 = .05$, and for Sex, $F(3, 1121) = 10.70$, $p < .000$, Pillai´s trace = .03, $\eta^2 = .03$. When looking separately on each of the subscales, the trend was the same with significant differences in two out of three subscales. ANOVAs showed effects of Country for PU, $F(2, 1123) = 10.88$, $p < .000$, $\eta^2 = .02$, as well as for EL, $F(2, 1123) = 18.80$, $p < .000$, $\eta^2 = .03$. Also for the Sex factor the same two subscales showed effects, $F(1, 1123) = 21.14$, $p < .000$, $\eta^2 = .02$, and $F(1, 1123) = 13.27$, $p < .000$, $\eta^2 = .01$, for PU and EL, respectively. For both scales women showed higher scores than men. Post-hoc tests of analyses of variance (ANOVAs) with Bonferroni adjusted alpha levels were conducted to reveal

significant differences among countries. For the PU scale the Slovenian sample scored significantly higher than the two other two samples (ps < .001), and for the EL scale the Swedish sample scored higher than the others (ps < .000).

Zumbo´s (1999) DIF-concept measures the effects of group and the interaction over and above the total-scale score. The test for statistical significance follows a hierarchy of steps for entering variables into the model. In the first step the total score variable is entered, at the second step the grouping variable (e. g., Sex) is included. In the last step, the interaction term between the variables is included in the first and second stage entered. The last step in the analysis describes whether the difference between the group responses of an item changes over the latent variable continuum. With the results from the chi-squared test for logistic regression from the first and third step, one can calculate the significant level through subtracting the chi-squared value in step three from the value in the first step. The differences in chi-squared value can then be compared to its distribution function with 2 degrees of freedom (3–1 = 2 *df*). To calculate the *R*-squared level, the same procedure is used as for chi-squared calculation. For an item to be classified as displaying DIF, two criteria must be met: First, the chi-squared must have a *p* value less or equal to .01. Secondly, the effect size measure must have an *R*-squared value of at least .035. In this study the effect-size criteria suggested by Jodoin & Gierl (2001) is used when quantifying the magnitude of DIF, that is, DIF is negligible for effect-size values below .035, moderate for levels between .035 and .070, and large for levels above .070.

Results of the DIF calculations are depicted in Figures 1 and 2. Number of items with values equal to or higher than .035 for each country comparison and for each scale are shown in Figure 1. Number of critical items for each country comparison and sex are shown in Figure 2.

Figure 1 reveals that the number of items demonstrating DIF is quite large for the total-score measure of the ESCQ instrument. The largest proportion of biased items was found in the Croatian-Swedish comparison, a noticeably smaller proportion in the Croatian-Slovenian comparison, and with the Slovenia-Sweden comparison in between. The same relationships are observed for PU and EL subscales, the MR subscale showing a small deviation, such that a somewhat larger proportion is found for Croatia-Slovenia than for Slovenia-Sweden. It should be noticed also that, generally, the MR scale shows larger DIF proportions than the other two subscales. In Figure 2 it is shown that the proportions of DIF are quite equal for men and women in all country comparisons with a slight male predominance for Croatia-Sweden and Slovenia-Sweden and with a somewhat larger proportion of DIF for women as compared to men in the Croatian-Slovenian comparison. A comparison of sex in each subscale did not reveal any distinct pattern.

Possible effects of sex were examined also within each national sample. Results are shown in Table 3. The total-score measure shows that there are very few biased items in each national version of the instrument. Similar calculations for the subscales showed that for both Croatia and Slovenia there were no biased items at all in the EL and MR scales.

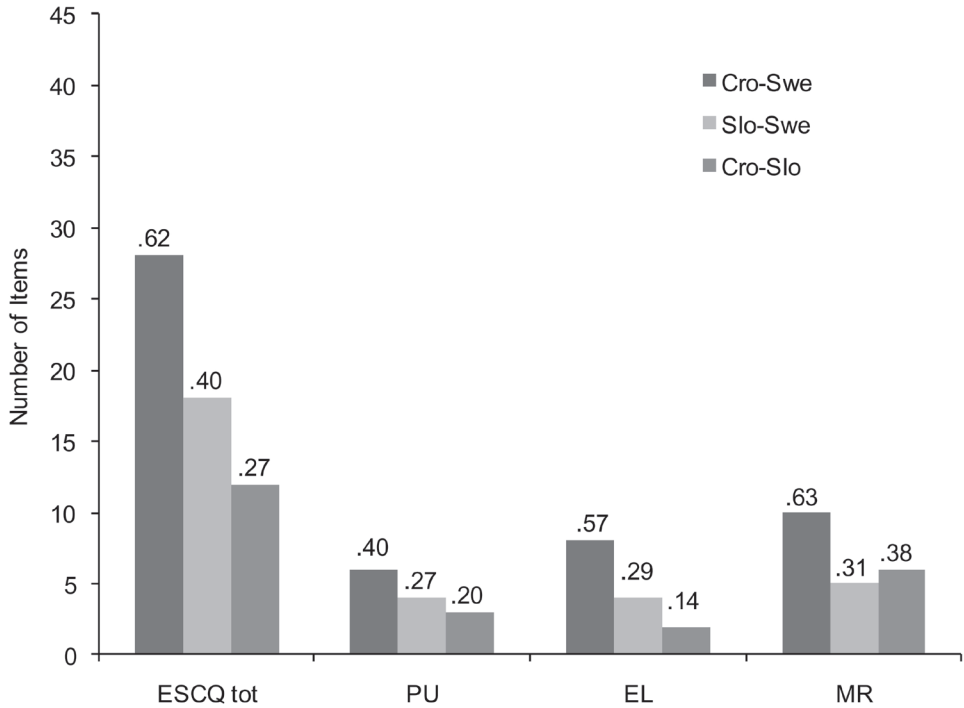

Figure 1. Number of DIF items as a function of country comparison and ESCQ scale. Proportion of DIF items of total number of items in the scale is shown above each bar.

Table 3. *Number of DIF items in comparisons of sex differences for each ESCQ scale and country*

	ESCQ Total Scores	PU	EL	MR
Croatia	3	1	0	0
Slovenia	2	0	0	0
Sweden	2	2	0	1

Note. PU = Perceive and Understand (15 items); EL = Express and Label (14 items); MR = Manage and Regulate (16 items).

In Table 4 are listed the five items with the highest DIF values (all over .07) for each country comparison together with the five items with the lowest effect values of those items that are below the DIF criterion (all below .035) in all country comparisons. This table does not suggest any simple pattern for understanding the differences among the country comparisons, except that items from all three subscales are represented, and that some items appear in more than one comparison. Of the five items in the Croatian-Swedish comparison only one item was unique

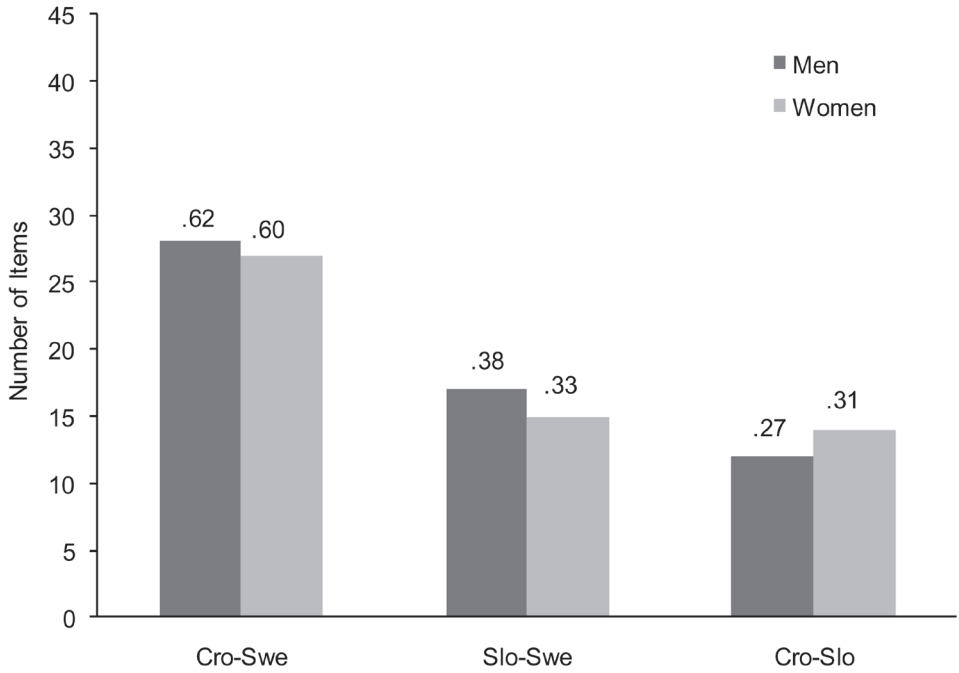

Figure 2. Number of DIF items for total ESCQ scores as a function of country comparison and gender. Proportion of DIF items is shown above each bar.

for that comparison. Only one item was unique also for the Croatian-Slovenian and Slovenian-Swedish comparisons, respectively. No item occurs in all comparisons, however. It should be noted that three of the five items shown in the right-hand column belong to the MR scale.

Discussion

As shown by the statistical analyses of the ESCQ scores, significant overall differences were obtained for both countries and sex. More specifically, the differences among countries were found for the PU (Perceive and Understanding) and the EL (Express and Label) subscales. Slovenia showed higher scores than Croatia and Sweden in the PU scale, and Sweden showed higher scores than Croatia and Slovenia in the EL scale. Our rather rough predictions about cultural differences in the introductory section, that is, that Croatia and Slovenia should be relatively close, with Sweden deviating from both countries but possibly less from Slovenia, are not fully in agreement with the results, although our predictions were intended for the total scores in the first place. Overall, women scored higher than men as expected,

Table 4. *The five items with the highest effect values in the DIF analysis on total ESCQ scores in each country comparison, and the five items with the lowest effect values below the DIF criterion (right-hand column)*

	Croatia-Sweden	Croatia-Slovenia	Slovenia-Sweden	All Comparisons Items with Lowest DIF
1	When I don´t like a person, I find ways to let him/her know (MR,7)	I can easily think of a way to approach a person I like (EL,16)	I can say that I know a lot about my emotional state (EL,24)	People can tell what mood I am in (EL,28)
2	When I see how someone feels, I usually know what has happened to him (PU,14)	I study and learn best, when I am in a good mood and happy (MR,11)	When I don´t like a person, I find ways to let him/her know (MR,7)	If I really want to, I will solve a problem that may seem insoluble (MR,12)
3	I study and learn best, when I am in a good mood and happy (MR,11)	I am able to tell the difference if my friend is sad or disappointed (PU,15)	I am able to tell the difference if my friend is sad or disappointed (PU,15)	There is nothing wrong with how I usually feel (MR,30)
4	I have found it easy to display fondness for a person of the opposite sex (EL,41)	When I see how someone feels, I usually know what has happened to him (PU,14)	If I observe a person in the presence of others, I can determine precisely her/his emotions (PU,25)	I usually understand why I feel bad (EL,32)
5	If I observe a person in the presence of others, I can determine precisely her/his emotions (PU,25)	I can say that I know a lot about my emotional state (EL,24)	I am able to maintain a good mood even if something bad happens (MR,1)	I try to keep up a good mood (MR,33)

Note. Scale origin and item number is given within parenthesis.

although it should be noted that in two of the seven countries presented in the papers by Faria et al. (2006) and Takšić et al. (2009) there were differences in the opposite direction. There were no interactions between sex and country or between sex and scale in the present analyses, but with all seven countries involved Takšić et al. (2009) reported significant Sex x Country interactions for the PU and EL scales.

What is striking in the present set of data is that the differences among countries in terms of mean scores are quite small, the difference between men and women being three times as large as the difference among countries. Normally, as in cross-cultural research on values (e. g., Schwartz & Rubel, 2005) it is the other way around, cultural effects are much larger than sex difference effects. This is true also here if the results

on the separate scales are considered. Judging from the η^2-measure in the ANOVA calculations, effects are generally larger for country than for sex. The weak cultural effect is anyway indicated by the lack of significant results for the MR subscale and for total ESCQ scores. In the Faria et al. (2006) and Takšić et al. (2009) studies the present three countries were compared with Portugal, Finland, Spain and Japan. The only country standing out from the others with respect to total scores is Japan. Such a result suggests that the ESCQ instrument is not sensitive enough to detect European cultural variation in emotional skills and competence, especially if data are based on such a rather homogenous group of participants as the university students in this study. It could also be argued that making cross-cultural comparisons with as few countries as here is quite risky, as the full range of emotional competence might be missed (Schwartz & Rubel).

The analysis of differential item functioning provides quite a different picture of the obtained data. As shown in Figures 1–2 and Tables 3–4, item bias is frequent in all scales, and the proportion items which are biased in the scales is varying from .14 to .63. The extent of bias in the ESCQ instrument is thus considerable, as judged by the present analysis. Furthermore, the extent of the bias seems to be a function of which countries that are compared. For all scales the Croatian-Swedish comparison gives rise to the highest proportions of item bias, followed by the Slovenian-Swedish comparison with proportions at a clearly lower level, and by the Croatian-Slovenian comparison showing proportions at the lowest level. Interestingly, these data are more in line with our hypothesis about cultural differences than the mean scores data. Somewhat surprisingly, though, Croatia and Slovenia differed more than expected, and Sweden and Slovenia differed less than expected. However, the Schwartz and Rubel studies (2005) on values, such as benevolence, tradition, security, and conformity, and where Croatia, Slovenia and Sweden are included, suggests a similar pattern for these three countries.

Not surprisingly, the observed pattern for countries was obtained also for sex. Croatian men compared with Swedish men and Croatian women compared with Swedish women yielded the highest proportions of item bias, followed by corresponding comparisons for Slovenia-Sweden and Croatia-Slovenia. Note, however, that very little item bias was observed when sex comparisons were performed within each country. Thus, the existence of item bias in the ESCQ instrument is likely to be primarily related to cultural differences or differences in translations of the items, rather than to sources inherent in the instrument.

The inclusion of Table 4 in this paper is primarily intended as an illustration of part of the results of the DIF procedure. It seems quite clear, however, that it will be pretty difficult to make conclusions about cultural differences or translation errors just by looking at the items in the table. The picture looks complex with some items occurring in more than one comparison and with no clear pattern among the subscales. Admittedly, Table 4 comprises quite few items, and a more thorough look

at all DIF items might give a better hint. Also, the inclusion of low-DIF items in the table suggests one hypothesis that could be examined better in the total material, that is, it looks as if low-DIF items have a simpler structure than the high-DIF items, and that all of them requires judgements only about the respondent. Most of the high-DIF items require judgment about situations where somebody else is involved. Thus, with increased item complexity the likelihood of DIF due to cultural differences might increase. However, before strong conclusions about cultural differences in the ESCQ instrument can be made it is necessary to consider alternative explanations.

One such explanation of high relevance here is related to the translation procedures and the question if those procedures might be different for different country comparisons. In the present case the original Croatian version of the instrument was translated to English (Takšić et al., 2001) and from English to Swedish. Quite a lot of effort was put into the Croatian-English version by professional linguistic people, who were native speaking Croats and fluent in English. Presumably, high knowledge of Croatian language was the case also in the Slovenian translation. Back-translation procedures were used for both the Slovenian and Swedish versions. Still, it is possible that some subtle differences in meaning have appeared in the various translated versions. It is also likely that back-translation and communication with the author of the instrument is easier for a Slovenian researcher than for a Swedish researcher due to general familiarity of the Croatian language and culture. When clusters of countries are compared, as in the present paper, it seems important to confirm translations not only in relation to the original country, but also to the other participating countries. Obviously, in future studies more attention has to be directed to the translation of the instrument (see Hambleton, 2001; van de Vijver & Hambleton ,1996; van de Vijver and Leung, 1997, for valuable guidelines).

In addition to bias related to items and the translation of the instrument, biases related to method may occur. The method for collecting data it is quite simple and straightforward and participants do not need much instruction. In the present study the procedure employed seems to have been carried out in a similar fashion in all three countries, although differences are likely to exist concerning, for example, test environment, way of presenting the study and explanation of how to use the scale. It is important, of course, that scalar equivalence is established in cross-cultural research (van de Vijever & Leung, 1997). There is high reliance on the test experience of the associated researchers, but a test manual describing how ESCQ should be distributed and managed may nevertheless be of value for reducing some of the error variance occurring during data collection. Our impression, though, is that the method is quite robust and reliable and not likely to be the cause of the DIF effects.

It is important that the participants in different countries are comparable with respect to background factors such as education. Here the samples are very homogenous as the participants are university students. Such a sample is a good choice in the present context, because students also are quite acquainted with test situations and with receiving verbal instructions in a group setting. It is possible, of course, that

students from different countries may differ in terms of university requirements and high school qualifications, but such differences are probably small in comparison with the effects of the general level of education. Thus, we don't consider the present choice of samples to be a problem. Nevertheless, future national studies should extend the choice of samples to include groups with varying cultural and educational backgrounds. Hopefully, such a national routine would also contribute to making international comparisons less biased with respect to choice of samples.

The DIF variation could definitely contribute irrelevant construct variance to the ESCQ instrument. Although the studies by Faria et al. (2006) and Takšić et al. (2009) showed good consistency in factor structure among the seven countries, there is certainly room for improvement. Furthermore, internal reliabilities are generally high, only the MR scale shows lower values, close to the .70 criterion. This lower reliability seems to be of some importance, because, as shown in Figure 1, the proportions of DIF are somewhat higher in the MR scale than in the PU and EL scales, and the relation among the three countries is changed, as well. Still, the same picture remains, that is, DIFs are plenty, and the number of DIFs for Croatia and Sweden is larger than for Slovenia and Sweden, and smallest for Croatia and Slovenia.

The large number of DIFs, more than half the number of items in the ESCQ total scale for the Croatian-Swedish comparison, could very well be a consequence of a too liberal criterion for flagged items. Different criteria exist in different models (e. g., Clauser & Mazor, 1998; Myers, Wolfe, Feltz, & Penfield, 2006; Swaminathan & Rogers, 1990; Zumbo, 1999). In the original paper by Zumbo the effect size criterion was set to 0.13. But this criterion is considered by Jodoin and Gierl (2001) to be too strict. They point out that there has been a lack of investigations of the effect size levels. Jodoin and Gierl evaluated the effect size measure and came up with a less restrictive threshold criterion, which has been used in the present study and in other studies (e. g. Escorial & Navas, 2007; Jette, Haley, Ni, Olarsch, & Moed, 2008). Although the choice of criterion level is no matter to take easy, the criterion used here is not likely to change the pattern of DIF much compared to other criteria in use.

If DIF is observed, what should be done? According to Zumbo (1999) it is not a good idea to get rid of such items. First, dropping items may limit too much the domain that is of interest. Secondly, the fact that an item is flagged for DIF does not mean that this item necessarily is biased. One has to follow up such items with various item analyses. On the other hand, if an item is not flagged there is no bias. In the present case, it will be necessary to inspect all DIF items carefully to begin with, bearing in mind the constructs emanating from the original EI model of the instrument. Different kinds of item analyses can then be performed, as, for example, checking uniform and nonuniform items, and using iterative procedures (e. g., van de Vijver & Leung, 1997). Replicating the DIF analyses on other samples from the same countries, and extending the analyses to the four other countries included in the Faria et al. (2006) study are other steps to be taken in future studies.

In conclusion, this study demonstrates the importance of performing analyses

of differential item functioning in a test instrument, even if that instrument has been proven to possess good psychometric qualities with respect to validity and reliability. Especially important are such analyses in the context of cross-cultural studies. Validation of the instrument, including DIF analyses, must be considered again whenever new cultural groups enter into the research. This study thus illustrates some of the problems to be faced by the instruments presently available for the study of emotional intelligence. Fortunately, most of these problems are possible to deal with.

References

Arar, Lj., & Molander, B. (1996). Croatian norms for familiarity, emotionality, motor activity, and memorability of 444 actions. Godišnjak Odsjeka za psihologiju [*Annual Report of Department of Psychology]*, 4-5, 7-51.

Asçi, H., Fletcher, R. B., & Çağlar, E. (2009). A differential item functioning analysis of the PSDQ with Turkish and New Zealand/Australian adolescents. *Psychology of Sport and Exercise, 10*, 12-18.

Bann, C. M., Iannacchione, V. G., & Sekscenski, E. S. (2005). Evaluating the effect of translation on Spanish speakers' ratings of Medicare. *Health Care Financing Review, 26*, 51-65.

Bar-On, R. (1997). *The Emotional Intelligence Inventory (EQ-i): Technical manual*. Toronto, Canada: Multi-Health Systems.

Biehl, M., Matsumoto, D., Ekman, P., Hearn, V., Heider, K., Kudoh, T., & Ton, V. (1997). Matsumoto and Ekman's Japanese and Caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior*, *21*, 3-31.

Brandt, M. E., & Boucher, J. D. (1986). Concepts of depression in emotion lexicons of eight cultures. *International Journal of Intercultural Relations*, *10*, 321-346.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement: Issues and Practice, 17*, 31-44.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face*. New York: Garland.

Escorial, S., & Navas, M. J. (2007). Analysis of the gender variable in the EPQ-R scales using differential item functioning techniques. *Educational and Psychological Measurement, 67*, 990-1001.

Faria, L., Lima Santos, N., Takšić, V., Räty, 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.*

Föllesdal, H., & Hagtvet, K. A. (2009). Emotional intelligence: The MSCEIT from the perspective of generalizability theory. *Intelligence, 37,* 94-105.

Hambleton, R. (2001). The next generation of the ITC test translation and adaptation guidelines. *European Journal of Psychological Assessment, 17,* 164-172.

Helm, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist*, *47*, 1083-1101.

Holmström, S. (2008). *Workplace stress measured by Job Stress Survey and relationships to musculoskeletal complaints*. Doctoral dissertation, Umeå University, Umeå, Sweden.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Jette, A. M., Haley, S. M., Ni, P., Olarsch, S., & Moed, R. (2008). Creating a computer adaptive test version of the late-life function and disability instrument. *The Journals of Gerontology: Series A, Biological Sciences and Medical Sciences, 63*, 1246-56.

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.

Kafetsios, K. (2004). Attachment and emotional intelligence abilities across the life course. *Personality and Individual Differences*, *37*, 129-145.

Kristjansson, E., Aylesworth, R., McDowell, I., & Zumbo, B. D. (2005). A comparison of four methods for detecting differential item functioning in ordered response items. *Educational and Psychological Measurement, 65*, 935-953.

Levy, R. I. (1973). *Tahitians*. Chicago: University of Chicago Press.

Lopes, P. N., Salovey, P., & Straus, R. (2003). Emotional intelligence, personality, and the perceived quality of social relationships. *Personality and Individual Differences*, *35*, 641-659.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224-253.

Matsumoto, D. (2001). Culture and emotion. In D. Matsumoto (Ed.), *The handbook of culture and psychology* (pp. 171-194). New York: Oxford University Press.

Matsumoto, D., & Juang, L. (2004). *Culture and psychology* (3rd ed.). London: Thomson Wadsworth.

Matsumoto, D., Wallbott, H. G., & Scherer, K. R. (1987). Emotions in intercultural communication. In M. K. Asante & W. B. Gudykunst (Eds.), *Handbook of international and cultural communication*. Newsbury Park, CA: Sage.

Mauro, R., Sato, K., & Tucker, J. (1992). The role of appraisal in human emotions: A cross-cultural study. *Journal of Personality and Social Psychology*, *62*, 301-317.

Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59*, 507-536.

Mayer, J. D., & Salovey, P. (1997). What is emotional intelligence? In P. Salovey & D. Sluyter (Eds.), *Emotional development and emotional intelligence: Educational implications* (3-31). New York: Basic Books.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) user's manual*. Toronto, Canada: Multi-Health Systems, Inc.

Mayer, J. D., Salovey, P., & Caruso, D. R. (2008). Emotional intelligence: New ability or eclectic traits? *American Psychologist*, *63*, 503-517.

Molander, B., & Arar, Lj. (1998). *Norms for 439 action events: Familiarity, emotionality, motor activity, and memorability.* Scandinavian Journal of Psychology, 39, 275-300.

Molander, B., & Arar, Lj. (2000). *Cross-cultural aspects on norms for action events.* International Journal of Psychology, 35, 354.

Murphy, K. R. (Ed.). (2006). *A critique of emotional intelligence: What are the problems and how can they be fixed?* London: LEA.

Myers, N. D., Wolfe, E. W., Feltz, D. L., & Penfield, R. D. (2006). Identifying differential item functioning of rating scale items with the Rasch model: An introduction and an application. *Measurement in Physical Education and Exercise Science*, *10*, 215-240.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory*. New York: McGraw-Hill.

Roberts, R. D., Schulze, R., Zeidner, M., & Matthews, G. (2005). Understanding, measuring, and applying emotional intelligence: What have we learned? What have we missed? In R. Schulze & R. D. Roberts (Eds), *Emotional intelligence: An international handbook* (pp. 311-341). Cambridge, MA: Hogrefe.

Russell, J. A. (1991). Culture and the categorization of emotions. *Psychological Bulletin*, *110*, 426-450.

Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, *115*, 102-141.

Scherer, K. R. (1997). The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology*, *73*, 902-922.

Scherer, K. R., Summerfield, A., & Wallbott, H. (1983). Cross-national research on antecedents and components of emotion: A progress report. *Social Science Information*, *22*, 355-385.

Schwartz, S. H., & Rubel, T. (2005). Sex differences in value priorities: Cross-cultural and multimethod studies. *Journal of Personality and Social Psychology, 89*, 1010-1028.

Smith, L. L. (2002). On the usefulness of item bias analysis to personality psychology. *Personality and Social Psychology Bulletin, 28*, 754-763.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.

Takšić, V. (1998). *Validacija konstrukta emocionalne inteligencije*. [Validation of the Emotional Intelligence Construct]. Unpublished doctoral dissertation, University of Zagreb, Croatia.

Takšić, V., Mohorić, T., Extremera, N., Fernández-Berrocal, P., Faria, L., Lima Santos, N., Räty, H., Molander, B., Holmström, S., Jansson, J., Avsec, A., & Toyota, H. (2009). *Confirmatory factor analysis of the Emotional Skills and Competences Questionnaire (ESCQ): Cross-cultural validation*. Unpublished manuscript.

Takšić, V., Tkalčić, M., & Brajković, S. (2001, July). *Emotional intelligence: An empirical validation of the construct*. Poster presented at the VIIth European Congress of Psychology, London.

Toyota, H., Morita, T., & Takšić, V. (2007). Development of a Japanese version of the Emotional Skills and Competence Questionnaire. *Perceptual and Motor Skills*, *105*, 469-476.

van de Vijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist, 1*, 89-99.

van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis of comparative research. In J. Berry, Y. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology: Vol. 1* (pp. 257-300). Needham Heights, MA: Allyn & Bacon.

Zeidner, M., Roberts, R. D., & Matthews, G. (2008). The science of emotional intelli-

gence: Current consensus and controversies. *European Psychologist*, *13*, 64-78.
Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modelling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense, Canada.

Prispelo/Received: 29.5.2009
Sprejeto/Accepted: 9.7.2009