Differential item functioning analysis of the Bar-On Emotional Quotient Inventory (EQ-i) across gender and age groups: an iterative hybrid ordinal logistic regression

Funcionamiento diferencial de los ítems del Inventario de Cociente Emocional de Bar-On (EQ-i): una regresión logística ordinal híbrida iterativa

Elahe Allahyari[1]
Department of Epidemiology and Biostatistics, School of Health,Medical Taxicology and Drug Abuse Research Center, Birjand University of Medical Sciences.

Fecha recepción: 4/08/2019 • Fecha aceptación: 20/11/2020

Para citar este artículo: Allahyari, E. (2020). Differential item functioning analysis of the Bar-On Emotional Quotient Inventory (EQ-i) across gender and age groups: an iterative hybrid ordinal logistic regression. *Psychologia*, *14*(2), 95-105
doi: 10.21500/19002386.4894

Abstract

There has been a heated debate on emotional intelligence (EI) and, more particularly, on the Bar-On Emotional Quotient Inventory (EQ-i) measuring all dimensions of emotional intelligence. To ensure measurement equivalence of EQ-i, the present article evaluated whether statements phrased in EQ-i questionnaire have equivalent meaning across respondents, regardless of their sex and age group membership. For 2,078 participants, three EI subscale (item 50 in reality testing, items 4 and 19 in stress tolerance, and items 7, 52, and 82 in interpersonal) for age groups had clinically significant Differential item functioning (DIF). So previous observed associations between EI and age might be misleading and deserve further study after removing or replacing DIF items.

Keywords: Differential item functioning (DIF), Emotional intelligence (EI), Gender, Age, The Bar-On Emotional Quotient Inventory (EQ-i)

1 elaheh.allahyari@gmail.com. https://orcid.org/0000-0003-3859-4820

Resumen

En medio del acalorado debate sobre la Inteligencia Emocional, este estudio retoma el Inventario de Cociente Emocional Bar-On (EQ-i), que mide todas las dimensiones de este constructo psicológico. Con el fin de comprobar la equivalencia de medición de EQ-i, se comprueba si las declaraciones formuladas en el cuestionario EQ-i tienen un significado equivalente entre los encuestados, independientemente de su sexo y grupo de edad. Se aplicó a los 2078 participantes las tres subescalas de IE. Se halló un funcionamiento diferencial de los ítems (DIF) clínicamente significativo. Por lo tanto, las asociaciones observadas anteriormente entre la IE y la edad pueden ser espurias y merecen un estudio adicional después de eliminar o reemplazar los elementos DIF.

Palabras clave: Funcionamiento diferencial de ítems (DIF), Inteligencia emocional (IE), Género, Edad, Inventario de cociente emocional Bar-On (EQ-i)

Introduction

The emotional quotient (EQ) or emotional intelligence (EI) was included all skills such as self-control, zeal, persistence, and the ability to motivate oneself. These skills enable people to behave at appropriate times and ways (Milhoan, 2007). For the greater part of the 20th century, studying emotional quotient has extensively increased in scientific circles as well as in the lay public because of the role it plays in person's effectiveness and success (Goleman, 1995). But EI has learned capability-based competence and affected by life experience (Weisinger, 1998). So measuring EI and improving weakness gives people more chance to use their hereditary talents or intelligence quotient (IQ) for success in life and work. To measure EI construct, various instruments are designed.

In 1935, the first instrument was designed to measure socially intelligent behavior in two subscales included Comprehension and Picture Arrangement (Doll, 1935). Then non-intellective factors were mentioned on intelligent behavior. And researchers gradually began to shift their attention to general intelligence: "The capacity of the individual to act purposefully" (Wechsler, 1958). For this purpose, they combine the emotional (intrapersonal intelligence) and social (interpersonal intelligence) components and introduce the Bar-On model which is one of the three major conceptual models in the Encyclopedia of Applied Psychology (Spielberger, 2004). The Bar-On Emotional Quotient Inventory (EQ-i) included the five key components of emotional intelligence from 1965 to the present (Bar-On, 2006). Bar-On selected 133 of 1000 items by factor analysis and evaluated this self-report questionnaire in 1996. Instrument validation was continued across different cultures. So that, there is more than 30 translation of this questionnaire now (Bar-On, 2006).

Ensuring acceptable validity and reliability is important before making comparisons among individuals or groups by any psychological test. But, the concept of validity is not especially clear in social and behavioral science. In these fields, the validity of obtained scores from a questionnaire inferred within content-related validity, criterion-related validity, and construct-related validity (Onwuegbuzie, Daniel, & Collins, 2009). The differential item functioning (DIF) will be a threat to construct-related validity as will threaten a valid interpretation of group differences (Clauser & Hambleton, 1994). DIF is present when an item has a different probability or likelihood to be endorsed for different groups of examinees, after controlling for EQ-i ability (Teresi, 2006).

Item response theory (IRT) analyses were used to detecting DIF items in Wong and Law Emotional Intelligence Scale (WLEIS), Schutte Self-Report Emotional Intelligence Test (SEIT), and Trait Emotional Intelligence Questionnaire (TEIQue) (Cho, Drasgow, & Cao, 2015; Karim, 2010). But there is currently a lack of studies surrounding DIF analysis of the Bar-On Emotional Quotient Inventory. Because the presence of DIF items can explain conflicting results in studies and reveal necessity of further research to replace or remove these items. Therefore, the aim of the present article was to evaluate whether is observed differences between

age and gender groups real or not depend on DIF items in the measurement process. So, DIF items of the Bar-On Emotional Quotient Inventory were detected by the hybrid OLR/IRT model.

One of the best techniques for detecting DIF was the ordinal logistic regression (OLR), which compared following regression models to examine ordinal polytomous items for DIF (Paul K Crane, Gibbons, Jolley, & van Belle, 2006).

$$\text{logit}\,[P(Y\leq k|g,\theta)]=\beta_{0k}+\beta_1\theta \qquad k=0,1,\ldots,m-1 \quad \text{(Model 1)}$$

$$\text{logit}\,[P(Y\leq k|g,\theta)]=\beta_{0k}+\beta_1\theta+\beta_2 g \qquad k=0,1,\ldots,m-1 \quad \text{(Model 2)}$$

$$\text{logit}\,[P(Y\leq k|g,\theta)]=\beta_{0k}+\beta_1\theta+\beta_2 g+\beta_3\theta g \qquad k=0,1,\ldots,m-1 \quad \text{(Model 3)}$$

In these formulas, m, θ, and g were assumed to be the number of domains, ability score, and grouping variable, respectively. There were two measures for DIF effect size, including Crane, van Belle and Larson criterion[2] (CvBL) and McFadden pseudo R-square[3] (ΔR^2) (P. K. Crane, van Belle, & Larson, 2004). The present work takes into account McFadden pseudo-R^2 more than 0.070 as large in both uniform and non-uniform DIF, and CvBL more than 0.01 as a practically meaningful uniform DIF.

One of the potential limitations of logistic regression DIF detection was the reliance on the observed sum score as the ability criterion. Therefore, if the test contains biased items, then a biased measure of the ability variable will be used for investigating DIF. So, DIF in one item can cause other items to show "pseudo-DIF" (Groenvold & Petersen, 2005). To overcome this potential limitation of OLR procedure, lordif package in R3.1.3 software incorporates the Rasch trait score rather than the sum of score ability and using an iterative procedure to detect DIF items (Choi, Gibbons, & Crane, 2011).

In order to do this, the first stage of DIF detection was done by the ordinal logistic regression model when the ability parameter is estimated by the IRT model for all groups combined. But, the separate IRT ability parameters for each group will be used for items that identified DIF in the previous step. Estimated IRT ability parameter of DIF items for one group is separately done in these steps by using the data of just that group and missing for all other groups. In all these stages, the OLR procedure determines DIF items again by new IRT ability parameters of DIF items and combined ability parameter of non-DIF items. This procedure will be repeated until the same items are flagged for DIF in two consecutive stages. Accordingly, despite existing biased items, the ability parameter will be unbiased for investigating DIF.

Material and Methods

In this article, those who were at least able to read and write in Persian can enter the study. They will provide explanations about the benefits of this questionnaire and its application. Then, informed consent is taken from the people who tended to the program. Participants were asked their demographic characteristics of age and gender as well as 90-item responses of EQ-i questionnaire. Data were collected from health centers, universities, and markets in 7 cities of Iran (Shiraz, Esfahan, Birjand, Tabriz, Sanandaj, Zahedan, and Gilan). The intention of this study was to recruit 300 persons (50 from health centers, 50 from markets, 100 from university students equal to girl and boy, and 100 university staff) between 16 and 65 years old in each city (total sample 2100). Therefore, we will have an acceptable ratio of women and men in all age groups which includes most of the ethnicities (Persians, Azerbaijanis, Lurs, Kurds, Baloch, and Turkmens) and social class (businessman, housewife, and government's employee). Before the analysis, 22 incomplete questionnaires were eliminated from the study, so we have 2078 persons available.

Results

Item responses were available from 2078 participants (833 Persians,210 Azerbaijanis, 249 Lurs, 243 Kurds, 271 Baloch, and 272 Turkmens) after eliminating the incomplete questionnaires. The data were clustered into 1096 non-academic (521 female, 575 male) and 982 academic degrees (522 female, 460 male). Respondents were between 16 and 65 years old, and the mean total age was 28.72 years (SD=7.76) in

2 CvBL=|[β1 (Model 1) - β1 (Model 2)]/β1 (Model 1)|

3 ΔR2(uniform DIF)=1-ln[L(Model 2)]/ln[L(Model 1)], ΔR2(non-uniform DIF)=1-ln[L(Model 3)]/ln[L(Model 2)]

female and 28.99 years (SD=7.89) in males. There were 345 (178 female, 167 male) subjects that were 20 years or younger, 923 (457 female, 466 male) were between 21 and 30 years old, and 810 (408 female, 402 male) were older than 30 years.

Total Cronbach alpha coefficient was 0.93, total coefficient omega was 0.94, and Moment Pearson Correlations were lower than 0.001 in all items. Root Mean Squared Error of Approximate (RMSEA), Standardized Root Mean Squared Residual (SRMR), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI) were 0.038, 0.056, 0.722, and 0.707 among all participants. These CFA results were respectively 0.041, 0.064, 0.693, and 0.677 for females and 0.039, 0.056, 0.714, and 0.70 for males. According to the cut-of values recommended, RMSEA and SRMR among all participants, and for each of the gender groups were in acceptable range, but TLI and CFI were always lower than 0.95 (Brown, 2014; Hu & Bentler, 1999).

DIF Analysis across Gender groups

Table 1 shows DIF detection results by the hybrid OLR/IRT model across gender groups (male vs female). This table reveals that with the exception of stress tolerance and impulse control all subscales have DIF (21 of 90 items). Eight subscales have one DIF item according to the χ^2 statistic (one for uniform, six for non-uniform and, and one for both of them), but none of their DIF magnitudes were significant (CvBL, between 0.0012 and 0.0015; ΔR^2, between 0.0008 and 0.0028). Other five subscales had more than one DIF items. In both self-regard and independence subscales, two items had a significant χ^2 statistic, compared with three items of assertiveness, empathy, and flexibility subscales. In all of these DIF items, McFadden pseudo-R^2 were between 0.0007 and 0.0072, whereas the CvBL was significant in item 54 of self-regard ($\Delta\beta_1$=0.0124) and item 3 of independence subscale ($\Delta\beta_1$=0.0156).

Table 1

The results of the hybrid OLR/IRT DIF analysis on the Bar-On Emotional Quotient Inventory (EQ-i) across men and women

	Non-uniform		uniform		
	χ^{2a}(P value)	ΔR^{2b}	χ^{2a}(P value)	ΔR^{2b}	$\Delta\beta_1{}^{c}$
Intrapersonal					
Emotional Self-Awareness					
66. Even when upset, I am aware of what`s happening to me.	**15.140 (0.0001)**	0.0026	0.002 (0.9633)	0.0000	0.0000
Assertiveness					
60. I am unable to express my ideas to others.	**13.830 (0.0002)**	0.0024	0.0131 (0.9089)	0.0000	0.0008
75. Others think that I lack assertiveness.	**15.138 (0.0000)**	0.0038	**8.030 (0.0046)**	0.0013	0.0092
90. It`s difficult for me to stand up for my rights.	**9.375 (0.0022)**	0.0014	**7.675 (0.0056)**	0.0012	0.0014
Self-Regard					
9. I believe that I can stay on top of tough situations.	**12.530 (0.0004)**	0.0023	**15.138 (0.0000)**	0.0034	0.0015
54. I generally expect things will turn out all right, despite setbacks from time to time.	**10.490 (0.0012)**	0.0019	**7.809 (0.0052)**	0.0015	**0.0124**
Self-Actualization					
56. I am impatient.	**15.140 (0.0001)**	0.0023	0.784 (0.3760)	0.0001	0.0006

	Non-uniform		uniform		
	χ^{2a}(P value)	ΔR^{2b}	χ^{2a}(P value)	ΔR^{2b}	$\Delta\beta_1{}^{c}$
Independence					
3. I prefer to join in which I'm told pretty much what I do.	1.866 (0.1719)	0.0003	**13.070 (0.0003)**	0.0022	**0.0156**
33. I prefer others to make decisions for me.	**5.386 (0.0203)**	0.0010	**6.764 (0.0093)**	0.0013	0.0052
Interpersonal					
Interpersonal Relationship					
7. I try to see things as they really are, without fantasizing or daydreaming about them.	**15.140 (0.0001)**	0.0028	0.028 (0.8679)	0.0000	0.0000
Empathy					
29. My friends can tell me intimate things about themselves.	0.2086 (0.6479)	0.0000	**12.530 (0.0004)**	0.0022	0.0006
44. I care what happens to other people.	**15.140 (0.0001)**	0.0028	0.684 (0.4082)	0.0001	0.0005
74. It is hard for me to see people suffer.	**15.138 (0.0000)**	0.0034	1.174 (0.2786)	0.0002	0.0011
Social Responsibility					
58. If I could get away with breaking the law in certain situations, I would.	**12.120 (0.0005)**	0.0020	3.431 (0.0640)	0.0006	0.0003
Adaptability					
Problem Solving					
31. When facing a problem, the first thing I do is stop and think.	**12.530 (0.0004)**	0.0022	2.586 (0.1078)	0.0004	0.0007
Flexibility					
12. It is difficult for me to begin new things.	**5.038 (0.0248)**	0.0008	**11.240 (0.0008)**	0.0018	0.0061
57. I am able to change old habits.	**4.019 (0.0450)**	0.0007	**4.872 (0.0273)**	0.0009	0.0057
87. It would be hard for me to adjust if I were forced to leave my home.	**15.138 (0.0000)**	0.0072	1.340 (0.2471)	0.0002	0.0037
Reality Testing					
65. I try to get as much as I can out of those things that I enjoy.	**11.780 (0.0006)**	0.0021	2.279 (0.1311)	0.0004	0.0059
General Mood					
Optimism					
83. I keep in touch with friends.	1.737 (0.1875)	0.0003	**6.051 (0.0139)**	0.0011	0.0012
Happiness					
62. I am fun to be with.	**4.936 (0.0263)**	0.0009	**4.642 (0.0312)**	0.0008	0.0015

Bold numbers represent the items showing uniform or non-uniform DIF

[a] value of the difference in -2 log likelihood of the Models 1 and 2, and Models 2 and 3 for testing uniform and non-uniform DIF, respectively

[b] the R^2 difference between the Models 1 and 2, and Models 2 and 3 for testing uniform and non-uniform DIF, respectively

[c] Crane, van Belle, and Larson criterion or $|[\beta_1 \text{ (Model 1)} - \beta_1 \text{ (Model 2)}]/\beta_1 \text{ (Model 1)}|$

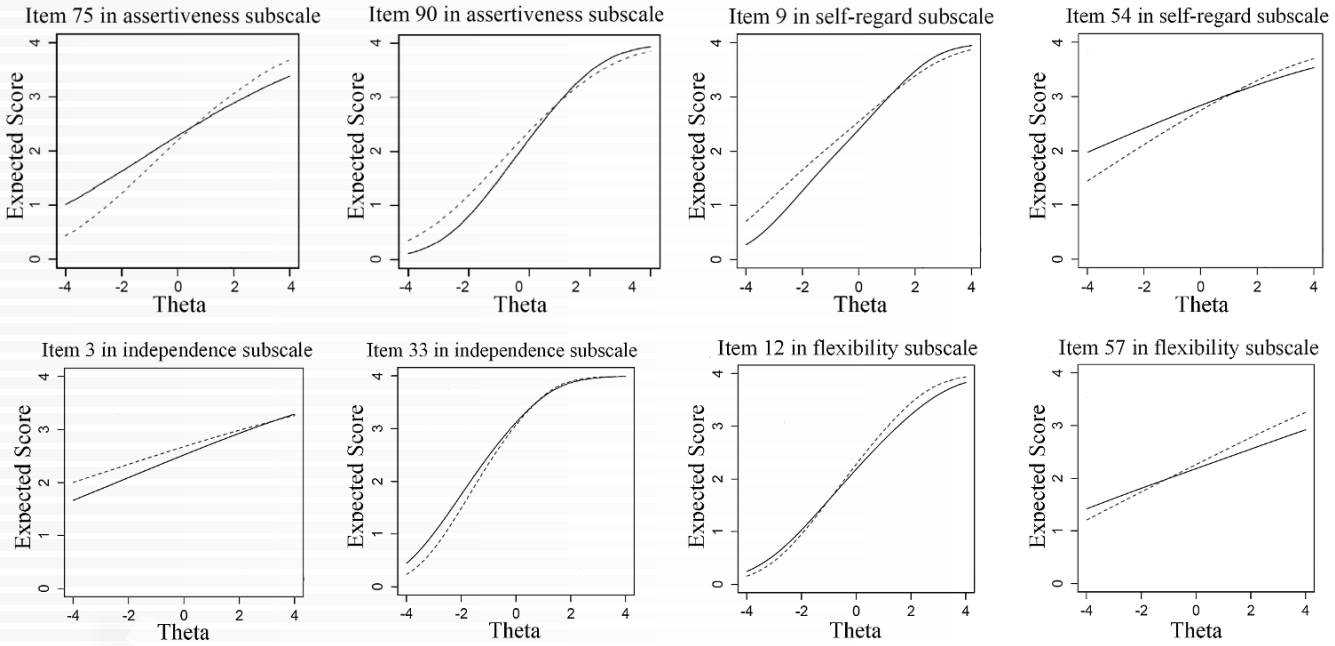

Fig. 1 test characteristic curve for DIF items for female (solid lines) and male (dashed lines) according to the hybrid OLR/IRT DIF analysis

When more than one item of dimension had uniform DIF, effect size may be canceled out at the domain level by another uniform DIF items in the opposite direction or may be increased at the domain level by another uniform DIF items in the favorable direction. However, only item 29 had uniform DIF ($\Delta\beta_1$=0.0006 and ΔR^2=0.0022) in empathy subscale, each subscale in the other four had 2 uniform DIF items which can affect each other. To assess the reaction in the DIF items visually, Figure 1 contains item score function for uniform DIF items in assertiveness, self-regard, independence, and flexibility subscales. The solid and dashed lines present female and males, respectively.

In the assertiveness subscale, items 75 and 90 not only had non-significant DIF magnitudes in table 1 ($\Delta\beta_1$=0.0092, 0.0014 and ΔR^2=0.0013, 0.0012) but also cancel each other out according to figure 1. In the self-regard and independence subscales, two uniform DIF items (items 9 and 54 for self-regard, items 3 and 33 for independence) had opposite direction. And in the flexibility subscale, DIF in items 12 and 57 was in the same direction but had not high magnitude to make significant additive effect.

DIF Analysis across Age Groups

DIF detection results across age groups are provided in Table 2. As the table clearly shows, 12 of 15 subscales had significant χ^2 statistic: the five of them had only one DIF item (one for uniform, two for non-uniform, and two for both of them), and the other seven had more than one. In the subscales that had one DIF item, only item 50 of reality testing shows a significant change in β_1 (CvBL=0.0119), whereas none of other four subscales had a significant $\Delta\beta$ (range 0.0009-0.0087) and ΔR^2 (range 0.0012-0.0082). In the subscales that had more than one DIF item, only the change in β_1 for item 60 of assertiveness and item 42 of flexibility.

Table 2.

The results of the hybrid OLR/IRT DIF analysis on the Bar-On Emotional Quotient Inventory (EQ-i) across age groups

	Non-uniform		uniform		
	χ^{2a}(P value)	ΔR^{2b}	χ^{2a}(P value)	ΔR^{2b}	$\Delta\beta_1{}^{c}$
Intrapersonal					
Self-Actualization					
56. I am impatient.	**5.483 (0.0192)**	0.0012	**15.140 (0.0001)**	0.0028	0.0087
Assertiveness					
30. When I disagree with someone, I am able to say so.	**15.140 (0.0001)**	0.0031	0.034 (0.8548)	0.0001	0.0002
45. It`s hard for me to say »no« when I want to.	**15.138 (0.0000)**	0.0104	**15.138 (0.0000)**	0.0047	0.0066
60. I am unable to express my ideas to others.	**8.426 (0.0037)**	0.0019	**9.550 (0.0020)**	0.0021	**0.0129**
75. Others think that I lack assertiveness.	**8.377 (0.0038)**	0.0018	**6.989 (0.0082)**	0.0016	0.0033
90. It`s difficult for me to stand up for my rights.	**6.465 (0.0110)**	0.0014	**6.653 (0.0099)**	0.0014	0.0044
Independence					
3. I prefer to join in which I am told pretty much what I do.	**11.020 (0.0009)**	0.0024	0.279 (0.5976)	0.0002	0.0016
Interpersonal					
Interpersonal Relationship					
7. I try to see things as they really are, without fantasizing or daydreaming about them.	0.3347 (0.5629)	0.0002	**13.830 (0.0002)**	0.0029	0.0026
22. People do not understand the way I think.	**15.138 (0.0000)**	0.0055	3.602 (0.0577)	0.0010	0.0038
37. I tend to fade out and lose contact with what happens around me.	**6.946 (0.0084)**	0.0016	1.147 (0.2842)	0.0004	0.0001
52. I get carried away with my imagination and fantasies.	**4.041 (0.0444)**	0.0010	**15.138 (0.0000)**	0.0066	0.0093
67. I tend to exaggerate.	**15.138 (0.0000)**	0.0147	0.424 (0.5151)	0.0002	0.0013
82. It is fairly difficult for me to express feelings.	**15.138 (0.0000)**	0.0050	**15.140 (0.0001)**	0.0032	0.0021
Empathy					
14. I am good at understanding the way other people feel.	0.0171 (0.8960)	0.0000	**5.270 (0.0217)**	0.0014	0.0036
89. I avoid hurting other people's feelings.	0.926 (0.3359)	0.0004	**15.138 (0.0000)**	0.0037	0.0042
Social Responsibility					
13. I like helping people.	**5.483 (0.0192)**	0.0017	**5.781 (0.0162)**	0.0018	0.0072
58. If I could get away with breaking the law in certain situations, I would.	**5.847 (0.0156)**	0.0014	**15.138 (0.0000)**	0.0047	0.0024
73. I am able to respect others.	3.051 (0.0807)	0.0010	**9.743 (0.0018)**	0.0024	0.0093
Adaptability					
Flexibility					
12. It is difficult for me to begin new things.	0.568 (0.4509)	0.0003	**8.195 (0.0042)**	0.0018	0.0072
27. It is difficult for me to change my opinion about things, once they are made.	**4.949 (0.0261)**	0.0012	**5.905 (0.0151)**	0.0014	0.0085
42. It is easy for me to adjust to new conditions.	0.817 (0.3660)	0.0003	**15.138 (0.0000)**	0.0041	**0.0144**

	Non-uniform		uniform		
	χ^{2a}(P value)	ΔR^{2b}	χ^{2a}(P value)	ΔR^{2b}	$\Delta\beta_1{}^{c}$
57. I am able to change old habits.	**15.138 (0.0000)**	0.0053	1.148 (0.2839)	0.0005	0.0041
Reality Testing					
50. I do not get that excited about my interests.	0.863 (0.3530)	0.0004	**15.138 (0.0000)**	0.0040	**0.0119**
Stress Management					
Stress Tolerance					
4. I can handle stress without getting nervous.	**9.550 (0.0020)**	0.0023	**12.532 (0.0004)**	0.0028	0.0003
19. I do hold up well under stress.	0.106 (0.7450)	0.0001	**11.489 (0.0007)**	0.0026	0.0031
34. I feel that it`s hard for me to control my anxiety	**15.138 (0.0000)**	0.0062	**4.872 (0.0273)**	0.0012	0.0052
49. I know how to keep calm in difficult situations.	**15.138 (0.0000)**	0.0061	1.045 (0.3067)	0.0004	0.0016
64. It`s hard for me to face unpleasant things.	**15.138 (0.0000)**	0.0042	0.108 (0.7429)	0.0001	0.0006
Impulse Control					
55. I feel comfortable with my body.	**15.138 (0.0000)**	0.0037	0.470 (0.4932)	0.0002	0.0002
General Mood					
Optimism					
23. I do not get along well with others.	**15.138 (0.0000)**	0.0082	**5.184 (0.0228)**	0.0014	0.0009
Happiness					
17. It is hard for me to smile.	2.869 (0.0903)	0.0009	**15.138 (0.0000)**	0.0039	0.0014
47. I am satisfied with my life.	**13.830 (0.0002)**	0.0031	**7.844 (0.0051)**	0.0019	0.0068
62. I am fun to be with.	**10.340 (0.0013)**	0.0023	**5.586 (0.0181)**	0.0014	0.0048
77. I get depressed.	**10.080 (0.0015)**	0.0023	0.010 (0.9223)	0.0000	0.0004

Bold numbers represent the items showing uniform or non-uniform DIF

[a] value of the difference in -2 log likelihood of the Models 1 and 2, and Models 2 and 3 for testing uniform and non-uniform DIF, respectively

[b] the R^2 difference between the Models 1 and 2, and Models 2 and 3 for testing uniform and non-uniform DIF, respectively

[c] Crane, van Belle, and Larson criterion or $|\beta_1$ (Model 1) - β_1 (Model 2)/β_1 (Model 1)$|$

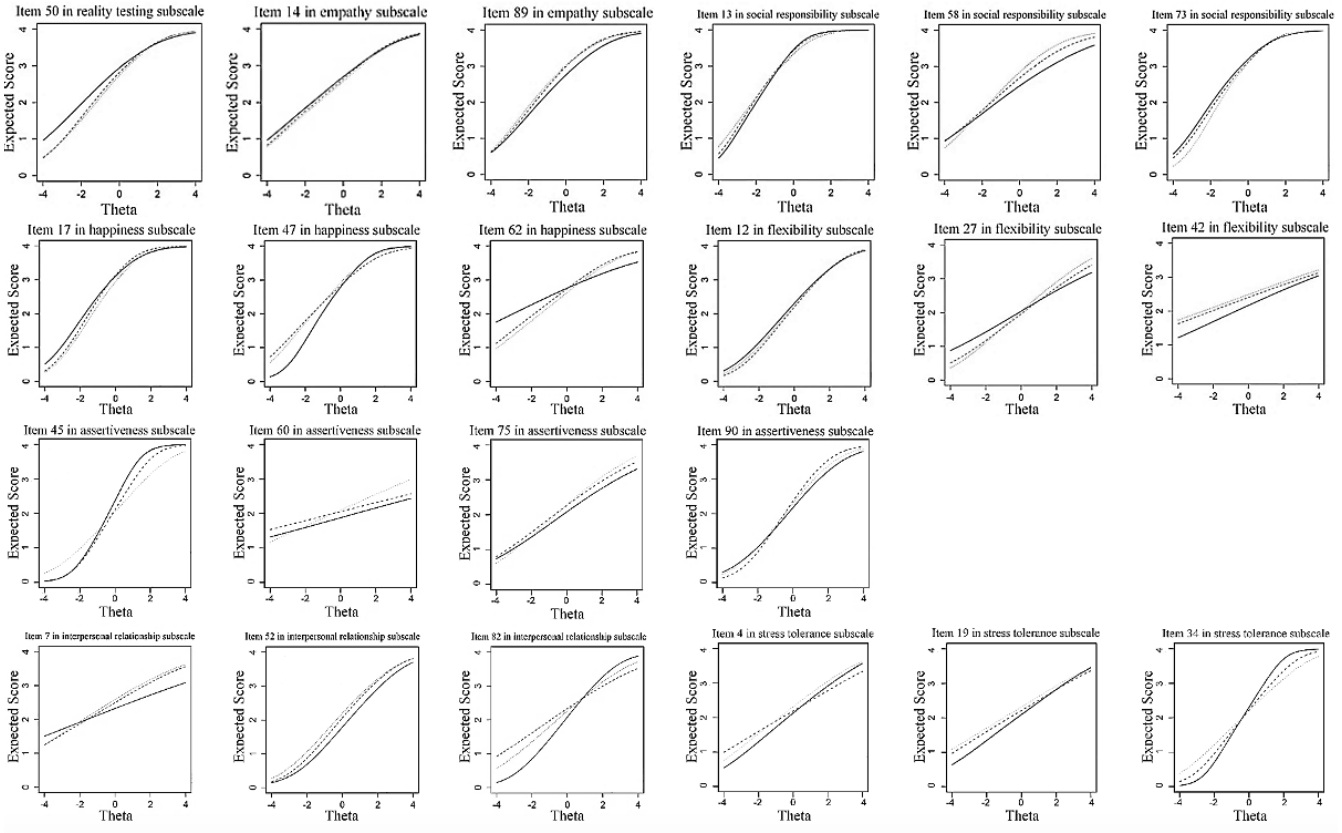

Fig. 2 test characteristic curve for DIF items for <=20 (solid lines), 21-30 (dashed lines) and >=31 (dotted lines) according to the hybrid OLR/IRT DIF analysis

0.0003 and 0.0093) and McFadden pseudo-R^2 (between 0.0010 and 0.0147) were not significant in the others. So, item score function should be drawn to show DIF direction in reality testing subscale and reaction in assertiveness, empathy, social responsibility, flexibility, stress tolerance, happiness, and interpersonal relationship subscales.

Figure 2 clearly portrays item score functions across age groups of <=20, 21-30, and >=31 by solid, dashed, and dotted lines, respectively. As the figure clearly shows, 20 years or younger people scored their reality testing higher than others. But, two items in empathy had opposite direction and cancelled out each other. Items 13 and 73 of social responsibility and items 47 and 62 of happiness subscale also cancel out each other. Therefore, items 58 and 17 which had not significant magnitude cannot affect age comparisons in these subscales.

Interpretation of findings was not so easy in assertiveness, flexibility, interpersonal relationship, and stress tolerance subscales. The direction of DIF items in flexibility subscale for items 12 and 27 was opposite with high magnitude item 42 as well as in assertiveness subscale for items 60, 75, and 90 with item 45. However uniform DIF had not high magnitude but in items 7, 52, and 82 of interpersonal relationship and 4 and 19 of stress tolerance subscale go in one direction in figure 2.

Discussion

The present work reveals that uniform DIF in item 50 misguided researchers on disparities when investigating the association between reality testing and age. Also, in comparing age groups, valid interpretation of group differences depended on uniform DIF in items 7, 52, and 82 of interpersonal relationship and items 4

and 19 of stress tolerance subscales. These findings may be one of following mentioned reasons.

The growing "current concerns" is responsible for more daydreams reported, while there is an inverse relationship between the time spent in daydreaming and the time spent on information processing (Kunzendorf & Wallace, 2000). Therefore, older people may exaggerate in making an evaluation of their interpersonal ability because of their attention-switching tasks, slow information processing, and fewer current concerns. Hence, it could explain why older people tend to evaluate their interpersonal skills higher than the younger counterparts in items "I try to see things as they really are, without fantasizing or daydreaming about them" and "I get carried away with my imagination and fantasies" where second item was scored inversely.

On the other hand, adolescents also tends to hide their real self and feelings behind an empty smile, forth fabrications, etc., because of the expectation of their parents, friends, and others (Elliott, 1982). They also try to present new images of the self. Therefore, the youth scored lower in the item "It`s fairly difficult for me to express feelings", whereas in items "I do not get that excited about my interests" young participants evaluated themselves higher, where both items were scored inversely.

Furthermore, elderly persons expected more negative implications and consequences during their lives because they spend a great deal of time thinking about loss of support, difficulty in making new friends, financial difficulties, being alone, etc. (Hansson, Jones, Carpenter, & Remondet, 1987). Thus, this counterproductive cognitive activity may explain why older people scored higher than younger ones in items "I can handle stress without getting nervous" and "I do hold up well under stress".

A problem with short scales, such as Bar-On EQ-i questionnaire, is the difficulty of identifying which item is causing the DIF because DIF in one item can cause a biased measure of test ability (Groenvold & Petersen, 2005). Hence, one of the strengths of this study is that lordif package uses an unbiased reduced test as the ability indicator in the final step, because of the latent variable item response theory and removed biased items from ability variable iteratively (Choi et al., 2011). On the other hand, this study focused on psychological studies and deciding what constitutes significant DIF depending on DIF thresholds. However, for detecting significant DIF, the present article is not limited to a significant effect size measure but has considered the interaction of the DIF items in one dimension and its effect on mean comparisons too.

Another strong point of our study is that these results could be recommended for the whole Iranian population. Forasmuch as, the sample included a variety of Iranian ethnic compositions (Persians, Turkic, Kurds, Baloch) so the Persian and Turkic groups had the first and second largest ethnicities, respectively. However, as far as we know, assessing measurement equivalence between or within cultures deserve further study to evaluate whether people in different cultures interpret statements phrased in Bar-On Emotional Quotient Inventory equally. Factors, such as the location and level of education, can affect people's perceptions because further research is also needed to test the measurement equivalence of the EQ-i instrument across these factors.

Conclusion

Although some studies have been published on the validity and reliability of the Bar-On EQ-i questionnaire, to the best of our knowledge, this is the first study that is based on DIF analysis. Dawada and Hart reported good item homogeneity and internal consistency in EQ-i subscales, but group difference is real when statements phrased in questionnaire have same interpretation across respondents irrespective of their group membership (Dawda & Hart, 2000). The present work reveals that the observed difference between groups on the composites and even total EQ-i score were dependent on DIF items in their subscales. Therefore, the relationships between interpersonal intelligence, adaptability, stress management, and total EQ-i with age might be misleading because of significant DIF in items 4 7, 19, 50, 52, and 82.

References

Bar-On, R. (2006). The Bar-On model of emotional-social intelligence (ESI). *Psicothema, 18 Suppl*, 13-25.

Cho, S., Drasgow, F., & Cao, M. (2015). An investigation of emotional intelligence measures using item response theory. *Psychol Assess, 27*(4), 1241.

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of statistical software, 39*(8), 1.

Clauser, B. E., & Hambleton, R. K. (1994). [Differential Item Functioning, P. W. Holland, H. Wainer]. *Journal of Educational measurement, 31*(1), 88-92.

Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical care, 44*(11), S115-S123.

Crane, P. K., van Belle, G., & Larson, E. B. (2004). Test bias in a cognitive test: differential item functioning in the CASI. *Stat Med, 23*(2), 241-256. doi: 10.1002/sim.1713

Dawda, D., & Hart, S. D. (2000). Assessing emotional intelligence: reliability and validity of the Bar-On Emotional Quotient Inventory (EQ-i) in university students. *Personality and individual differences, 28*(4), 797-812. doi: Doi 10.1016/S0191-8869(99)00139-7

Doll, E. A. (1935). A genetic scale of social maturity. *American Journal of Orthopsychiatry, 5*(2), 180.

Elliott, G. C. (1982). Self-esteem and self-presentation among the young as a function of age and gender. *Journal of youth and Adolescence, 11*(2), 135-153.

Goleman, D. (1995). *Emotional intelligence*. New York, NY: Bantam Books.

Groenvold, M., & Petersen, M. A. (2005). The role and use of differential item functioning (DIF) analysis of quality of life data from clinical trials. *Assessing quality of life in clinical trials*, 195-208.

Hansson, R. O., Jones, W. H., Carpenter, B. N., & Remondet, J. H. (1987). Loneliness and adjustment to old age. *The International Journal of Aging and Human Development, 24*(1), 41-53.

Karim, J. (2010). An item response theory analysis of Wong and Law emotional intelligence scale. *Procedia-Social and Behavioral Sciences, 2*(2), 4038-4047.

Kunzendorf, R. G., & Wallace, B. (2000). *Individual differences in conscious experience* (Vol. 20): John Benjamins Publishing.

Milhoan, P. L. (2007). *Emotional intelligence competencies of department chairs in the West Virginia State Community College System and their faculty members' perceptions of organizational climate*: West Virginia University.

Onwuegbuzie, A. J., Daniel, L. G., & Collins, K. M. T. (2009). A meta-validation model for assessing the score-validity of student teaching evaluations. *Quality & Quantity, 43*(2), 197-209. doi: 10.1007/s11135-007-9112-4

Spielberger, C. (2004). *Encyclopedia of applied psychology*: Academic Press.

Teresi, J. A. (2006). Overview of quantitative measurement methods. Equivalence, invariance, and differential item functioning in health applications. *Med Care, 44*(11 Suppl 3), S39-49. doi: 10.1097/01.mlr.0000245452.48613.45

Wechsler, D. (1958). The measurement and appraisal of adult intelligence.

Weisinger, H. (1998). *Emotional intelligence at work*. San Francisco, CA: Jossey-Bass Publishers.