Psychometric Study of the CES-D: Factor Analysis and DIF

Bruno D. Zumbo, Michaela N. Gelin, & Anita M. Hubley
University of British Columbia

A copy of the full paper can be found on the www at:
www.educ.ubc.ca/faculty/zumbo/ins2001/

Abstract: We report on a psychometric study of the Center for Epidemiologic Studies Depression (CES-D) scale with 600 community-dwelling adults between the ages of 17 and 87 years. The mean age for males is 46 years (N=310) and 42 years for females (N=290). We propose and test a one-factor measurement model with confirmatory factor analysis that takes into account method effects. The method effects represent the distinction between positively and negatively worded items. Also, we studied gender based differential item functioning (DIF) using a method proposed by Zumbo (1999). These DIF analyses were followed-up by nonparametric IRT DIF and differential test functioning. Our results indicate that the proposed measurement model fits and hence helps one understand the disparate literature on the factorial structure of the CES-D. This one factor model was also completely invariant (including method effects) across genders. With regard to the item level analyses investigating the DIF, "crying" and "eating" displayed gender DIF. This item-level DIF translates to substantial effects in scale score interpretation. This is the first study on the CES-D to have modeled the method effects in a one-factor measurement model, tested these method effects across genders, and to have examined gender DIF using Zumbo's method.

- The CES-D (Radloff, 1977) is a widely used self-report measure developed for use in studies exploring the epidemiology of depressive symptomatology in the general population.
- Few studies have examined both the item and scale level psychometric properties of the CES-D with a large community-dwelling sample.
- Used in numerous studies to: (a) compare the prevalence of depressive symptomatology between men and women, (b) select a non-depressive sample for a research study, or (c) split a sample into depressed and non-depressed groups for comparison on some other variable of interest.
- Community-Dwelling sample obtained in northern British Columbia as part of a survey with the Institute for Social Research and Evaluation at the University of Northern British Columbia:

Report

Your present age?

Sex	Mean	N	Std. Deviation	Minimum	Maximum
female	42.19	290	13.44	18	87
male	46.05	310	12.07	17	82
Total	44.19	600	12.88	17	87

Analyses:

- Confirmatory factor analysis of two models (see Figures 1 and 2) were tested with LISREL 8.30. Polychoric correlation matrix with generally weighted least squares estimation (and the asymptotic covariance matrix) was used.

- Model 2 (allowing for method effect due to item wording) fits well and is statistically better fitting than Model 1, (i.e., testing the nested models: Chi-squared df=6 of 205.2, p=0.000).

- As predicted by the psychometric work of Zimmerman, Zumbo, & Lalonde (1993) the reliability for Model 1 (i.e., when ignoring the correlated errors) will be inflated by the number of correlated errors and their magnitude. The six error correlations among the four positively worded items ranged from 0.21 to 0.37.

- Model 2 was fit for males and females using a simultaneous multi-group confirmatory factor analysis (LISREL 8.30) using the same methods as described above. It was found that the model fit equally well for males and females.

- Next, an item analysis was computed to investigate whether item level differential item functioning is present for males and females. As recently shown by Zumbo (in press) item-level DIF will not necessarily manifest itself in scale level analyses such as factor analysis. Zumbo's (1999) ordinal logistic regression method to detect differential item functioning was used.

- No items showed non-uniform DIF. Only the "eating" and "crying" items (#2 and #17) showed statistically significant uniform gender DIF. Females were 2.24 times more likely to respond with a higher item score on the "eating" item than males and 9.3 times more likely to respond with a higher item score on the "crying" item than males – in each case scoring in a more depressive fashion.

- Figures 3 and 4 show the non-parametric item response theory (IRT) analysis (TestGraf: Ramsey, 2000, 1991) used to understand the DIF. (Group 1= males and Group 2 = females)

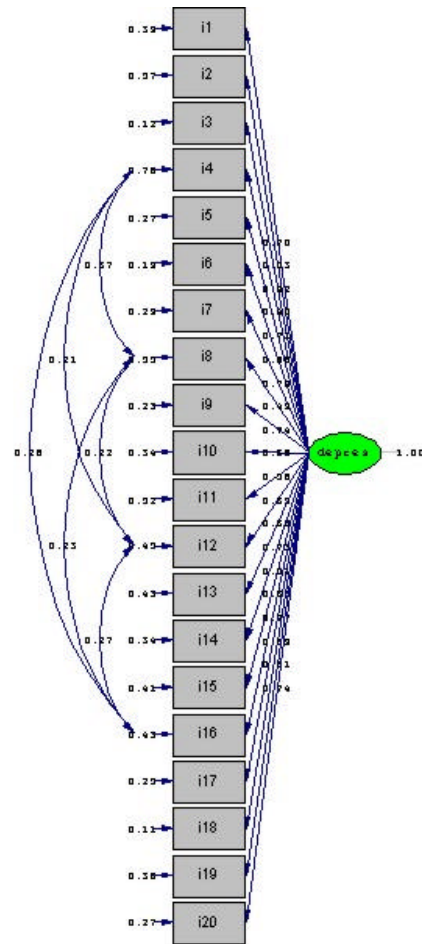# Figure 1.   Model 1: Single-factor model
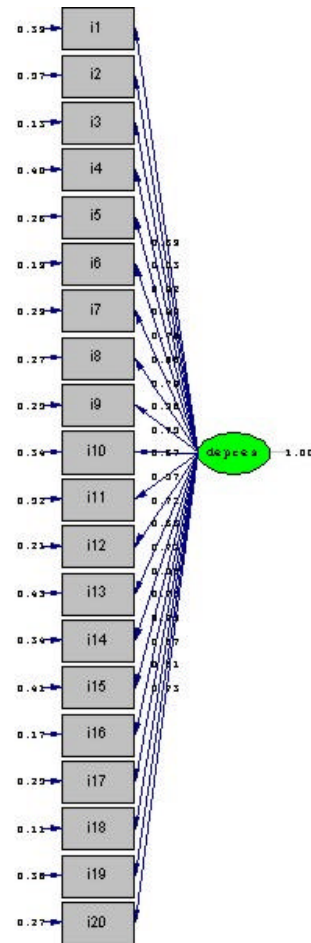
Root mean square error of approximation (RMSEA)=0.086
P-value for the test of close fit (RMSEA <0.05)= 0.00 ∴ Not an adequate fit of the model.
Reliability of 0.968

Figure 2. Model 2: Single-factor model allowing for a method effect

Root mean square error of approximation (RMSEA)=0.051
P-value for the test of close fit (RMSEA <0.05)= 0.33   ∴ Adequate fit of the model.
Reliability of 0.962

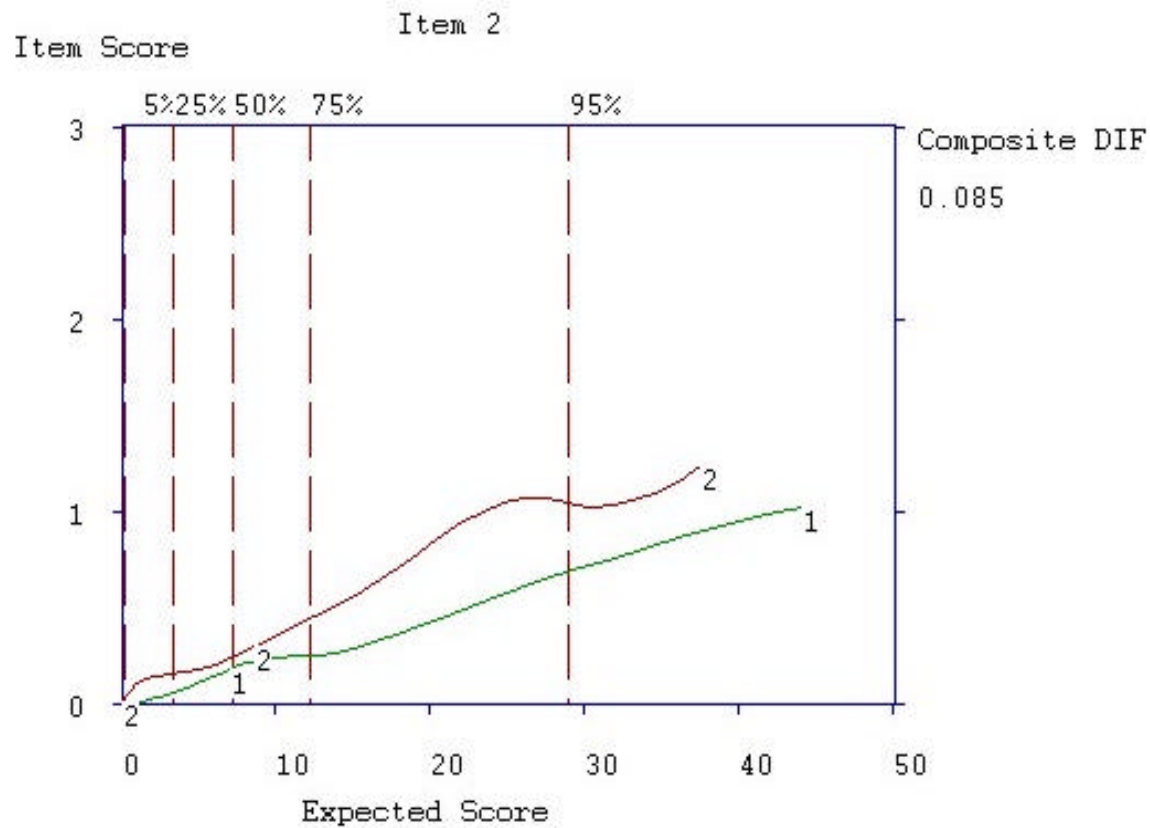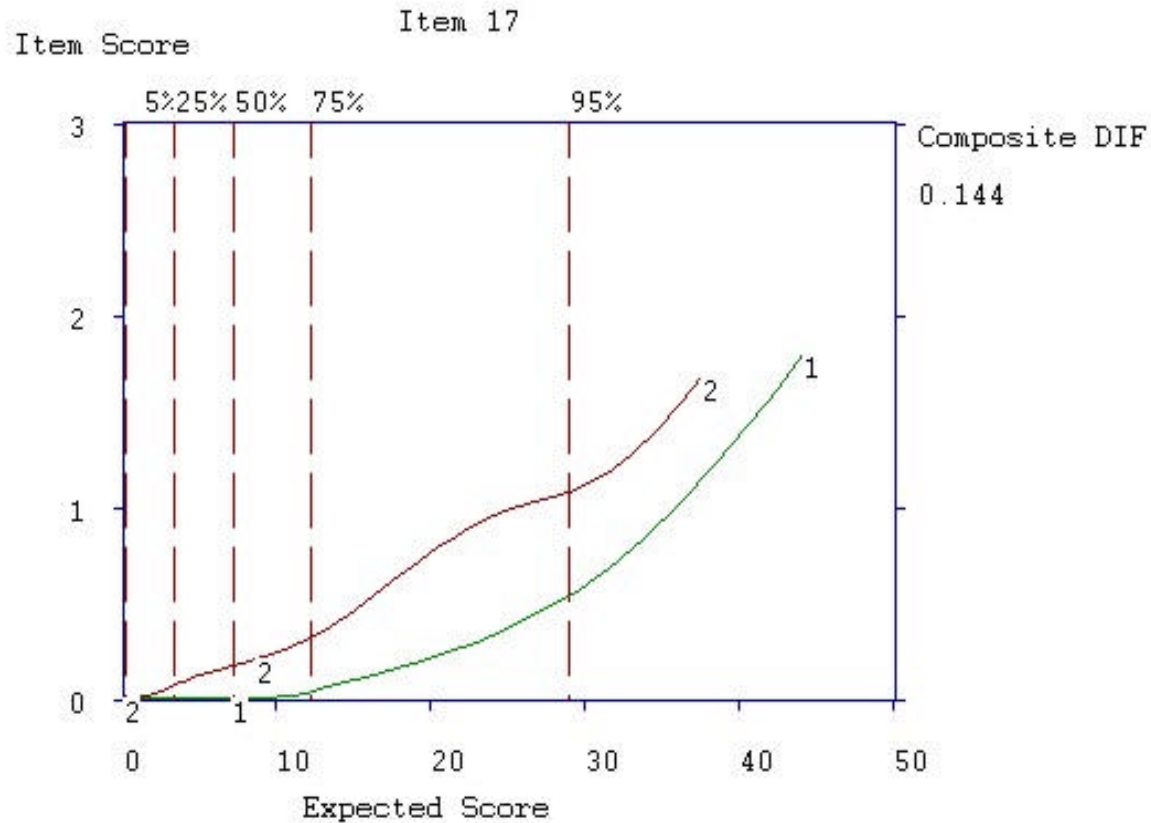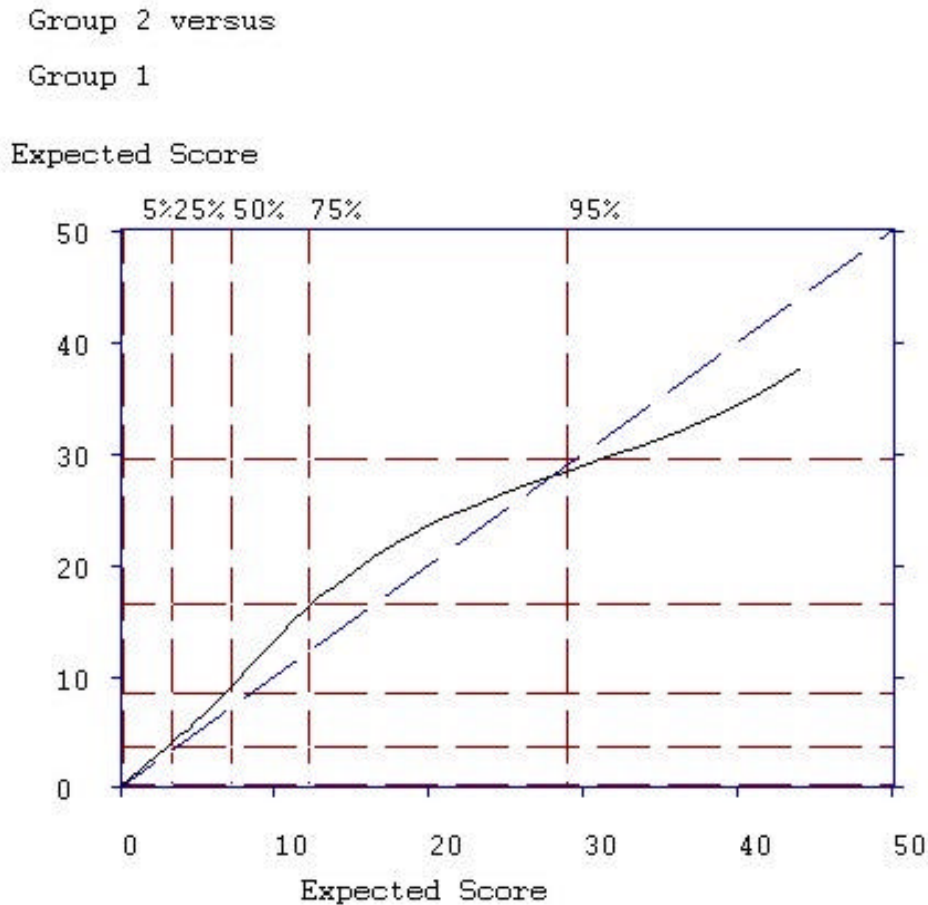# Figure 3. Nonparametric DIF Plots for the Two DIF Items

Figure 3 (cont'd.).

Note: Because the scale scores are clustered toward the lower end of the scale, the curves above the 95th percentile should not be considered in the conclusions (too little statistical information at that level of the continuum).

Figure 4. Nonparametric Differential Test Functioning Plot.

Note: The vertical axis = scores by females; horizontal axis = scores by males. Also, because the scale scores are clustered toward the lower end of the scale, the curve above the 95th percentile should not be considered in the conclusions (too little statistical information at that level of the continuum).

The CES-D Items.

For each statement, circle the number (see the guide below) to indicate how often you felt or behaved this way **during the past week**.

        0 = rarely or none of the time (less than 1 day)
        1 = some or a little of the time (1-2 days)
        2 = occasionally or a moderate amount of time (3-4 days)
        3 = most or all of the time (5-7 days)

	not even 1 day	1-2 days	3-4 days	5-7 days
1. I was bothered by things that usually don't bother me.	0	1	2	3
2. I did not feel like eating; my appetite was poor.	0	1	2	3
3. I felt that I could not shake off the blues even with help from my family or friends.	0	1	2	3
4. I felt that I was just as good as other people.	0	1	2	3
5. I had trouble keeping my mind on what I was doing.	0	1	2	3
6. I felt depressed.	0	1	2	3
7. I felt that everything I did was an effort.	0	1	2	3
8. I felt hopeful about the future.	0	1	2	3
9. I thought my life had been a failure.	0	1	2	3
10. I felt fearful.	0	1	2	3
11. My sleep was restless.	0	1	2	3
12. I was happy.	0	1	2	3
13. I talked less than usual.	0	1	2	3
14. I felt lonely.	0	1	2	3
15. People were unfriendly.	0	1	2	3
16. I enjoyed life.	0	1	2	3
17. I had crying spells.	0	1	2	3
18. I felt sad.	0	1	2	3
19. I felt that people dislike me.	0	1	2	3
20. I could not get "going".	0	1	2	3

Note: Items 4, 8, 12, and 16 were reverse coded