ORIGINAL PAPER

Rhonda Small · Judith Lumley · Jane Yelland · Stephanie Brown

The performance of the Edinburgh Postnatal Depression Scale in English speaking and non-English speaking populations in Australia

■ **Abstract** *Background* The Edinburgh Postnatal Depression Scale (EPDS) has been widely used to assess maternal depression following childbirth in a range of English speaking countries, and increasingly also in translation in non-English speaking ones. It has performed satisfactorily in most validation studies, has proved easy to administer, is acceptable to women, and rates of depression in the range of 10–20% have been consistently found. *Methods* The performance of the EPDS was compared across different population samples in Australia: (i) Women born in Australia or in another English speaking country who completed the EPDS in English as part of the 1994 postal Survey of Recent Mothers (SRM) 6–7 months after birth ($n = 1166$); (ii) Women born in non-English speaking countries who also completed the EPDS in English in the same survey ($n = 142$); and (iii) Women born in Vietnam ($n = 103$), Turkey ($n = 104$) and the Philippines ($n = 106$) who completed the EPDS 6–9 months after birth in translation in the Mothers in a New Country Study (MINC) study (total $n = 313$). The pattern of item responses on the EPDS was assessed in various ways across the samples and internal reliability coefficients were calculated. Exploratory factor analyses were also conducted to assess the similarity in the factor solutions across the samples. *Results* The EPDS had good construct validity and item endorsement by women was similar across the samples. Internal reliability of the scale was also very satisfactory with Cronbach's alpha for each sample being ≥8. Between 39 and 46% of the variance in each of the three main samples was accounted for by one principal factor 'depression' (6–7 items loading), with two supplementary factors 'loss of enjoyment' (2 items loading) and 'despair/self-harm' (2–3 items loading) accounting for a further 20–25% of the variance. Alternative one and two factor solutions also showed a great deal of consistency between the samples. *Conclusions* The good item consistency of the EPDS and the relative stability of the factor patterns across the samples are indicative that the scale is understood and completed in similar ways by women in these different English speaking and non-English speaking population groups. With the proviso that careful translation processes and extensive piloting of translations are always needed, these findings lend further support to the use of the EPDS in cross-cultural research on depression following childbirth.

■ **Key words** postnatal depression – Edinburgh Postnatal Depression Scale (EPDS) – cross-cultural research – psychometrics

Introduction

The measurement of depression across cultures has been much discussed in the last 20 years [e.g. 1–6] and issues raised about the validity of simply translating depression tools developed in one culture for use in another, a critical question being: Are standard depression assessment tools measuring the same thing in different cultures? The Edinburgh Postnatal Depression Scale (EPDS) [7] is a self-report measure widely used in research to assess probable maternal depression following childbirth. It has been used in a range of English speaking countries as well as in translation in non-English speaking ones. The EPDS has performed satisfactorily in most validation studies, in English and in translation, the scale has proved easy to administer, is acceptable to women and rates of depression in the range of 10–20% have been consistently found [8–16].

R. Small (✉) · J. Lumley · J. Yelland · S. Brown
Mother and Child Health Research
La Trobe University
251 Faraday Street
Carlton (VIC) 3053, Australia
E-Mail: r.small@latrobe.edu.au

This paper seeks to extend what is known about the performance of the EPDS, by analysing data from two Australian studies with largely representative samples, comprising women from widely divergent language and cultural backgrounds who completed the EPDS in English or in translation.

Methods

■ The studies

Two studies provide the data for this paper. The Survey of Recent Mothers (SRM) was a population-based postal survey of all women who gave birth in a 2 week period in 1993 in the state of Victoria, Australia ($n = 1366$) [17, 18]. The Mothers in a New Country (MINC) study was a companion interview study of immigrant women from Vietnam, Turkey and the Philippines who gave birth in Melbourne, Victoria between 1994 and 1996 ($n = 318$), with interviews conducted by bicultural interviewers in the language of women's choice [19, 20]. In both studies, women completed the EPDS – in English or in translation – providing a measure of probable clinical depression. On comparison with routinely collected data on all births in Victoria, both studies had largely representative obstetric samples in terms of parity, method of birth and infant birthweight: in SRM compared with all women giving birth, and in MINC compared with Vietnamese, Turkish and Filipino women giving birth, during the respective study periods [17, 19]. The MINC sample was also representative of all Vietnamese, Turkish and Filipino women giving birth for maternal age and marital status. As might be expected with a postal survey however, women under 25, single women and women born overseas of non-English speaking backgrounds were under-represented in SRM [17].

■ The samples

All participants with complete responses for the EPDS in both SRM ($n = 1310$) and MINC ($n = 313$) were included, forming five samples for comparison of the performance of the EPDS:

(i) Women born in Australia or in another English speaking country who completed the EPDS in English as part of SRM: *SRM–ESB, n = 1168*;
(ii) Women born in non-English speaking countries who also completed the EPDS in English in the same survey: *SRM–NESB: n = 142*.

Women participating in MINC who completed the EPDS mostly in translation:

(iii) Women born in Vietnam: *Vietnamese: n = 103*;
(iv) Women born in Turkey: *Turkish: n = 104*; and
(v) Women born in the Philippines: *Filipino: n = 106*.

For each of the analyses, the MINC sample was also considered as a whole ($n = 313$) for comparison with the two SRM samples.

■ The scale

The EPDS is a ten-item self-report measure specifically designed for use in the postpartum period. The use of most standardised depression measures has proved problematic at this time because of the presence of a range of somatic items (e.g. changes in sleep patterns, appetite, fatigue/lethargy) which many mothers will experience after the birth of a baby and which cloud the identification of women who are depressed. A frequent outcome of the use of such measures is an inflated proportion of women assessed as depressed [21]. In developing the EPDS, Cox describes the delib-

erate exclusion or adaptation of items, which might be inappropriately endorsed by women and a focus instead on the feeling states associated with depression [22]. Thus the item which relates to sleep on the EPDS states specifically '*I have felt so miserable that I have had trouble sleeping*,' thereby avoiding the difficulty caused by the fact that the usual trouble women have in sleeping at this time is due to the baby waking.

In the MINC study, the EPDS was translated into Vietnamese, Turkish and Tagalog using forward and back translations, a process of committee review, consultation with ethnic psychiatrists and piloting with monolingual and bilingual recent mothers in each community [23]. In response to each item on the EPDS, women score from 0 to 3, with total scores ranging from 0 to 30.

■ The research questions addressed

The performance of the EPDS in culturally diverse samples and in different languages (English, Vietnamese, Turkish and Tagalog) was examined to determine how similarly women from a range of backgrounds responded to and interpreted items on the scale. Six research questions were addressed in comparing the samples:

1. How similar are the score distributions on the EPDS?
2. Are there significant differences in the proportion of women scoring on each item?
3. How similar are the items on which women score most frequently?
4. How good is the internal reliability of the scale?
5. Is the underlying structure of the EPDS (as revealed by factor analysis) comparable?
6. Is the degree of similarity of factor patterns a function of the method used or of the data themselves?

Analysis

The performance of the EPDS across the samples was compared using a range of strategies. First the distribution of scores in the samples was examined (means, medians and standard deviations). The proportion of women reporting some level of symptomatology (defined as a score >0 on any of the ten items) was calculated and compared for all items of the EPDS in each sample using Chi-square as a test of differences in proportions between the samples.

Second, the internal consistency of scale performance in each sample was assessed, considering both item-total and inter-item Spearman's correlations. Cronbach's alpha co-efficients were calculated as summary measures of the internal reliability of the EPDS for all samples.

Third, an unrestricted exploratory factor analysis was conducted, using principal components analysis with varimax rotation, to assess and compare the factor structure of the scale in each sample. Given the potential for correlations between factors (quite likely given the common correlation between depression and anxiety) an oblique rotation was also undertaken and the factor patterns compared. This approach also served to explore research question 6 above: if the factor solutions remain stable with different rotation methods, the factor patterns can more readily be assumed to represent patterns in the data themselves, rather than being artefacts of the particular method adopted [24]. Comparison of the underlying factor structure for each sample was undertaken using Scree plots to examine the variance contributions of the factors derived in each sample. This initial factor analysis was followed by a decision to extract one, two and three factors in turn for each sample, in order to compare the final factor loadings across the samples for each solution derived, and to compare the findings with those of previous studies.

All analyses were undertaken using SPSS 9.0 [25] and Epi Info [26].

Table 1 EPDS score distribution statistics in all samples

Statistic	Viet	Turkish	Filipino	MINC	SRM–NESB	SRM–ESB
Mean	6.6	9.2	6.6	7.5	8.4	7.1
Standard deviation	4.5	5.1	4.0	4.8	5.7	5.1
Median	6.0	8.0	6.0	7.0	7.5	6.0
Range of scores	0–21	0–24	0–19	0–24	0–25	0–30

Results

■ Distribution of EPDS scores

The means and standard deviations, medians and the range of scores obtained on the EPDS in all samples are shown in Table 1.

■ Item analyses

In Table 2 the proportion of women in each sample scoring >0 on each item of the EPDS is shown. Of a total of 60 comparisons made within the MINC sample and between the MINC sample combined and the two SRM samples (Vietnamese:Turkish; Vietnamese:Filipino; Turkish:Filipino; MINC:SRM–NESB; MINC:SRM–ESB; SRM–NESB:SRM–ESB for each item), there were statistically significant differences in proportions on 15 ($P < 0.01$).

Table 3 places the items for each sample in descending order of frequency by which women scored >0 and demonstrates considerable similarity between the samples.

All items in all samples were significantly correlated with total scores (Table 4), with one exception, Item 10 in the Vietnamese sample (where all but two women scored zero). Inter-item correlations for each sample were also calculated (data not shown): in the combined MINC and SRM samples, all items were signif-

Table 2 Percentage of women reporting some level of symptomatology on each item of the EPDS (i.e. scoring >0) across all samples

EPDS Item*	Viet %	Turkish %	Filipino %	MINC sample overall %	SRM–NESB %	SRM–ESB %
1 Funny	39.8	36.5	17.9	30.6	44.4	38.0
2 Enjoy	36.9	39.4	31.8	36.0	43.0	34.7
3 Blame	67.0	84.6	59.8	70.4	73.2	82.0
4 Worry	63.1	77.9	67.3	69.4	65.4	67.9
5 Scared	53.4	59.6	61.3	58.1	49.3	41.7
6 Coping	75.7	76.9	69.2	74.9	80.3	80.4
7 Sleep	65.0	61.5	55.1	60.5	54.9	38.2
8 Sad	57.3	71.2	52.3	59.2	62.7	58.9
9 Crying	60.2	50.9	46.7	52.9	37.3	32.2
10 Harm	1.9	7.7	9.3	6.4	14.8	6.8

* EPDS Item statements (abbreviation): 1, I have been able to laugh and see the funny side of things (Funny); 2, I have looked forward with enjoyment to things (Enjoy); 3, I have blamed myself unnecessarily when things went wrong (Blame); 4, I have felt worried and anxious for no very good reason (Worry); 5, I have felt scared or panicky for no very good reason (Scared); 6, Things have been getting on top of me (Coping); 7, I have been so unhappy that I have had difficulty sleeping (Sleep); 8, I have felt sad or miserable (Sad); 9, I have been so unhappy that I have been crying (Crying); 10, The thought of harming myself has occurred to me (Harm)
Significant differences by item: Item 1, V:F $\chi^2 = 16.7$, $P < 0.001$; T:F $\chi^2 = 13.2$, $P < 0.001$; MINC:NESB $\chi^2 = 8.2$, $P = 0.004$; Item 2, no significant differences; Item 3, V:T $\chi^2 = 8.7$, $P = 0.003$; T:F $\chi^2 = 13.3$, $P < 0.001$; MINC:ESB $\chi^2 = 20.6$, $P < 0.001$; Item 4, V:T $\chi^2 = 5.4$, $P = 0.002$; Item 5, MINC:ESB $\chi^2 = 27.5$, $P < 0.001$; Item 6, no significant differences; Item 7, MINC:ESB $\chi^2 = 50.7$, $P < 0.001$; NESB:ESB $\chi^2 = 14.9$, $P < 0.001$; Item 8, T:F $\chi^2 = 7.9$, $P = 0.005$; Item 9, MINC:NESB $\chi^2 = 9.4$, $P = 0.002$; MINC:ESB $\chi^2 = 46.0$, $P < 0.001$; Item 10, MINC:NESB $\chi^2 = 8.4$, $P = 0.004$; NESB:ESB $\chi^2 = 11.6$, $P < 0.001$

Table 3 Items on the EPDS in descending order by the proportion of women scoring >0 on each item in each sample

Vietnamese	Turkish	Filipino	MINC	SRM–NESB	SRM–ESB
Coping	Blame	Coping	Coping	Coping	Blame
Blame	Coping	Worry	Blame	Blame	Coping
Sleep	Worry	Scared	Worry	Worry	Worry
Worry	Sad	Blame	Sleep	Sad	Sad
Crying	Sleep	Sleep	Sad	Sleep	Scared
Sad	Scared	Sad	Scared	Scared	Sleep
Scared	Crying	Crying	Crying	Funny	Funny
Funny	Enjoy	Enjoy	Enjoy	Enjoy	Enjoy
Enjoy	Funny	Funny	Funny	Crying	Crying
Harm	Harm	Harm	Harm	Harm	Harm

Table 4 EPDS item/total score correlations (Spearman's rho) across the samples

EPDS item	Vietnamese	Turkish	Filipino	MINC	SRM–NESB	SRM–ESB
1. Funny	**0.73**	**0.51**	**0.41**	**0.56**	**0.58**	**0.62**
2. Enjoy	**0.76**	**0.53**	**0.35**	**0.54**	**0.60**	**0.61**
3. Blame	**0.68**	**0.51**	**0.67**	**0.63**	**0.57**	**0.67**
4. Worry	**0.62**	**0.60**	**0.67**	**0.63**	**0.78**	**0.78**
5. Scared	**0.67**	**0.72**	**0.69**	**0.68**	**0.66**	**0.68**
6. Coping	**0.75**	**0.67**	**0.68**	**0.70**	**0.62**	**0.74**
7. Sleep	**0.68**	**0.69**	**0.69**	**0.67**	**0.70**	**0.69**
8. Sad	**0.79**	**0.67**	**0.73**	**0.74**	**0.76**	**0.79**
9. Crying	**0.72**	**0.61**	**0.62**	**0.62**	**0.61**	**0.66**
10. Harm	0.17	**0.42**	**0.40**	**0.38**	**0.46**	**0.35**
Correlations in bold above are significant at $P < 0.001$						
Mean item-total correlation	0.66	0.59	0.59	0.61	0.63	0.65

Table 5 Exploratory factor analysis: comparison of eigenvalues and amounts of variance explained in each sample for three factors extracted

Sample	Factor 1 'depression'	Factor 2 'loss of enjoyment'	Factor 3 'despair/self-harm'	Total amount of variance explained by the three factors
Vietnamese				
Eigenvalues	4.67	1.07	1.02	
Variance explained	46.7%	10.7%	10.2%	67.6%
Turkish				
Eigenvalues	3.61	1.36	1.06	
Variance explained	36.1%	13.6%	10.6%	60.3%
Filipino				
Eigenvalues	3.71	1.27	1.00	
Variance explained	37.1%	12.7%	10.0%	60.8%
MINC				
Eigenvalues	3.94	1.16	1.00	
Variance explained	39.4%	11.6%	10.0%	61.0%
SRM–NESB				
Eigenvalues	4.17	1.51	0.97	
Variance explained	41.7%	15.1%	9.7%	66.5%
SRM–ESB				
Eigenvalues	4.76	0.93	0.91	
Variance explained	47.6%	9.3%	9.1%	61.0%

icantly correlated with all other items on the scale, a sign of good construct validity across the samples.

Finally, the internal reliability of the EPDS demonstrated very good internal scale reliability in all samples, with Cronbach's alpha co-efficients ≥ 0.80 (range 0.80–0.87).

■ Exploratory factor analysis

Figure 1 shows the scree plots produced with prinicpal components analysis for all six samples. The similarity in the plots is striking, with almost identical patterns of variance distribution (represented by the eigenvalues) among factors in all samples.

The scree plots were also used to determine the number of factors to be extracted, using a combination of two approaches: Catell's 'eigenvalues > 1' rule and examination of the scree plots to determine the point at which the slope approaches zero (the 'elbow' in the plot) [27]. In all of the samples, one factor clearly accounted for a large proportion of the variance, followed by eigenvalues for the second and third factors 'hovering' around one, (with three just under one) and the elbows in each of the plots also occur-

ring at about these points. Various possibilities could be entertained: three, two or one factor solutions.

Table 5, displays for each sample the eigenvalues and the contribution made to the total variance by the first three factors identified. In each sample, these three factors explain >60% of the variance, exceeding the 50% recommended for a meaningful factor solution [28].

A description of how EPDS items loaded on the three factors for each sample is given in Table 6. For clarity, only loadings >0.3 (those generally agreed to be considered meaningful [28]) are shown. The pattern of items loading on the three factors appears least consistent between the three groups in the MINC sample, although in all three samples items 1 and 2 load together, as do items 3, 4 and 5, and items 9 and 10 load together for the Turkish and Filipino samples, but not the Vietnamese. The pattern of loadings for the Turkish and Filipino samples is actually very similar, with only items 6 and 7 'switching' between factor 1 and 2. The Vietnamese sample on the other hand, has a rather different range of items loading on each factor compared with the other two samples. It is possible however, that these differences are simply

Table 6 Item loadings on factors 1–3 in each sample

EPDS Items	Factor 1	Factor 2	Factor 3
Vietnamese			
1. Funny	**0.84**		
2. Enjoy	**0.83**		
3. Blame	**0.32**	**0.69**	
4. Worry		**0.76**	
5. Scared	0.34	**0.69**	
6. Coping	**0.75**	0.33	
7. Sleep	0.41	**0.62**	
8. Sad	**0.74**	0.40	
9. Crying	**0.71**		
10. Harm			**0.97**
Turkish			
1. Funny			**0.87**
2. Enjoy			**0.70**
3. Blame	**0.66**	−0.50	
4. Worry	**0.68**		
5. Scared	**0.73**		
6. Coping	**0.65**		
7. Sleep	**0.70**	0.38	
8. Sad	**0.50**		0.39
9. Crying		**0.75**	
10. Harm		**0.79**	
Filipino			
1. Funny			**0.65**
2. Enjoy			**0.85**
3. Blame	**0.71**		
4. Worry	**0.75**		
5. Scared	**0.83**		
6. Coping	0.30	**0.60**	0.39
7. Sleep	0.35	**0.69**	
8. Sad	**0.57**	0.49	
9. Crying		**0.83**	
10. Harm		**0.58**	
MINC			
1. Funny		**0.82**	
2. Enjoy		**0.78**	
3. Blame	**0.78**		
4. Worry	**0.70**		
5. Scared	**0.75**		
6. Coping	**0.47**	0.46	0.32
7. Sleep	**0.55**		0.54
8. Sad	**0.50**	0.46	0.32
9. Crying		0.40	**0.67**
10. Harm			**0.80**
SRM–NESB			
1. Funny			**0.87**
2. Enjoy			**0.86**
3. Blame	**0.73**		
4. Worry	**0.78**		
5. Scared	**0.65**	0.41	
6. Coping	**0.81**		
7. Sleep		**0.61**	0.38
8. Sad	0.47	**0.65**	
9. Crying		**0.75**	
10. Harm		**0.75**	
SRM–ESB			
1. Funny		**0.85**	
2. Enjoy		**0.81**	
3. Blame	**0.73**		
4. Worry	**0.80**		
5. Scared	**0.71**		
6. Coping	**0.56**	0.50	0.43
7. Sleep	**0.53**	0.40	
8. Sad		**0.48**	
9. Crying	0.45		**0.58**
10. Harm	0.44		**0.88**

due to random variation given the smallish sample sizes for each country of birth group.

Comparing the total MINC sample with the two SRM samples, there is much greater consistency in the loading patterns for each factor across the samples. All three have items 3, 4 and 5 loading on factor 1, items 1 and 2 load together on either factor 2 or 3, and items 9 and 10 load together, also on factor 2 or 3 in each sample. It needs to be remembered that the order of the factors depends on their eigenvalues and that for factors 2 and 3 these are mostly quite similar, so that 'switches' in item loadings are less significant between these two factors in the different samples.

Undertaking an oblique rotation of the factors to allow for potential correlation between factors [29] made only small differences to the pattern of factor loadings in any of the samples, with similar final solutions obtained to those obtained by orthogonal rotation. These data are therefore not presented, but they indicate that the underlying factor patterns are quite stable and likely to be a product of actual patterns in the data, rather than of the method of rotation.

■ One and two-factor solutions

When the initial factor analysis of the 13-item version of the EPDS was done by Cox, three items were removed to provide what was assumed to be a 10-item uni-dimensional measure of depression [22]. Subsequent factor analytic studies of the EPDS have resulted in a range of one, two and three-factor solutions being extracted (see Table 7).

To compare the findings in our samples with this previous work, and because of the somewhat subjective nature of decisions about how many factors to extract, the effect of extracting one and then two factors for each sample, was also examined (data not shown).

In our samples, the one-factor solution – representing a uni-dimensional model of depression – showed remarkable consistency in terms of the strength of item loadings (all >0.4) in all the samples, with only two exceptions: item 10 in the Vietnamese sample and item 2 in the Filipino sample which both loaded lower than 0.3. Given that only two Vietnamese women scored on item 10 and that item 2 had proved problematic in the Filipino translation, [23] these discrepancies are perhaps not surprising.

Similarly, a two-factor solution also demonstrated remarkable consistency across the samples, with items 3–9 mostly loading on factor 1 'depression'; and items 1, 2 and 10 'despair/self-harm' mostly loading on factor 2.

Discussion

Neither the Mothers in a New Country study nor the Survey of Recent Mothers was designed to contribute

Table 7 Published factor analyses of the EPDS

	Cox [22] 13-item pre-cursor to the English EPDS (n = 57?)	Pop et al. [34] 10-Item Dutch EPDS (n = 293) 4 weeks pp	Astbury et al. [35] 10-item English EPDS (n = 771) 6–9 months pp	Guedeney and Fermanian [37] 10-Item French EPDS (n = 87) Up to 4 months (ave of 7 weeks pp)	Des Rivieres-Pigeon et al. [38] 10-item French EPDS (Canada) (n = 224) 3–5 weeks pp	Berle et al. [36] 10-item Norwegian EPDS (n = 411) 6–12 weeks pp	Teissedre and Chabrol [39] 10-item French EPDS (n = 722) 4–6 weeks pp	Chabrol and Teissedre [40] 10-item French EPDS (n = 299) 2–3 days pp	Adouard et al. [41] 10-item French EPDS (n = 60) 28–34 weeks gestation
Factor extraction method used	Principal components	Principal components	Principal components	Principal components	Principal components	Principal components	Principal components	Principal components	Principal components
Variance accounted for by extracted factors	Factor 1: 46%	Factor 1: 38.5% Factor 2: 16.5% Factor 3: 11.1%	Factor 1: 38.9% Factor 2: 5.2%	Factor 1: 40.6% Factor 2: 12.7%	Factor 1: Factor 2: [variance accounted for not reported]	Factor 1: 46.6%	Factor 1: 40%	Factor 1: 28% Factor 2: 18% Factor 3: 17%	Factor 1: } Factor 2: } 62%
Rotation method used	? not stated	Orthogonal (Varimax)	Oblique (Oblimin)	Orthogonal (Varimax)	Oblique	Oblique	Orthogonal (Varimax)	Orthogonal (Varimax)	Orthogonal
Factor loadings for rotated factor solution	**Factor 1** ("depression") 10 items loaded highly **Factor 2** 2 irritability items and 1 mothering item loaded highly (these were subsequently dropped from the scale, resulting in the 10-Item EPDS now in common use	**Factor 1** ("depressive feelings") Item 1: 0.57 Item 2: 0.54 Item 7: 0.61 Item 8: 0.88 Item 9: 0.74 Item 10: 0.31 **Factor 2** ("cognitive anxiety") Item 3: 0.66 Item 4: 0.57 Item 5: 0.79 Item 6: 0.44	**Factor 1** ("depression") Item 1: 0.46 Item 2: 0.59 Item 6: 0.46 Item 7: 0.63 Item 8: 0.74 Item 9: 0.68 Item 10: 0.56 **Factor 2** ("anxiety") Item 3: 0.49 Item 4: 0.73 Item 5: 0.75	**Factor 1** ("depressive symptoms") Item 3: 0.55 Item 4: 0.71 Item 5: 0.78 Item 6: 0.62 Item 7: 0.60 Item 9: 0.69 **Factor 2** ("depressive mood") Item 1: 0.78 Item 2: 0.84 Item 8: 0.63 Item 10: 0.35	**Factor 1** ("symptoms of depression") Item 1 Item 2 (loadings not reported) Item 7 Item 8 Item 9 Item 10 **Factor 2** ("symptoms of anxiety") Item 3 Item 4 Item 5 Item 6	**Factor 1** ("depression") Item 1: 0.73 Item 2: 0.65 Item 3: 0.66 Item 4: 0.70 Item 5: 0.67 Item 6: 0.70 Item 7: 0.73 Item 8: 0.80 Item 9: 0.72 Item 10: 0.38	**Factor 1** ("depression") Item 1: 0.64 Item 2: 0.56 Item 3: 0.64 Item 4: 0.69 Item 5: 0.68 Item 6: 0.61 Item 7: 0.63 Item 8: 0.77 Item 9: 0.70 Item 10: 0.30	**Factor 1** ("anxiety") Item 3: 0.66 Item 4: 0.69 Item 5: 0.82 Item 6: 0.69 Item 7: 0.74 **Factor 2** ("depressive mood/ self harm") Item 8: 0.72 Item 9: 0.86 Item 10: 0.60 **Factor 3** ("anhedonia") Item 1: 0.76 Item 2: 0.78	**Factor 1** ("depression & other disorders, e.g. anxiety") Item 3: 0.63 Item 4: 0.55 Item 5: 0.79 Item 6: 0.73 Item 7: 0.52 Item 9: 0.61 Item 10: 0.77 **Factor 2** ("depression") Item 1: 0.92 Item 2: 0.69 Item 7 0.63 Item 8: 0.80 Item 9: 0.61
Confirmatory factor analysis					**Factor 1** ("symptoms of depression") Item 1: 0.75 Item 2: 0.78 Item 6: 0.47 Item 7: 0.86 Item 8: 0.91 Item 9: 1.0 Item 10: 0.95 **Factor 2** ("cognitive anxiety") Item 3: 0.47 Item 4: 1.0 Item 5: 1.0			Confirmatory factor analysis conducted, but final factors and item loadings not reported	
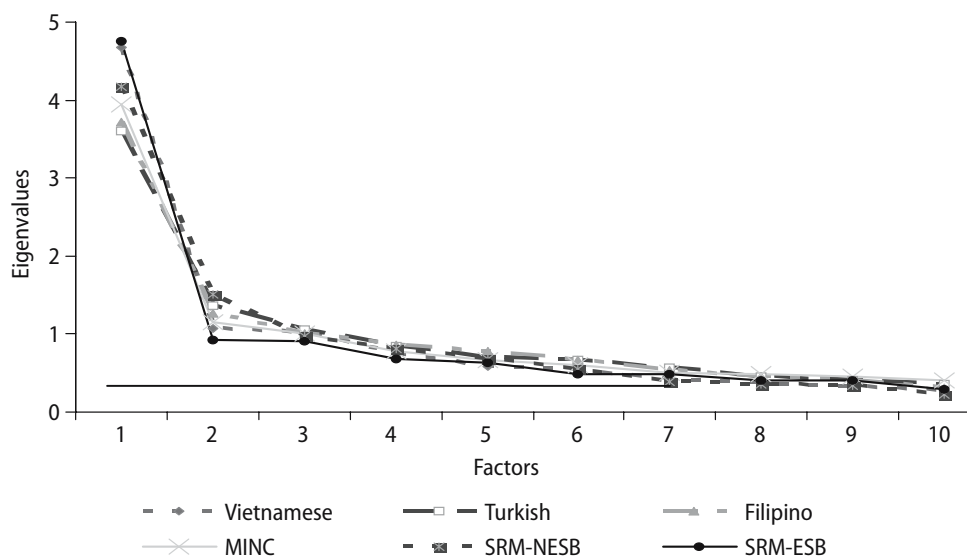
to the evidence about the validity of the EPDS in comparison to psychiatric interview in the way of traditional validation studies, but the findings presented in this paper can make a contribution of another kind. The strengths of the two studies for this investigation are their largely representative population samples and the significant involvement of women from a variety of backgrounds. By examining how the EPDS performs in the different populations represented in the two studies, it has been possible to investigate differences and similarities in the responses to the scale for women from very different language and cultural backgrounds, including when the EPDS is completed in English and in translation. This is the first time such an analysis has been reported.

It needs to be said that the strength of the approach taken here is best seen when similarities are found, despite the differences in women's backgrounds. Given a great deal of consistency, it is reasonable to assume that the EPDS is completed and understood in broadly similar ways by women regardless of their backgrounds. We cannot necessarily assume, however, that what the scale measures, completely captures the construct of depression for each culture. There may be domains of symptom expression missed, even if the ones covered by the scale appear cross-culturally stable and relevant.

Where differences are uncovered between samples, these are less readily explained. Any number of competing possibilities exist, including: a less than adequate scale translation (despite the care taken); cultural differences in symptom expression or preparedness to disclose feelings in a research interview or postal questionnaire, other differences in the nature of participating samples; differences in scale administration (postal in SRM, completed scale collected by the interviewer in MINC), and so on. It seems important to emphasise this point because of the temptation to interpret differences in findings in purely cultural terms, when other explanations might be equally plausible – a criticism that can reasonably be applied to much research into cultural differences [30–32].

What is revealed by the analyses of the EPDS from the data presented in this paper is a remarkable degree of similarity and consistency in women's responses to the scale across very different samples. Similar items were most commonly and least commonly 'endorsed' by women, patterns of inter-item and item-total correlations were broadly similar and the scale's internal reliability is more than adequate in all of the samples. The factor analyses from MINC and SRM reported here identified broadly similar underlying components to the scale across these culturally and linguistically divergent samples, and these patterns remained stable when one, two or three factors were extracted, and when a different rotation method allowing for factors to be correlated, was applied.

There are nine previously published reports of factor analyses undertaken with the EPDS (Table 7). Cox et al. reported a factor analysis of the 13-item predecessor to the EPDS, [22] and the first factor analysis of the 10-item EPDS was reported by Pop et al. in 1992 [34] using a Dutch version of the scale in a large follow up study on the prevalence of depression and thyroid dysfunction after childbirth. Astbury et al. [35] reported the first factor analysis of the English 10-item EPDS in a large Australian population sample ($n = 771$). A factor analysis was undertaken with a Norwegian version of the EPDS in a study of women attending postnatal visits ($n = 411$, Berle et al. [36]). Five factor analytic studies of French versions of the EPDS have now also been published. Guedeney and Fermanian conducted a study of 87 women in Paris (half were judged probably depressed by maternal and child health nurses and half were randomly selected) [37]. More recently, factor analyses of

the EPDS have been reported from Quebec (Des Rivieres-Pigeon et al.) [38] involving a selected sample of 224 women recruited from a perinatal health promotion program for women of low socio-economic status; there are two studies by Teissedre and Chabrol (*n* = 772 [39] and *n* = 299 [40]); and Adouard et al. [41] report a factor analysis from a small study (*n* = 60) using the EPDS in pregnancy.

Some methodological issues regarding factor analysis require comment. As Brislin points out, the first requirement for any correlational study is an adequate and representative sample in order to avoid problems of measurement error [29]. The Astbury et al. Australian study is the only previous study reporting a factor analysis of the EPDS that has utilised an unselected and largely representative population sample [35]. Second, there is little agreement about how to calculate appropriate sample sizes for factor analytic studies. Brislin's 'rough guide' to sample size is to square the number of variables and add 50, or as a bare minimum to have a sample size of at least 10 times the number of variables to be analysed [29]. For the 10-item EPDS this would mean a sample size of at least 100–150. Small samples may result in incorrect estimation of both the number of factors and the structure of the factors and the lack of theoretical basis for any of the rules of thumb about sample size has been noted [24]. Whatever guide is followed, at least three of the published studies shown in Table 7 [22, 37, 41] are likely to have been too small for reliable assessment of the factor structure of the EPDS.

Despite these methodological issues, and the fact that English and translated versions of the EPDS have been used, that time-points at which the EPDS was administered have been widely divergent and that two studies utilised a different method of factor rotation, the factor patterns found in these previous studies of the 10-item EPDS can be seen as quite similar. In most of the studies the exploratory factor analyses produced one principal component or factor that accounted for around 40% of the variance. After rotation, items commonly loading highly on this factor were: items 7, 8, 9 and 10 labelled 'depression' or 'depressive symptoms/feelings' and those loading together on either the second or third factors, accounting for less than 5–18% of the variance each, were: items 3, 4 and 5 often called 'anxiety'. Items 1, 2 (sometimes labelled 'anhedonia') and Item 6 loaded somewhat less consistently in the different studies. The differences seen in the number of factors extracted, the exact factor structures and the subsequent naming of factors are just as likely to result from some of the methodological issues described above, and possibly to the differing (and somewhat subjective) decisions by study investigators about the number of factors to extract, as they are to result from real differences in scale performance in the different studies. Our own findings indicated that one, two or three factor solutions all showed considerable consistency across the studied samples.

Conclusion

This analysis of the performance of the EPDS in five samples of women from culturally and linguistically diverse backgrounds in Australia provides little evidence to suggest the existence of major differences in the way Australian-born and immigrant women have responded to the scale. These findings lend further support to the use of appropriately translated and carefully piloted versions of the EPDS in cross-cultural research on depression following childbirth.

References

1. Littlewood R (1990) From categories to contexts: a decade of the 'new cross-cultural psychiatry'. Br J Psychiatry 156:308–327
2. Kleinman A (1977) Depression, somatization and the new 'cross-cultural psychiatry'. Soc Sci Med 11:3–10
3. Leff J (1990) The 'new cross-cultural psychiatry': a case of the baby and the bathwater. Br J Psychiatry 156:305–307
4. Kumar R (1994) Postnatal mental illness: a transcultural perspective. Soc Psychiatry Psychiatr Epidemiol 29:250–264
5. Canino G, Lewis-Fernandez R, Bravo M (1997) Methodological challenges in cross-cultural mental health research. Transcult Psychiatry 34:163–184
6. O'Hara MW (1994) Postpartum depression: identification and measurement in a cross-cultural context. In: Cox J, Holden J (eds) Perinatal psychiatry. The use and misuse of the Edinburgh Postnatal Depression Scale. Gaskell, London
7. Cox JL, Holden J, Sagovsky R (1987) Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 150:782–786
8. Murray L, Carothers AD (1990) The validation of the Edinburgh Postnatal Depression Scale on a community sample. Br J Psychiatry 157:288–290
9. Boyce PM, Stubbs R, Todd A (1993) The Edinburgh Postnatal Depression Scale: validation in an Australian sample. Aust NZ J Psychiatry 27:472–476
10. Thompson WM, Harris B, Lazarus J Richards C (1998) A comparison of the performance of rating scales used in the diagnosis of postnatal depression. Acta Psychiatr Scand 98:224–227
11. Jadresic E, Araya R, Jara C (1995) Validation of the Edinburgh Postnatal Depression Scale in Chilean postpartum women. J Psychosom Obstet Gynaecol 16:187–191
12. Lee DTS, Yip SK, Chiu HFK, Leung TYS, Chan KPM, Chau IOL, Leung HCM, Chung TKH (1998) Detecting postnatal depression in Chinese women. Validation of the Chinese version of the Edinburgh Postnatal Depression Scale. Br J Psychiatry 172:433–437
13. Benvenuti P, Ferrara M, Niccolai C, Valoriani V, Cox JL (1999) The Edinburgh Postnatal Depression Scale: validation for an Italian sample. J Affect Disord 53:137–141
14. Wickberg B, Hwang CP (1996) The Edinburgh Postnatal Depression Scale: Validation on a Swedish community sample. Acta Psychiatr Scand 94:181–184
15. Ghubash R, Abou-Saleh MT, Daradkeh TK (1997) The validity of the Arabic Postnatal Depression Scale. Soc Psychiatry Psychiatr Epidemiol 32:474–476
16. Eberhard-Gran M, Eskild A, Tambs K, Opjordsmoen S, Samuelsen SO (2001) Review of validation studies of the Edinburgh Postnatal Depression Scale. Acta Psychiatr Scand 104:243–249

17. Brown S, Lumley J (1997) The 1993 Survey of Recent Mothers: Issues in survey design, analysis and influencing policy. Int J Qual Health Care 9:265–275
18. Brown S, Lumley J (1998) Maternal health after childbirth: results of an Australian population based survey. BJOG 105:156–161
19. Small R, Lumley J, Yelland J, Liamputtong Rice P (1998) Shared antenatal care fails to rate well with women of non-English speaking backgrounds. Med J Aust 168:15–18
20. Small R, Lumley J, Yelland J (2003) Cross-cultural experiences of maternal depression: associations and contributing factors for Vietnamese, Turkish and Filipino immigrant women in Victoria, Australia. Ethnicity Health 8:189–206
21. Harris B, Huckle P, Thomas R, John S, Fung H (1989) The use of rating scales to identify postnatal depression. Br J Psychiatry 154:813–817
22. Cox JL (1986) Postnatal Depression A Guide for Health Professionals. Churchill Livingstone, Edinburgh
23. Small R, Lumley J, Yelland J, Rice PL, Cotronei V, Warren R (1999) Cross-cultural research: trying to do it better. 2. Enhancing data quality. Aust NZ J Public Health 33:390–395
24. Fayers PM, Machin D (1998) Factor analysis. In: Staquet MJ, Hays RD, Fayers PM (eds) Quality of Life Assessment in Clinical Trials: Methods and Practice. OUP, Oxford, pp 191–223
25. SPSS for Windows (1998) Version 9.0 for Windows. Chicago, Illinois: SPSS Inc
26. Dean AG, Dean JA, Coulombier D, et al (1994) Epi Info, Version 6: A word processing, database and statistics program for epidemiology on microcomputers. Centers for Disease Control and Prevention, Atlanta, Georgia
27. Floyd FJ, Widaman KF (1995) Factor analysis in the development and refinement of clinical assessment instruments. Psychol Assess 7:286–299
28. Streiner DL (1994) Figuring out factors: The use and misuse of factor analysis. Can J Psychiatry 39:135–140
29. Brislin RW (1973) Factor analysis. In: Brislin RW, Lonner E, Thorndike RM (eds) Cross Cultural Research Methods. John Wiley and Sons, New York 255–287
30. Ahmad WIU (1996) The trouble with culture. In: Kelleher D, Hillier S (eds) Researching Cultural Differences in Health. Routledge, London, pp 190–219
31. Stubbs P (1993) 'Ethnically sensitive' or 'anti-racist'? Models for health research and service delivery. In: Ahmad WIU (ed) 'Race' and Health in Contemporary Britain. Open University Press, Buckingham, pp 34–50
32. Smaje C (1996) The ethnic patterning of health: New directions for theory and research. Sociol Health Illness 18:139–171
33. Watkins D (1989) The role of confirmatory factor analysis in cross-cultural research. Int J Psychol 24:685–701
34. Pop VJ, Komproe IH, van Son MJ (1992) Characteristics of the Edinburgh Postnatal Depression Scale in The Netherlands. J Affect Disord 26:105–110
35. Astbury J, Brown S, Lumley J, Small R (1994) Birth events, birth experiences and social differences in postnatal depression. Aust J Public Health 18:176–184
36. Berle JØ, Aarre TF, Mykeltun A, Dahl AA, Holsten F (2003) Screening for postnatal depression. Validation of the Norwegian version of the Edinburgh Postnatal Depression Scale, and assessment of risk factors for postnatal depression. J Affect Disord 76:151–156
37. Guedenay N, Fermanian J (1998) Validation study of the French version of the Edinburgh Postnatal Depression Scale (EPDS): new results and psychometric properties. Eur Psychiatry 13:83–89
38. Des Rivières-Pigeon C, Séguin L, Brodeur J-M, Perreault M, Boyer G, Colin C (2000) L' Échelle de Dépression Postnatale d' Édimbourg: validité au Québec auprès de femmes de statut socio-économique faible. Revue Canadienne de Santé Mentale Communautaire 19:201–214
39. Teissedre F, Chabrol H (2004) Étude de l'EPDS (Échelle post-natale d'Édinbourg) chez 859 mères: dépistage des mères à risque de developper une dépression du post-partum. L'Éncephale 30:376–381
40. Chabrol H, Teissedre F (2004) Relation between the Edinburgh Postnatal Depression Scale scores at 2–3 days and 4–6 weeks postpartum. J Reprod Infant Psychol 22:33–39
41. Adouard F, Glangeaud-Freudenthal NMC (2005) Validation of the Edinburgh Postnatal Depression Scale (EPDS) in a sample of women with high risk pregnancies in France. Arch Womens Ment Health 8:89–95