Journal of Adolescence 36 (2013) 181–190

Contents lists available at SciVerse ScienceDirect

Journal of Adolescence

journal homepage: www.elsevier.com/locate/jado

The factorial invariance of the CES-D during adolescence: Are symptom profiles for depression stable across gender and time?

Marjolein Verhoeven [a,*], Michael G. Sawyer [b,1], Susan H. Spence [c,2]

[a] Department of Educational Sciences, Utrecht University, P.O. Box 80144, 3508 TC Utrecht, The Netherlands
[b] Research and Evaluation Unit, Children, Youth and women's Health Service, 72 King William Road, North Adelaide, South Australia 5006, Australia
[c] 2.17, Bray Centre, Nathan Campus, Griffith University, QLD 4111, Australia

Keywords:
Depression
Factorial invariance
Adolescence
Gender
CES-D

A B S T R A C T

This study examined the factorial invariance of the Center for Epidemiologic Studies of Depression Scale (CES-D) across gender and time during adolescence. The factor structure of the CES-D was compared at four annual measurement waves in a community sample of 2650 Australian adolescents. Confirmatory factor analyses showed that the factor structure of the CES-D was generally invariant across gender and time. However, gender differences were found on three items: for all waves the item 'I had crying spells' was a stronger indicator for depressive affect in females than males. On the final three waves the item 'people were unfriendly' loaded significantly higher on the factor 'Interpersonal Relations' for males than females. On Wave 2 and 3 males interpreted the item 'everything I did was an effort' with a positive connotation, whereas females interpret it with a negative association. These gender-differences are discussed from both a theoretical and a methodological perspective.

In both the DSM-IV and ICD-10, the symptom criteria for Major Depressive Disorders are the same for males and females (APA, 2000; WHO, 1992), assuming that the presentation of depression is equivalent across genders. Furthermore, although DSM-IV identifies symptoms such as somatic complaints, irritability, and social withdrawal as being particularly common in children (APA, 2000; p. 354), the same criteria are generally utilized to identify Depressive Disorders in all age groups. Similar issues apply to self-report questionnaires designed to identify and track changes in the level of depressive symptoms experienced by individuals in the general population (e.g., Radloff, 1977). Typically, individual items on questionnaires are summed and the total score is used to rate levels of depressive symptoms. Although some questionnaires provide different cutoff scores for classifying severe depression in males and females, this approach does not address the possibility that profiles of depressive symptoms may vary for different gender or age groups.

If valid comparisons are to be made across genders, or over time using results from a measure of depressive symptoms, then it is essential that the psychometric characteristics of the measurement instrument are the same for males and females, and at different points in time. In particular, it is important to establish that the instrument is factorially invariant across gender and time in terms of the number of underlying factors, the items that load upon those factors and the pattern of relationships among the underlying factors (Byrne, Baron, & Campbell, 1993). If construct measurement varies across gender

* Corresponding author. Tel.: +31 (0) 30 253 7681; fax: +31 (0) 30 253 4601.
E-mail addresses: J.C.T.Verhoeven@uu.nl (M. Verhoeven), michael.sawyer@adelaide.edu.au (M.G. Sawyer), s.spence@griffith.edu.au (S.H. Spence).
[1] Tel.: +61 (08) 8161 7207.
[2] Tel.: +61 7 3735 6427; fax: +61 7 3735 7507.

0140-1971/$ – see front matter
http://dx.doi.org/10.1016/j.adolescence.2012.10.007

and/or age, comparisons between different groups and time points cannot be unambiguously interpreted, as different constructs may have been measured. Despite the fundamental importance of factorial invariance, few studies have examined this issue for depressive symptoms, especially in adolescents (Motl, Dishman, Birnbaum, & Lytle, 2005).

There are, however, indications that depression might be best described by different symptoms for males and females. To illustrate, depressed girls reported more guilt, body image dissatisfaction, self-blame, self-disappointment, concentration problems, difficulty working, sadness/depressed mood, sleep problems, fatigue, and health worries than depressed boys in a study among adolescents referred to a pediatric depression clinic. Depressed boys, on the other hand, tended to report higher rates of anhedonia, depressed morning mood, and morning fatigue (Bennett, Ambrosini, Kudes, Metz, & Rabinovich, 2005). A study by Fu-I and Wang (2008) among clinically depressed children (5–9 years old) and adolescents (10–17 years old) established that female adolescents had lower self-esteem, whereas males showed more decreased concentration. There is also evidence for gender differences in symptoms of depression in non-clinical populations. A previous study using a self-report questionnaire on depressive symptoms reported that in a community based sample of adolescents, males more commonly reported symptoms such as social withdrawal, insomnia, and somatic preoccupation, whereas females tended to endorse crying, body image distortion, self-blame, self-dislike and loss of appetite (Baron & Joly, 1988). Similar results were found in a study on a mildly to moderately depressed sample of 18–19 years-old students (Hammen & Padesky, 1977).

In addition, there are indications that profiles of depression symptoms may change over time. To illustrate, the DSM-IV identifies certain depressive symptoms as being particularly common in children, compared to adolescents and adults (APA, 2000; p. 354). For example, Fu-I and Wang (2008), found that compared to clinically depressed children (5–9 years), clinically depressed adolescents (10–17 years) reported significantly more depressed mood, lower self-esteem, and more concentration problems. Differences in the patterns of depressive symptoms are also evident during the adolescent years. Kovacs, Obrosky, and Sherrill (2003) found that, from early to late adolescence, reduced appetite was increasingly likely across time as a symptom of depression both in girls and boys and was more prevalent among girls. In contrast, negative body image became less likely as a symptom of depression with age. There was also an age-related rise in the prevelance of hypersomnia as a depressive symptom among adolescents.

Not only do the adolescent years coincide with known increases in the prevalence of depressive symptoms (e.g., Garber, Keiley, & Martin, 2002), significant cognitive and neuropsychological development also take place during this developmental phase. This has the potential to change the ways in which adolescents construct complex and generalized perceptions of 'self' (Ashman & Dawson, 2002; Harter, 1999, 2003; Huttenlocher & Dabholkar, 1997). Frontal lobe development, the capacity for more integrative thought, and the ability to generalize beyond specific times and events all progress well into adolescence and may underlie the emergence of more stable self-perceptions (Cole & Martin, 2005). In addition, changes in autonomy and the way in which adolescents perceive and experience relationships (within the family, peers, and romantic; e.g., Connolly & Goldberg, 1999; Steinberg, 2001), biological changes associated with puberty and maturation including sexual maturation and interest (Ellis, 2004; Graber & Sontag, 2006), and cognitive development changes in perspective taking (Eisenberg, Cumberland, Guthrie, Murphy, & Sheppard, 2005), all occur at a rapid pace during adolescence. These cognitive developmental changes are likely to influence the way adolescents experience and report on depressive symptoms and thus impact upon the factorial invariance of the measurement tool.

Moreover, it is well established that gender differences in developmental trajectories of depression emerge in adolescence, with increases in depression in females versus stability or decreases in males (e.g., Garber et al., 2002; Ge, Conger, & Elder, 2001). It could be that symptom profiles of depression develop differently in males and females during adolescence. For example, Fu-I and Wang (2008) found that gender differences in depressive symptoms of clinically depressed children did not emerge until adolescence, with females having a lower self-esteem and males reporting more difficulties concentrating. In addition, following a sample of 183 clinically referred children from childhood (ages 8–13) to young adulthood (up to age 21), Kovacs et al. (2003) detected gender-differences in the development of depressive symptom patterns. They found that irritability as a symptom of depression in girls was age-related: depressed girls were most likely to experience irritability during mid-adolescence (13–15 years), and less so as they got older, whereas irritability appeared as a consistent feature of depressed males' experiences. During mid-adolescence, girls were at peak risk for attempting suicide, whereas the peak rate of suicide attempts for boys occurred during late adolescence (16–18 years).

Despite these findings of gender-, and time differences in the occurrence of specific symptoms of depression, little is known about the extent to which these symptoms are actually indicators of depression in males and females across adolescence. The present study examines the factorial invariance of the Center for Epidemiologic Studies of Depression Scale (CES-D) in adolescents. The CES-D is a 20 item questionnaire describing a wide range of depressive symptoms which was developed to identify depression among adults in the general population (Radloff, 1977). Several studies have examined the factor structure of the CES-D in adults, generally identifying four specific factors described as depressive affect, well being, somatic symptoms, and interpersonal relations that together load on the common factor Depression (Motl et al., 2005; Radloff, 1977; Shafer, 2006).

Although originally developed to identify depressive symptoms among adults, the CES-D has been widely used to examine depressive symptoms in adolescents (Garrison, Addy, Jackson, McKeown, & Waller, 1991; Rushton, Forcier, & Schectman, 2002). Studies examining the validity and reliability of the CES-D have concluded that this instrument is adequate for screening purposes for depression in this age-range (Garrison et al., 1991; Manson, Ackerson, Dick, Baron, & Fleming, 1990; Roberts, Andrews, Lewinsohn, & Hops, 1990; Schoenbach, Kaplan, Grimson, & Wagner, 1982). Recently, Philips et al. (2006) examined 12 different, previously reported factor-structures of the CES-D in a large sample of 7th Grade students and

concluded that the four-factor model reported by Radloff (1977) and others (see Shafer, 2006) had the strongest support. They did not compare this factor-model for males and females. However, despite good evidence that the factor-structure of the CES-D in adolescents is comparable to that in adults, little is known about whether the factor structure is invariant across gender and time in adolescents.

To the best of our knowledge, only three previous studies have investigated the longitudinal and gender-invariance of the CES-D in adolescents. Roberts et al. (1990) examined the factor structure of the CES-D in 4 cohorts of students in Grade 9–12. The study identified a four-factor model among adolescents and found factor loadings of the CES-D items were equivalent for males and females with the exception of two items: 'I had crying spells' and 'lost appetite'. 'I had crying spells' had a higher loading for females; the item 'lost appetite' had a lower loading for females. This indicates that crying was more characteristic of depression in girls, whereas loss of appetite was more characteristic of depression in boys. In addition, they examined the factorial invariance of de CES-D across a period of one month, and found the CES-D to be invariant within this short time frame.

Dick, Beals, Keane, and Manson (1994) examined gender invariance of the CES-D in a minority sample of American Indian adolescents with a widespread age range (13–20 years of age). They identified a three-factor solution to be the most appropriate, with the items of the depressive affect factor and the somatic symptoms factor comprising a single factor. There were no gender differences in the factor loadings of the CES-D items, or in the interrelatedness of the three factors.

More recently, Motl et al. (2005) examined the gender- and longitudinal invariance of the CES-D. They identified the high-order model with four first-order and 1 second-order factor. The factor structure coefficients were invariant across gender at the age of 13, and the authors decided to combine males and females in one sample to examine factorial invariance over time. For this combined sample of adolescents, they found longitudinal invariance of the CES-D over a three year period. Although this study provides us with insight in the factorial invariance of the CES-D in early adolescence across gender and time, no information regarding the longitudinal cross-gender factorial invariance of the CES-D was provided.

The present study will examine the factorial invariance of the CES-D in a large, normative sample of 2650 Australian adolescents, starting from a CES-D factor structure comprised of four first-order factors with an underlying second-order factor as reported by Motl et al. (2005). The aim of the current study is two-fold; firstly to replicate findings regarding the factorial invariance of the CES-D of previous studies, but using a large cohort, and secondly to expand existing knowledge by examining the longitudinal cross-gender factorial invariance of the CES-D across four annual measurement waves.

Method

Participants

The *beyondblue* Schools Research Initiative is a longitudinal study of Australian secondary school students which evaluated the effectiveness of a universal intervention designed to reduce depressive symptoms among high-school students (see also Sawyer et al., 2010; Spence et al., 2005). Prior to survey administration, active consent was obtained from 5634 adolescents enrolled in Year 8 in 50 participating secondary schools and their parents. The original study involved a randomized controlled trial examining the impact of an intervention designed to prevent the development of depression. The 50 schools were matched and paired in terms of SES and enrollment size and then randomly allocated to either the intervention or control group. In line with the ethics approval and written informed consent, when a clinically high score was detected the guidance officer or counselor at the school was advised of the situation.

The present study included data from the 25 control group schools only, and included 2650 Year 8 adolescents ($M_{age} = 13.02$ years at baseline, SD = .82), representing approximately 64% of adolescents from the total Year 8 school enrollment in control schools. Data were collected at four time points; baseline in May, 2003 and October, 2004, October, 2005 and October, 2006. Data at one or more waves were missing for 1240 (46%) adolescents. At baseline, 93.4% were born in Australia, 3.2% identified with Aboriginal or Torres Strait Islander background, and 71% lived with both of their parents (Table 1).

Table 1
Gender, ethnicity, and family status for our sample.

	n	%
Gender		
Male	1273	48
Female	1377	52
Australian origin		
Yes	2461	93.4
No	173	6.6
Identifies with aboriginal/torres strait island		
Yes	83	3.2
No	2539	95.8
Family status		
Lives with both parents	1834	71
Live apart/divorced	651	25.2
One or both parents passed away	46	1.8
Different	52	2

Measure

Center for Epidemiological Studies Depression Scale (CES-D: Radloff, 1977)

The CES-D consists of 20 items describing a wide range of depressive symptoms (see Appendix 1 for a full description). Respondents rate their experience of each symptom in the past week on a 0–3 scale from "Rarely or none of the time (less than 1 day)" to "Most or all of the time (5–7 days)". The total summed score can range from 0 to 60. Internal consistency (Crohnbach's alpha) was good for the present sample, ranging from .78 to .80 for females and from .79 to .82 for males.

Statistical procedures

Confirmatory factor analyses were conducted using MPlus version 6.0 (Muthén & Muthén, 2011). The goodness of fit of the models was assessed by multiple criteria: the chi-square likelihood ratio statistic, root mean squared error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker Lewis index (TLI). A RMSEA value less than .08 and CFI and TLI values greater than .90 are considered an acceptable fit (Hu & Bentler, 1999; Kline, 2005). We used the full information maximum likelihood (FIML) estimation within MPlus version 6.0 (Muthén & Muthén, 2011) to handle missing data.

Analyses were conducted in three stages. First, data for males and females were examined separately for each of the four measurement waves to establish baseline models of the factor structure of the CES-D. Second, the invariance of the factor structure was tested across gender for each of the four measurement waves. Third, the invariance of the factor structure of the CES-D was tested across time for males and females separately.

Stage 1: Determine baseline models

Before examining factorial invariance across different groups (e.g., gender, age), it is important to establish a well-fitting baseline model for each group separately (Byrne et al., 1993). The factor model with four first-order and one underlying second-order factor as proposed by Motl et al. (2005; Fig. 1) was tested separately for goodness-of-fit to male and female data for each of the four measurement Waves. First-order factor loadings were assessed in the pattern-matrix Lambda (LY), second-order factor loadings in Beta (BE). Once the baseline model was determined for each gender, at each of the four waves the similarity of item measurements (i.e., factor loadings) and theoretical structure (i.e., relatedness between depressive factors) was tested.

Stage 2: Factorial invariance across gender

A multi-group model was used to examine whether the first-order factor loadings and second-order factor loadings were similar for males and females for each of the four measurement waves. The test for gender differences involved the comparison of a model in which the factor loadings for the female sample were allowed to be different from the factor loadings for the male sample (unconstrained model), to a model in which these loadings were constrained to be equal across gender (constrained model). A significant ($p < .01$) difference in the chi-square of these two models would indicate that there were significant differences in the factor structures for males and females.

Because a chi-square test is sensitive to sample size, we also examined the $\Delta\chi 2/\Delta df$ ratio. A ratio greater than 5 is considered evidence of differences in the factor structure for males and females (Rosay, Gottfredson, Armstrong, & Harmon, 2000). If both the chi-square test and the ratio $\Delta\chi 2/\Delta df$ indicated differences in the factor structure for males and females, subsequent analyses were conducted to examine which of the factor loadings gave rise to these differences. One by one, the first-order factor loadings were constrained to be equal across gender. When non-invariant first-order factor loadings were identified and unconstrained, the same procedure was followed for factor loadings on the higher-order factor.

Stage 3: Factorial invariance across time

Longitudinal invariance was examined for males and females separately. For both genders, a model was constructed with 16 first-order factors (i.e., 'Depressive Affect', 'Well Being', 'Somatic Symptoms', and 'Interpersonal Relations' for each measurement wave) and four higher order factors ('General Depression' for each measurement wave). The four higher-order factors of 'General Depression' were allowed to correlate across time. However, the first-order factors were not allowed to correlate across time (Motl et al., 2005). Subsequent steps examined whether the factor loadings and the theoretical structure were invariant across time.

Results

Baseline models

The model outlined in Fig. 1 was examined for males and females separately at each of the four measurement waves. Model fit indices showed that the higher-order, four factor model (Motl et al., 2005; Radloff, 1977) fitted the data reasonably well for both genders on each occasion with χ^2 ranging from 540.74 to 969.86 ($df = 166$), RMSEA between .05 and .07, CFI between .91 and .96, and TLI between .90 and .95.

However, the residual variance of Depressive Affect was negative on T1, T2, and T3 for females, and on all four waves for males. Thorough examination learned that there were no misspecifications in our models. As model complexity might be the

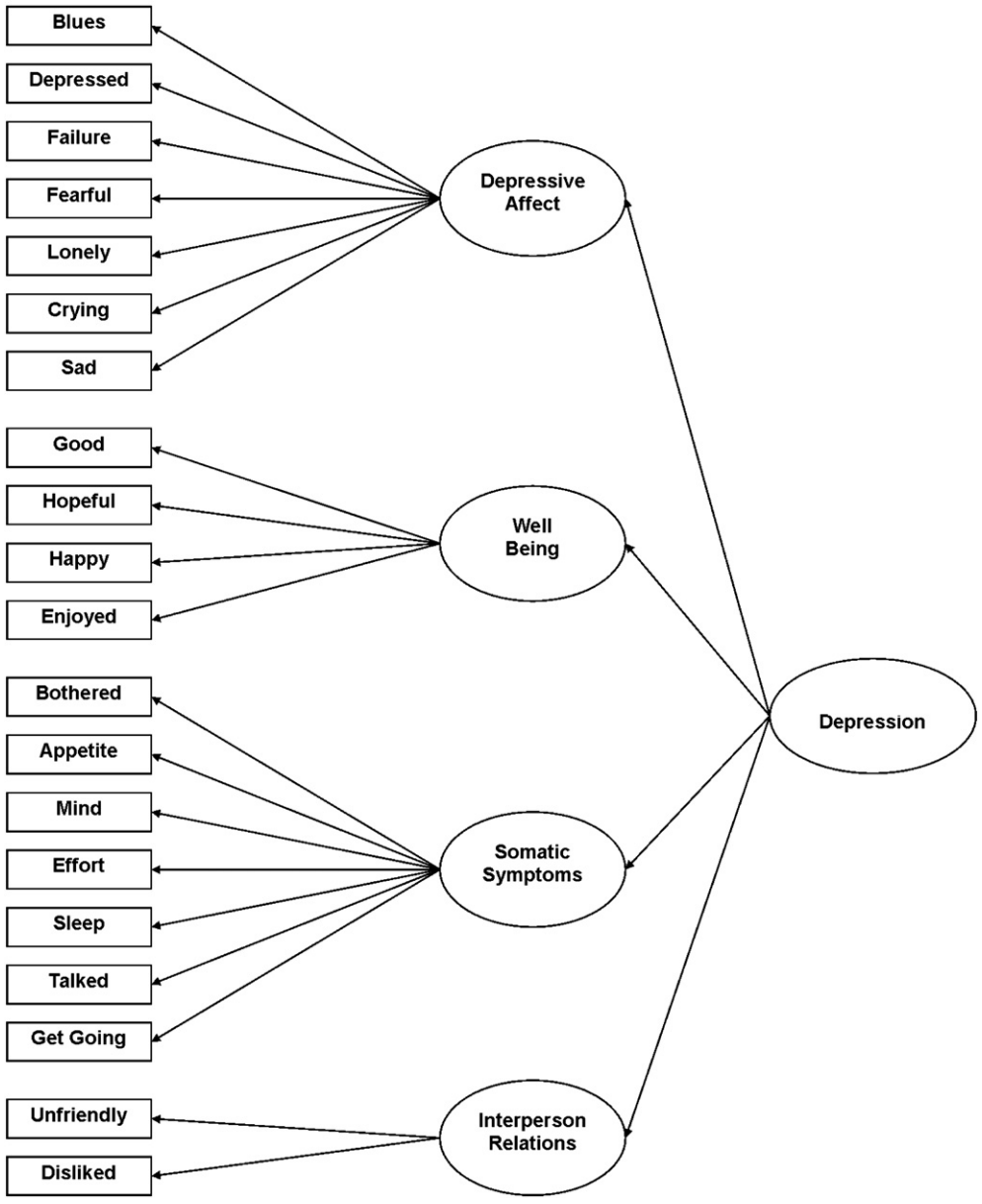

Fig. 1. Baseline factor structure of the CES-D.

cause of negative residual variances (Dillon, Kumar, & Mulani, 1987), we decided to run a simpler model by omitting the higher-order factor of Depression. This leads to a four-factor model in which the factors of Depressive Affect, Well being, Somatic Symptoms, and Interpersonal Relations were allowed to correlate. Model fit indices of these four-factor models were acceptable (Table 2) and negative residual variances no longer occurred. However, the relatively high correlations between Somatic Symptoms and Depressive Affect ($r > .83$; see Table 4) suggested that a three-factor model might describe our data better. These three-factor models showed acceptable fit (χ^2 ranging from 639.18 to 1031.83 (df = 167), RMSEA between .05 and .07, CFI between .90 and .95, and TLI between .89 and .94). However, the fit-indices for the four-factor models (Table 2) were slightly better. Comparison between the fit of the two models was also guided by the Akaike Information Criteria (AIC), for which the model with the lowest AIC reflects the best balance of goodness of fit and parsimony (Akaike, 1987). AIC's were lower for the four-factor model (ranging from 30026.25 to 62894.69) than for the 3-factor models (ranging from 30138.24 to 62960.49). This also indicates that the four-factor models had a better fit with the data. We therefore decided to continue with the four-factor model resembling the standard factor structure of the CES-D specified by Radloff (1977). All following results are based on this four-factor model.

Table 2
Summary of fit statistics for the four factor models for males and females at wave 1–4.

		Fit of model				
		χ^2	df	RMSEA	CFI	TLI
Males	Wave 1	961.72	164	.06	.91	.90
	Wave 2	812.38	164	.06	.93	.92
	Wave 3	832.02	164	.07	.93	.92
	Wave 4	547.58	164	.06	.94	.93
Females	Wave 1	740.45	164	.05	.95	.94
	Wave 2	761.06	164	.05	.95	.94
	Wave 3	754.18	164	.06	.95	.95
	Wave 4	537.67	164	.05	.96	.95

The four-factor model fitted the data well for both males and females at all four waves (Table 2) and did not produce negative residual variance. Factor loadings and correlations between the four factors of the unconstrained four-factor models are presented in Table 3 and Table 4 respectively. Modification indices for these four-factor models indicated that at Wave 1 for both males and females, the item 'effort' cross-loaded on the factor 'Well Being' ($\lambda_{Males} = .47$, $p < .001$; $\lambda_{Females} = .38$, $p < .001$). For males this cross-loading was also statistically significant on Wave 2 ($\lambda = .34$, $p < .001$) and Wave 3 ($\lambda = .25$, $p < .001$). Consistent with the procedure described by Byrne (1989), this cross-loading was not maintained in the subsequent tests of factorial invariance.

Multigroup models: factorial invariance across gender

For each of the four waves, multigroup models were used to examine whether the factor structure of the CES-D was invariant across gender. First, the four-factor structure of the CES-D was assessed to see if it was equivalent for males and females (unconstrained four-factor multigroup model). Table 4 shows that all four unconstrained multigroup models showed acceptable fit measures, indicating that the four-factor structure of the CES-D was similar for males and females at each of the four measurement waves.

In order to test factorial invariance across gender, the factor loadings were constrained to be equal for males and females. These cross-gender constraints reduced the extent to which the model fitted data from all four measurement waves (Table 5). Further examination of these constraints revealed that on all four measurement waves, the factor loading for the item 'I had crying spells' was significantly higher on the factor 'Depressive Affect' for females than for males (Table 3). Also, the item 'I felt

Table 3
Standardized and unstandardized (standard errors) factor loadings for the CES-D items for males and females based on the unconstrained four factor model.

		Wave 1				Wave 2				Wave 3				Wave 4			
		Males		Females		Males		Females		Males		Females		Males		Females	
Depressive affect																	
1	Blues	.62	.77 (.04)	.71	.80 (.03)	.70	.80 (.03)	.78[a]	.88 (.03)	.71	.78 (.03)	.83[a]	.89 (.03)	.74	.81 (.04)	.80	.95 (.03)
2	Depressed	.77	1.00	.80	1.00	.81	1.00	.83	1.00	.82	1.00	.87	1.00	.82	1.00	.84	1.00
3	Failure	.71	.91 (.04)	.71	.75 (.03)	.69	.81 (.03)	.72	.75 (.03)	.75	.85 (.03)	.77	.76 (.02)	.70	.77 (.04)	.69	.68 (.03)
4	Fearful	.60	.69 (.03)	.61	.60 (.03)	.66	.71 (.03)	.66	.64 (.03)	.74	.77 (.03)	.67	.63 (.02)	.65	.70 (.04)	.66	.66 (.03)
5	Lonely	.77	.96 (.04)	.75	.88 (.03)	.75	.92 (.03)	.76	.85 (.03)	.77	.66 (.04)	.81	.90 (.03)	.79	1.01 (.04)	.74	.89 (.04)
6	Crying	.58	.55 (.03)	.70[a]	.78 (.03)	.54	.44 (.02)	.73[a]	.84 (.03)	.58	.49 (.03)	.75[a]	.81 (.03)	.59	.54 (.03)	.73[a]	.86 (.04)
7	Sad	.78	.93 (.03)	.83	.94 (.03)	.80	.93 (.03)	.85	.97 (.03)	.83	.95 (.03)	.87	.94 (.02)	.83	.97 (.04)	.85	.99 (.03)
Well being																	
8	Good	.55	.80 (.05)	.52	.69 (.04)	.52	.74 (.05)	.60	.79 (.04)	.59	.83 (.05)	.65	.87 (.04)	.64	.93 (.06)	.65	.90 (.04)
9	Hopeful	.54	.72 (.04)	.46	.60 (.04)	.54	.72 (.04)	.54	.69 (.04)	.61	.79 (.04)	.60	.76 (.04)	.68	.92 (.05)	.66	.84 (.04)
10	Happy	.80	1.00	.86	1.00	.82	1.00	.87	1.00	.82	1.00	.86	1.00	.80	1.00	.88	1.00
11	Enjoyed	.78	1.01 (.04)	.82	.98 (.03)	.81	1.00 (.04)	.84	1.03 (.03)	.83	1.05 (.04)	.87	1.07 (.03)	.87	1.12 (.05)	.88	1.04 (.03)
Somatic symptoms																	
12	Get going	.70	1.00	.74	1.00	.72	1.00	.75	1.00	.71	1.00	.76	1.00	.69	1.00	.70	1.00
13	Sleep	.58	.90 (.05)	.62	.97 (.05)	.57	.88 (.05)	.62	.96 (.05)	.65	1.05 (.06)	.60	.88 (.05)	.62	1.05 (.07)	.57	.90 (.06)
14	Bothered	.51	.64 (.04)	.58	.74 (.04)	.54	.65 (.04)	.62	.77 (.04)	.60	.74 (.04)	.63	.78 (.04)	.61	.77 (.05)	.64	.83 (.05)
15	Appetite	.43	.56 (.04)	.51	.73 (.04)	.50	.61 (.04)	.56[a]	.81 (.04)	.60	.73 (.04)	.60	.81 (.04)	.66	.86 (.05)	.55	.80 (.06)
16	Mind	.53	.89 (.05)	.62	.94 (.04)	.50	.83 (.05)	.58	.85 (.04)	.55	.91 (.06)	.59	.83 (.04)	.54	.91 (.07)	.59	.86 (.06)
17	Effort	.22	.40 (.06)	.18	.27 (.05)	.29	.50 (.06)	.43	.60 (.04)	.45	.75 (.06)	.51	.67 (.04)	.42	.67 (.06)	.53	.76 (.05)
18	Talked	.52	.84 (.05)	.58	.82 (.04)	.57	.86 (.05)	.59	.78 (.04)	.60	.87 (.05)	.59	.73(.04)	.68	1.00 (.06)	.60	.81 (.05)
Interpersonal relations																	
19	Unfriendly	.75	.82 (.03)	.74	.74 (.03)	.75	.85 (.05)	.66[a]	.63 (.04)	.78	.87 (.04)	.66[a]	.59 (.03)	.71	.81 (.04)	.59[a]	.53 (.04)
20	Disliked	.86	1.00	.87	1.00	.85	1.00	.90	1.00	.86	1.00	.89	1.00	.83	1.00	.88	1.00

Note: All factor loadings are significantly different from zero ($p < .001$).
[a] Factor loading differs from that in males at $p < .001$.

Table 4
Standardized and unstandardized (standard errors) correlations between factors of depressive symptoms for females and males in the unconstrained four factor model at wave 1–4.

	1. Depressive affect		2. Well being		3. Somatic symptoms		4. Interpers. relations	
Wave 1								
1. Depressive affect			−.52	−.28 (.02)	.91	.36 (.02)	.79	.42 (.02)
2. Well being	−.76	−.52 (.03)			−.40	−.20 (.02)	−.41	−.27 (.03)
3. Somatic symptoms	.93	.52 (.03)	−.68	−.38 (.02)			.73	.35 (.02)
4. Interpers. relations	.81	.59 (.03)	−.61	−.45 (.03)	.76	.45 (.03)		
Wave 2								
1. Depressive affect			−.50	−.27 (.02)	.93	.37 (.02)	.87	.42 (.02)
2. Well being	−.76	−.53 (.03)			−.41	−.20 (.02)	−.43	−.25 (.02)
3. Somatic symptoms	.92	.56 (.03)	−.67	−.39 (.03)			.79	.34 (.02)
4. Interpers. relations	.72	.53 (.03)	−.54	−.37 (.03)	.66	.41 (.03)		
Wave 3								
1. Depressive affect			−.51	−.28 (.02)	.93	.39 (.03)	.83	.43 (.03)
2. Well being	−.78	−.56 (.03)			−.41	−.20 (.02)	−.38	−.23 (.03)
3. Somatic symptoms	.92	.59 (.03)	−.72	−.42 (.03)			.79	.35 (.02)
4. Interpers. relations	.79	.60 (.03)	−.61	−.42 (.03)	.73	.45 (.03)		
Wave 4								
1. Depressive affect			−.51	−.27 (.03)	89	.34 (.03)	.86	.37 (.03)
2. Well being	−.75	−.44 (.03)			−.38	−.17 (.02)	−.39	−.20 (.03)
3. Somatic symptoms	.89	.45 (.03)	−.67	−.34 (.03)			.78	.30 (.03)
4. Interpers. relations	.74	.42 (.03)	−.53	−.31 (.03)	.69	.34 (.03)		

Note: Females are below diagonal, males are above diagonal. All factor correlations are significantly different from zero ($p < .001$).

that I could not shake off the blues' showed a stronger loading on the 'Depressive Affect' factor for females than for males at waves 2 and 3. In addition, on wave 2, 3, and 4 the item 'people were unfriendly' had a significantly higher loading on the interpersonal factor for males than for females (Table 3). At wave 2 the item 'I had poor appetite' had a significantly higher loading on Somatic Symptoms for females than males.

Factorial invariance across time

To test longitudinal invariance, a model was constructed for males and females separately with 16 factors (i.e., 'Depressive Affect', 'Well Being', 'Somatic Symptoms', and 'Interpersonal Relations' for each measurement wave). Stability paths were estimated for each factor from T1 to T2, T2 to T3, and T3 to T4. In addition, within each measurement wave the four factors were allowed to correlate. These models showed reasonably acceptable fit for both males and females (Table 6). Constraining the factor loadings to be equal across the four measurement waves did not reduce the model's fit for males or females, indicating that the four-factor structure was invariant from early to middle adolescence for both males and females.

Discussion

This paper investigated whether the factor structure of the CES-D was invariant across gender and time from early to middle adolescence in a large sample of young people. Consistent with previous studies, the four-factor model of the CES-D

Table 5
Summary of test for gender invariance of the four factor model at wave 1–4.

		Fit of model					$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$
		χ^2	df	RMSEA	CFI	TLI			
Wave 1									
1.	Unconstrained four factor model	1702.18	328	.06	.93	.92			
2.	Pattern of item loadings invariant	1865.70	348	.06	.93	.92	163.52	20	8.18
3.	Item loadings invariant except item 6	1794.73	347	.06	.93	.92	92.55	19	4.87
Wave 2									
1.	Unconstrained four factor model	1573.44	328	.06	.94	.93			
2.	Pattern of item loadings invariant	1876.34	348	.06	.93	.92	302.90	20	15.15
3.	Item loadings invariant except item 1, 6, 15 & 19	1654.67	344	.06	.94	.93	81.24	16	5.08
Wave 3									
1.	Unconstrained four factor model	1586.20	328	.06	.94	.93			
2.	Pattern of item loadings invariant	1867.39	348	.07	.93	.92	281.19	20	14.06
3.	Item loadings invariant except item 1, 6 & 19	1662.27	345	.06	.94	.93	76.07	17	4.47
Wave 4									
1.	Unconstrained four factor model	1085.25	328	.05	.95	.94			
2.	Pattern of item loadings invariant	1243.32	348	.06	.94	.94	158.07	20	7.90
3.	Item loadings invariant except 6 & 19	1148.65	346	.05	.95	.94	63.40	18	3.52

Note: item 1 = 'blues'; item 6 = 'crying'; 15: 'effort' item 19 = 'unfriendly'; DA = Depressive Affect; WB = Well Being.

Table 6
Summary of test for longitudinal invariance of the four factor model for males and females.

		Fit indices					$\Delta\chi^2$	Δdf	$\Delta\chi^2/\Delta df$
		χ^2	df	RMSEA	CFI	TLI			
Males									
1.	Comparison model	7717.65	3044	.04	.88	.87			
2.	Item loadings invariant	7816.55	3092	.04	.87	.87	98.9	48	2.06
Females									
1.	Comparison model	7885.12	3044	.03	.90	.90			
2.	Item loadings invariant	8069.04	3092	.03	.90	.90	183.92	48	3.83

was broadly supported for both genders and at each time point. However, there is some evidence of slight variability with respect to specific questionnaire items at particular time points.

Factorial invariance across gender

At all four measurement waves the item 'I had crying spells' had a higher loading on the factor 'Depressive Affect' for females than males. This is consistent with previous findings (Byrne et al., 1993; Roberts et al., 1990). As suggested by Byrne et al. (1993), it may reflect the impact of social values which make it more acceptable for females than males to acknowledge crying in response to depressive affect. Alternatively, it may reflect a genuine difference in the way that depression is manifested across the genders. Either way, this effect has the potential to make crying spells a better indicator of depressive affect in females than in males. Similarly the item 'I felt that I could not shake off the blues' was a stronger indicator of 'Depressive Affect' for females than for males at waves 2 and 3 for girls than for boys.

In addition, on the last three waves, the item 'People were unfriendly' was a better indicator of depression (i.e., manifested in 'Interpersonal Relations') for males than for females. There are various explanations that could be proposed for this effect, but they can only be speculative. It is possible that depressed young males are particularly sensitive to interpersonal difficulties compared to their non-depressed counterparts. Alternatively, the other symptoms of depression manifested in young males may place them at increased risk for receiving unfriendly responses from others, particularly if peers are less tolerant of depression among males than females. A recent study indeed found that depressive symptoms can make a child seem more vulnerable and thereby make the child an easy target for unfriendly responses from others (Fekkes, Pijpers, Fredriks, Vogels, & Verloove-Vanhorick, 2006; Reijntjes, Kamphuis, Prinzie, & Telch, 2010). Future studies should examine whether this association between being depressed and peer relationship difficulties is moderated by the child's gender.

Another notable finding was that modification indices at the first wave showed that the item 'I felt everything that I did was an effort' cross-loaded positively on the factor 'Well Being' in addition to the 'Somatic Symptoms' factor, for adolescent males and females. This cross-loading became weaker over time, but was statistically significant on the second and third measurement waves for males. Thus, it appeared that males, particularly when younger, tended to interpret this item ambiguously, in that it was not only reflecting a somatic symptom of depression, but was also being seen by boys in general as suggesting that the experience of effort was a positive thing, associated with better well being. The ambiguity in the item is understandable, although females did not tend to make this dual interpretation. The finding is, however, consistent with result of previous studies (Dick et al., 1994; Motl et al., 2005; Phillips et al., 2006). It has been suggested that adolescents might perceive the term 'effort' as a positive quality in relation to being goal oriented (Motl et al., 2005).

Factorial invariance across time

The results from the present study provide support for the hypothesis that generally the symptom manifestation of depression is relatively stable across time in younger adolescents, and indeed is equivalent to the way in which depressive symptoms are presented in adults. Our results show strong evidence of factorial invariance over time for the individual questionnaire items and are in line with previous findings (Motl et al., 2005). This means that the pattern and strength of factor loadings of individual symptoms upon the four factors of depression, namely 'Depressive Affect', 'Well Being', 'Somatic Symptoms', and 'Interpersonal Relations' showed a consistent patterns and strength of associations over time.

Limitations

Before considering the implications of the findings of this study, some limitations should be mentioned. First, the retention rate of the current study was an issue, particularly by the final wave, although this is a common problem in longitudinal research of this type. Approximately 46% of the adolescents who participated at the first wave failed to take part in at least one of the following measurement waves. It is possible that those who failed to continue in this study were at elevated risk for developing depression and the possibility that the factor structure may have differed for those who dropped out versus those who were retained cannot be discounted. In line with this, it would be interesting to examine longitudinal and gender invariance of the CES-D in a clinical sample of adolescents. Second, the data rely on self reports of symptoms and as such are subject to a range of reporting biases and other sources of error which limit the conclusions that can be drawn.

Conclusion

The current study extended previous knowledge regarding the longitudinal factorial invariance of the CES-D by testing for the equivalence of the CES-D factor structure across gender and time for a large normative sample of Australian adolescents across a period of four years. Findings from the present study are of substantial importance to our understanding of the manifestation of depression in adolescence, and of the factorial structure of the CES-D in particular. First, the factorial structure of the CES-D in the adolescent sample generally mirrored that found in adults, with four related factors of 'Depressive Affect', 'Well Being', 'Somatic Symptoms' and 'Interpersonal Relations'. Furthermore, with only one exception ('I felt that everything I did was an effort'), the individual symptoms that loaded on these four factors generally reflected the pattern of results found in adults. This suggests that the CES-D, which was originally developed for adults, is also an appropriate measurement instrument for depression throughout adolescence.

Second, although this factor structure fitted the data for both males and females, gender differences related to the functioning of CES-D items nonetheless existed for a very small number of items ('I had crying spells'; 'I felt that I could not shake off the blues'; 'I felt that everything I did was an effort'; 'People were unfriendly'). It is recommended that researchers and clinicians working with adolescent samples take into account gender differences in the way in which these symptoms relate to the construct of depression. For example, for adolescent females self-reported 'crying' may be a better marker of high levels of depressive symptoms for than it is for males, whereas having the feeling that 'people are unfriendly' may be a better indicator of depression in adolescent males. It should be noted, however, that although there were slight gender differences in factor structure, the original proposed four-factor structure (Radloff, 1977) provided a relatively good fit to the data for both genders, on all occasions.

In conclusion, although there were some minor differences in factor loadings across the genders, results from the present study suggest that the profile of symptoms identified by the CES-D is very similar for male and female adolescents, and for adolescents of different ages. This provides support for the approach used in DSM-IV and ICD-10, and commonly used self-report questionnaires which assume that the symptom profile of depression is invariant across gender and time.

Acknowledgments

The authors wish to thank the beyondblue state project teams, school action teams, and staff and students from the participating schools. Funding for the project was provided by beyondblue: the national depression initiative and the National Medical and Research Council (NHMRC). Neither funding body had any role in the study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Appendix 1

Item content of the CES-D.

	Item content	Key word[a]
1.	I felt that I could not shake off the blues even with help from my family and friends	Blues
2.	I felt depressed	Depressed
3.	I thought my life had been a failure	Failure
4.	I felt fearful	Fearful
5.	I felt lonely	Lonely
6.	I had crying spells	Crying
7.	I felt sad	Sad
8.	I felt I was just as good as other people	Good
9.	I felt hopeful about the future	Hopeful
10.	I was happy	Happy
11.	I enjoyed life	Enjoyed
12.	I was bothered by things that usually don't bother me	Bothered
13.	I did not feel like eating; my appetite was poor	Appetite
14.	I had trouble keeping my mind on what I was doing	Mind
15.	I felt that everything I did was an effort	Effort
16.	My sleep was restless	Sleep
17.	I talked less than usual	Talked
18.	I could not 'get going'	Get Going
19.	People were unfriendly	Unfriendly
20.	I felt that people disliked me	Disliked

[a] *Note*: Key words are used in the text, tables, and figure of the manuscript to indicate a particular item of the CES-D.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317–332.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*. Text Revision (4th ed.). Washington, DC: American Psychiatric Association.

Ashman, S. B., & Dawson, G. (2002). Maternal depression, infant psycho-biological development, and risk for depression. In S. H. Goodman, & I. H. Gotlib (Eds.), *Children of depressed parents: Mechanism of risk and implications of treatment* (pp. 37–58). Washington, D.C.: American Psychological Association.

Baron, P., & Joly, E. (1988). Sex-differences in the expression of depression in adolescence. *Sex Roles, 18*, 1–7.
Bennett, D. S., Ambrosini, P. J., Kudes, D., Metz, C., & Rabinovich, H. (2005). Gender differences in adolescent depression: do symptoms differ for boys and girls? *Journal of Affective Disorders, 89*, 35–44.
Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer Verlag.
Byrne, B. M., Baron, P., & Campbell, T. L. (1993). Measuring adolescent depression: factorial validity and invariance of the beck depression inventory across gender. *Journal of Research on Adolescence, 3*, 127–143.
Cole, D. A., & Martin, N. C. (2005). The longitudinal structure of the children's depression inventory: testing a latent trait-state model. *Psychological Assessment, 17*, 144–155.
Connolly, J., & Goldberg, A. (1999). Romantic relationships in adolescence: the role of friends and peers in their emergence and development. In W. Furman, B. B. Brown, & C. Feiring (Eds.), *The development of romantic relationships in adolescence* (pp. 266–290). New York: Cambridge University Press.
Dick, R. W., Beals, J., Keane, E. M., & Manson, S. M. (1994). Factorial structure of the CES-D among American Indian adolescents. *Journal of Adolescence, 17*, 73–79.
Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: comments on the causes of and solutions to Heywood cases. *Psychological Bulletin, 101*, 126–135.
Eisenberg, N., Cumberland, A., Guthrie, I. K., Murphy, B. C., & Shepard, S. A. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *Journal of Research on Adolescence, 15*, 235–260.
Ellis, B. J. (2004). Timing of pubertal maturation in girls: an integrated life history approach. *Psychological Bulletin, 130*, 920–958.
Fekkes, M., Pijpers, F. I. M., Fredriks, A. M., Vogels, T., & Verloove-Vanhorick, S. P. (2006). Do bullied children get ill, or do ill children get bullied? A prospective cohort study on the relationship between bullying and health-related symptoms. *Pediatrics, 117*, 1568–1574.
Fu-I, L., & Wang, Y. P. (2008). Comparison of demographic and clinical characteristics between children and adolescents with major depressive disorder. *Revista Brasileira de Psiquiatria, 30*, 124–131.
Garber, J. J., Keiley, M. K., & Martin, N. C. (2002). Developmental trajectories of adolescents' depressive symptoms: predictors of change. *Journal of Consulting and Clinical Psychology, 70*, 79–95.
Garrison, C. Z., Addy, C. L., Jackson, K. L., McKeown, R. E., & Waller, J. L. (1991). The CES-D as a screen for depression and other psychiatric disorders in adolescents. *Journal of American Academy of Child and Adolescent Psychiatry, 30*, 636–641.
Ge, X., Conger, R. D., & Elder, G. H. (2001). Pubertal transition, stressful life events, and the emergence of gender differences in adolescent depressive symptoms. *Developmental Psychology, 37*, 404–417.
Graber, J. A., & Sontag, L. M. (2006). Puberty and girls' sexuality: why hormones are not the complete answer. *New Directions for Child and Adolescent Development, 112*, 23–38.
Hammen, C. L., & Padesky, C. A. (1977). Sex differences in the expression of depressive responses to the beck depression inventory. *Journal of Abnormal Child Psychology, 86*, 609–614.
Harter, S. (1999). *The construction of the self: A developmental perspective. Distinguished contributions in psychology*. New York: Guilford Press.
Harter, S. (2003). The development of self-representations during childhood and adolescence. In M. R. Leary, & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 610–642). New York: Guilford Press.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Huttenlocher, P. R., & Dabholkar, A. S. (1997). Developmental autonomy of prefontal cortex. In N. A. Krasnegor, G. R. Lyon, & P. S. Goldman-Rakic (Eds.), *Development of the prefrontal cortex* (pp. 69–83). Baltimore: Brookes.
Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.
Kovacs, M., Obrosky, D. S., & Sherrill, J. (2003). Developmental changes in the phenomenology of depression in girls compared to boys from childhood onward. *Journal of Affective Disorders, 74*, 33–48.
Manson, S. M., Ackerson, L. M., Dick, R. W., Baron, A. E., & Fleming, C. M. (1990). Depressive symptoms among American Indian adolescents: psychometric characteristics of the Center for Epidemiologic Studies Depression Scale (CES–D). *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 231–237.
Motl, R. W., Dishman, R. K., Birnbaum, A. S., & Lytle, L. A. (2005). Longitudinal invariance of the Center for Epidemiologic Studies-Depression Scale among girls and boys in middle school. *Educational and Psychological Measurement, 65*, 90–108.
Muthén, L. K., & Muthén, B. O. (2011). *Mplus (Version 6.11)*. Los Angeles: Muthén & Muthén [Computer software and manual].
Phillips, G. A., Shadish, W. R., Murray, D. M., Kubik, M., Lyle, L. A., & Birnbaum, A. S. (2006). The center for Epidemiologic Studies Depression Scale with a young adolescent population: a confirmatory factor analysis. *Multivariate Behavioral Research, 41*, 147–163.
Radloff, L. S. (1977). The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
Reijntjes, A., Kamphuis, J. H., Prinzie, P., & Telch, M. J. (2010). Peer victimization and internalizing problems in children: a meta-analysis of longitudinal studies. *Child Abuse & Neglect, 34*, 244–252.
Roberts, R. E., Andrews, J. A., Lewinsohn, P. M., & Hops, H. (1990). Assessment of depression in adolescents using the Center for Epidemiologic Studies Depression Scale. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 122–128.
Rosay, A. B., Gottfredson, D. C., Armstrong, T. A., & Harmon, M. A. (2000). Invariance of measures of prevention program effectiveness: a replication. *Journal of Quantitative Criminology, 16*, 341–367.
Rushton, J. L., Forcier, M., & Schectman, R. M. (2002). Epidemiology of depressive symptoms in the national longitudinal study of adolescent health. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*, 199–205.
Sawyer, M. G., Pfeiffer, S., Spence, S. H., Bond, L., Graetz, B., Kay, D., et al. (2010). School-based prevention of depression: a randomised controlled study of the beyondblue schools research initiative. *Journal of Child Psychology and Psychiatry, 51*(2), 199–209.
Schoenbach, V. J., Kaplan, B. H., Grimson, R. C., & Wagner, E. H. (1982). Use of a symptom scale to study the prevalence of a depressive syndrom in young adolescents. *American Journal of Epidemiology, 116*, 791–800.
Shafer, A. B. (2006). Meta-analysis of the factor structures of four depression questionnaires: Beck, CES-D, Hamilton, and Zung. *Journal of Clinical Psychology, 62*, 123–146.
Spence, S. H., Burns, J., Boucher, S., Glover, S., Graetz, J., Kay, D., et al. (2005). The beyondblue schools research initiative: the conceptual framework and intervention. *Australasian Psychiatry, 13*(2), 159–164.
Steinberg, L. (2001). We know some things: parent–adolescent relationships in retrospect and prospect. *Journal of Research on Adolescence, 11*, 1–19.
World Health Organization. (1992). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines*. Geneva, Switzerland: World Health Organization.