Title	Validating the Beck Depression Inventory-II for Hong Kong Community Adolescents
Author(s)	Byrne, BM; Stewart, SM; Lee, PWH
Citation	International Journal of Testing, 2004, v. 4 n. 3, p. 199-216
Issued Date	2004
URL	http://hdl.handle.net/10722/43488
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Validating the Beck Depression Inventory–II for Hong Kong Community Adolescents

Barbara M. Byrne
School of Psychology
University of Ottawa

Sunita M. Stewart
University of Texas Southwestern Medical Center at Dallas

Peter W. H. Lee
Department of Psychiatry
The University of Hong Kong

The primary purpose of this study was to test for the validity of a Chinese version of the Beck Depression Inventory–II (C–BDI–II) for use with Hong Kong community (i.e., nonclinical) adolescents. Based on a randomized triadic split of the data ($N = 1460$), we conducted exploratory factor analysis on Group1 ($n = 486$) and confirmatory factor analysis (CFA) within the framework of structural equation modeling on Groups 2 ($n = 487$) and 3 ($n = 487$); the second CFA served as a cross-validation of the determined factor structure. Factor analytic results, based on a 4-factor structure that comprised 1 2nd-order general factor of Depression and 3 1st-order factors representing Negative Attitude, Performance Difficulty, and Somatic Elements, replicated those reported previously for Canadian (Byrne & Baron, 1993), Swedish (Byrne, Baron, Larsson, & Melin, 1995), and Bulgarian (Byrne, Baron, & Balev, 1998) nonclinical adolescents. Based on this cross-validated factor structure, findings related to internal consistency reliability, stability over a 6-month time lag, and relations with relevant external criteria provided strong support for the valid use of the C–BDI–II in measuring depressive symptoms for Hong Kong community adolescents.

Requests for reprints should be sent to Barbara M. Byrne, 3301 Diamond Key Court, Punta Gorda, FL 33955–4656. E-mail: bmbyrne@swfla.rr.com

Given the many pressures, anxieties, and uncertainties inherent in our present-day world, it is not surprising that people in all societies and walks of life are experiencing the debilitating effects of depression. Indeed, the World Health Organization (2001) identified depression as one of the leading causes of disability worldwide. Pertinent to adolescents, in particular, the prevalence of depression is now well-established (Lewinsohn, Hops, Roberts, Seeley, & Andrews, 1993), with symptom patterns reportedly similar to those of adults (Lewinsohn, Pettit, Joiner, & Seeley, 2003). Moreover, even those afflicted with subclinical depression can exhibit significant negative psychosocial behavior (Lewinsohn, Solomon, Seeley, & Zeiss, 2000).

Fortunately, depression among adolescents is treatable (Curry, 2001), and use of psychometrically sound and appropriate screening instruments can identify potential cases of depression quickly and inexpensively. Unfortunately, the availability of such instruments in Eastern countries in general and in Asia in particular is extremely limited (see, e.g., F. M. Cheung, Leong, & Ben-Porath, 2003). Thus, psychometric information related to an appropriate screening instrument, specifically adapted for use in Chinese populations, would have high utility not only for practitioners, but also for researchers for whom they would serve a pivotal role in empirical investigations of adolescent depressive disorders. The intent of the study reported in this article is to provide this information as it relates to the Chinese version (C–BDI–II; Chinese Behavioral Sciences Society, 2000) of the Beck Depression Inventory–II (BDI–II; Beck, Steer, & Brown, 1996) for use with Hong Kong adolescents.

A review of the literature bearing on adolescent depression reveals the BDI (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) to be the favored assessment measure (Dozois, Dobson, & Ahnberg, 1998; Leung & Wong, 2003). However, despite widespread use of the BDI with both clinical and nonclinical populations, and extensive reporting on its psychometric properties (see, e.g., Beck, Steer, & Garbin, 1988), only one study prior to 1993 (Tanaka & Huba, 1984) had actually tested hypotheses related to its factor structure; all other validity studies had used only exploratory factor analytic procedures, albeit with widespread varying results (Dozois et al., 1998). Recognizing a need to validate use of the BDI with nonclinical adolescents, Byrne and Baron (1993) tested and cross-validated the instrument using data from three independent samples of Canadian adolescents. The results of their study revealed the data to be most appropriately represented by a four-factor model that comprised one higher order factor of general Depression and three lower order factors that represented Negative Attitudes, Performance Difficulty, and Somatic Elements. Validity of this model of BDI structure has subsequently been tested for Swedish (Byrne et al., 1995) and Bulgarian (Byrne et al., 1998) adolescents and its invariance tested across gender (Byrne, Baron, & Balev, 1996; Byrne, Baron, & Campbell, 1993, 1994; Byrne, Baron, Larsson, & Melin, 1996), across the three cultural groups noted previously (Byrne & Campbell, 1999), and across Canadian English and French cultural groups (Byrne & Baron, 1994).

Subsequent to the work of Byrne, Baron, and colleagues, the BDI–II (Beck et al., 1996) was developed and made available. Essentially, the BDI–II was designed to parallel refinements in the definition of depression over the years and to address recommendations that the instrument be made more compatible with criteria set forth in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–IV*; American Psychiatric Association, 1994; Dozois et al., 1998). Although the BDI and BDI–II are consistent in being composed of 21 items, several changes were implemented in the latter. First, several items were reworded in the interest of clarity. Second, content designed to tap sleep and appetite changes were modified to reflect both increases and decreases in this behavior. Third, 4 new items assessing agitation, worthlessness, concentration difficulty, and loss of energy were added, while 4 old items assessing body image, work difficulty, weight loss, and hypochondria were deleted. Finally, 2 items were relocated to a different area of the inventory (Beck et al., 1996).

The C–BDI–II was adapted from the original BDI–II by the Chinese Behavioral Sciences Society and made available in 2000 through The Psychological Corporation. Although psychometric information has been reported for the Chinese version of the original BDI (Chan, 1991; Shek, 1990, 1991), to date, no empirical studies have yet been published in English-language journals regarding the C–BDI–II (Leong, Okazaki, & Tak, 2003). Thus, our study is the first to provide this critically needed information. By doing so, it addresses the concern that use of an assessment measure, merely on the basis of its translation into the language of choice, is clearly inadequate (F. M. Cheung, Leon, & Ben-Porath, 2003); substantially more empirically tested information is needed regarding its construct validity and other psychometric data.

Given that items comprising the C–BDI–II were translated from English into Chinese for use with populations in China, the instrument represents an adaptation of the BDI–II. As such, the C–BDI–II realistically stands as a new instrument, thereby requiring that both the validity of its factor structure and the soundness of its psychometric properties be tested statistically (see, e.g., Hambleton, Yu, & Slater, 1999; Tanzer, 1995). As van de Vijver and Hambleton (1996) so cogently noted, just because a measuring instrument has demonstrated adequate validity and reliability in one culture, it cannot be assumed that these same psychometric properties will prevail in another culture; such evidence needs to be empirically derived (see, e.g., Tanzer & Sim, 1999). This, then, was the focus of the study described here wherein we report findings from several approaches to testing for the validity of the C–BDI–II for use with Hong Kong community adolescents. Specifically, we test for (a) a conceptually meaningful factor structure using both exploratory factor analyses (EFA) and confirmatory factor analyses (CFA), followed by cross-validation of this determined factor structure using CFA; (b) internal consistency reliability related to the total scale, as well as to each factor or subscale, at each of two time points; (c) stability of both the lower and higher order factors over a 6-month

time lag; (d) invariance of both the lower and higher order factor loadings, and the measurement error variances over time; and (e) conceptually meaningful correlations between depression and related external criteria.

METHOD

Sample and Procedure

As part of a larger study designed to assess cognitive concomitants of suicidality and depressed mood, data used in the study presented here comprised C–BDI–II responses for 1460 (males, $n = 687$; females, $n = 773$) adolescents attending 11 Hong Kong schools. In light of our large sample, together with our intent to test aspects of BDI structure from different perspectives that involved various analytic procedures, we preferred to work with data that were complete both within and across time. More important, however, our decision to implement listwise deletion of missing data was based on two vital criteria: (a) the amount of missing data across two time points was less than 9%; and (b) the data were missing completely at random (see Little & Rubin, 1987), as evidenced by nonsignificant results derived from the Generalized Least Squares test of homogeneity of covariance matrices representing complete and incomplete data, $\chi^2(4774, N = 1,602) = 3019.20$, $p = 1.0$. Participants ranged in age from 14 to 18 years ($Mdn = 15$ years), and sampling was designed to incorporate three indexes of variation in this culture: geographical location, socioeconomic status, and range of academic abilities. Final average participation was 80%, with all students being ethnic Chinese.

Measures were administered to all participants on two occasions with a time lag of 6 months. Informed consent was obtained from students, and participation was voluntary. The study was approved by the Ethics Committee of the Faculty of Medicine at the University of Hong Kong.

Instrumentation

The C–BDI–II, consistent with the BDI–II, is composed of 21 items that are structured on a 4-point scale ranging from 0 (*symptom not present*) to 3 (*symptom strongly present*), with resulting summary scores ranging from 0 to 63. Item content for the BDI–II (and, relatedly, for the C–BDI–II) correspond to criteria for the diagnosis of depressive disorders as specified in the *DSM–IV* (American Psychiatric Association, 1994). As such, the instrument is used to assess the severity of depressive symptomatology. Internal consistency reliability has been reported for the total BDI–II scale for both college students ($\alpha = .93$, Beck et al., 1996; $\alpha = .91$, Dozois et al., 1998) and clinical outpatients ($\alpha = .92$, Beck et al., 1996).

Specific to our study, only 20 of the 21 C–BDI–II items were used in tapping depressive symptoms for high school adolescents. Item 21, designed to assess

changes in sexual interest, was considered to be objectionable by several school principals, and the item was subsequently deleted from the inventory.

Data Analyses

Analyses were conducted in several stages using the EQS 6.1 (Bentler, 2004) program; each is now fully described.

Stage 1.   The data were first randomly split into three independent groups and then tested using one of two factor analytic approaches. For Group 1 ($n = 486$), we applied EFA using equal prior instant communalities extraction with direct oblimin rotation. Considering sample size and case–variable ratio (Gorsuch, 1983), together with the tendency for oblique solutions to yield loadings that are lower than those from orthogonal solutions (Kerlinger, 1984), a value of .35 was determined as a viable cutpoint for judging the saliency of factor loadings. Using the three-factor structure determined to be most appropriate for Group 1, and heeding the recommendations of Byrne and Baron (1993) that BDI structure is most appropriately represented as a hierarchical structure, we tested next for the validity of a four-factor second-order structure for Group 2 ($n = 487$) using CFA within the framework of structural equation modeling. More specifically, the model specified the three lower order factors determined in the EFA, albeit with the addition of a single higher order factor of Depression. Several criteria were used in determining the goodness of fit to the data for this hypothesized structure; these included the Comparative Fit Index (CFI; Bentler, 1990), the standard root mean squared residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA), along with its related 90% confidence interval (Steiger, 1990). (Evaluative criteria related to each is detailed later in the Results section.) Finally, using CFA again, the best-fitting model for Group 2 was cross-validated with data representing Group 3 ($n = 487$). All subsequent testing of the data were based on the full sample ($N = 1460$).

Stage 2.   Based on the final best-fitting model from Stage 1, we determined the internal consistency reliability, at two time points, for the total scale and for each of the three lower order factors. Relatedly, we also determined stability of the higher order factor and the three lower order factors by examining the latent factor correlation between each with its Time 2 counterpart.

Stage 3.   Stability of both the higher and lower order factor loadings was assessed using analysis of covariance structures to test for their invariance across the two time points.

Stage 4. As a final validity check of our postulated structure of the C–BDI–II, we considered it important to examine correlations between depression and theoretically linked external criteria. Based on theory and empirical research, scores on depression should correlate with risk factors as well as with environmental precipitants. Cognitive theories of depression (see, e.g., Beck, Rush, Shaw, & Emery, 1979) suggest that underlying beliefs, such as the tendency to interpret events to support negative predictions (cognitive distortions) and to attribute negative events to stable causes (hopelessness), are central to the development of depressed mood. Suicidal ideation and intent also have been proposed as having depression as a core underlying concomitant (Lewinsohn, Rohde, & Seeley, 1996). For this reason, we expected that hopelessness, cognitive distortions, and suicidal ideation would show strong correlations with depression scores.

Self-efficacy—an important protective variable according to the Western literature—was also expected to correlate with depression scores (Bandura, 1997). In contrast to the variables cited previously, however, we anticipated this relation to be weaker for at least two reasons. First, self-efficacy may have an indirect effect on depressive symptoms, thereby influencing behaviors that decrease the likelihood of increased environmental stress (Bandura, 1997). Second, cross-cultural theory suggests that, in collective cultures (e.g., Hong Kong), beliefs that emphasize internal sense of personal worth, efficacy, and control may be less salient than in individualistic cultures (Markus & Kitayama, 1994); relatedly, they should therefore be weakly protective against depressed mood. Given our adolescent data, we also expected some variance to be explained by environmental conditions such as peer acceptance, parental acceptance, and school performance. However, because the effects of these conditions may be filtered through the pathway of cognitions, we expected their correlations with depression to be weaker than the correlation between cognitions (specifically cognitive distortions and hopelessness) and depression.

Accordingly, Pearson product-moment correlations were examined between total scores on the C–BDI–II and scores on the Hopelessness Scale for Children (Kazdin, Rodgers, & Colbus, 1986), the Children's Negative Cognitive Errors Questionnaire (Leitenberg, Yost, & Carroll-Wilson, 1986), a 4-item Suicidality Scale (Lewinsohn et al., 1996), and the General Self-Efficacy Scale (Schwarzer, 1992). In addition, correlations were examined between C–BDI–II total scores and students' responses to questions related to peer acceptance (e.g., "My friends care about me and accept me the way I am"), parental acceptance (e.g., "My parents are caring and understanding"), and school performance. The mean score for each of these variables was based on self-reported responses structured on a 5-point Likert scale that ranged from 1 (*strongly disagree*) to 5 (*strongly agree*) for peer and parental acceptance and from 1 (*very bad*) to 5 (*very good*) for school performance.

RESULTS

Stage 1 Analyses

EFA.   These analyses, performed on Group 1 (n = 486) data, yielded a three-factor solution that was remarkably similar to the three lower order factor structures reported by Byrne and colleagues (1993, 1994, 1995, 1996, 1998, 1999) in their many studies of the BDI for adolescents. With the exception of 4 items, all remaining items loaded substantially on three factors that could be labeled appropriately as Negative Attitude, Performance Difficulty, and Somatic Elements (consistent with Byrne et al.). All factor loadings are reported in Table 1; the underlined and italicized factor loadings represent items considered to load on an inappropriate factor. More specifically, it would be more reasonable that Items 1 (Sadness) and 10 (Crying) load on the Negative Attitude factor, despite evidence of their substantial loadings on the Performance Difficulty factor. Likewise, Item 15 (Energy) and Item 20 (Fatigue) should more reasonably load on the Somatic Elements factor, rather than the Performance Difficulty factor.

Taking into account conceptual meaningfulness, together with previous empirical research bearing on the factor structure of the BDI, we considered it both ap-

TABLE 1
Exploratory Factor Analysis Three—Factor Solution

Item Descriptor	Factor 1—Negative Attitude	Factor 2—Performance Difficulty	Factor 3—Somatic Elements
1. Sadness		*0.489*	
2. Pessimism	0.321		
3. Past failure	0.551		
4. Loss of pleasure		0.455	
5. Guilty feelings	0.535		
6. Punishment feelings	0.599		
7. Self-dislike	0.469		
8. Self-criticalness	0.402		
9. Suicidal thoughts	0.508		
10. Crying		*0.341*	
11. Agitation		0.697	
12. Loss of interest		0.535	
13. Indecisiveness		0.488	
14. Worthlessness	0.533		
15. Loss of energy		*0.589*	
16. Changes in sleeping pattern			0.465
17. Irritability		0.711	
18. Changes in appetite			0.384
19. Concentration difficulty		0.466	
20. Fatigue		*0.646*	

propriate and important to adjust the loadings of these 4 items in the specification of our postulated structure of the C–BDI–II to be tested next using CFA. We turn now to these results.

CFA 1.   Based on data from Group 2 (*n* = 487)—the calibration sample—we tested for the validity of C–BDI–II structure as described previously and schematically portrayed in Figure 1. Results revealed an exceptionally well-fitting model to the sample data. Goodness-of-fit statistics related to the test of this hypothesized model are shown in Table 2.
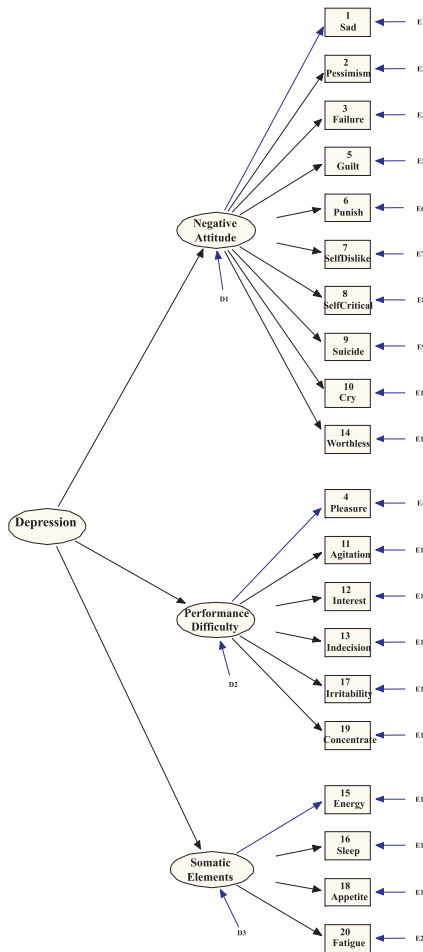

FIGURE 1   Hypothesized model of factorial structure for the Chinese version of the Beck Depression Inventory–II.

TABLE 2
Hypothesized Model of C–BDI–II Structure: Goodness-of-Fit Statistics

Group	χ^2	df	S-Bχ^2	*CFI	*RMSEA	90% *RMSEA CI	SRMR
Calibration[a]	329.56	168	254.11	.95	.03	.02, .04	.04
Validation[a]	375.01	168	289.58	.94	.04	.03, .05	.05

Note. C–BDI–II = Chinese Beck Depression Inventory–II; S-Bχ^2 = Satorra-Bentler scaled statistic; *CFI = robust Comparative Fit Index; *RMSEA = robust root mean square error of approximation; CI = confidence interval; SRMR = standard root mean squared residual.
[a]$N = 487$.

By convention, the chi-square statistic and its related degrees of freedom are reported in the analysis of covariance structures. The chi-square likelihood ratio test assesses the magnitude of discrepancy between the hypothesized (or fitted) covariance matrix and the sample covariance matrix. However, at least two weaknesses associated with the chi-square test are now well-known and therefore preclude its use as a viable evaluative indicator of goodness of fit. First, given small to moderate discrepancies of the data from normality, the chi-square test will reject the hypothesized model (West, Finch, & Curran, 1995). Second, given its extreme sensitivity to sample size, the chi-square test will reject a model on the basis of very small discrepancies from the model that may be of no theoretical or practical substance (Bentler & Bonett, 1980; West et al., 1995). As a consequence, other goodness-of-fit statistics that take a more pragmatic approach to the model evaluation process have been developed and recommended in reporting results for analysis of covariance structures. As noted earlier and heeding the recommendations of Hu and Bentler (1998), those included in this article are the CFI, the SRMR, and the RMSEA.

Before describing these practical indexes of fit, it is important to note that the values reported in Table 2 represent what the EQS program terms as "robust statistics." That is, they have been scaled (or corrected) to take into account some non-normality in the data. Given that our data include responses from community (i.e., nonclinical) adolescents, evidence of both skewness and kurtosis is certainly not unexpected and indeed is consistent with other BDI studies of community samples in general and adolescents in particular. Thus, this aspect of the data is of little concern. However, given these distributional characteristics, what is important is that analyses be based on the correct statistics—that is, they are based on algorithms designed to take this non-normality into account. With EQS 6.1, researchers have the option of using robust statistics, rather than the regular statistics, in addressing this problem. Interpretation of all analytic work in this study was based on the robust statistics.

The S-Bχ^2 statistic reported in Table 2 represents the Satorra-Bentler chi-square (Satorra & Bentler, 1988), which serves as a correction for the chi-square statistic when distributional assumptions are violated. It has been shown to be the most reli-

able test statistic for evaluating covariance structure models under various distributions and sample sizes (Hu, Bentler, & Kano, 1992). In the study presented here, all CFA analyses were based on the S-Bχ^2, rather than on the uncorrected chi-square statistic.

The *CFI represents the robust version of the CFI in that its computation is based on the S-Bχ^2 statistic. It ranges in value from 0 to 1.00, with a value of .95 serving as the rule-of-thumb cutpoint of acceptable fit (see Hu & Bentler, 1999). The *RMSEA is a robust version of the usual RMSEA and takes into account the error of approximation in the population and asks the question, "How well would the model, with unknown but optimally chosen parameter values, fit the population covariance matrix if it were available?" (Browne & Cudeck, 1993, pp. 137–138). This discrepancy, as measured by the *RMSEA, is expressed per degree of freedom, thus making it sensitive to model complexity; values less than .05 indicate good fit, and values as high as .08 represent reasonable errors of approximation in the population. For completeness, we also include the 90% confidence interval provided for *RMSEA. Finally, the SRMR is the average standardized residual value derived from fitting the hypothesized variance covariance matrix to that of the sample data. Its value ranges from 0 to 1.00, with a value less than .08 being indicative of a well-fitting model (Hu & Bentler, 1999).

In reviewing Table 2, it is clearly evident that the hypothesized model of C–BDI–II structure, as portrayed in Figure 1, represented a very well-fitting model for the calibration sample of adolescents. Furthermore, all parameters were statistically significant, and there was no indication that the model could be further improved, substantially, through the incorporation of additional parameters. Of interest now, then, is the extent to which the model is replicated over a second independent sample.

CFA 2.   As indicated in Table 2, testing of the hypothesized model for Group 3—the validation sample—once again yielded a very well-fitting model, and all parameters were statistically significant. From these results, we concluded that the hypothesized model of C–BDI–II structure, as shown in Figure 1, represented data for nonclinical Hong Kong adolescents extremely well.

All subsequent analyses used in determining the psychometric properties of the instrument were based on the full sample of adolescents ($N = 1460$). A summary of descriptive statistics related to this sample is presented in Table 3.

Stage 2 Analyses

Internal consistency.   Internal consistency reliability coefficients, as computed for Cronbach's coefficient alpha, are reported in Table 4 for both Time 1 and Time 2. Internal consistency of the total scale score for overall depression was high at Time 1 ($\alpha = .94$), but slightly lower, albeit still substantial, at Time 2 ($\alpha = .91$).

TABLE 3
C–BDI–II Descriptive Statistics for Full Sample of Hong Kong Adolescents

Item Descriptor	M		SD		Skewness		Kurtosis	
	T1	T2	T1	T2	T1	T2	T1	T2
1. Sadness	0.75	0.71	0.88	0.86	0.70	0.81	−0.83	−0.60
2. Pessimism	0.46	0.49	0.72	0.69	1.56	1.41	1.98	1.82
3. Past failure	0.90	0.79	0.91	0.89	0.37	0.60	−1.33	−1.03
4. Loss of pleasure	0.50	0.47	0.71	0.68	1.54	1.49	0.71	2.24
5. Guilty feelings	0.55	0.46	0.76	0.71	1.44	1.66	1.80	2.50
6. Punishment feelings	0.71	0.55	0.90	0.86	1.33	1.59	0.94	1.68
7. Self-dislike	0.56	0.52	0.86	0.81	1.44	1.46	1.12	1.20
8. Self-criticalness	0.66	0.54	0.93	0.80	1.37	1.60	0.89	2.08
9. Suicidal thoughts	0.30	0.26	0.54	0.54	1.89	2.28	4.03	5.56
10. Crying	0.53	0.42	0.98	0.84	1.64	2.03	1.24	2.97
11. Agitation	0.92	0.92	0.86	0.82	0.73	0.66	−0.26	−0.77
12. Loss of interest	0.50	0.51	0.67	0.70	1.31	1.37	1.71	1.68
13. Indecisiveness	0.63	0.61	0.69	0.71	0.98	1.06	0.95	1.01
14. Worthlessness	0.52	0.51	0.79	0.75	1.38	1.36	0.99	1.03
15. Loss of energy	0.76	0.79	0.82	0.82	0.81	0.77	−0.11	−0.14
16. Changes in sleeping pattern	0.91	0.96	0.81	0.77	0.47	0.37	−0.57	−0.46
17. Irritability	0.56	0.63	0.69	0.74	1.14	0.99	1.02	0.45
18. Changes in appetite	0.61	0.60	0.86	0.82	1.38	1.38	1.11	1.29
19. Concentration difficulty	0.89	0.86	0.82	0.81	0.61	0.60	−0.26	−0.38
20. Fatigue	0.90	0.92	0.70	0.73	0.46	0.53	0.19	0.73

Note. $N = 1,460$. C–BDI–II = Chinese Beck Depression Inventory–II; T1 = Time 1; T2 = Time 2.

TABLE 4
Internal Consistency and Stability of the C–BDI–II for Full Sample[a] of
Hong Kong Adolescents

C–BDI–II Subscales	Internal Consistency Coefficients		Stability Coefficients[a] Across Time 1 and Time 2
	Time 1	Time 2	
Negative attitude	0.83	0.84	0.73
Performance difficulty	0.79	0.82	0.66
Somatic elements	0.68	0.67	0.69
Depression (total score)	0.94	0.91	0.74

Note. $N = 1,460$. C–BDI–II = Chinese Beck Depression Inventory–II.
[a]Stability coefficients represent latent factor correlations.

Indeed, these Time 1 results are quite consistent with those reported by Beck et al. (1996) and Dozois et al. (1998) for the BDI–II. Although internal consistency for the Somatic Elements subscale was somewhat weaker than for the Negative Atti-

tude and Performance Difficulty subscales, it nonetheless exhibited fairly good re-liability. Indeed, the somewhat lower alpha coefficient for the Somatic Elements factor may be due to the relative rarity of somatic symptoms in the normal popula-tion. Along with depressed mood, even only one or two somatic symptoms in-crease the likelihood of clinical levels of depressive disorder.

Stability.   To assess the constancy of the three lower order factors, as well as the higher order general factor of depression, we examined latent factor correla-tions across time. As indicated in Table 4, both the higher order factor of Depres-sion and the Negative Attitude factor exhibited fairly good stability, with the re-maining factors somewhat less so. Indeed, depressed mood, as measured by the C–BDI–II, is likely to have both state and trait features. Furthermore, some degree of drift in the measure over a 6-month period would not be unexpected, given the responsiveness of mood to external stressors that vary over time. However, the pro-pensity to respond to stress with depressive symptoms may be more of a trait mea-sure, thereby accounting for a degree of stability over time. For example, Tems and colleagues (Tems, Stewart, Skinner, Hughes, & Emslie, 1993) reported that, even following remission of their disorder, children and adolescents who had been hos-pitalized reported higher levels of depressive symptoms than did their nonhospitalized controls; albeit, these levels were lower than when they were ini-tially hospitalized.

Stage 3 Analyses

In addition to estimating the stability of the factors themselves, we considered it important also to assess the stability of hierarchically structured factor loadings. To this end, we tested for the invariance of all factor loadings estimated at Time 1 across the time lag of 6 months. As such, we first tested for the overall fit of the 4-factor model at Time 1 and again at Time 2; no equality constraints were im-posed. The fit of this model (*CFI = .93), as shown in Table 5, was slightly less than the cutoff point suggested by Hu and Bentler (1999), but nonetheless indica-tive of a well-fitting model that included two time points of data.

We next proceeded to test for the invariance of the lower order factor loadings. As such, the eight-factor model (Time 1 and Time 2) was again estimated, but this time with equality constraints placed on all lower order factor loadings across Time 1 and Time 2. Determining evidence of invariance involves testing and com-paring the difference in fit for a series of nested models. For example, comparison of a model in which no constraints are imposed (Model 1) with one in which equal-ity constraints are specified for all lower order factor loadings (Model 2) would constitute a nested model comparison. Taking the difference between the two chi-square values and their degrees of freedom renders a chi-square difference value ($\Delta\chi^2$) that is distributed as a chi-square statistic, with degrees of freedom

TABLE 5
Tests for Invariance of C–BDI–II Factor Structure Across Time:
Goodness-of-Fit and Comparative Statistics

Model	χ^2	df	S-Bχ^2	*CFI	Δdf	ΔS-Bχ^{2a}	Δ*CFI
1. No constraints	2233.21	722	1738.66	.929	—	—	—
2. All lower order factor loadings constrained equal	2285.11	739	1781.12	.928	17	41.82*	.001
3. All lower and higher order factor loadings constrained equal	2291.73	742	1787.07	.927	20	48.36*	.002
4. All factor loadings and error variances constrained equal	2458.12	762	1914.70	.920	40	175.16*	.009

Note. C–BDI–II = Chinese Beck Depression Inventory–II; S-Bχ^2 = Satorra-Bentler scaled statistic; *CFI = robust Comparative Fit Index.
[a]Represents corrected value.
*$p < .001$.

equivalent to the difference in degrees of freedom (Δdf) between the two models. A significant $\Delta\chi^2$ value argues for evidence of noninvariance between the two models. Although these difference values have typically been reported for invariance studies, the fact remains that they are both equally as sensitive to sample size and non-normality as the chi-square statistic itself. Analogously, the same comparisons can be made based on the ΔS-B$\chi^2_{(\Delta df)}$; albeit, a correction to the value is needed, as this difference is not distributed as chi-square (Bentler, 2004). Nonetheless, until the recent simulation study of nested model comparisons involving various practical indexes of fit (G. W. Cheung & Rensvold, 2002), there has been no alternative statistic that could be used to replace the chi-square (and S-Bχ^2) difference test. Based on their findings, G. W. Cheung and Rensvold proposed that a difference in CFI values equal to or less than .01 can rightfully serve as viable evidence of invariance. Given that the large size of our sample would most certainly trigger a significant ΔS-Bχ^2, regardless of whether the item scores were totally invariant across time, we based conclusions of invariance on the Δ*CFI criterion, together with overall goodness of fit.

Accordingly, given an overall fit of .93 and a minimal change in overall fit between Model 1 and Model 2 (Δ*CFI = .001), we concluded that the lower order factor loadings were invariant across time. Likewise, comparison of Model 3, in which both the lower and higher order loadings were constrained equal across time, with Model 1—the unconstrained model—yielded a Δ*CFI of .002, thereby once again indicating evidence of invariance across time. Moreover, goodness of fit for the whole model remained virtually unchanged from that for Model 2 (*CFI = .93).

In a final, extremely stringent assessment of invariance, we tested for the equality of measurement error variances across time; this test reflects on the reliability of the items. Once again, results yielded a Δ*CFI of .009, thereby providing good

evidence of invariance across time. Although the goodness of fit dropped a notch to .92, this two time point model still represented an adequately fitting model. In fact, given that all factor loadings, as well as all error covariances, were constrained equal, it is rather astounding that the model fits as well as it does!

Stage 4 Analyses

From a practical perspective, it is important to know the extent to which scores from a measuring instrument correlate with theoretically related external criteria. For example, with respect to the measurement of depression, it is of interest to practitioners to know the extent to which scores from the BDI–II correlate with relevant cognitive variables such as hopelessness, cognitive distortions, suicide ideation, and self-efficacy. For practitioners working with adolescents, in particular, it may also be important to know how depression relates to peer acceptance, parental acceptance, and actual school performance. Based on structure of the C–BDI–II, as postulated in the study presented here, correlations with these external criteria are reported in Table 6.

As indicated earlier, correlations between depression and hopelessness, cognitive distortions, and suicide would be expected to be high. In contrast, correlations between depression and self-efficacy, peer acceptance, parental acceptance, and school performance would be expected to be lower. Indeed, a review of the results presented in Table 6 would seem to fulfill these expectations quite adequately.

SUMMARY AND DISCUSSION

The primary focus in the study reported here was to validate use of the C–BDI–II with Hong Kong community adolescents. To this end, we identified, tested, and

TABLE 6
Correlation Between the C–BDI–II and External Criteria for Hong Kong Adolescents[a]

External Criteria	Depression
Self efficacy	−0.19
Hopelessness	0.57
Cognitive distortions	0.51
Suicidal ideation	0.40
Parental acceptance	−0.25
Peer acceptance	−0.27
School performance	−0.23

Note.   C–BDI–II = Chinese Beck Depression Inventory–II.
[a]Pearson product-moment correlation coefficients.

cross-validated the factor structure of the C–BDI–II for three independent samples of randomly split data using EFA and CFA. Consistent with previous studies of the BDI for nonclinical adolescents (Byrne et al., 1998; Byrne et al., 1993; Byrne et al., 1995), our findings revealed the data for Hong Kong adolescents to be best represented by a hierarchically structured model defined by three lower order factors comprising Negative Attitude, Performance Difficulty, and Somatic Elements and one higher order factor representing general Depression (see Figure 1). Not only was this hypothesized model exceptionally well-fitting, but examination of the modification indexes for evidence of possible misspecification suggested no viably sound rationale for further modifying the existing parameterization.

Provided with evidence of a sound factor structure, we proceeded next to test for its internal consistency reliability and for its stability across two time points. Typically, only total score internal consistency is reported, and in this regard, findings from our study were consistent with those reported for the BDI–II (Beck et al., 1996; Dozois et al., 1998). In the interest of completeness, however, we also tested for internal consistency related to each of the three factors (Negative Attitude, Performance Difficulty, and Somatic Elements). In this instance, values were at least acceptable, with the weakest findings being associated with the Somatic Elements factor. However, the fact that somatic symptoms tapped by the 4 items comprising this scale are rarely found in the normal population likely accounts for its somewhat lower reliability.

Likewise, we considered overall depression score stability to be quite acceptable, particularly given the potential for external stress to intervene over the 6-month test–retest of C–BDI–II responses. On the other hand, some stability would nonetheless be expected, as individuals are likely to have enduring trait-like tendencies in their mood responsiveness. As predicted, overall depression correlated most strongly with cognitive variables and suicidality and less so, albeit still significantly, with self-efficacy, peer and family acceptance, and school performance. These findings were consistent with those of an earlier study (Stewart, Betson, Lam, Chung, Ho, & Chung, 1999) that demonstrated congruity between the correlates of depressed mood reported in Western studies and those found for Hong Kong adolescents.

We believe that our study has addressed two concerns cited in the literature. First, Ehrenberg, Cox, and Koopman (1990) emphasized the need for research directed toward the early detection of depression for adolescents. Given that such detection is dependent on the use of psychometrically sound instrumentation, we are hopeful that our findings regarding the C–BDI–II are of substantial interest and utility to researchers and practitioners concerned with the well-being of Chinese adolescents. Second, F. M. Cheung et al. (2003) noted that research on depression in East Asia has relied heavily on translated assessment measures, with the BDI being the instrument of choice. Undoubtedly, this trend is likely to continue with respect to the BDI–II. However, translated instruments, regardless of their recon-

structive precision, address only the linguistic equivalence of an assessment measure (see, e.g., Leung & Wong, 2003; van de Vijver & Hambleton, 1996). Beyond this basic requirement lies the need to establish their construct validity and psychometric properties before claims of complete adaptation for use with a particular population can be made. To this end, we consider the findings from the study described in this article to have made important inroads toward the goal of establishing a sound adaptation of the C–BDI–II for use with Chinese adolescents in general and Hong Kong adolescents in particular.

ACKNOWLEDGMENT

REFERENCES

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: Guilford.

Beck, A. T., Steer, R., & Brown, G. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: The Psychological Association.

Beck, A. T., Steer, R., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8,* 77–100.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P. M. (2004). *EQS 6.1: Structural equations program manual*. Encino, CA: Multivariate Software Inc.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M., & Baron, P. (1993). The Beck Depression Inventory: Testing and cross-validating an hierarchical structure for nonclinical adolescents. *Measurement and Evaluation in Counseling and Development*, *26*, 164–178.

Byrne, B. M., & Baron, P. (1994). Measuring adolescent depression: Tests of equivalent factorial structure for English and French versions of the Beck Depression Inventory. *Applied Psychology: An International Review*, *43*, 33–47.

Byrne, B. M., Baron, P., & Balev, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian adolescents. *Personality and Individual Differences*, *21*, 641–651.

Byrne, B. M., Baron, P., & Balev, J. (1998). The Beck Depression Inventory: A cross-validated test of factorial structure for Bulgarian adolescents. *Educational and Psychological Measurement*, *58*, 241–251.

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.

Byrne, B. M., Baron, P., & Campbell, T. L. (1994). The Beck Depression Inventory (French version): Testing for gender-invariant factorial structure for nonclinical adolescents. *Journal of Adolescent Research*, *9*, 166–179.

Byrne, B. M., Baron, P., Larsson, B., & Melin, L. (1995). The Beck Depression Inventory: Testing and cross-validating a second-order factorial structure for Swedish nonclinical adolescents. *Behaviour Research and Therapy*, *33*, 345–356.

Byrne, B. M., Baron, P., Larsson, B., & Melin, L. (1996). Measuring depression for Swedish nonclinical adolescents: Factorial validity and equivalence of the Beck Depression Inventory across gender. *Scandinavian Journal of Psychology*, *37*, 37–45.

Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross-cultural Psychology*, *30*, 557–576.

Chan, D. W. (1991). The Beck Depression Inventory: What difference does the Chinese version make? *Psychological Assessment*, *3*, 616–622.

Cheung, F. M., Leong, F. T. L., & Ben-Porath, Y. S. (2003). Psychological assessment in Asia: Introduction to the special section. *Psychological Assessment*, *15*, 243–247.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Chinese Behavioral Sciences Society. (2000). *The Chinese version of the Beck Depression Inventory, Second Edition. Licensed Chinese translation*, The Psychological Corporation. New York: Harcourt Brace.

Curry, J. (2001). Specific psychotherapies for childhood and adolescent depression. *Biological Psychiatry*, *49*, 1091–1100.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment*, *10*, 83–89.

Ehrenberg, M. F., Cox, D. N., & Koopman, R. F. (1990). The prevalence of depression in high school students. *Adolescence*, *13*, 127–134.

Gorsuch, R. (1983). *Factor analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hambleton, R. K., Yu, J., & Slater, S. C. (1999). Fieldtest of the ITC guidelines for adapting educational and psychological tests. *European Journal of Psychological Assessment*, *15*, 270–276.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424–453.

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.

Hu, L.-T., Bentler, B. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362.

Kazdin, A. E., Rodgers, A., & Colbus, D. (1986). The Hopelessness Scale for Children: Psychometric characteristics and concurrent validity. *Consulting and Clinical Psychology*, *54*, 241–245.

Kerlinger, F. N. (1984). *Liberalism and conservatism: The nature and structure of social attitudes*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Leitenberg, H., Yost, L., & Carroll-Wilson, D. (1986). Negative cognitive errors in children: Questionnaire development, normative data, and comparisons between children with and without self-reported symptoms of depression, low self-esteem, and evaluation anxiety. *Journal of Consulting and Clinical Psychology*, *54*, 528–536.

Leong, F. T. L., Okazaki, S., & Tak, J. (2003). Assessment of depression and anxiety in East Asia. *Psychological Assessment*, *15*, 290–305.

Leung, P. W. L., & Wong, M. M. T. (2003). Measures of child and adolescent psychopathology in Asia. *Psychological Assessment*, *15*, 268–279.

Lewinsohn, P. M., Hops, H., Roberts, R. E., Seeley, J. R., & Andrews, J. A. (1993). Adolescent psychopathology I: Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *Journal of Abnormal Psychology*, *102*, 133–144.

Lewinsohn, P. M., Pettit, J. W., Joiner, T. E., & Seeley, J. R. (2003). The symptomatic expression of Major Depressive Disorder in adolescents and young adults. *Journal of Abnormal Psychology*, *112*, 244–252.

Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1996). Adolescent suicidal ideation and attempts: prevalence, risk factors, and clinical implications. *Clinical Psychology: Science and Practice*, *3*, 25–36.

Lewinsohn, P. M., Solomon, A., Seeley, J. R., & Zeiss, A. M. (2000). Clinical implications of subthreshold depressive symptoms. *Journal of Abnormal Psychology*, *109*, 345–351.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Markus, H. R., & Kitayama, S. (1994). The cultural construction of self and emotion: Implications for social behavior. In S. Kitayama & H. Markus (Eds.), *Emotion and culture* (pp. 89–133). Washington, DC: American Psychological Association.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *American Statistical Association 1988 proceedings of the business and economics section* (pp. 308–313). Alexandria, VA: American Statistical Association.

Schwarzer, R. (1992). *Self-efficacy: Thought control of action*. Washington, DC: Hemisphere.

Shek, D. T. L. (1990). Reliability and factor structure of the Chinese version of the Beck Depression Inventory. *Journal of Clinical Psychology*, *46*, 35–43.

Shek, D. T. L. (1991). What does the Chinese version of the Beck Depression Inventory measure in Chinese students—General psychopathology or depression? *Journal of Clinical Psychology*, *47*, 381–390.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.

Stewart, S. M., Betson, C., Lam, T. H., Chung, S. F., Ho, H. H., & Chung, T. F. C. (1999). The correlates of depressed mood in adolescents in Hong Kong. *Journal of Adolescent Health*, *25*, 27–34.

Tanaka, J. S., & Huba, G. J. (1984). Confirmatory hierarchical factor analysis of psychological distress measures. *Journal of Personality and Social Psychology*, *46*, 621–635.

Tanzer, N. K. (1995). Cross-cultural bias in Likert-type inventories: Perfect matching factor structures and still biased? *European Journal of Psychological Assessment*, *11*, 194–201.

Tanzer, N. K., & Sim, C. Q. E. (1999). Adapting instruments for use in multiple languages and cultures: A review of the ITC guidelines for test adaptations. *European Journal of Psychological Assessment*, *15*, 258–269.

Tems, C. L., Stewart, S. M., Skinner, J. R., Hughes, C. W., & Emslie, G. (1993). Cognitive distortions in depressed children and adolescents: Are they state dependent or traitlike? *Journal of Clinical Child Psychology*, *22*, 316–326.

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

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.

World Health Organization. (2001). *World Health Report, 2001. Burden of mental and behavioral disorders.* Retrieved September 30, 2002, from http://www.who.int/whr/2001/main/en/chapter2/0002e1.htm