A STUDY OF THE SPANISH VERSION OF THE BECK DEPRESSION INVENTORY – II

by

CECILIA SIERRA HEREDIA

B. Psych. Honours, Universidad Nacional Autónoma de México, 2001

A THESIS SUBMITED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Measurement, Evaluation & Research Methodology Program, Department of Educational and Counselling Psychology, and Special Education, Faculty of Education)

> We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA April 2004

© Cecilia Sierra Heredia, 2004

Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Sierra Heredia, Cecilia Name of Author (please print) 23/04/2004

Date (dd/mm/yyyy)

Title of Thesis: A study of the Spanish version of the Beck Depression Inventory-II				
Degree:	M.A.		Year:	2004
Department of ECPS The University of British Columbia Vancouver, BC Canada				

Abstract

Depression is the second most important health issue for individuals aged 15 to 44 years worldwide (WHO, 2001b). One of the most widely used scales to assess depression is the Beck Depression Inventory (BDI-II; Beck, Steer & Brown, 1996). While the BDI-II is also available in Spanish (The Psychological Corporation, 1996), there are only two studies (Novy, Stanley, Averill & Daza, 2001; Penley, Wiebe & Nwosu, 2003) that have examined its psychometric properties. No research has been conducted on the Spanish BDI-II with a Mexican sample or with a student sample. The present research (a) examined the psychometric properties of the Spanish BDI-II with 202 Spanish-speaking Mexican students and (b) explored the interpretation of each item made by a subset of 18 participants using a Talk-Aloud Protocol. The Spanish BDI-II was found to have an essentially unidimensional factor structure. The alpha coefficient was .89 for the total sample with corrected item-total correlations ranging from .36 (loss of interest in sex) to .63 (self-dislike). No statistically significant gender differences were found on the total score and gender differential item functioning was not found in any of the Spanish BDI-II items. When a Talk-Aloud Protocol was used, a number of issues related to the translation of the items and the test format were raised. The implications of these findings for the use of the Spanish BDI-II with Mexican participants and recommendations for future revisions of the test were discussed.

Abstract	ii
Table of Contents	iii
List of Tables	v
List of Figures	vi
Acknowledgments	vii
Introduction	1
Prevalence and Impact of Depression on Health	1
Beck Depression Inventory	2
Spanish versions of the BDI and BDI-II	2
Overview and Research Questions	4
Method	5
Study I: An examination of the psychometric properties of the	
Spanish BDI-II	5
Participants	5
Materials	6
Procedure	7
Study II: Examination of Participants' Responses to the Spanish BDI-II	
Using a Talk-Aloud Protocol	7
Participants	7
Materials	8
Procedure	9
Results	10

Study I	10
Study II	12
Discussion	17
References	22
Author's Note	27
Appendix: Literature Review	35

iv

List of tables

v

Table number	Title	Page
Table 1	Means, standard deviations and factor loadings for the BDI-II items	28
Table 2	Corrected item-total correlation for the total sample, alpha-if-item-	
	deleted for the total sample and subsamples	. 30
Table 3	Differential item functioning: Chi-squared test, p-values and effect	*
	size R-squared	32

List of figures

Figure number	Title		
Figure 1	Scree Plot of the BDI-II total score		
- - 	· · · · · · · · · · · · · · · · · · ·		
	,		
•			
		· · ·	

.

Acknowledgments

To Mexico's Consejo Nacional de Ciencia y Tecnología (CONACYT) for funding this project.

To the Faculty of Education at UBC for supporting this project through a Graduate Student Research Grant.

To the students of the Universidad del Valle de México (UVM) for sharing their time and thoughts.

To the personnel at the Universidad del Valle de México (UVM) for their support in the data collection stage.

To my thesis supervisor, Anita Hubley, for being an outstanding mentor and a true friend.

To Bruno Zumbo, for sharing your passion about statistics, education and life in general.

To Kim Schonert-Reichl, for having such a positive and professional attitude.

To my peers at the Adult Development and Psychometrics lab, the MERM program and the staff at the ECPS department for the help, suggestions and above all, the smiles.

To Rodrigo, for your unconditional love and support. Don't be surprised if I love you for all that you are.

To my parents, Francisco and Lourdes, and to my brother Pablo, always so close, always a source of inspiration.

To my family and friends. Just as it takes a village to raise a child, it took a village to write this thesis.

Introduction

Prevalence and Impact of Depression on Health

According to the World Health Organization (WHO), 25% of individuals develop one or more mental or behavioural disorders at some stage of life, in both developed and developing countries (WHO, 2001a). Unipolar depression is the most prevalent and common of all the mental disorders (Judd & Kunovac, 1997). An estimated 121 million people currently suffer from depression worldwide and about one in five adults will suffer from major depression in their lifetime (WHO, 2001b). There is a significantly higher prevalence of depression in females, which is estimated to be about 1.5 to 3.0 times that expressed by males. There is also an increased risk of first onset of major depression in women (Kaelber et al., 1995). The impact of depression is such that the WHO Global Burden of Disease (GBD) Study ranked it as the single most burdensome disease worldwide in terms of total disability-adjusted life years among individuals in the middle years of life (i.e., ages 15-44 years) (WHO, 2001b). For developed and developing regions, unipolar major depression is the most burdensome disease for women; for men it ranks first in developing countries and third in developed countries, following alcohol use and road traffic accidents (Murray & Lopez, 1996). Although the core feelings of depression, such as loss, emptiness, and helplessness, seem to be experienced worldwide, other symptoms of depression and their personal and social implications seem to differ across cultures (Tsai & Chentsova-Dutton, 2002). It is acknowledged that the expression of the symptoms and prevalence of depression may vary among cultures within the same country and among different countries (Sartorius, 1986).

Beck Depression Inventory

The Beck Depression Inventory (BDI) is the most widely used self-report measure of depression in the world (Dozois, 2003) and has been used in a number of cross-cultural studies of depression (Byrne & Campbell, 1999; Novy, Santley, Averill & Daza, 2001; Penley, Wiebe & Nwosu, 2003). The BDI (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) was revised in 1996 and renamed the Beck Depression Inventory-II (BDI-II; Beck, Steer & Brown, 1996). Studies assessing the psychometric properties of the English version of the BDI-II have been conducted with Canadian (Dozois et al., 1998) and American non-clinical samples of undergraduate students (Steer, Ball, Ranieri, & Beck, 1999; Whisman, Perez, & Ramel, 2000) and with a qualitative methodology (Barroso & Sandelowski, 2001). A Spanish version of the BDI-II is available from the test publisher, The Psychological Corporation, but the BDI-II manual (Beck et al., 1996) does not provide any details about the procedures used in the translation of, or any psychometric evidence for, this version.

Spanish versions of the BDI and BDI-II

Studies of the psychometric properties of the Spanish version of the BDI have been conducted with Argentinean (Bonicatto, Dew & Soria., 1998; Richaud & Sacchi, 2002) and Mexican samples (Jurado, Villegas, Méndez, Rodríguez, Loperena, & Varela, 1998; Suarez-Mendoza, Cardiel, Caballero-Uribe, Ortega-Soto, & Marquez-Marin, 1997; Torres-Castillo, Hernández-Malpica & Ortega-Soto, 1991). Using principal components analysis, the studies conducted with Argentinean samples have reported four-factor (using varimax and promax rotations; Richaud & Sacchi, 2002) and two-factor structures (using a varimax rotation; Bonicatto et al., 1998). The studies with Argentinean samples

also reported good internal consistency (alpha coefficient=.87), evidence of concurrent validity (correlation of z-scores from the BDI and the DEP subscale of the SCL90, r=.68), evidence of construct validity (correlation of standardized scores on the Beck Hopelessness Scale and the BDI, r=.43), and evidence of criterion validity (higher scores for women, younger participants, and participants with lower education and income levels; Bonicatto et al., 1998). The studies conducted with the Mexican samples included participants diagnosed with rheumatoid arthritis (Suarez-Mendoza et al., 1997) and cardiovascular disease (Torres-Castillo et al., 1991), as well as students and individuals from the general population (Jurado et al., 1998). The authors report evidence of good concurrent validity (Suarez-Mendoza et al., 1997; Torres-Castillo et al., 1991), good testretest reliability over a 48 hours interval (r=.89), and a high corrected intraclass correlation coefficient of .77 (Torres-Castillo et al., 1991), high internal consistency (alpha coefficient=.87) and a three-factor structure (with a non-specified factor analysis using varimax rotation; Jurado et al., 1998).

Studies evaluating the psychometric properties of the Spanish version of the BDI-II have been conducted in the U.S. with adults born in the U.S., South and Central America, and Mexico who were diagnosed with an anxiety disorder (Novy et al., 2001) and a medical sample of Mexican American patients with end-stage renal disease (Penley et al., 2003). Validity evidence was provided in the form of discriminant validity (correlation of .60 between the BDI-II and the Beck Anxiety Inventory; Novy et al, 2001), high correlations between the total scores on the Spanish and the English versions of the BDI-II (r= .94, Novy et al., 2001; r=.70, Penley et al., 2003), good internal consistency (alpha coefficient=.94, Novy et al., 2001; alpha coefficient=.92, Penley et al.,

2003), and corrected item-total correlations between .33 and .72 (Penley et al., 2003). Using confirmatory factor analysis, Penley et al. (2003) tested a model with two firstorder factors of depression (*Cognitive* and Somatic-Affective) and a single second-order factor of depression. Evidence of an adequate fit to the data in this study was found.

Overview and Research Questions

When a test is used with a sample from a different target population or culture or when the test has been translated, adapted, modified, or revised, it is important to provide evidence supporting the reliability and validity of the scores and inferences, respectively, from that test. In the case of the BDI-II, evidence obtained from the English version of the test should not be assumed to apply to the Spanish version. Likewise, psychometric evidence obtained with the Spanish version of the BDI should not be assumed to apply to the Spanish version of the BDI-II. Furthermore, the psychometric properties of the Spanish BDI-II must be evaluated when used with a different culture (e.g., Mexicans, Mexican-Americans, Argentineans) or target population (e.g., clinical vs. non-clinical).

Although the Spanish version of the BDI-II is available (The Psychological Corporation, 1996) and its use has been examined for populations like Spanish-speaking Mexican Americans living in the U.S. (Novy et al., 2001; Penley et al., 2003), there does not appear to be any research with the Mexican population. Information about the psychometric properties of the Spanish BDI-II is required in order to determine the appropriateness of its use with Mexican samples. This research conducted two studies that looked at the psychometric properties of the Spanish BDI-II (Study I) and to the participants' responses to the Spanish BDI-II Using a Talk-Aloud Protocol (Study II). The research questions for Study I investigated the factor structure and internal consistency of the Spanish version of the BDI-II, the presence of differential item functioning due to gender in the responses to any of the individual items, and the presence of gender differences on the overall scores of the Spanish version of the BDI-II. For Study II, the research questions addressed the interpretation of each item on the Spanish version of the BDI-II made by Mexican students, and the presence of differences in the interpretation of the meaning that may not appear in the mean item scores. Also, Study II investigated whether men and woman interpret the items in the same way or, if not, what are the main differences and on which items, and do these explain any differences in the responses to particular items by gender? Do any differences appear here that are not apparent in the mean item scores?

Method

Study I: An examination of the psychometric properties of the

Spanish BDI-II

Participants

1. A. B. C.

A total of 209 male and female students from Mexican Valley University (Universidad del Valle de Mexico, UVM) San Rafael' campus in Mexico City participated in this study. Seven participants were not included in the final analysis because over half of the BDI-II was not answered or the participant did not include his/her gender.

All but two of the participants self-identified their citizenship as Mexican. Two participants left this item blank. Of the 202 participants included in the study, 65.8%

(n=133) of the participants were women. The mean age of the sample was 20.54 years (SD=3.10) and ranged from 17 to 42 years.

All the participants were undergraduate students. The mean number of years of undergraduate education was 1.64 (SD=1.09). Most participants were majoring in Psychology (25.2%), Communication (24.3%), Law (13.9%), Marketing (8.9%) and Tourism (8.9%).

Only 2.5% of the participants worked full-time, 15.3% regularly worked part-time and 15.3% occasionally worked part-time. The majority (96%) of the participants have never been married. Nearly 15% of the participants had an annual household income of less than \$10,740 CAD, which is equivalent to less than 6250 Mexican pesos per month or less than \$7,500 US per year; 15.3% of the participants had an annual household income between \$10,740.-:\$21,420 CAD, which is equivalent to 6250 to 12500 Mexican pesos per month or between \$7,500 to \$15,000 US per year; and 25.7% of the participants had an annual household income over \$21,420 CAD, equivalent to more than 12500 Mexican pesos per month or more than \$15,000 US per year¹.

All the participants reported Spanish as their first language, both for themselves and their parents. English was the second language for 57.4% of participants. *Materials*

Materials consisted of an informed consent sheet, the Spanish version of the Beck Depression Inventory-II (BDI-II; Beck & Steer, 1996), and a demographic sheet that included items on citizenship, gender, age, major, years in that major, employment, marital status, household income, parents' language, participants' native language, second and third language. All materials and instructions were given in Spanish.

¹ These equivalences were calculated based on the exchange rates available in June 2003.

The BDI-II is a revised version of the Beck Depression Inventory (Beck, Ward, Mendelson, Mock & Erbaugh, 1961), which is the most widely used self-report measure of depression (Dozois, 2003). The purpose of the BDI-II is to assess symptoms corresponding to criteria for diagnosing depressive disorders listed in the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM–IV; APA, 1994). The intended population for the BDI–II consists of individuals who are 13-80 years of age. The BDI–II can be administered to groups or individuals in 5-10 minutes.

The BDI-II consists of 21 items. Consistent with DSM-IV, participants select the statement in each item that best applies to them over the past two weeks. Total scores are derived by summing weights (which range from (0) *not present* to (3) *severe* for each of the four sentences under each item) corresponding to the statements endorsed over the 21 items. The responses are summed to yield a total score that ranges from 0 to 63.

Procedure

This study was conducted in classroom units of students. Participants read and signed an informed consent form, answered a personal demographic form and then completed the BDI-II. Each group session lasted approximately 30 minutes.

\$1.21

Study II: Examination of Participants' Responses to the Spanish BDI-II

Using a Talk-Aloud Protocol

Participants

A separate group of 18 male and female university students from Mexican Valley University (Universidad del Valle de Mexico, UVM) San Rafael' campus in Mexico City

participated in Study II. The participants for this study met the same criteria as the subjects in the Study I.

All the participants self-identified their citizenship as Mexican. Of the 18 participants, 9 were women. The mean age was 20.39 years (SD=1.61) and ranged from 18 to 23 years. All the participants were undergraduates. The mean number of years of undergraduate education was 2.33 (SD=.97). Most participants were majoring in Law (7) and Marketing (3); although Graphic Design, Psychology, Communication, International Affairs and Architecture were also selected.

Only 5 students regularly worked part-time and 4 occasionally worked part-time. The majority (17) of the participants have never been married. The household income of 10 of the participants was more than \$10,740 CAD yearly, this is equivalent to less than 6250 Mexican pesos per month or less than \$7,500 US per year. Two participants reported an annual household income between \$10,740 – \$21,420 CAD, which is equivalent to 6250 to 12500 Mexican pesos per month or between \$7,500 to \$15,000 US per year; 8 participants reported an annual household income over \$21,420 CAD, equivalent to more than 12500 Mexican pesos per month or more than \$15,000 US per year.

All the participants reported Spanish as their first language, both for themselves and their parents. English was the second language for 15 of the participants. *Materials*

This study used the same materials as described in Study I, but the informed consent sheet contained specific details regarding the interviews (e.g. Talk-Aloud Protocol and audio-taping).

Procedure

For this study, the participants completed an informed consent sheet, a personal demographic sheet and then completed the Spanish version of the BDI–II. A Talk Aloud Protocol (TAP, also known as Thinking-Out-Loud or Thinking-Aloud protocol), was conducted individually and audio-taped. Specific questions were asked as the participants completed the BDI-II about the meaning(s) that each item had for the respondents, the degree of difficulty they had in ascertaining what each item or statement was supposed to mean, and their reasons for responding to each item as they did (Barroso & Sandelowski, 2001). Each interview lasted approximately 45 minutes.

In TAP, participants are encouraged to verbalize their thoughts regarding a specific task while they complete it (Ericcson & Simon, 1993). These verbalizations are considered to reflect the inner mechanisms involved in the resolution of that task. TAP has been used to study higher level processes in reading (e.g., predictions, inferences or schema elaborations) (Olson, Duffy & Mack, 1984), reading comprehension (Yuill, Oakhill, & Parkin, 1989 in Whitney & Budd, 1996), problem-solving tasks (Montague, 1993), and in the evaluation of questionnaires such as the General Social Survey (Bailey & Marsden, 1999).

Messick (1989) also suggested that Talk Aloud Protocols (or protocol analysis, as he called it) can provide evidence about the validity of a test score's because the analysis of information processing could provide evidence for construct validity in tests. This evidence is based in the fact that Talk Aloud Protocols provide an analysis of the process underlying item performance.

Results

Study I: Psychometric properties of the Spanish BDI-II

The mean score on the BDI-II for the total sample was 9.90 (SD=7.88). No statistically significant differences were found between men and women, t (197) = -1.2, p=.23, and the effect size (d= 0.18) is considered small according to Cohen's (1992) criteria.

An exploratory factor analysis using principal axis factoring was conducted on the 21 items of the BDI-II. According to Gorsuch's (1983) guideline of 5 to 10 cases per item, the sample size for the present study (n= 202) was considered adequate for the 21-item BDI-II. The data met the criteria for Bartlett's Test of Sphericity, X^2 (210) = 1354, p<0.001 and the Kaiser-Meyer-Olkin criteria for sampling adequacy, KMO= .89 (Pett, Lackey & Sullivan, 2003).

An examination of the eigenvalues and scree plot indicated an essentially unidimensional factor structure for the Spanish BDI-II, (see Figure 1). The first factor had an eigenvalue of 6.88 and explained 32.8% of the variance in participants' responses. The ratio of the first to the second eigenvalue was 4.69.

Insert Figure 1 about here

When the factor analysis was conducted separately for males and females, essential undimensionality was found for both. For males, the ratio of the first to the second eigenvalue was 3.33; for females it was 4.69. This finding supports the use of a total score on the BDI-II. Table 1 presents the means, standard deviations and factor loadings for the 21 items of the BDI-II with the total sample, as well as the factor loadings for males and females separately. The correlation between the factor loadings for males and females was r=.32, p=.165, which indicates that the two groups showed little consistency in the magnitude of the loadings for each item.

Insert Table 1 about here

Internal consistency reliability estimates using coefficient alpha were .88 for the males, .89 for the females and .89 for the total sample. None of the alpha-if-item-deleted coefficients were higher than the alpha coefficient for the total sample (r=.89), which supports retaining all of the items. For the male group, only item 10 (crying) had a slightly higher alpha-if-item-deleted coefficient (.89); whereas for the female group, only item 11 (agitation) had a slightly higher alpha-if-item-deleted coefficient (.90). These results, on their own, are not substantial enough to suggest revisions to these items. Corrected item-total correlations ranged from .36 (loss of interest in sex) to .63 (self-dislike). Table 2 presents the corrected item-total correlations for the total sample as well as the alpha-if-item-deleted coefficients for the total sample and for the male and female subsamples.

Insert Table 2 about here

Because the structure of the BDI-II was determined to be essentially unidimensional, a differential item functioning (DIF) analysis by gender was conducted on the BDI-II items. DIF is present when males and females, who have been matched on their BDI-II total score, show different probabilities of endorsing any given BDI-II item. DIF is a necessary but not sufficient condition for item bias. Thus, if the BDI-II items do not display DIF one can assume that these items are not biased by gender in this sample. To explore DIF, ordinal logistic regression was conducted with each of the 21 BDI-II items. According to Zumbo (1999), an item that displays DIF will have a two-degree-of-freedom Chi-squared test with a p-value less than or equal to 0.01 (set at this level because of the multiple hypothesis tested) *and* a Zumbo-Thomas effect size (R-squared) measure of at least 0.130. As seen on Table 3, none of the items on the Spanish BDI-II displayed DIF.

Insert Table 3 about here

Study II: Examination of Participants' Responses to the Spanish BDI-II

Using a Talk-Aloud Protocol

The analysis of the Talk-Aloud Protocols was conducted in Spanish. Specific comments were selected and translated to English in order to illustrate the findings discussed in this section.

The eighteen participants who completed the BDI-II using a Talk-Aloud Protocol (TAP) raised a number of issues that could be divided into: (a) issues particular to the Spanish BDI-II, and (b) issues about the BDI-II test format.

(a) Issues particular to the Spanish BDI-II.

Two items (item 6 - punishment and 14 - worthlessness) on the Spanish BDI-II raised comments from the participants. On item 6 (punishment), only two of the 18 respondents indicated that they felt they were being punished (response option 1). The option "I feel I may be punished" was translated into Spanish as "*siento que quizá esté siendo castigado(a) por la vida*" with the phrase "by life" added to the end of the original sentence. Although the options in the Spanish version mentioned that life was (or was

not) punishing the respondent, six of the respondents (four men and two women) explained that they saw this punishment as sent by God as a consequence of their past actions. There does not seem to be an obvious reason for the addition of the phrase "by life" to the Spanish version and this addition could be a potential source for differences between versions. The six participants who mentioned God, selected the response option 0 ("I don't feel I'm being punished"). Two of these participants also mentioned that what they really believed is that punishment would be a consequence of one's actions.

With respect to item 14, three women and three men expressed disagreement with the words used in the options. This is the only item for which problems with language were mentioned. Whereas in the English version, the words used were "worthless" and "worthwhile", in the Spanish version, the words applied were "*inservible*" or "*valioso*" y "*útil*"; concepts that in the participants' words: were more commonly employed with material objects rather than with people.

It seems as though the Spanish equivalent of the English words would be "*inútil*" or "*devaluado*" for "worthless", "*valioso*" for "worthwhile" and "*útil*" for "useful". So, respondents were able to understand the general meaning of the options but still felt that the wording was incorrect.

(b) Issues about the BDI-II test format

Since these issues where not particular to the outcomes of the translation process, they could be generalized to the non-Spanish versions of the BDI-II. However, specific evidence regarding each version is needed in order to confirm these concerns. Three main issues arose involving the BDI-II test format: (a) links between the items made by

the participants, (b) the response options from each item and their place in a continuum, and (c) the two-week period (particularly regarding items 9 and 12).

Some respondents linked the items related to energy, changes in sleep and tiredness; they also mentioned other elements as causes for their answers such as illness (e.g. colds), smoking, alimentation, a change or an increase in their activities such as end of term work load, weather (since it was the beginning of summer) and irritability. The relevance of these links is that they were not always formulated when the participants selected an answer for the items related to energy (15), tiredness (20) or changes in sleep (16). Participants would assess themselves as having less energy without considering their increase in activities, which is qualitatively different than the loss of energy experienced in depression (that is not accompanied by changes in the level of activity).

Twelve participants (six men and six women) raised questions about the response options from items and the ordering of these responses in the continuum from 0 to 3. Items 6 (punishment), 10 (crying) and 16 (changes in sleep) appeared to be particularly problematic in this regard. With respect to item 6, two respondents explicitly mentioned that the response options 1 and 3 looked the same to them. This could be explained by the fact that the options use only a slight variation of another to distinguish between those options.

In terms of item 10, six participants (three men and three women) described how some circumstances would make them feel like crying even though they would not because of the people they were with or the circumstances in which the situation arose. When participants mentioned other people as the reason they did not cry, it was because they did not want to upset these individuals by crying or because they were people to

whom they did not feel close enough to feel comfortable crying in their presence. Participants also described how, after avoiding crying in one event (for the reasons described above), they would cry in response to a stimulus that in other circumstances would not have triggered tears, and acknowledged that avoiding crying with the first event was the main reason for crying with the second one. From the 18 interviewees, four men selected the option number three for this item, compared with one woman. The three male participants described how response option 3 ("I feel like crying but I can't") meant that someone could feel like crying because they were feeling sadness but it was not enough sadness to make them actually cry in that particular situation. Participants seemed to make a conscious effort to hold the tears and were also aware of the feeling and the context; they described how "I cry more than I used to" (response option 1) was not suitable for them because they did not cry more than in the past, but did not select "I don't cry any more than I used to" (response option 0) because they were aware of the change in their feelings. It seems as if participants would select "I cry more than I used to" (which is scored with 1) if the context were different. It is important to consider if, for this sample at least, the response option "I feel like crying but I can't" is not appropriately scored with 3 (the highest option) and if the difference in total scores due to selecting response options 1 or 3 for this question is fair considering that this could be due to problems with the rank ordering of the options.

As for item 16 (changes in sleep) seven of the 18 participants had difficulty deciding what constituted sleeping "somewhat" versus "a lot" more or less than usual. Most of them seemed to select a response by quantifying each option, although these quantities might overlap (e.g. "somewhat" ranged from 1-5 hours and "a lot" ranged from

3-10 hours). One of the participants even suggested that the options should be more specific. Because respondents had difficulty inferring or understanding the elements of some response options, they were forced to "develop their own idiosyncratic meaning for them" (Podsakoff, MacKenzie, Lee & Podsakoff, 2003, p. 883) which could threaten the validity of the test by adding systematic measurement error.

Another issue about the test format that arose was the two-week period mentioned in the BDI-II instructions and how the participants incorporated this element into their decision process. This issue seemed to be particularly relevant for items 9 (suicide) and 12 (loss of interest). The BDI-II instructions ask participants to rate their answers based on how they have been feeling during the past two weeks. In spite of the fact that participants read the instructions aloud as part of the TAP, some participants seemed to consider events made in the remote past when selecting a response option. Although some participants mentioned the two-week period spontaneously, when this did not occur, it was raised by the researcher after they selected their response. Their responses fall into one of two categories: some reconsidered their options and some kept the same choice.

The answers to item 9 (suicide) seemed to be particularly influenced by the twoweek time period. Although 13 participants endorsed option 0, five participants (three women and two men) mentioned that they had suicidal ideations in the past. One woman mentioned the two week period herself and, based on that, selected option 0. Two participants (one man and one woman) selected response option 1 (having thoughts of suicide but didn't intended to carry them out) and also confirmed their option when reminded of the two-week period. Two other participants (a man and a woman) changed their options from 1 to 0 when reminded of the two-week period, and then explained that, in the past, they have thought about committing suicide but did not attempt it.

Regarding item number 12 (loss of interest) eight participants did not describe any loss of interest and the other 10 participants described less interest in other persons or activities (seven of them selected response option 1, two selected response option 2, and one selected response option 3). When questioned about their decision process, these participants described events or facts that seemed to happen in the past but still stayed with their selected response option when reminded of the two-week period.

Discussion

This study provides preliminary evidence of the psychometric properties of the BDI-II Spanish version and it appears to be the first study to examine these properties with a student sample and with Mexican participants. The essentially unidimensional factor structure found in this study supports the use of a total score for the BDI-II. The internal consistency estimate of .89 for the sample of Mexican undergraduate students is very good. The alpha coefficient is slightly lower than those reported in previous studies using the Spanish version with clinical samples (i.e., alpha = .94 in Novy et al., 2001; alpha = .92 in Penley et al., 2003) and consistent with the alphas reported for the English version of the BDI-II with student samples (i.e., alpha = .93 in Beck et al., 1996; alpha = .87 in Dozois et al., 1998; alpha = .89 in Whisman et al., 2000). Future research that examines the test-retest reliability for the Spanish BDI-II is needed.

The corrected item-total correlations obtained in the present study were similar to those reported by Beck et al. (1996), Dozois et al. (1998) and Whisman et al. (2000) for

student samples using the English version of the BDI-II. In the present study with Mexican students, item 21 (loss of interest in sex) displayed the lowest (.36), and item 7 (self-dislike) displayed the highest ,corrected item-total correlation (.63). Interestingly, these items were also found to have the lowest and highest corrected item-total correlations in the above cited studies with students. Specifically, the present results can be compared to .27 (loss of interest in sex) and .74 (self-dislike) for the Canadian student sample described in Beck et al. (1996) and .28 (loss of interest in sex) and .63 (selfdislike) for the U.S. students in Whisman et al. (2000). The relatively high item-total correlations found in the present study indicate a strong relationship between each item and the total score. The similarity of these results to other studies suggests that the Spanish version functions similarly to the English BDI-II when used with student samples.

观察员

- 10

No statistically significant gender differences were found on the total score for the Spanish BDI-II. This is contrary to Beck et al. (1996) and Steer et al. (1999), who reported that females scored significantly higher than males (indicating more depressive symtomatology) in student and clinical samples with the English BDI-II. However, it is in line with Dozois et al. (1998) who did not find significant mean differences due to gender for a sample of Canadian students with the English BDI-II, and Penley et al. (2003), who did not find gender differences with a clinical sample using the English and Spanish versions of the BDI-II. Furthermore, the factor structure of the responses did not differ by gender in this sample. In addition, differential item functioning was not found for any of the items. This means that the items on the Spanish BDI-II functioned the same way for males and females in the present study. Nonetheless, further research is needed to

understand the inconsistency in the gender difference findings on the BDI-II total scores and to examine differential item functioning with different samples.

In summary, the psychometric findings from the present study support the use of the Spanish BDI-II with Spanish speaking Mexican students.

In addition to the main psychometric findings, a Talk-Aloud Protocol was used to explore the interpretation of each item made by a second sample of 18 Mexican students. The appropriateness of the rank order of the response options, was raised with items 6 (punishment) and 10 (crying). Participants described difficulties in distinguishing between the response options for item 6 due to the subtlety of the wording. For the item 10, four men selected option 3 and described what could be seen as a different approach to the rank order of the response options. These men seemed to perceive the response option 3 (*I feel like crying but I can't*) as a middle point between the response option 0 (*I don't cry anymore than I used to*) and response option 1 (*I cry more than I used to*) because it reflected their feelings about the appropriateness of crying in certain contexts without necessarily suggesting it was a more severe expression of symptomatology.

A related issue to the ranking occurred with item 16 (changes in sleep) wherein participants felt the need to quantify each option (in terms of hours) in order to select a response ,but the quantities posed to each option seemed to overlap across participants. These are all issues that need to be considered in any future revisions to the BDI-II and are likely not limited to the Spanish version of the measure.

Another issue that became apparent when the Talk-Aloud Protocol was used was the tendency for participants to forget about, or ignore, the two-week period of reference when completing the measure. This finding is similar to that reported by Barroso and

4. 医原门耳鼻前肌

18:2

Sandelowski (2001). In their study, the participants also evaluated events that occurred further in the past than the prior two weeks. If the participant completes the BDI-II without considering the two-week period consistently, the very goal of the test is affected and problems with validity arise. Future revisions of the BDI-II should instruct the test administrators to explicitly remind the test takers of the two-week period throughout the test or the test could include written reminders after the 7th and 14th items.

A final issue raised in the Talk-Aloud Protocols was related to the translation of the BDI-II from English to Spanish. Item 6 (punishment) included the phrase "by life" in the response options on the Spanish version. It is not clear why the phrase was added. Several participants mentioned God or referred to past actions when talking about this item. However, all of these participants ultimately selected 0 (I don't feel I'm being punished) and it is not clear if they really did not feel punished or if they did not agree with the phrase "by life" in the other response options. Item 14 (worthlessness) raised some comments with regard to the appropriateness of the vocabulary used. Specifically, the Spanish version used words typically used to describe objects rather than people. Because the test manual (Beck et al., 1996) provides no details of the translation procedure or the composition or cultural background of the translator(s) involved, it is not possible to know if the issues raised by the participants are particular to the Mexican sample or would be shared by participants from other Spanish speaking countries. Future revisions of the Spanish version of the BDI-II might consider not only the words used in item 14 but might generally use translation procedures that would explicitly consider the words deemed appropriate by Spanish speakers with different cultural backgrounds.

- 44.55 21.50 22

 $\Phi_{i,R}$

In summary, the issues raised when a Talk-Aloud Protocol was used suggest that both the Spanish BDI-II as well as the English BDI-II could be further improved in future revisions.

Finally, while the present study generally supports the use of the Spanish BDI-II with Spanish speaking Mexican students and contributes to our knowledge about the psychometric properties of the Spanish BDI-II and some issues related to its administration and the wording of particular items, there is one key limitation to the study that is worth noting. Because the participants in this study all live in one city and belong to a student population, the sample does not represent all adults or even all students and this limits the generalizability of the study findings. Lack of representativeness in terms of ethnicity and geographic distribution has been criticized in previous studies with the BDI-II (Farmer, 2001). Thus, the inclusion of participants from the general community and clinical populations as well as from different cities and rural areas in Mexico

References

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*, (4th ed.). Washington, D.C.: Author.
- Bailey, S. & Marsden, P.V. (1999). Interpretation and interview context: Examining the
 General Social Survey name generator using cognitive methods. *Social Networks*, 21, 287 309.
- Barroso, J., & Sandelowski, M. (2001). In the field with the Beck Depression Inventory. Qualitative Health Research, 11, 491-504.
- Beck, A.T., Steer, R.A., & Brown, G.K. (1996). BDI-II manual. San Antonio, TX: The Psychological Corporation.
- 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.
 - Bonicatto, S., Dew, A.M., & Soria, J.J. (1998). Analysis of the psychometric properties of the Spanish version of the Beck Depression Inventory in Argentina. *Psychiatry Research, 79*, 277-285.
 - Byrne, B., & 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, 555-574.

Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.

- Dozois, D., Dobson, K.S., & Ahnberg, J.L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment*, 10, 83-89.
- Dozois, D. (2003). The psychometric characteristics of the Hamilton Depression Inventory. Journal of Personality Assessment, 80, 31-40.

Ericsson, K.A. & Simon, H.A. (1993). Protocol analysis: Verbal reports as data. Cambridge: The MIT Press.

Farmer, R. (2001). Review of the Beck Depression Inventory - II. In J.C. Impara & B.S. Plake (Eds.) *The fourteenth mental measurement yearbook*. (pp. 123-126). Lincoln, NB: The Buros Institue of Mental Measurements.

Gorsuch, R.L. (1983). Factor analysis. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Judd, L. L., & Kunovac, J.L. (1997). Diagnosis and classification of depression. In A. Honig &

H.M. Van Praag (Eds.). Depression. Neurobiological, psychopathological and therapeutic advances: Vol.3. Wiley Series on Clinical and Neurobiological Advances in Psychiatry, (pp. 3-15). Chichester: John Wiley & Sons.

Jurado, S., Villegas, E., Méndez, L., Rodríguez, F., Loperena, V. & Varela, R. (1998). La estandarización del Inventario de Depresión de Beck para los residentes de la ciudad de México [Standardization of the Beck Depresión Inventory for Mexico City' residents]. Salud Mental, 21, 26-31.

Capatrio L

2,4

Kaelber, C,T., Moul, D.E., & Farmer, M.E. (1995). Epidemiology of depression. In E.E.
Beckham & W.R. Leber (Eds.). *Handbook of depression*, (2nd ed), (pp. 3-35). New York: The Guilford Press.

Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement*, (3rd ed). (pp. 13-103). New York: American Council on Education & Macmillan Publishing Company.

Montague, M. (1993) Middle school students' mathematical problem-solving: An analysis of think-aloud protocols. *Learning Disability Quarterly*, 16, 19 – 32.

Murray, C.J.L. & Lopez, A.D. (1996). The Global Burden of Disease in 1990: Final results and their sensitivity to alternative epidemiological perspectives, discount rates, age-weights

and disability weights. In C.J.L. Murray & A.D. Lopez (Series Eds., Vol. Ed.). *The Global Burden of Disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: Vol. 1. Global Burden of Disease and Injury Series.* (pp. 247-293). Cambridge, MA: Harvard School of Public Health on behalf of the World Health Organization and the World Bank.

- Novy, D. Stanley, M. Averill, P., & Daza, P. (2001). Psychometric comparability of English and Spanish language measures of anxiety and related affective symptoms. *Psychological Assessment, 13*, 347-355.
- Olson, G.M., Duffy, S.A. & Mack, R.L.(1984). Thinking-Out-Loud as a Method for studying Real-Time Comprehension Processes. In D.A. Kieras, M.A. Just (Ed.), *New methods in reading comprehension research*. New Jersey: Lawrence Erlbaum Associates.

Henley, J.A., Wiebe, J.S., & Nwosu, A. (2003). Psychometric properties of the Spanish Beck

Depression Inventory-II in a medical sample. Psychological Assessment, 15, 596-577.

- Pett, M.A., Lackey, N.R., Sullivan, J.J. (2003). Making sense of factor analysis. The use of factor analysis for instrument development in health care research. Thousand Oaks, CA: Sage Publications.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88, 879-903.
- Richaud, C., & Sacchi, C. (2002). Adaptacion del inventario de Depresion de Beck a sujetos argentinos normales / Adaptation of Beck Depression Inventory to a sample of Argentine general population. *Revista Iberoamericana de Diagnostico y Evaluacion Psicologica*, 12, 11-17.

- Sartorius, N. (1986). Cross-cultural research on depression. *Psychopathology*, 19, suppl 2, 6-11.
- Steer, R., Ball, R., Ranieri, W.F., & Beck, A.T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, 55, 117-128.
- Suarez-Mendoza, A.A., Cardiel, M.H., Caballero-Uribe, C.V., Ortega-Soto, H.A., & Marquez-Marin, M. (1997). Measurement of depression in Mexican patients with rheumatoid arthritis: Validity of the Beck Depression Inventory. *Arthritis Care and Research, 10*, 194-199.

The Psychological Corporation. (1996). *BDI-II. Spanish translation*. San Antonio, TX: Author. Torres-Castillo, M., Hernández-Malpica, E. & Ortega-Soto, H.A. (1991). Validez y

- [Validity and reproducibility of the Beck Depression Inventory in a cardiac care hospital]. *Salud Mental, 14*, 1-6.
- Tsai, J.L., & Chentsova-Dutton, Y. (2002). Understanding depression across cultures. In I.H. Gotlib & C.L. Hammen (Eds.) *Handbook of depression*. (pp 467-491). New York: The Guilford Press.
- Whisman, M., Perez, J., & Ramel, W. (2000). Factor structure of the Beck Depression
 Inventory-Second edition (BDI-II) in a student sample. *Journal of Clinical Psychology*, 56, 545-551.
- Whitney, P. & Budd, D. (1996). Think-Aloud Protocols and the study of comprehension. Discourse Processes 21, 341 – 351.

World Health Organization. (2001a). Mental and neurological disorders. *Fact Sheet No.* 265. Retrieved July 29, 2003, from http://www.who.int/inf-fs/en/fact265.html

World Health Organization. (2001b). The World Health Report 2001. Mental health: New understanding, new hope. France: Author.

 Zumbo, B. D. (1999). A Handbook on the theory and methods of Differential Item Functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa ON: Directorate of Human Resources Research and Evaluation, Department of National Defense.

"行动影响和如何"

:- '

经济利用品牌

Author's Note

This study was supported by a scholarship from Mexico's National Council for Science and Technology (CONACYT) and a Graduate Student Research Grant to Cecilia Sierra Heredia from the Faculty of Education at the University of British Columbia. The author would like to thank the personnel at the Universidad del Valle de México (UVM) for their support in the data collection stage.

计合件输出性

CHARLES (LA

Table 1

	Symptom	Mean	SD	Factor Loading Total Sample	Factor Loading Females	Factor Loading Males
1.	Sadness	.26	.52	.64	.627	.678
2.	Pessimism	.25	.51	.53	.551	.503
3.	Past failure	.28	.63	.53	.528	.521
4.	Loss of pleasure	.59	.72	.60	.631	.612
5.	Guilty feelings	.59	.59	.54	.531	.576
6.	Punishment feelings	.17	.37	.45	.467	.448
7.	Self-dislike	.40	.80	.69	.732	.644
8.	Self-criticalness	.56	.84	.60	.693	.415
9.	Suicidal thoughts	.25	.49	.51	.460	.587
10.	Crying	.52	.92	.45	.548	.240
11.	Agitation	.66	.75	.36	.290	.458
12.	Loss of interest	.54	.65	.57	.570	.587
13.	Indecisiveness	.52	.72	.54	.485	.647
14.	Worthlessness	.20	.52	.68	.723	.616
15.	Loss of energy	.65	.69	.57	.575	.584
16.	Changes in sleep	.84	.77	.54	.493	.654
17.	Irritability	.48	.66	.49	.537	.381
18.	Changes in appetite	.75	.78	.49	.545	.348

Means, Standard Deviations and Factor Loadings for the BDI-II items

ئى. ئەخمەش يە
19.	Concentration	.64	.72	.57	.529	.623
	difficulty					
20.	Tiredness – fatigue	.58	.66	.56	.601	.464
21.	Loss of interest in sex	.19	.51	.36	.374	.291

. .

Table 2

Corrected Item-Total Correlation for the total sample, Alpha-If-Item-Deleted for the total sample and subsamples.

Alpha-if Alpha-if Alpha-if Corrected Symptom item-deleted itemitem-deleted Item-total Total deleted Female correlation Sample Male Sample Sample 1. Sadness . .59 .88 .87 .89 2. Pessimism .49 .89 .88 .89 3. Past failure ..49 .88 .88 .89 4. Loss of pleasure .87 .56 .88 .89 5. Guilty feelings .50 .88 .87 .89 · 6. Punishment feelings .88 .89 .42 .89 7. Self-dislike .63 .88 .87 .88 8. Self-criticalness .57 .88 .88 .89 9. Suicidal thoughts .47 .89 .87 .89 10. Crying .44 .89 .89 .89 Agitation 11. .36 .89 .87 .90 Loss of interest .53 .88 .87 12. .89 13. Indecisiveness .53 .88 .87 .89 Worthlessness 14. .61 .88 .87 .89 Loss of energy 15. .54 .88 .87 .89 16. Changes in sleep .51 .88 .87 .89

Irritability	.47	.89	.88	.89	
Changes in appetite	.48	.89	.88	.89	
Concentration difficulty	.55	.88	.87	.89	
Tiredness – fatigue	.53	.88	.88	.89	
Loss of interest in sex	.32	.89	.88	.89	
	Irritability Changes in appetite Concentration difficulty Tiredness – fatigue Loss of interest in sex	Irritability.47Changes in appetite.48Concentration difficulty.55Tiredness – fatigue.53Loss of interest in sex.32	Irritability.47.89Changes in appetite.48.89Concentration difficulty.55.88Tiredness – fatigue.53.88Loss of interest in sex.32.89	Irritability.47.89.88Changes in appetite.48.89.88Concentration difficulty.55.88.87Tiredness – fatigue.53.88.88Loss of interest in sex.32.89.88	Irritability .47 .89 .88 .89 Changes in appetite .48 .89 .88 .89 Concentration difficulty .55 .88 .87 .89 Tiredness – fatigue .53 .88 .88 .89 Loss of interest in sex .32 .89 .88 .89

Table 3

Differential Item Functioning by gender: Chi-squared test, p-values and Effect Size R-squared

······································	X^2	p-value	R-squared effect
Symptom			size
1. Sadness	8.65	0.013	0.038
2. Pessimism	1.07	0.585	0.006
3. Past failure	0.51	0.774	0.003
4. Loss of pleasure	6.49	0.039	0.024
5. Guilty feelings	1.46	0.482	0.006
6. Punishment feelings	4.61	0.100	0.032
7. Self-dislike	0.70	0.703	0.003
8. Self-criticalness	4.74	0.094	0.017
9. Suicidal thoughts	2.69	0.261	0.010
10. Crying	2.73	0.255	0.012
11. Agitation	6.12	0.047	0.028
12. Loss of interest	2.37	0.306	0.010
13. Indecisiveness	3.02	0.220	0.012
14. Worthlessness	0.65	0.722	0.003
15. Loss of energy	6.39	0.041	0.020
16. Changes in sleep	2.16	0.340	0.008
17. Irritability	7.42	0.024	0.033
18. Changes in appetite	2.74	0.254	0.011
19. Concentration difficulty	1.90	0.386	0.007

20. Tiredness – fatigue	0.45	0.798	0.002
21. Loss of interest in sex	0.44	0.801	0.003

Note. All Chi-squared tests had two degrees of freedom.





Factor Number

Appendix

Epidemiology of Depression

Psychiatric Definition

Depression is considered a mood disorder, characterized by sadness, loss of interest in activities, and decreased energy (Boniccato, Dew, Zaratiegui, Lorenzo & Pecina, 2001). Although these symptoms may be present in the general population and are considered normal under some circumstances (e.g., mourning), the severity and duration of symptoms and their effects on the daily functioning of the individual distinguish normal levels of depressive feelings from clinical levels of depression (Kaelber, Moul & Farmer, 1995).

The International Statistical Classification of Diseases and Related Health Problems (ICD-10) incorporates depression among the mental and behavioural disorders. A general description of these disorders states that they are characterized by some combination of abnormal thoughts, emotions (mood), behaviour, and relations with others (World Health Organization [WHO], 2001a). According to ICD-10, lowering of mood, decreased capacity for enjoyment, interest and concentration, reduction of energy and activity, changes in sleep, diminished appetite, lower levels of self- esteem and selfconfidence, and ideas of guilt and worthlessness are all symptoms of a depressive episode. This system typifies depression as mild, moderate or severe according to the number of symptoms expressed by the individual and the changes in his/her everyday behaviour as follows: two or three symptoms for mild depression (with distress, but not impairment in conducting daily activities), four or more symptoms for moderate depression (including expressing great difficulty in conducting ordinary tasks), and most of the symptoms for severe depression (accompanied by suicidal thoughts or acts) (WHO,1992).

The American Psychiatric Association (APA) has developed its own classification for mental ailments, named the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). According to the DSM-IV, a Major Depressive Episode is best characterized by: a period of at least two weeks during which there is either the loss of interest or pleasure in nearly all activities or depressed mood (i.e., sadness or irritability) (APA, 1994). The criteria for diagnosis specify that at least four of the following symptoms must be also present, regardless of age:

- Significant weight loss or weight gain or changes in appetite
- Changes in sleep (i.e., insomnia or hypersomnia)
- Changes in psychomotor activity (i.e., increase or retardation observable by others)
- Fatigue or decreased energy
- Feelings of worthlessness or excessive guilt
- Difficulty thinking, concentrating or making decisions
- Recurrent thoughts, plans or attempts of suicide.

Significant impairment in social or personal areas of functioning and/or clinically significant distress must arise as well. Two important issues in diagnosing depression are that the symptoms should not be caused by direct physiological effects of a substance (e.g., medication) or a general medical condition and that they should not be better accounted for by bereavement (APA, 1994).

A Major Depressive Disorder is characterized by the occurrence of one or more major depressive episodes, without a history of mania, mixed or hypomanic episodes, substance-induced mood disorders, or a mood disorder due to a general medical condition (APA, 1994).

Impact of Depression on Health

The impact of depression is such that the WHO Global Burden of Disease (GBD) Study ranked it as the single most burdensome disease worldwide in terms of total disability-adjusted life years (DALYs) among individuals in the middle years of life (i.e., ages 15-44 years). For developed and developing regions, unipolar major depression is the most burdensome disease for women; for men, it ranks first in developing countries and third in developed countries (following alcohol use and road traffic accidents) (Murray & Lopez, 1996). Among mental and behavioural disorders, depression accounts for the largest percentage (12%) of disability (WHO, 2001a).

The GBD methodology was developed to estimate the impact that illness has in each individual's life and at a population level. An important concept in this system is the "disability-adjusted life year" (DALY) that combines information on the impact of premature death and disability. One DALY can be conceived of as "one lost year of healthy life, and the burden of disease as a measurement of the gap between current health status and an ideal situation where everyone lives into old age free of disease and disability" (WHO, 2001a, p.25). Unipolar depressive disorders generally account for 4.4% of the total DALYs. For males, it accounts for 3.4% of DALYs compared to 5.5% for women. In the age group between 15 to 44 years, unipolar depressive disorders rank second in the causes of DALYs (preceded only by HIV/AIDS) and account for 8.6%. Depressive disorders are always among the top seven causes of DALYs whether considering the total population or dividing by age or gender (WHO, 2001a). *Comorbidity*

Besides the disability induced by depression itself, other illnesses or behaviours (both physiological and psychological) are known to be related to depression. Suicide is known to be linked with depression as most individuals who commit suicide are clinically depressed as well (Murray & Lopez, 1996). Symptoms of depression can precede or accompany alcohol dependence (APA, 1994). Lifetime alcohol abuse has a high comorbidity with major depression in countries such as Germany, Korea, Puerto Rico, New Zealand, the U.S., and Canada. Comorbidity of depression with drug abuse also tends to be present in the latter three countries (Lépine, 2001). Smoking or nicotine dependence has also been linked with depression (Lerman et al., 1996). Depression is an important predictor of smokers' success in quitting smoking (Hall, Munoz, Reus, & Sees, 1993). In addition, those smokers with a history of depression who succeed in quitting smoking are at increased risk to develop another episode of major depression (Glassman et al., 1990). Compared to non-depressed women, females diagnosed with a major depressive disorder smoke at a higher rate during their heaviest smoking periods, have a greater number and severity of withdrawal syndromes, and demonstrate lower selfefficacy in managing food intake (Borrelli et al., 1999). Coronary heart disease (CHD) has also been linked with depression. High rates of persistent depression are present in patients with CHD (Hance, Carney, Freedland, & Skala, 1996); more severe levels of depression are related to nonfatal CHD events in women and to both nonfatal and fatal CHD events for men (Ferketich, Schwartzbaum, Frid, & Moescherberger, 2000).

Prevalence

Among the health issues present worldwide, psychological problems represent a major portion. According to the World Health Organization (WHO), 25% of individuals develop one or more mental or behavioural disorders at some stage of life in both developed and developing countries (WHO, 2001b). In 2001, the WHO estimated that some 450 million people suffered from a mental or behavioural disorder (WHO, 2001a). The most prevalent and common of all the mental disorders is unipolar depression (Judd & Kunovac, 1997). An estimated 121 million people currently suffer from depression worldwide and about one in five adults will suffer from major depression in their lifetime (WHO, 2001b).

Surveys conducted with community samples in the U.S. have found that up to 20% of adults and 50% of children report depressive symptoms during recall periods of one week to six months. Highest scores are reported among the youngest and the oldest respondents, with lower scores for middle aged participants, producing a U-shape distribution. In the U.S., at least one out of every six adults has met the criteria for either minor, major or recurrent brief depression (Kessler, 2002).

Gender and Epidemiology

While the overall prevalence of mental and behavioural disorders appears to be equivalent for men and women (WHO, 2001a), gender does seem to play an important role in the epidemiology of depression. According to the DSM-IV-TR, women are at significantly greater risk than men to experience a major depressive episode (APA, 2000). There is a significantly higher prevalence of depression in females, which is estimated to be about 1.5 to 3.0 times that expressed by males. There is also an increased

risk of first onset of major depression in women (Kaelber et al., 1995). Some symptoms of depression (e.g., low sex drive, hypersomnia, fear of "being poor", and decrease in food intake) have been found to be gender-related, regardless of the incidence of depression (Lange, Thalbournee, Houran & Lester, 2002). In children, research has not shown gender differences in depression. However, by late adolescence (i.e., after the age of 15), girls are twice as likely as boys to experience depressive symptoms (Nolen-Hoeksema & Girgus, 1994). The higher prevalence of depression in women could be related to the greater risk of first onset in women, a greater recurrence of depression in women, longer depressive episodes in women than men, or a combination of these factors (Nolen-Hoeksema, 2002).

Depression in Mexico

The Mexican Republic, or Mexico, is composed of 32 states and one federal district (Mexico City). Like most of the Latin American countries, Mexico inherited Spanish as its official language from Spain, the European conqueror, although it has been an independent country since 1821 (Cosío Villegas et al., 1973). In 2001, WHO estimated the population of Mexico to be 100,367,000 (WHO, 2003a)¹. Mexico City, the most populated metropolis, had 16,674,160 residents in 1995 (INEGI & SEMARNAP, 1997). Using INEGI statistics (2001), the gender ratio in Mexico was 95 males for 100 females. The mean age was 22 years for men and 23 years for women (INEGI, 2001). Whereas life expectancy at birth in 2001 was 74.2 years (71.6 years for males and 76.7

¹ As a note of comparison, Canada's population in 2001 was 31,014,000. (WHO, 2003b)

years for females), the healthy life expectancy (at birth) was 63.8 years (62.6 years for men and 65.0 years for women) (WHO, 2003a).²

Although core feelings of depression, such as loss, emptiness, and helplessness, seem to be experienced worldwide, other symptoms of depression and their personal and social implications seem to differ across cultures (Tsai & Chentsova-Dutton, 2002). It is acknowledged that the expression of the symptoms and prevalence of depression may vary among cultures within the same country and among different countries (Sartorius, 1986). Somatization, or the experience of physical symptoms attributable to underlying psychological problems, is relatively unusual in Western populations but very common in many non-Western groups, such as Asian communities (Thakker & Ward, 1998).

In Latin American countries, approximately 16.5 million people suffered from depression in the 1980s. Prevalence/rates oscillated between 9.8% for the Dominican Republic to 53% in Costa Rica. The rate for Mexico was 19% (García-Alvarez, 1986). It is estimated that, in 2020, unipolar major depression will continue to be the leading cause of disability burden in developing countries (Lépine, 2001).

In Mexico, the first National Survey on Addictions (NSA) was conducted countrywide in 1988 with a probabilistic sample of urban residents. The National Survey on Mental Health (NSMH) was gathered using a probabilistic sub sample of individuals between 18 and 65 years from the NSA and included a Spanish version of the Center for Epidemiologic Studies Depression Scale (CES-D) to assess depressive symptomatology. The results, comparable to those obtained in other countries, show a prevalence of 3.0% to 4.6% for women and 1.3% to 3.0% for men, for depression in the past six months

² The Canadian life expectancy at birth was 79.3 years (76.6 years for males and 81.9 years for females) and the healthy life expectancy (at birth) was 69.9 years (68.2 years for men and 71.6 years for women) in 2001 (WHO, 2003b).

(Caraveo et al., 1994). The gender ratio for depression is reported to be 1.8 women for each man (Caraveo-Anduaga et al., 1996).

A study conducted in 1995 among 1,932 residents of Mexico City ages 18 to 65 years used the Composite International Diagnostic Interview 1.1 (CIDI) (Caraveo-Anduaga, Colmenares, & Saldivar, 1999a). Results showed that 28.7% of the adult population suffered a psychiatric illness at some point in their lives; with 4% having suffered a severe illness. With a prevalence of 7.9%, depressive episodes were the second most prevalent illness, preceded only by alcohol abuse. The comorbidity of depressive episodes with other psychopathologies is 67.2% although, for this sample, anxiety disorders, substance dependence and manic states have higher comorbidities with one or more psychopathologies. For the participants in this study, the prevalence of depressive episodes during participants' lives was:7:9% (10.0% for women and 5.2% for men). The incidence of depressive episodes in the 12 months prior to the interview was 1.3% overall (0.9% for men and 1.6% for women) (Caraveo-Anduaga, Colmenares & Saldivar, 1999b).

For the participants in the study conducted in 1995, the average age of onset for depression was 26.6 years with no significant difference in age of onset between males and females. Among depressed individuals, mild depressive episodes were slightly more frequent among women (22.6%) than men (17.1%) whereas moderate episodes were more prevalent for men (51.8%) than women (46.6%). No significant gender differences were found for severe episodes of depression (30.8% total) (Caraveo-Anduaga et al., 1999b).

In a study of college students from Mexico City, no significant gender differences were found on scores from the Beck Depression Inventory (BDI) (Rocha & Ortega-Soto, 1995). The students scored as non-depressed with an average score of 7.7.

Because depression has become a major public health problem both worldwide and in Mexico, there is an increased need for treatment and for measures in Spanish that detect its symptoms adequately. The use of a measure that can support valid and reliable interpretations will have a positive impact on the case formulation, treatment, planning and outcome evaluation (Dozois, 2003).

Issues in the Cross – Cultural Study of Depression Cultural Issues in Mental Health Assessment

As noted, the prevalence of mental ailments has been acknowledged in almost every country in the world; standard procedures and terminology are now used by health professionals worldwide. However, the diagnosis, treatment and prognosis of those afflicted by such mental ailments are strongly influenced by their culture and society.

According to Marsella (1979), the main sources of cultural influence in mental disorders are: (a) cultural variations in stress inducers such as social change, acculturation, life events, goal-striving discrepancy, role discrimination or role conflict stress, (b) cultural variations in the definition of abnormality, and (c) personality structure and dynamics affecting expressions of normal and abnormal behavior.

Differences in cultural norms can also affect the equivalence of translated tests. Social desirability and the openness to discuss certain topics can affect the choices that individuals from different cultures make when responding to a test, leading them to avoid answering specific items or modifying their answers according to what is considered acceptable in their society (Kristjansson, Desrochers & Zumbo, 2003). Therefore, it is essential to take in account the cultural context when studying mental ailments and to apply a principle of cultural relativism to the study of these phenomena.

Cross – Cultural Research of Depression Measures

One of the mental disorders that has been studied cross-culturally is depression. Along with schizophrenia, depression has been the topic of several research projects in settings as diverse as the U.K., Switzerland, Canada, Ghana, Japan and Iran (Tanaka– Matsumi, 2001).

Although depression has a significant genetic component, cultural factors also seem to play an important role in its expression (Thakker & Ward, 1998). Explanations for differences in the prevalence and manifestations of depression have been related to some aspects of family structure (e.g., social support provided by extended families, close mother–child relationships) or mourning rituals (e.g., expressions of grief) (Berry et al., 1992).

Some of the core symptoms of depression (such as sad affect and loss of enjoyment) have been found both in Western and non–Western cultures. However, symptoms related to somatization are acknowledged to be culture–bound. This type of symptomatology is relatively unusual in Western countries, but is very common in many non-Western populations (including Asian communities such as Turkey). One explanation for this discrepancy is the socially accepted expression of negative emotions found in differing countries (Thakker & Ward, 1998). Such differences in the manifestation of depressive symptoms may result into a lack of conceptual equivalence in tests intended to assess depression (Kristjansson et al., 2003). An international project conducted by WHO in 1983 is considered to be a prototypical cross-cultural study of depression. Its goal was to test the feasibility of using standardized instruments across several countries. In total, 573 patients from Switzerland, Canada, Japan and Iran were tested with the Schedule for Standardized Assessment of Depressive Disorders (WHO/SADD). This inventory examined 39 symptoms of depression and reported a reliability coefficient of .96. The results from the study included the presence (in 76% of the depressed patients) of symptoms such as sadness, anxiety, lack of energy, and loss of interest or ability to concentrate. Cross–cultural variations (both within and across cultures) were also present and included both somatic complaints and obsessions (Tanaka–Matsumi, 2001).

Manson, Shore and Bloom (1985) worked on the development of the American Indian Depression Schedule-Hopi version by conducting interviews with members of the Hopi tribe in their native language. This work lead to the creation of an instrument related to the DSM–III and to Hopi forms of illness and ideas about depression.

According to Marsella (1979), cross–cultural epidemiological studies in depression can be improved with procedures such as: (a) an emic determination of mental disorder categories for different cultures, (b) the establishment of symptom frequency and duration baselines for both normal and abnormal populations, (c) the use of factor analysis to produce symptom patterns, and (d) the study of culturally relevant definitions of mental disorders for each culture with similar research methods. These issues have been raised in other inquiries about the manifestation of depression in different countries (Byrne & Campbell, 1999; Kojima et al., 2002; Novy, Santley, Averill & Daza, 2001;

Penley, Wiebe & Nwosu, 2003) and could provide considerations for future research in the study of depression across cultures.

Validity and Translated Tests.

Paunonen and Ashton (1998) outline three reasons to translate (or adapt) a measure developed in a different language. The first reason is the lack of local measures of the construct in the targeted culture. Second, is an interest in examining the generalizability of the inventory and the construct it measures. The third reason is an interest in exploring whether personality is manifested the same way across cultures.

When a test is used with a different population than the one for which it was first designed, the problem of generalizability arises. Because the original validity evidence was obtained with a specific population, the generalization of those findings to a new one cannot be taken for granted. Specifically, ecological generalizability should be demonstrated when using a measure in different cultural settings (Messick, 1989).

In the validation process, test developers have the responsibility of specifying the intended interpretation of the scores, the population, and the definition of the construct measured by the test (APA, AERA, & NCME, 1999, Standard 1.2). Test users are also responsible for providing additional evidence whenever they intend to use the test in a way that has not been validated previously (AERA, APA, & NCME, 1999, Standard 1.4).

This concern for the generalizability of inferences made from a test score has been addressed by some authors (Byrne & Campbell, 1999; Byrne & Watkins, 2003) as one of measurement invariance or measurement equivalence. Generalizability of inferences means that items from a test are perceived and interpreted in exactly the same way across different samples. Measurement invariance is present when the tests have equivalent factor loadings and equivalent factorial structure (Byrne & Watkins, 2003). Measurement invariance is especially relevant for cross-cultural research that employs measures originally developed in another language and for another culture (Byrne & Campbell, 1999).

When equivalence is not present, this may be due to defective translations that result in bias³, although it should be noted that bias can be present even in a test that is used for two populations with a common language, but different cultural backgrounds (Byrne & Watkins, 2003). Lack of semantic equivalence, differences in cultural norms, or lack of conceptual equivalence can all be responsible for the bias found in a test that has been translated or adapted (Kristjansson et al., 2003). When this is known to be the case, the instrument should be modified or new norms should be developed for each population in order to enhance its validity (Lee & Sue, 2001).

Back translation of a test traditionally was considered to be a good technique to determine the equivalence of two versions (Geisinger, 1994). Back translation procedures involve first translating the test into a different language, and then translating the translated version back into the original language. A comparison is then made between the original version of the test and the "back-translated" version to study their similarities and differences. However, it is known now that this technique only studies literal meanings and leaves aside the general equivalence of the original item and its translation (Kristjansson et al., 2003).

A process that involves more steps in the translation process was suggested by authors like Geisinger (1994). Besides the translation (the first step), this new process

³ Bias is said to occur when "deficiencies in a test itself or the manner in which it is used result in different meanings for scores earned by members of different identifiable subgroups" (AERA, APA & NCME, 1999, p. 74)

incorporates a review of the new version of the measure, an adaptation of it on the basis of the comments of the reviewers, a pilot testing and a field testing of the new instrument, a new standardization of the scores and also a new process of validation.

When a measure has been translated, two techniques are now used to study translation measurement invariance: scale-level analysis and item-level analysis. In scalelevel analysis, factor analysis is used to test the factorial invariance of the test. In itemlevel analysis, one item at a time is analyzed to detect differential item functioning (DIF) (Zumbo, 2003). When DIF is detected, the assumption of local independence is violated; this assumption requires that the measurement properties of items remain the same regardless of external variables such as culture or gender (Lange et al., 2002).

Cross-cultural applicability is achieved when an instrument is known to possess the same reliability, criterion-related validity and construct validity (Paunonen & Ashton, 1998). Only when measurement invariance is demonstrated between the original and the new versions (translated or adapted) can cross-cultural research be conducted.

Beck Depression Inventory

The Beck Depression Inventory (BDI) is the most widely used self-report measure of depression (Dozois, 2003), and has been used in a number of cross-cultural studies of depression (Byrne & Campbell, 1999; Novy et al., 2001; Penley et al., 2003).

The Beck Depression Inventory (BDI) was introduced in 1961 (Beck, Ward, Mendelson, Mock & Erbaugh, 1961) and a second version, the BDI-1A was published in 1978. The purpose of the BDI was to assess the presence and severity of depression in individuals already diagnosed as depressed and in normal populations of both adolescents and adults (Dozois, Dobson & Ahnberg, 1998). Good internal consistency, test-retest

reliability, and construct and factorial validity were found with the BDI, but there were concerns about its content validity and the presence of gender bias in some items (Arbisi, 2001; Dozois et al., 1998). In the BDI-IA, 15 of the 21 items were re-worded and double negatives were eliminated. This revised version also became a popular measure of depression (Bedi, Maraun, & Chrisjohn, 1998).

The BDI was again revised in 1996 and renamed the Beck Depression Inventory-II (BDI-II; Beck, Steer & Brown, 1996). The BDI-II is a 21-item, self-report measure of depression that can be administered in five to ten minutes to individuals 13 years of age and older. Its purpose is to assess symptoms corresponding to the criteria for diagnosing depressive disorders listed in the DSM-IV (APA, 1994). In line with the DSM-IV, the BDI-II asks for symptom ratings over the past two weeks. Items from the BDI related to weight loss, body image, change, somatic preoccupation and work difficulty were eliminated in the BDI-II. The BDI-II was based on items from the BDI-IA, although all but three items were reworded in some way. New items included in the BDI-II address psychomotor agitation, concentration difficulties, sense of worthlessness, and loss of energy. Also, the items for weight gain or loss and for changes in sleep patterns (i.e., insomnia or hypersonnia) were modified so both options were included in a single item for each phenomenon (Arbisi, 2001; Farmer, 2001). The BDI-II can be administered to groups or individuals, and it can be scored by hand or computer. It is available in English and Spanish.

的感到的

The BDI-II total score is derived by summing the weights corresponding to the statements endorsed over all 21 items. The four statements under each item range from 0 (*not present*) to 3 (*severe*). Total scores range from 0 to 63.

Psychometric properties of the English Versions of the BDI-II

へお外交目に対し

法律

According to the test manual (Beck et al., 1996), two samples were used to evaluate the psychometric characteristics of the BDI-II. The first sample included 500 outpatients, mostly White, with an average age of 37.2 years (SD= 15.91) from the U.S. The second sample included 120 predominantly White Canadian college students, with an average age of 19.58 years (SD = 1.84). Both samples have been described as nonrepresentative in terms of ethnicity and geographic distribution (Farmer, 2001). The BDI-II scores displayed very high internal consistency coefficients (over .90 for both samples). The test-retest reliability coefficient across the period of a week was .93 among a small subsample of 26 outpatients from one clinic site.

In terms of the convergent validity of the BDI-II, the authors reported high correlations with tests such as the Hamilton Psychiatric Rating Scale for Depression (r=.73), the Beck Hopelessness Scale (r=.68), the BDI-IA (r=.93), and the SCL – 90 – R Depression subscale (r=.89) (Farmer, 2001; Steer, Ball, Ranieri & Beck, 1997). As for discriminant validity, the BDI-II has low correlations with the Hamilton Rating Scale for Anxiety (r=.47) but higher correlations with the SCL-90-R Anxiety subscale (r=.71). Steer et al. (1997) attributed this correlation to the underlying dimension of negative affectivity shared by anxiety and depression. A maximum-likelihood confirmatory factor analysis performed with the college students sample produced a two factor model: *cognitive – affective* and *somatic*. Both total scores and factor scores had high alpha coefficients (.89 overall, .72 for somatic and .87 for cognitive-affective), showing adequate internal consistency (Whisman, Perez & Ramel, 2000).

ج].

Studies assessing the psychometric properties of the BDI-II have been conducted with Canadian (Dozois et al., 1998) and American samples (Steer, Ball, Ranieri, & Beck, 1999; Whisman et al., 2000). A study conducted with a sample of undergraduate psychology students at the University of Calgary (Dozois et al., 1998) obtained evidence of high internal consistency with an alpha coefficient of .91. A high correlation (r=.93) between the scores on the BDI and the BDI-II was also reported. A factor analysis with a maximum likelihood extraction method and an oblique rotation showed a two-factor structure which they labeled *cognitive-affective* and *somatic-vegetative*; a confirmatory factor analysis found that this structure had a good fit for the data.

A two-factor structure for the BDI-II was also found with American non-clinical undergraduate students (Whisman et al., 2000). After conducting a principal-axis exploratory factor analysis, two factors were found and labeled as *cognitive-affective* and *somatic*.

4.55

An exploratory factor analysis conducted with BDI-II scores from a clinical sample of depressed outpatients also found a two-factor structure; the factors were labeled *somatic-affective* and *cognitive* (Steer et al., 1999). Gender differences, showing higher mean scores for women, were found with this clinical sample (Steer et al., 1999), but not with Canadian students (Dozois et al., 1998).

Beck et al. (1996) corrected item-total correlations ranged from .39 to .70 for the clinical sample and from .27 to .74 for the non-clinical sample. An iterated principal-factor analysis using a promax rotation showed a two-factor structure for the clinical sample. The factors were *somatic-affective* and *cognitive*. When an iterated principal-factor analysis using a promax rotation was conducted with the scores from the student

(non-clinical) sample, a two-factor structure was also found, but the factors were labelled *cognitive-affective* and *somatic* in this case.

When a test is used with a sample from a different target population or culture, or when the test has been translated, adapted, modified, or revised it is important to provide evidence supporting the reliability and validity of the scores and inferences, respectively, obtained using that test. In the case of the BDI-II, evidence obtained from earlier versions of the test (i.e., BDI or BDI-IA) should not be assumed to apply to the new version. Likewise, psychometric evidence obtained with the English version of the BDI-II should not be assumed to the Spanish version of the BDI-II. These generalizations should not be made particularly when the samples included in the studies are described as nonrepresentative in terms of ethnicity and geographic distribution (Farmer, 2001). Furthermore, the psychometric properties of the Spanish BDI-II must be evaluated when used with a different culture (e.g., Mexicans, Mexican-Americans, Argentineans) or target population (e.g., Clinical vs. Non Clinical).

Non-English versions of the BDI and BDI-II

One of the main goals of many articles involving the translation and adaptation of inventories to other languages is to assess their psychometric properties. The crosscultural comparison of those properties often constitutes a secondary goal. Studies have been conducted with the Bulgarian, Swedish, Portuguese and Spanish versions of the BDI; as well as with the Arabic, Japanese and Spanish versions of the BDI-II. First, a review of the studies involving the BDI will be provided.

Byrne and Campbell (1999) conducted a confirmatory factor analysis of the BDI given to Bulgarian, Canadian, and Swedish non-clinical high school students. A common

three-factor structure (negative attitude, performance difficulty, and somatic elements) was found. The inventory had been translated to Swedish and Bulgarian by independent research teams that conducted pilot testing for the Bulgarian version and cross-validation by bilingual translators for the Swedish version.

A psychometric study of the Portuguese version of the BDI was conducted with Brazilian college students (Gorenstein, Andrade, Guerra, Chei Tung & Artes, 1999). The scores on this test showed a high internal consistency (alpha coefficient=.86) and itemtotal correlations between .18 and .58. A principal components factor analysis with varimax rotation was conducted and produced a three-factor structure. The factors were labeled *low self-esteem*, *cognitive-affective* and *somatic*. When the factor analysis was conducted separately for women and men, two-factor structures were found for each gender (labeled *affective* and *somatic*).

Studies of the psychometric properties of the Spanish version of the BDI have been found with Argentinean (Bonicatto et al., 1998; Richaud & Sacchi, 2002) and Mexican populations (Jurado et al., 1998; Suarez-Mendoza et al., 1997; Torres-Castillo, Hernández-Malpica & Ortega-Soto, 1991).

STATES OF

1200

Richaud & Sacchi (2002) conducted a principal components analysis using varimax and promax rotations of a Spanish version of the BDI, that the authors translated themselves with a non-clinical Argentinean sample. A four-factor structure was found which they labeled *mildest levels of depression*, *affective state and performance*, *physical complains* and *cognitive rumination*.

Another study conducted with an Argentinean non-clinical sample used the Spanish version of the BDI published by the Psychological Corporation. The internal

consistency of the BDI was high (alpha coefficient=.87). A principal component analysis using a varimax rotation found a two-factor structure; the factors were labeled *cognitive-affective* and *somatic* (Bonicatto et al., 1998).

Suarez-Mendoza et al. (1997) conducted a study with 35 participants diagnosed with rheumatoid arthritis and reported evidence of good concurrent validity, higher average scores for patients diagnosed with depression by a psychiatrist and lower average scores for patients not diagnosed with depression.

In a study with 96 Mexican patients diagnosed with a cardiovascular disease, Torres-Castillo et al. (1991) found a statistically significant difference between the scores of depressed and non-depressed patients as diagnosed by a psychiatrist, good testretest reliability over a 48 hours interval (r=.89), and a high corrected intraclass correlation coefficient (.77).

Jurado et al. (1998) used a Spanish version of the BDI adapted according to the procedures suggested by Geisinger (1994). With a sample of 1508 students and individuals from the general population in Mexico City, internal consistency was high (alpha coefficient=.87). A factor analysis with varimax rotation showed a three-factor structure, although the authors did not provide any labels for the factors.

The studies with the BDI-II involve Arabic, Japanese and Spanish versions of the BDI-II. In an Arabic sample of undergraduate students, a three-factor model (*cognitive-affective, overt emotional upset* and *somatic complains*) was found for the BDI-II when a principal axes factor analysis with oblique rotation was used (Al–Musawi, 2001). A confirmatory factor analysis was also conducted with this data and the author found evidence of good fit.

 ~ 1

The two-factor model (*cognitive* and *somatic-affective*) found by Beck et al. (1996) for the BDI-II with a clinical sample, was present in the Japanese sample (Kojima et al., 2002) when a principal component analysis was conducted.

In a study with 98 bilingual adults (fluent in English and Spanish) born in the US, South and Central America and Mexico, Novy et al. (2001) found evidence of good internal consistency for both the English and Spanish versions of the BDI-II (alpha coefficients=.94 for both). Discriminant validity was assessed through the correlation between the BDI-II and the Beck Anxiety Inventory. The correlation was .58 for the English version and .60 for the Spanish version. The correlation between the total scores on the Spanish and the English versions of the BDI-II was r=.94.

Penley et al. (2003) reported evidence of good internal consistency (alpha coefficient=.92) in a medical sample of 122 Mexican American patients with end-stage renal disease. A subsample of participants (n=23) answered both the English and Spanish versions of the BDI-II. The correlation between the Spanish and the English total score of the BDI-II was r=.70, with no statistically significant differences between Spanish and English BDI-II total scores. Corrected item-total correlations were found between .33 (for suicidal thoughts) and .72 (for worthlessness). A confirmatory factor analysis was performed to a model with two first-order factors of depression and a single second-order factor of depression and evidence of an adequate fit to the data in this study was found. *Cultural Issues in Assessing Depression with the BDI and BDI-II*

For both the BDI and the BDI-II, research has found similar phenomena in different countries such as age-related differences like omission of specific items with

the Japanese participants (Kojima et al., 2002) or an inverse relationship between age and depressive symptoms in the Brazilian respondents (Gorenstein et al., 1999).

In spite of this evidence of generalizability for the BDI and BDI-II, we can imagine the presence of cultural differences, not only in their translations but also in their results. One of the findings for the Japanese translation is linked to cultural factors. According to the authors (Kojima et al., 2002), sexual matters in Japan are not discussed openly. Therefore, they were not surprised when they found that the item most likely to be missed by older women in their sample was the one that addressed "loss of interest in sex". This phenomenon echoes Kristjansson et al. (2003) description of the culturally bound effects of social desirability and the openness to discuss certain topics. Avoiding answering specific items or modifying certain answers, according to what is considered wacceptable in their society, can be some of the participants' responses to the test'items that differ from one culture to another. However no conclusive arguments could be stated by Kojima et al. (2002) due to the limited amount of information that could be derived from the scores. Research with the BDI has encouraged the incorporation of qualitative techniques to the study of depression inventories to "partly close the gaps between meaning and measurement" (Barroso & Sandelowski, 2001, p.502).

References

- Al-Musawi, N.M. (2001). Psychometric properties of the Beck Depression Inventory-II with university students in Bahrain. *Journal of Personality Assessment*, 77, 568-579.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*, (4th ed.). Washington, D.C.: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*, (4th ed, Text revision). Washington, D.C.: Author.
- Arbisi, P. (2001). Review of the Beck Depression Inventory II. In J.C. Impara & B.S. Plake (Eds.) *The fourteenth mental measurement yearbook*. (pp.121-123). Lincoln, NB: The

Buros Institute of Mental Measurements

- Barroso, J., & Sandelowski, M. (2001). In the field with the Beck Depression Inventory. *Qualitative Health Research*, 11, 491-504.
- 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.
- Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *BDI-II manual*. San Antonio, TX: The Psychological Corporation.
- Bedi, R.P., Maraun, M.D. & Chrisjohn, R.D. (2001). A multisample item response theory analysis of the Beck Depression Inventory-1A. *Canadian Journal of Behavioural Science*, 33, 176-187.

-5

- Berry, J., Poortinga, Y., Segall, M., & Dasen, P. (Eds). (1992). *Cross-cultural psychology: Research and applications*. (pp. 1-14, 356-377). New York: Cambridge University Press.
- Bonicatto, S., Dew, A.M., & Soria, J.J. (1998). Analysis of the psychometric properties of the Spanish version of the Beck Depression Inventory in Argentina. *Psychiatry Research*, 79, 277-285.
- Borrelli, B., Marcus, B.H., Clark, M.M., Bock, B.C., King T.K., & Roberts, M. (1999). History of depression and subsyndromal depression in women smokers. *Addictive Behaviours*, 24, 781-794.
- Byrne, B., & 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, 555-574.
- Byrne, B., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-cultural Psychology*, 34, 155-175.
- Caraveo, J., Medina-Mora, M.E., Rascón, M.L., Villatoro, J., López, E.K., Juárez, F., Gómez, M., & Martínez, N.A. (1994). Características psicopatológicas de la población urbana adulta en México. Resultados de una encuesta nacional en hogares [Psychopathological characteristics of the Mexican urban adult population. Results from a household national survey]. México: Anales del Instituto Mexicano de Psiquiatría.
- Caraveo, J., Medina-Mora, M.E., Rascón, M.L., Villatoro, J., Martínez-Velez, A., & Gómez,
 M. (1996). La prevalencia de los trastornos psiquiátricos en la población urbana adulta en
 México [Prevalence of psychiatric disorders in the Mexican urban adult population].
 Salud Mental, 19, 14-21.

- Caraveo-Anduaga, J., Colmenares, E., & Saldívar, G. (1999a). Morbilidad psiquiátrica en la ciudad de México: Prevalencia y comorbilidad a lo largo de la vida [Psychiatric morbidity in Mexico City: Lifelong prevalence and comorbidity]. Salud Mental, Número Especial, 62-67.
- Caraveo-Anduaga, J., Colmenares, E., & Saldívar, G. (1999b). Estudio clínico epidemiológico de los trastornos depresivos [A clinical epidemiological study of depressive disorders].
 Salud Mental, 22, 7-17.
- Cosío Villegas, D., Bernal, I., Moreno Toscano, A., González, L., Blanquel, E., & Meyer, L.
 (1973). *Historia mínima de México* [Minimum history of Mexico]. México: El Colegio de México.
- Dozois, D., Dobson, K.S., & Ahnberg, J.L. (1998). A psychometric evaluation of the Beck
 Depression Inventory-II. *Psychological Assessment*, 10, 83-89. Retrieved August 13,
 2003, from the PsycARTICLES database
- Dozois, D. (2003). The psychometric characteristics of the Hamilton Depression Inventory. Journal of Personality Assessment, 80, 31-40.
- Farmer, R. (2001). Review of the Beck Depression Inventory II. In J.C. Impara & B.S. Plake (Eds.) *The fourteenth mental measurement yearbook*. (pp. 123-126). Lincoln, NB: The Buros Institue of Mental Measurements.
- Ferketich, A.K., Schwartzbaum, J.A., Frid, D.J., & Moescherberger, M.L. (2000). Depression as an antecedent to heart disease among women and men in the NHANES I study. *Archives of Internal Medicine*. 160, 1261-1268.

García-Alvarez, R. (1986). Epidemiology of depression in Latin America. *Psychopathology*, *19, suppl 2*, 22-25.

59

TE CONVERNIT

n Ái

Glassman, A.H., Helzer, J.E., Covey, L.S., Cottler, L.B., Stetner, F., Tipp, J.E., & Johnson, J. (1990) Smoking, smoking cessation, and major depression. *The Journal of the American Medical Association*, 264, 1546-1150.

- Geisinger, K. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6, 304-312.
- Gorenstein, C., Andrade, L., Guerra, A.H., Chei Tung, T., & Artes, R. (1999). Psychometric properties of the Portuguese version of the Beck Depression Inventory on Brazilian college students. *Journal of Clinical Psychology*, 55, 553-562.
- Hall, S.M., Muñoz, R.F., Reus, V.I. & Sees, K.L. (1993). Nicotine, negative affect, and depression. *Journal of Consulting and Clinical Psychology*, *61*, 761-767.
- Hance, M., Carney, R.M., Freedland, K.E. & Skala, J. (1996). Depression in patients with coronary heart disease? A 12-month follow-up. *General Hospital Psychiatry*, 18, 61-65.
- INEGI & SEMARNAT. (1997). *Estadísticas del medio ambiente* [Environmental statistics]. Distrito Federal. Retrieved July 29,2003, from http://www.inegi.gob.mx/difusion/ español/fietab.html
- INEGI. (2001). Estados Unidos Mexicanos. XII Censo General de Población y Vivienda, 2000. Tabulados básicos y por entidad federativa. Bases de datos y tabulados de la muestra censal [United States of Mexico. Population and households' general census. Basic estimates and federate entities]. Retrieved July 29, 2003, from http://www.inegi.gob.mx/difusion/ español/fietab.html
- Judd, L. L., & Kunovac, J.L. (1997). Diagnosis and classification of depression. In A. Honig &H.M. Van Praag (Eds.). Depression. Neurobiological, psychopathological and

therapeutic advances: Vol.3. Wiley Series on Clinical and Neurobiological Advances in Psychiatry, (pp. 3-15). Chichester: John Wiley & Sons.

- Jurado, S., Villegas, E., Méndez, L., Rodríguez, F., Loperena, V. & Varela, R. (1998). La estandarización del Inventario de Depresión de Beck para los residentes de la ciudad de México [Standarization of the Beck Depresión Inventory for Mexico City' residents]. Salud Mental, 21, 26-31.
- Kaelber, C,T., Moul, D.E., & Farmer, M.E. (1995). Epidemiology of depression. In E.E.
 Beckham & W.R. Leber (Eds.). *Handbook of depression*, (2nd ed), (pp. 3-35). New York: The Guilford Press.
- Kessler, R.C. (2002). Epidemiology of depression. In I/H. Gotlib and C.L. Hammen (Eds.) *Handbook of depression*. (pp. 23-42). New York: The Guilford Press.
- Kojima, M., Furukawa, T.A., Takahashi, H., Kawai, M., Nagaya, T., & Tokudome, S. (2002).
 Cross-cultural validation of the Beck Depression Inventory-II in Japan. *Psychiatry Research, 110*, 291-299.
- Kristjansson, E.A., Desrochers, A., & Zumbo, B.D. (2003). Translating and adapting measurement instruments for cross-linguistic and cross-cultural research: A guide for practitioners. *Canadian Journal of Nursing Research*, 35, 127-142.
- Lange, R., Thalbourne, M.A., Houran, J., & Lester, D. (2002). Depressive response sets due to gender and culture-based differential item functioning. *Personality and Individual Differences*, 33, 937-954.
- Lee, J., & Sue, S. (2001). Clinical psychology and culture. In D. Matsumoto (Ed.), *The handbook of culture and psychology*. (pp. 287-305). New York: Oxford University Press.

- Lépine, J.P. (2001). Epidemiology, burden and disability in depression and anxiety. *Journal of Clinical Psychiatry*, 62, suppl 13, 4-9.
- Lerman, C., Audrain, J., Orleans, C.T., Boyd, R., Gold, K., Main, D., & Caporaso, N. (1996).
 Investigation of mechanisms linking depressed mood to nicotine dependence. *Addictive Behaviors*, 21, 9-19.
- Manson, S.M., Shore, J.S. & Bloom, J.D. (1985). The depressive experience in American
 Indian Communities: A challenge for psychiatric theory and diagnosis. In A. Kleinman &
 B. Good. *Culture and depression. Studies in the anthropology and cross-cultural psychiatry of affect and disorder.* (pp. 331-368). Berkeley, CA: University of California
 Press.
- Marsella, A. (1979). Cross-cultural studies of mental disorders. In A. Marsella, R. Tharp & T. Ciborowski (Ed.), *Perspectives on cross-cultural psychology*. (pp. 233-262). New York: Academic Press.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement*, (3rd ed). (pp. 13-103). New York: American Council on Education & Macmillan Publishing Company.
- Murray, C.J.L. & Lopez, A.D. (1996). The Global Burden of Disease in 1990: Final results and their sensitivity to alternative epidemiological perspectives, discount rates, ageweights and disability weights. In C.J.L. Murray & A.D. Lopez (Series Eds., Vol. Ed.). The Global Burden of Disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: Vol. 1. Global Burden of Disease and Injury Series. (pp. 247-293). Cambridge, MA: Harvard School of Public Health on behalf of the World Health Organization and the World Bank.

- Nolen-Hoeksema, S. & Girgus, J.S. (1994). The emergence of gender differences in depression during adolescence. *Psychological Bulletin*, 115, 424-443.
- Nolen-Hoeksema, S. (2002). Gender differences in depression. In I.H. Gotlib & C.L. Hammen (Eds.), *Handbook of depression*. (pp. 492-509). New York: The Guilford Press.
- Novy, D. Stanley, M. Averill, P., & Daza, P. (2001). Psychometric comparability of English and Spanish language measures of anxiety and related affective symptoms. *Psychological Assessment*, 13, 347-355.
- Paunonen, S. V., & Ashton, M.C. (1998). The structured assessment of personality across cultures. *Journal of Cross-cultural psychology*, 29, 150-170.
- Penley, J.A., Wiebe, J.S., & Nwosu, A. (2003). Psychometric properties of the Spanish Beck Depression Inventory-II in a medical sample. *Psychological Assessment*, 15, 596-577.
- Richaud, C., & Sacchi, C. (2002) Adaptacion del inventario de Depresion de Beck a sujetos argentinos normales / Adaptation of Beck Depression Inventory to a sample of Argentine general population. *Revista Iberoamericana de Diagnostico y Evaluacion Psicologica*, 12, 11-17.
- Rocha, H.M., & Ortega-Soto, H.A. (1995). La depresión en los estudiantes universitarios de la Escuela Nacional de Estudios Profesionales Plantel Aragón. Salud Mental, 18, 31-34.
 Sartorius, N. (1986). Cross-cultural research on depression. Psychopathology, 19, suppl 2, 6-11.
- Steer, R., Ball, R., Ranieri, W.F., & Beck, A.T. (1997). Further evidence for the construct validity of the Beck Depression Inventory-II with psychiatric outpatients. *Psychological Reports*, 80, 443-446.

- Steer, R., Ball, R., Ranieri, W.F., & Beck, A.T. (1999). Dimensions of the Beck Depression
 Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, 55, 117128.
- Suarez-Mendoza, A.A., Cardiel, M.H., Caballero-Uribe, C.V., Ortega-Soto, H.A., & Marquez-Marin, M. (1997). Measurement of depression in Mexican patients with rheumatoid arthritis: Validity of the Beck Depression Inventory. *Arthritis Care and Research, 10*, 194-199.
- Tanaka-Matsumi, J. (2001). Abnormal psychology and culture. In D. Matsumoto (Ed.), *The handbook of culture and psychology*. (pp. 265-286). New York: Oxford University Press.
- Thakker, J., & Ward, T. (1998). Culture and classification: The cross-cultural application of the DSM-IV. *Clinical Psychology Review*, 18, 501-529.
- Torres-Castillo, M., Hernández-Malpiča, En& Ortega-Soto, H.A. (1991). Validez y reproducibilidad del Inventario para Depresión de Beck en un hospital de cardiología [Validity and reproducibility of the Beck Depression Inventory in a cardiac care hospital]. Salud Mental, 14, 1-6.
- Tsai, J.L., & Chenstsova-Dutton, Y. (2002). Understanding depression across cultures. In I.H. Gotlib & C.L. Hammen (Eds.) *Handbook of depression*. (pp 467-491). New York: The Guilford Press.
- Whisman, M., Perez, J., & Ramel, W. (2000). Factor structure of the Beck Depression
 Inventory-Second edition (BDI-II) in a student sample. *Journal of Clinical Psychology*, 56, 545-551.
- World Health Organization. (1992). International statistical classification of diseases and related health problems, (10th rev.). France: Author.
World Health Organization. (2001a). The World Health Report 2001. Mental health: New understanding, new hope. France: Author.

World Health Organization. (2001b). Mental and neurological disorders. *Fact Sheet No.* 265. Retrieved July 29, 2003, from http://www.who.int/inf-fs/en/fact265.html

World Health Organization. (2003a). Selected health indicators for Mexico. Retrieved July 29, 2003, from http://www3.who.int/whosis/country/indicators/

cfm?country=MEX&language=english

World Health Organization. (2003b). Selected health indicators for Canada. Retrieved August 13, 2003, from http://www3.who.int/whosis/country/

indicators/cfm?country=can&language=english

Zumbo, B.D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language Testing*, 20, 136-147.

41.2