Multistudy Report

Validating the Enright Forgiveness Inventory – 30 (EFI-30)

International Studies

Robert Enright[1,4], Julio Rique[2], Romulo Lustosa[3], Jacqueline Y. Song[4], Mary Cate Komoski[5], Iffat Batool[6], Daniel Bolt[1], HyunJung Sung[1], S. T. Tina Huang[7], Younghee Park[8], Paul E. Leer-Salvesen[9], Thiago Andrade[2], Anam Naeem[6], Jocelyn Viray[10], and Edna Costuna[10]

[1]Department of Educational Psychology, The University of Wisconsin-Madison, Madison, WI, USA
[2]Department of Psychology, Federal University of Paraiba, João Pessoa, PB, Brazil
[3]Department of Foundations of Education, Federal University of Paraiba, João Pessoa, PB, Brazil
[4]International Forgiveness Institute Inc., Madison, WI, USA
[5]Department of Psychology and Communication, Texas A&M International University, Laredo, TX, USA
[6]Department of Psychology, Government College University, Lahore, Pakistan
[7]Department of Psychology, National Chung-Cheng University, Taiwan, ROC
[8]Department of Psychology, Duksung Women's University, Seoul, South Korea
[9]Department of Religion, Philosophy, and History, University of Agder, Norway
[10]Department of Psychology, Centro Escolar University, Manila, Philippines

Abstract: This study reports the process of item reduction of the Enright Forgiveness Inventory – EFI, a measure of interpersonal forgiveness, from 60 to 30 items for a more practical assessment of this construct. Data from the US were used in the creation of the new measure and applied to seven nations: Austria, Brazil, Israel, South Korea, Norway, Pakistan, and Taiwan. The question was: do the best EFI-30 items from the US have discriminative power in seven other cultures? Results provided the psychometric evidence for the reduced version of the EFI-30 across cultures. The discrimination values are positive, suggesting that the selected items have the sensitivity to differentiate accurately people with different degrees of forgiveness and good psychometric properties of internal consistency.

Keywords: Enright Forgiveness Inventory (EFI), interpersonal forgiveness, forgiveness assessment, counseling for forgiveness

This study presents psychometric evidence of a short version of the Enright Forgiveness Inventory (here and thereafter, EFI; Subkoviak et al., 1995) reporting data from the US and seven other countries: Austria, Brazil, Israel, Korea, Norway, Pakistan, and Taiwan. The EFI has become internationally available in several languages and applied to human development and mental health research and treatment (Baskin & Enright, 2004; Enright & Fitzgibbons, 2015; Hanke & Fisher, 2012; Lee & Enright, 2019; Rique & Camino, 2010).

Enright and The Human Development Study Group (1991) created the EFI to provide a psychological assessment for Enright's theory of interpersonal forgiveness (see Enright & Fitzgibbons, 2015). The theory defines forgiveness as a moral response of goodness from the perspective of victims, toward the offending person, after specific interpersonal injustices. Forgiveness is "a willingness to abandon one's right to resentment, negative judgment, and indifferent behavior toward one who unjust injured us while fostering the undeserved qualities of compassion, generosity, and even love toward him or her" (Enright et al., 1998, pp. 46–47). Central to this definition is that to forgive one should be able to separate offending persons from their behavior, seeing them as fully human and their actions as part of the human condition, without condoning those actions. This is generosity and compassion for others, a psychological response that requires victims to overcome negative feelings, thoughts, and behaviors while

cultivating positive feelings, thoughts, and behaviors toward offenders.

The EFI is designed to assess these six psychological stances of victims toward offenders: positive affect, negative affect, positive behavior, negative behavior, positive cognition, and negative cognition. Subkoviak et al. (1995) published the first empirical article using the EFI with US American university students and their middle-aged parents. The authors verified a significant negative relationship between a person's degree of forgiveness and a state of anxiety. They compared groups of people by age, degree of hurt, and kind of hurt in close relationships. Results showed that a person who was hurt a great deal by someone close to them but forgave to a larger degree than others, who were hurt a great deal by similar offenses, had significantly lower state anxiety. These findings were encouraging to apply Enright's theory and definition in many directions, including the theory supporting a therapeutic process model and a socio-cognitive developmental model (Enright & Fitzgibbons, 2015). For clinical psychology, mental health professionals began to apply the process model (Freedman & Enright, 2020) in forgiveness therapy and verified the effectiveness of the virtue to reduce anger and anxiety. There is now strong empirical evidence and understanding that forgiveness interventions, when given sufficient time beyond brief therapy, reduce residual resentment, anxiety, and depression that can have debilitating effects in physical health (e.g., Lee & Enright, 2019; Waltman et al., 2009) and in mental health (see, e.g., Akhtar & Barlow, 2018; Freedman & Enright, 1996; Reed & Enright, 2006). For researchers in human and social development, forgiveness has been a dependent or criterion variable in many correlational designs and clinical studies. For many professionals in the US and abroad, the EFI has become their measure of choice (McElroy-Heltzel et al., 2020). The measure is suitable to verify affective, behavioral, and cognitive changes in people who choose forgiveness for improvement of their well-being.

In the 1990s, theories and research on forgiveness went beyond the borders of US American scientific psychology and Enright and The Human Development Study Group pioneered an international project to expand studies on forgiveness to other nations. The EFI became central to research on forgiveness outside of the US and has since been translated from English (original language) into Brazilian-Portuguese, Chinese, German, Hebrew, Korean, Dutch, and other languages. Many international graduate students dedicated to studying interpersonal forgiveness attested to the reliability and validity of the EFI in Master's theses, doctoral dissertations, and empirical articles published in the US and abroad (Enright & Fitzgibbons, 2015; Worthington & Wade, 2020). However, across nations, researchers and counseling psychologists applying the EFI in their professions began to say that the measure was too long (i.e., at least 5 pages long depending on the language into which it has been translated). This is time-consuming and expensive when professionals have to duplicate several measures and participants needed to respond to the EFI (60 items) together with other assessments or do multiple assessments across time. The 60-item version was taking more than twice as long as the 30-item version possibly because of fatigue and having to contemplate so many issues. The authors then decided to conceive a short version of the EFI that could convey the same meaningful information on the degree of interpersonal forgiveness across the domains of affect, behavior, and cognition. In other words, we now have a smaller scale that also presents evidence of validation with the benefit of administration time.

To attend to the demand of the construction of the EFI-30, a stepwise process for conceptual and empirical analysis was developed and is reported here. The first decision was that the new EFI should maintain the theoretical model conceived by Enright and The Human Development Study Group (1991). In other words, data should fit the model that interpersonal forgiveness is the degree to which a person's responses express a decrease in negative affect, negative behavior, and negative cognition, and at the same time express an increase in positive affect, positive behavior, and positive cognition toward the other person who was unjust in specific offenses. In addition, the six first-order factors (Andrade, 2014; Rique Neto et al., 2009) should show a positive correlation with the 1-Item Forgiveness and no correlation with Crowne-Marlowe Social Desirability scale (1960; Enright & Rique, 2004). The second decision was to reduce each of the six subscales of the EFI from 10 to 5 items by following assumptions of the Item Response Theory (IRT; Drasgow et al., 1995) applied to data from the US sample and, finally, to test the power of discrimination of those same items in data from seven nations to verify if the best EFI-30 items from the US have discriminative power and maintain its good psychometric properties of internal consistency across cultures.

Method

Procedures

Subkoviak et al. (1995) collected data for the EFI (60 items) in the US. Following this initial study, researchers from Austria, Brazil, Israel, South Korea, Taiwan, and Norway were invited to translate and validate the EFI (60 items) into their language without modification of items or with a few modifications as possible from its original version in English. They followed a manual for translation that

required them to use the translation-back translation method and to follow similar procedures for data collection. More recently, Pakistan provided data already using the EFI (30 items). Participants in each nation, except for Norway, also completed Crowne-Marlowe's (1960) social desirability scale. To reduce the EFI from 60 to 30 items, the item analyses reported in this study were performed on the dataset from the USA and, after, the chosen items were tested in the international data collection from the seven nations: Austria, Brazil, Israel, South Korea, Norway, Pakistan, and Taiwan.

Procedures for Scale Administration

Data analyzed in this article was from a database regarding the EFI international project on forgiveness. The researchers recruited non-probabilistic convenience samples and data were collected in colleges and universities. Instructions asked students to volunteer for the study and then each student approached one parent with a request to participate. Standard instructions for researchers using the EFI across nations were: the EFI should be paper-and-pencil administered either in groups or individually, a sociodemographic questionnaire should always be the first in a package. The EFI, including the 1-Item Forgiveness as a final question in this inventory and the social desirability scale, were randomly placed to avoid the order effect.

Participants

Participants who provided data for item analysis were 372 Austrians [men (n = 186) and women (n = 186), average age of 37 years (SD = 15.56)], 200 Brazilians [men (n = 100) and women (n = 100), average age of 36 years (SD = 14.87)], 176 Israelis [men (n = 58) and women (n = 118), average age of 36 years (SD = 14.78)], 326 Koreans [men (n = 161) men and women (n = 165), average age of 34 years (SD = 15.50)], 141 Norwegians [men (n = 52) and women (n = 89), average age of 28 years (SD = 9.89)], 404 Pakistanis [men (n = 202) men and women (n = 202), average age of 36 years (SD = 15.84)], 339 Taiwanese [men (n = 157) and women (n = 182), average age of 33 years (SD = 14.28)], and 394 US Americans [men (n = 190) and women (n = 204), average age of 36 years (SD = 15.71)].

Instruments

Sociodemographic Information

This survey provided information on age and gender.

The Enright Forgiveness Inventory – EFI (60 Items)

The Enright Forgiveness Inventory – EFI (60 Items) is composed of two sections: A Front Page that gathers information about a specific situation of hurt from a particular person and the Inventory with 60 items (originally). In addition to the 60 items, the EFI has a pseudo-forgiveness scale and the 1-Item Forgiveness scale that follows the Inventory as independent measures for content validity. The description of each scale follows:

The EFI Front Page. This is the first page of the EFI. Instructions should accommodate particular counseling or research interest with a particular kind of sample. For example, a nonspecific set of instruction is:

> "We are sometimes hurt by people, whether in family, friendship, school, work, or other situations. We ask you to think of the **most recent experience** of someone hurting you **unfairly** and **deeply**. For a few moments, visualize in your mind the events of that interaction. Try to see **the person** and try to experience what happened."

The specific set of instructions is, for example, if counselors or researchers are assessing the partnering relationship, they might consider the following wording:

> "We are sometimes hurt by people, whether in family, friendship, school, work or other situations. We ask you now to think of one particular situation in which **your partner** hurt you the most **unfairly** and **deeply**. For a few moments, visualize in your mind once again the events of that interaction. Try to see your partner and try to experience what happened."

As another example, if counselors or researchers are assessing Adverse Childhood Experiences, they might consider the following wording:

> "We are sometimes hurt by people, whether in family, friendship, school, work, or other situations. We ask you now to think of **the one person** who hurt you the most **unfairly** and **deeply when you were a child**. For a few moments, visualize in your mind once again the events of that interaction. Try to see the person and try to experience what happened."

Following the instructions, the Front Page questions the degree of hurt in a 5-point scale (1 = *no hurt* to 5 = *a great deal of hurt*); the agent of hurt, that is, respondents should identify offenders in nonspecific situations (e.g., child, spouse, relatives, same-sex friend, friend of opposite sex, employer, and others); the status of the agent of hurt that asks whether the person is living (yes or no); time since the hurt (days, months, weeks, or years) and, finally, a final, open-ended question that asks respondents to describe the offense briefly. Data from the Front Page on the demographics of specific events of injustice will not be reported in this article that is about the psychometric evidence of item reduction for a version of the EFI 30-Item Inventory.

The EFI – Inventory (60 items). The inventory was originally composed of six subscales with 10 items each for positive affect (e.g., tenderness), negative affect (e.g., cold), positive behavior (e.g., show friendship), negative behavior (e.g., avoid), positive cognition (e.g., a good person), and negative cognition (e.g., a bad person) toward the offender. Items were randomly placed within the subscales. For each item, respondents indicate their agreement on a 6-point scale that ranges from 1 = *strongly disagree* to 6 = *strongly agree*. Each of the six subscale's scores range from 10 to 60 points, with a high score representing high forgiveness.

EFI Scoring System. The negative items of each subscale should be reversed into positive to have a score result per scale of affect, behavior, and cognition. That indicates *the absence of* negative feelings, behaviors, and cognitions, and *the presence of* positive feelings, behaviors, and cognitions.

EFI Administration. The EFI is self-administered and works well for both group and individual administrations; the EFI should be introduced as an Attitude Scale that measures how people think about a hurtful interpersonal offense. There is no mention of the word forgiveness; this is technically not a measure of attitude, but the word "attitude" is used so that participants cannot glean the demand characteristics of this assessment. Respondents should answer in the order of pages presented and the word "forgiveness" should not be mentioned in oral or written instructions to participants until respondents answer the 1-Item Forgiveness scale at the end. This latter instruction intends to prevent respondents from seeing the 1-Item Forgiveness scale before answering the Inventory. Respondents have reported that the instrument is easy to complete and testing time per person is about 25 min for the 60 items version but depends on the purpose of the administration.

Pseudo-Forgiveness Scale

This scale is commonly placed after the EFI is completed as the last five items to measure a degree of denial or condonation, that is used for internal validation of the EFI. Researchers should eliminate a participant's data from analysis when this scale has a score of 20 or higher. Clinicians might consider the information since a high pseudo-forgiveness score suggests that the respondent does not see an injustice or a serious problem with what occurred.

The 1-Item Forgiveness Scale

This is an independent measure used as a final question of the EFI. It asks: To what extent have you forgiven the person you rated on the attitude scale? Respondents rate their answer on a 5-point scale that ranges from 1 = *not at all* to 5 = *complete forgiveness*.

Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe, 1960)

This is a 33-item true-false scale that measures the extent to which a person is "faking good" on test items. This measure provides construct validity for the EFI. A social desirability total score ranges from 0 to 33.

Data Analysis

Procedures

The EFI data were tabulated by country in Microsoft Excel® software. The R environment (version 2.15.1) and the ltm package (Rizopoulos, 2006, 2015) were used to estimate the GRM model (Samejima, 1969) for each of the six subscales for each country. The fit of the model and its factorial invariance between the eight countries was verified in AMOS (v. 22, SPSS, @IBM Company, Chicago, IL). The estimator used was the Maximum Likelihood (ML). Univariate and multivariate normality was verified using the asymmetry ($|Sk| < 3.0$) and kurtosis ($|Ku| < 10$) criteria of Marôco (2014). The EFI variables showed $|Sk| < 1.3$ and $|Ku| < 1.5$ with no severe violations of normality. Pearson's correlation and Cronbach's α reliability analyses were performed in SPSS. In the Electronic Supplementary Material (ESM 1) readers find the inputs and outputs for the analyses conducted.

IRT analyses were performed in two ways. First, analysis was conducted to compare two models for best fit, one with constant discrimination between the items of the EFI (the most parsimonious model) and the other with variable discrimination between the items of the EFI (more complex model). The likelihood ratio, that is, the value that shows the proximity of the data to the IRT assumptions between the two models was used to decide which model was the most appropriate for the data. Second, analyses considered the χ^2 values of the frequencies of the observed response patterns in relation to the frequencies predicted by IRT. Values < 3.00 indicate that there is little difference between the observed response patterns and the response patterns expected by the model, thus a good fit (Drasgow et al., 1995).

The results indicated that the model with changeable discrimination presents the better fit, with only a few exceptions that will be highlighted in Tables 1 and 2. Besides that, the frequencies of observed response patterns were very close to zero, with all χ^2 values below 3.00, which indicated a good fit of data to the parameters of IRT. There were only minor exceptions (0.06% in Austria and Korea, in the items of negative affect; 0.8% in Austria in the items of negative behavior; and 0.1% in Brazil and Taiwan in the items of positive cognition and negative cognition).

Table 1. Parameters description of EFI-30 Items in the sample from the US

Subscale/Item	b_1	b_2	b_3	b_4	b_5	a
Negative Affect						
Unloving	−1.79	−1.10	−0.54	−0.23	0.36	2.89
Repulsed	−1.98	−1.28	−0.69	−0.39	0.15	3.07
Cold	−1.61	−0.87	−0.40	−0.03	0.64	4.22
Dislike	−1.63	−1.07	−0.51	−0.22	0.42	4.44
Disgust	−1.60	−1.00	−0.50	−0.20	0.42	2.94
Mean (*SD*)	−1.72 (0.50)	−1.06 (0.10)	−0.53 (0.10)	−0.21 (0.13)	0.40 (0.17)	3.51 (0.75)
Positive Affect*						
Warm	−1.02	−0.39	−0.06	0.48	1.20	4.00
Tender	−1.06	−0.33	0.03	0.50	1.34	4.00
Caring	−1.43	−0.73	−0.35	0.17	0.93	4.00
Affection	−1.03	−0.42	−0.07	0.37	1.12	4.00
Friendly	−1.30	−0.61	−0.23	0.33	1.12	4.00
Mean (*SD*)	−1.16 (0.80)	−0.49 (0.15)	−0.13 (0.15)	0.37 (0.13)	1.14 (0.10)	4.00 (0.00)
Negative Behavior						
Avoid	−1.31	−0.73	−0.11	0.19	0.77	3.80
Ignore	−1.54	−0.85	−0.33	0.01	0.73	3.62
Neglect	−2.11	−1.24	−0.60	−0.15	0.74	2.76
Not attend to him/her	−1.96	−1.24	−0.70	−0.24	0.69	2.36
Stay away	−1.37	−0.81	−0.27	0.07	0.71	3.64
Mean (*SD*)	−1.65 (0.5)	−0.97 (0.25)	−0.40 (0.24)	−0.02 (0.17)	0.73 (0.02)	3.23 (0.63)
Positive Behavior						
Show friendship	−1.76	−1.28	−0.89	−0.30	0.68	2.84
Lend him/her a hand	−1.85	−1.39	−1.04	−0.39	0.54	4.38
Establish good relations with him/her	−1.97	−1.36	−0.95	−0.32	0.69	3.28
Do a favor	−1.55	−1.07	−0.72	−0.16	0.73	4.26
Aid him/her when in trouble	−1.84	−1.44	−1.18	−0.58	0.28	4.09
Mean (*SD*)	−1.79 (0.50)	−1.30 (0.14)	−0.95 (0.17)	−0.34 (0.15)	0.58 (0.18)	3.77 (0.67)
Negative Cognition						
Horrible	−1.64	−1.12	−0.72	−0.34	0.26	4.16
Dreadful	−1.78	−1.15	−0.72	−0.42	0.27	4.35
Worthless	−1.73	−1.38	−1.15	−0.78	−0.01	3.59
A bad person	−1.64	−1.11	−0.82	−0.49	0.13	4.12
Disapprove of him/her	−1.56	−0.74	−0.08	0.23	1.02	2.48
Mean (*SD*)	−1.66 (0.80)	−1.10 (0.22)	−0.69 (0.39)	−0.35 (0.40)	0.33 (0.40)	3.74 (0.76)
Positive Cognition*						
Of good quality	−2.01	−1.50	−1.05	−0.32	0.74	3.61
A good person	−1.98	−1.40	−1.04	−0.48	0.62	3.61
Wish him/her well	−2.27	−1.61	−1.24	−0.62	0.28	3.61
Think favorably of him/her	−1.86	−1.24	−0.66	−0.06	0.91	3.61
Hope he/she succeeds	−2.21	−1.80	−1.39	−0.81	0.07	3.61
Mean (*SD*)	−2.06 (0.7)	−1.51 (0.21)	−1.07 (0.7)	−0.45 (0.28)	0.52 (0.34)	3.61 (0.00)

Note. *Model with constant discrimination. See the last paragraph on the procedures for analysis.

Items Selection for the EFI-30

Sample From the US

This study reports, first, the psychometric parameters of the applied IRT for item selection of the EFI using the EFI (60 items) data from the US original sample (Subkoviak et al., 1995). The goal was to reduce the number of items in each subscale from 10 to 5. The best items should show the highest discriminative power in each subscale and observe the Item Characteristic Curves to ascertain the best

Table 2. Means (*M*) and standard deviations (*SD*) of response thresholds for the EFI-30 from seven nations

	b_1	b_2	b_3	b_4	b_5	a
Nation/Scale	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Austria						
Negative Affect	−1.59 (0.42)	−0.90 (0.50)	−0.28 (0.44)	0.12 (0.55)	0.85 (0.69)	2.28* (0.00)
Positive Affect	−1.58 (0.38)	−0.94 (0.97)	−0.32 (0.24)	0.19 (0.41)	0.89 (0.59)	2.50 (0.49)
Negative Behavior	−1.88 (0.65)	−1.43 (0.55)	−0.85 (0.33)	−0.31 (0.28)	0.47 (0.27)	2.48 (1.47)
Positive Behavior	−2.16 (0.27)	−1.60 (0.39)	−1.11 (0.43)	−0.53 (0.59)	0.37 (0.69)	2.37 (0.55)
Negative Cognition	−1.82 (0.28)	−1.34 (0.39)	−0.83 (0.38)	−0.27 (0.50)	0.47 (0.61)	2.85* (0.00)
Positive Cognition	−1.97 (0.36)	−1.40 (0.41)	−0.90 (0.44)	−0.41 (0.42)	0.31 (0.48)	2.59 (1.46)
Brazil						
Negative Affect	−1.86 (0.12)	−1.15 (0.19)	−0.63 (0.27)	−0.39 (0.31)	0.44 (0.35)	2.91 (0.68)
Positive Affect	−1.30 (0.17)	−0.58 (0.08)	−0.35 (0.06)	0.23 (0.15)	1.20 (0.27)	3.09 (0.67)
Negative Behavior	−1.75 (0.37)	−0.78 (0.37)	−0.29 (0.34)	−0.02 (0.30)	0.87 (0.66)	2.60 (1.09)
Positive Behavior	−1.82 (0.14)	−1.13 (0.12)	−0.91 (0.22)	−0.32 (0.17)	0.82 (0.19)	2.76 (0.79)
Negative Cognition	−2.37 (0.23)	−1.37 (0.39)	−0.91 (0.49)	−0.63 (0.56)	0.46 (0.60)	3.14 (0.00)
Positive Cognition	−1.87 (0.22)	−1.19 (0.37)	−0.86 (0.35)	−0.31 (0.36)	0.86 (0.22)	3.59 (0.67)
Israel						
Negative Affect	−1.62 (0.53)	−0.77 (0.46)	−0.19 (0.26)	0.16 (0.20)	0.77 (0.17)	2.50 (0.68)
Positive Affect	−0.90 (0.17)	−0.27 (0.15)	0.05 (0.19)	0.59 (0.22)	1.14 (0.26)	2.82 (0.69)
Negative Behavior	−1.38 (0.39)	−0.72 (0.36)	−0.12 (0.30)	0.22 (0.29)	0.89 (1.09)	2.46 (0.37)
Positive Behavior	−1.33 (0.24)	−0.71 (0.34)	−0.29 (0.41)	0.27 (0.31)	1.01 (0.35)	3.21 (0.87)
Negative Cognition	−1.70 (0.23)	−1.12 (0.23)	−0.66 (0.20)	−0.33 (0.23)	0.20 (0.26)	2.45 (0.14)
Positive Cognition	−1.38 (0.12)	−0.89 (0.18)	−0.50 (0.23)	0.01 (0.27)	0.67 (0.30)	3.05 (0.95)
South Korea						
Negative Affect	−1.69 (0.18)	−0.87 (0.16)	−0.12 (0.27)	0.24 (0.33)	1.21 (0.45)	2.50 (0.68)
Positive Affect	−1.16 (0.21)	−0.23 (0.17)	0.23 (0.18)	0.95 (0.18)	1.88 (0.29)	2.82 (0.69)
Negative Behavior	−1.93 (0.26)	−0.94 (0.07)	−0.19 (0.13)	0.13 (0.16)	1.11 (0.20)	2.46 (0.37)
Positive Behavior	−1.74 (0.09)	−0.82 (0.14)	−0.40 (0.15)	0.30 (0.10)	1.44 (0.26)	3.21 (0.87)
Negative Cognition	−2.02 (0.32)	−1.40 (0.31)	−0.81 (0.26)	−0.47 (0.25)	0.65 (0.22)	2.45 (0.14)
Positive Cognition	−1.64 (0.14)	−0.79 (0.25)	−0.24 (0.35)	0.43 (0.42)	1.52 (0.56)	3.05 (0.95)
Norway						
Negative Affect	−2.13 (0.23)	−1.30 (0.30)	−0.61 (0.23)	−0.21 (0.27)	0.44 (0.40)	3.00 (1.25)
Positive Affect	−1.32 (0.45)	−0.55 (0.32)	−0.12 (0.27)	0.53 (0.22)	1.29 (0.15)	3.27 (0.74)
Negative Behavior	−1.82 (0.06)	−1.12 (0.11)	−0.72 (0.07)	−0.43 (0.13)	0.25 (0.11)	4.22* (0.00)
Positive Behavior	−1.86 (0.18)	−1.12 (0.20)	−0.59 (0.23)	−0.01 (0.30)	0.90 (0.14)	3.36* (0.00)
Negative Cognition	−2.41 (0.39)	−1.87 (0.21)	−1.38 (0.26)	−0.90 (0.27)	0.12 (0.50)	2.64 (1.36)
Positive Cognition	−2.38 (0.41)	−1.65 (0.29)	−1.16 (0.50)	−0.45 (0.46)	0.59 (0.39)	3.51 (0.91)
Pakistan						
Negative Affect	−1.41 (0.57)	−0.49 (0.20)	0.20 (0.21)	0.83 (0.34)	1.83 (0.75)	1.96 (0.95)
Positive Affect	−1.73 (0.90)	−0.73 (0.32)	−0.07 (0.18)	0.83 (0.66)	2.01 (1.30)	1.65 (1.07)
Negative Behavior	−1.11 (0.19)	−0.31 (0.18)	0.23 (0.19)	0.78 (0.13)	1.58 (0.10)	2.42 (0.58)
Positive Behavior	−1.30 (0.22)	−0.61 (0.17)	−0.07 (0.34)	0.49 (0.49)	1.28 (0.65)	2.56 (0.81)
Negative Cognition	−1.51 (0.45)	−0.76 (0.37)	−0.18 (0.30)	0.45 (0.30)	1.15 (0.32)	2.19 (0.50)
Positive Cognition	−1.51 (0.15)	−0.88 (0.20)	−0.34 (0.24)	0.26 (0.29)	0.99 (0.42)	2.54 (0.88)
Taiwan						
Negative Affect	−1.95 (0.36)	−0.96 (0.17)	−0.70 (0.15)	0.58 (0.17)	1.71 (0.23)	2.58 (0.79)
Positive Affect	−1.29 (0.15)	−0.31 (0.19)	0.28 (0.26)	1.00 (0.41)	1.94 (0.22)	2.60 (0.47)
Negative Behavior	−2.28 (0.29)	−1.00 (0.90)	0.30 (0.70)	0.69 (0.11)	1.88 (0.31)	2.04 (0.58)
Positive Behavior	−1.79 (0.26)	−0.99 (0.12)	−0.36 (0.12)	0.51 (0.24)	1.74 (0.26)	2.62 (0.86)
Negative Cognition	−2.40 (0.45)	−1.32 (0.14)	−0.43 (0.43)	0.30 (0.63)	1.40 (0.83)	2.21 (0.76)
Positive Cognition	−2.40 (0.27)	−1.40 (0.28)	−0.26 (0.45)	0.61 (0.52)	1.76 (0.52)	2.34 (0.77)

Note. *Model with constant discrimination. See the last paragraph on the procedures for analysis.

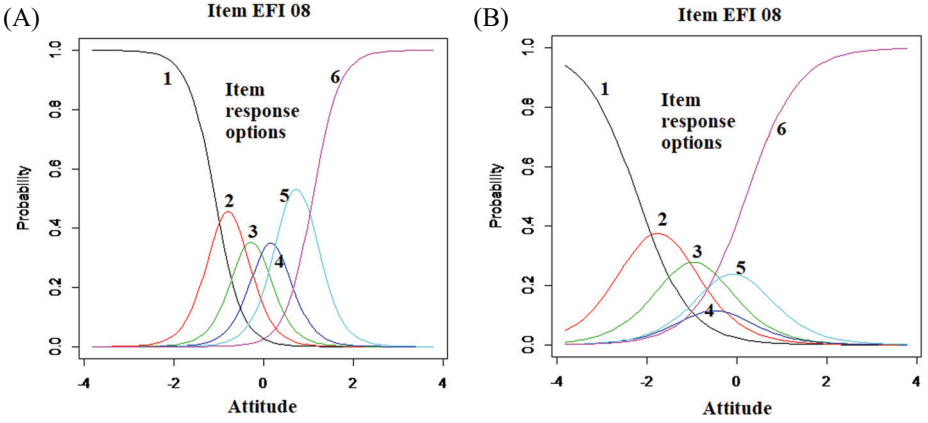

Figure 1. Example of characteristic curves of items. (A) Curve model with response alternatives do not overlap. (B) Curve model with response alternatives do overlap.

appropriateness of the response category. Figure 1 shows examples, (A) is a curve model that would be chosen because its response alternatives do not overlap, unlike curve (B). Finally, the same items should perform with high discriminative power in samples from the other countries.

Table 1 presents the parameters of the five selected items in each subscale from the US American sample. Eighty percent of the items' discriminations were > 3.0. The averages of the fifth response thresholds (b_5), which show the levels of difficulty of items' responses, were always below 1.0 theta, with the exception of positive affect.

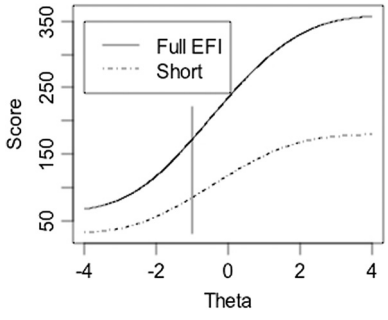

Figure 2. Characteristic curves for the long (60 items) and short (30 items) EFI.

The Selected 30 Items Tested in Samples from Seven Countries

The selected items from the US were tested in samples from the seven countries defined previously. The six-factor model showed an acceptable fit, $\chi^2(388) = 3{,}204.03$; $\chi^2/df = 8.26$; GFI = .90; AGFI = .88; CFI = .95; RMSEA = .056, p(RMSEA < .001) = 90% CI [.054, .057], with correlations between the residues of the items: "Wish him/her well" with "Hope he/she succeeds" and between items "Horrible" with "Dreadful". The constrained model with Measurement weights, Intercepts and Structural covariances among the eight countries investigated showed a statistically significant difference with the free parameter model, indicating that there is no factorial invariance that ensures the comparison between the countries studied, $\Delta\chi\lambda(168) = 974.07$; $p < .001$; $\Delta\chi i(378) = 5{,}485.70$; $p < .001$; $\Delta\chi cov(525) = 6{,}474.68$; $p < .001$).

The analysis followed the same procedures for IRT, but for parsimony and saving space, results will show the means and standard deviations from each subscale per country, which can be compared with the means from the US presented in Table 1.

Table 2 shows the means of discrimination and the response thresholds of the five items of the EFI-30 versions for each of the six subscales in the seven countries. All averages of discrimination were > 2.0, with 33% higher than 3.0. Regarding the difficulty level of items, 90% of the items in Korea and Taiwan had a response threshold of five (b_5) averages between 1.0 and 2.0. This pattern was not found in the other countries in the study, which presented 83% of the fifth response thresholds (b_5) < 1.0. The *SDs* also deserve to be highlighted, because the higher the *SD* then the more the subscale items vary in the level of difficulty, covering a larger amplitude of theta dimension. In this sense, results showed that more than 90% were below 0.7 *SD*.

Table 3 shows the reliability indexes assessed by Cronbach's α of the six subscales of the EFI-30 and Pearson's correlation with the 1-Item Forgiveness scale and Crowne-Marlowe's (1960) Social Desirability scale for content validity.

Finally, the intent of developing the EFI-30 was to answer the call from most users for a shorter scale rather than a longer one. Nevertheless, there will be contexts in which professionals would like to know how well individual respondents would do across assessments or studies comparing long and short forms of scales. The IRT analysis defines and tests characteristic curves for both forms (IRT

Table 3. Reliability and validity of the EFI-30 in eight nations

			Pearson's (*r*)	
Nation	Subscales of the EFI	α	1-item forgiveness	Social desirability
Austria (*N* = 372)	Positive Affect	.88	.760**	.007
	Negative Affect	.85	.743**	.045
	Positive Behavior	.83	.686**	.102*
	Negative Behavior	.88	.717**	.068
	Positive Cognition	.89	.715**	.010
	Negative Cognition	.89	.715**	−.026
Brazil (*N* = 200)	Positive Affect	.92	.599**	−.063
	Negative Affect	.88	.653**	.109
	Positive Behavior	.93	.679**	.099
	Negative Behavior	.87	.600**	.190**
	Positive Cognition	.91	.645**	.027
	Negative Cognition	.87	.601**	.048
Israel (*N* = 176)	Positive Affect	.91	.645**	−.002
	Negative Affect	.82	.591**	−.049
	Positive Behavior	.88	.590**	.000
	Negative Behavior	.85	.594**	−.033
	Positive Cognition	.91	.671**	−.011
	Negative Cognition	.90	.621**	.035
Korea (*N* = 326)	Positive Affect	.89	.593**	−.037
	Negative Affect	.86	.607**	−.011
	Positive Behavior	.91	.595**	−.015
	Negative Behavior	.86	.549**	.109*
	Positive Cognition	.90	.613**	−.085
	Negative Cognition	.87	.512**	−.101
Norway (*N* = 141)	Positive Affect	.92	.634**	–
	Negative Affect	.87	.628**	–
	Positive Behavior	.93	.674**	–
	Negative Behavior	.94	.658**	–
	Positive Cognition	.91	.647**	–
	Negative Cognition	.83	.650**	–
Pakistan (*N* = 404)	Positive Affect	.82	.421**	.012
	Negative Affect	.80	.497**	−.024
	Positive Behavior	.90	.537**	.001
	Negative Behavior	.89	.443**	−.028
	Positive Cognition	.89	.566**	.041
	Negative Cognition	.87	.486**	.049
Taiwan (*N* = 339)	Positive Affect	.88	.171**	.052
	Negative Affect	.88	.176**	.083
	Positive Behavior	.88	.155**	.077
	Negative Behavior	.81	.070	.115*
	Positive Cognition	.84	.152**	.057
	Negative Cognition	.81	.185**	.032
USA (*N* = 392)	Positive Affect	.94	.683**	−.042
	Negative Affect	.92	.689**	−.014
	Positive Behavior	.95	.627**	−.056
	Negative Behavior	.91	.610**	.044
	Positive Cognition	.93	.651**	−.030
	Negative Cognition	.90	.599**	−.022

Note. The social desirability scale was not administered in Norway. *$p < 0.05$; **$p < .001$.

true score equating). Figure 2 shows test characteristic curves for the long (60 items) and short (30 items) EFI and a vertical line that identifies how matching scores across the long and short form can be determined. Results also indicate a linear relationship between scores of the long- and short-EFI. A recommended linear regression coefficient for the long-form (60 items) on the short-form (30 items) to identify matching scores for when one does not have a correspondence table is EFI = 7.333 + 1.933 × Short-EFI. The new version short EFI (30 Items) and their Scoring Key are available for users in ESMs 2 and 3.

Discussion

Results of this study provided the psychometric evidence for the reduced version of the Enright Forgiveness Inventory (EFI-30) with data from eight countries. Table 1 presented the five items selected based on the established criteria. After presenting the choice of items in the first analysis using the US American sample, Table 2 sets out the means of discrimination and the response thresholds of the five items in each of the six subscales. Even with Hambleton et al. (1991) stating that it would be very difficult, or would be an exception, to find values greater than 2.00, this study showed that all means (*a*) were above this value. However, most of the considerations on discrimination focused on dichotomous items and the weighted cut-off points for these parameters need to increase according to the number of options in the scale studied (Baker & Kim, 2017). This is because, with the increase in the number of response options, there is a greater possibility that the respondents distribute their responses among the different options (Cavalcanti et al., 2016). Even with these observations, the discrimination values of the present study appear to be particularly positive. This suggests that the selected items have the sensitivity to differentiate accurately people with different degrees of forgiveness.

Regarding the response thresholds, it was verified that the thresholds were close to the value of −2.00, and considering the standard deviations, it is noticed that the items cover the majority of people with low scores in the subscales. At the other extreme, the results showed that the means of the last thresholds (b_5), for the most part, were < 1.00. This indicates that subscales have more difficulty to evaluating people with higher theta scores. To be more specific, as theta is standardized on the *z* score, it is possible to have an estimate of the percentage of people for whom the scale would have less reliability in the assessment. As the subscales estimate roughly even people with −2.00 theta, this indicates that their responses would be less precise in only about 2% of the people with the lowest degree of forgiveness. At the opposite extreme, the means of the subscales practically did not exceed 1.00 theta. This suggests that it would have less capacity to evaluate, approximately, the 15% of the people with the highest degrees of forgiveness. It is noteworthy for the b_5 thresholds from Taiwan and South Korea that these were the largest among the eight countries. It suggests that the EFI-30 is able to evaluate people with the greatest degree of forgiveness in these two countries. In particular, Taiwan also has presented the b_1 minors, having the highest breadth of evaluation between b_5 and b_1, that is, the EFI-30, implying that Taiwan can assess with more information the people with the greatest differences in their degree of interpersonal forgiveness.

Items on any scale hardly will have the same difficulty and discrimination levels. The more items a scale has, the greater is a chance of finding power to discriminate respondents along the continuous theta. Thus, reducing items on a scale impacts the power to discriminate items' response capacity, as seen in Figure 2. Another point that deserves to be highlighted is that subscales are limited to evaluate respondents with greater attitudes of forgiveness. That is reflected in the fact that the test will classify people with higher forgiveness responses within the same range of scores. In general, researchers using paper-and-pencil scales need to choose in which space of the continuous theta they propose to better discriminate people. It is not possible to have high discriminative power both at low and high scores at the same time. The greater the attempt to discriminate against respondents in a larger theta continuum, then the lower the slope of the characteristic curves of the items and the test will be.

The EFI scale has been extensively used in research and counseling. For research, the consequences of working with a short-version are mainly linked to the reduction of variability in people with a high score of forgiveness, which may have an impact on the decrease in covariance with other variables in a study. In relation to the field of counseling, it is necessary to understand that the EFI is relevant because there is the possibility that clients may be having difficulty positioning themselves in their responses for forgiveness, so the scale is particularly useful when it accurately assesses people with low degrees of forgiveness. However, perhaps a way to be able to deliver the smallest number of items, without losing discriminatory capacity in specific parts of the theta, is to use adaptive testing by computer so that the computer with a large bank of items delivers specific items to each person.

Results for internal reliability (Cronbach's α) and Pearson's correlations between the six subscales with the 1-Item Forgiveness scale and Crowne-Marlowe's (1960) Social Desirability scale for concurrent validity showed results in the expected direction in all countries with data available. The EFI-30 has valid six subscales with strong

internal consistency per subscale and positive correlations with the 1-item forgiveness scale and no correlation with the social desirability scale. A few exceptions occurred, such as the Taiwanese sample showing no correlation between positive behavior and 1-item forgiveness, and the sample from Austria showing a positive correlation between positive behavior and social desirability. The samples from Brazil, South Korea, and Taiwan showed positive correlations between negative behavior and social desirability. Note that Taiwan and South Korea showed the highest thresholds of item discrimination. As previously stated, the EFI-30 is able to evaluate people with the greatest degree of forgiveness in these two countries and these particular findings focused on the behavioral component of responses that is the closest researchers can get to observed behaviors in real life. Participants from these countries are probably informing that to reduce the degree of avoiding the other who offended unjustly and to start showing some friendship is desirable because particular cultural social norms apply to their behavioral expression of forgiveness. In all cases, that is a question for future research on the cultural sensitivity of the EFI across nations.

Mental health professionals and researchers in human development now have a reliable measure in its short- or long-forms to work within aiding clients to assess their degree of forgiveness toward a person or persons who acted unfairly toward that person. The EFI-30 is available in ESM 2 here and free for use. Researchers and counselors should keep in mind that forgiveness implies a Self in equilibrium. For a person to reduce, for example, negative judgments about their offender is important but not enough if social interactions with them raise anxiety. A decrease in negative feelings, which sometimes happen to let off some steam reducing internal pressure, is no indication of effective progress in the process if it happens unaccompanied by changes in the other stances. Therefore, counselors and researchers should interpret patterns of changes in the six stances of forgiveness to have a better understanding of the process.

Considering that more research is necessary to verify similarities and differences for the forgiveness process in therapy, education, human development, and cross-cultural validation among countries, this work has furnished the field with a parsimonious and valid tool for investigating people's interpersonal forgiveness across a wide variety of cultures.

Final Considerations

The results are satisfactory because they provide accurate alternatives for researchers and mental health professionals applying a scale of the degree of forgiveness in different countries. However, it is necessary to take into account some limitations of this study. Data analysis did not consider age, degree of hurt, if the offender was alive, how close victims were to offenders, socioeconomic level, educational background, peculiarities of cultures, and the reports of hurt among other relevant variables to the study of forgiveness. Therefore, subsequent studies that replicate the estimates may consider these specific variables. The EFI-30 should be tested against other measures of forgiveness for concurrent and divergent validity. However, those comparisons should preserve some level of similarity between operational definitions behind the assessment models, otherwise, researchers will be comparing different phenomena. If measures are translated across cultures, the process of translation and validation also should be observed to provide sounds results. Finally, the EFI-30 now is available with this article providing the opportunity for researchers from all nations to begin addressing further these limitations. The authors should be contacted for versions of the EFI-30 in future researchers' particular languages or to obtain manuals for translations and administration according to their specific research questions.

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1027/1015-5759/a000649

ESM 1. Inputs and outputs for the analyses conducted
ESM 2. EFI-30 scale
ESM 3. EFI-30 answer key

References

Akhtar, S., & Barlow, J. (2018). Forgiveness therapy for the promotion of mental well-being: A systematic review and meta-analysis. *Trauma, Violence, and Abuse, 19*(1), 107–122.

Andrade, T. F. (2014). *Refinamento Psicométrico da Escala de Atitudes para o Perdão – EFI* [Psychometric Refinement of the Attitude for Forgiveness Scale – EFI] (Unpublished master's thesis). Universidade Federal da Paraíba.

Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R*. Springer Nature. https://doi.org/10.1007/978-3-319-54205-8

Baskin, T., & Enright, R. D. (2004). Intervention studies on forgiveness: A meta-analysis. *Journal of Counseling and Development, 82*(1), 79–82. https://doi.org/10.1002/j.1556-6678.2004.tb00288.x

Cavalcanti, T. M., De Melo, R. L. P., De Medeiros, E. D., Santos, L. C. D. O., & Gouveia, V. V. (2016). Escala de Avaliação da Fadiga: funcionamento diferencial dos itens em regiões brasileiras [Fatigue Assessment Scale: Differential item functioning in Brazilian regions]. *Avaliação Psicológica, 15*(1), 105–113. https://doi.org/10.15689/ap.2016.1501.11

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349–354. https://doi.org/10.1037/h0047358

Drasgow, F., Levine, M. V., Tsien, S., Williams, B., & Mead, A. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement, 19*(2), 143–165. https://doi.org/10.1177/014662169501900203

Enright, R. D., & Fitzgibbons, R. P. (2015). *Forgiveness therapy: An empirical guide for resolving anger and restoring hope*. American Psychological Association.

Enright, R. D., Freedman, S., & Rique, J. (1998). The psychology of interpersonal forgiveness. In R. D. Enright & J. North (Eds.), *Exploring Forgiveness* (pp. 46–62). The University of Wisconsin Press.

Enright, R. D., & Rique, J. (2004). *Manual for the Enright Forgiveness Inventory*. Mind Garden.

Enright, R. D., The Human Development Study Group. (1991). The moral development of forgiveness. In W. Kurtiness & J. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 123–152). Erlbaum.

Freedman, S. R., & Enright, R. D. (1996). Forgiveness as an intervention goal with incest survivors. *Journal of Consulting and Clinical Psychology, 64*(5), 983–992. https://doi.org/10.1037/0022-006X.64.5.983

Freedman, S., & Enright, R. D. (2020). Review of the empirical research using Enright's process model of interpersonal forgiveness. In E. L. Worthington Jr. & N. G. Wade (Eds.), *Handbook of forgiveness* (2nd ed., pp. 266–276). Routledge.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.

Hanke, K., & Fisher, R. (2012). Socioeconomical and political correlates of interpersonal forgiveness: A three level meta-analysis of the Enright Forgiveness Inventory across 13 societies. *International Journal of Psychology, 48*(4), 514–526. https//doi.org/10.1080/00207594.2011.651086

Lee, Y., & Enright, R. D. (2019). A meta-analysis of the association between forgiveness of others and physical health. *Psychology & Health, 34*(5), 626–643. https://doi.org/10.1080/08870446.2018.1554185

Marôco, J. (2014). *Análise de equações estruturais: Fundamentos teóricos, software e aplicações (2a edição)* [Analysis of structural equations: theoretical foundations, software and applications (2nd ed.)]. Reportnumber.

McElroy-Heltzel, S., Davis, D. E., Ordaz, A. C., Griffin, B. J., & Hook, J. N. (2020). Measuring forgiveness and self-forgiveness: Descriptions, psychometric support, and recommendations for research and practice. In E. L. Worthington Jr. & N. G. Wade (Eds.), *Handbook of forgiveness* (2nd ed., pp. 74–84). Routledge.

Reed, G., & Enright, R. D. (2006). The effects of forgiveness therapy on depression, anxiety, and post-traumatic stress for women after spousal emotional abuse. *Journal of Consulting and Clinical Psychology, 74*(5), 920–929. https://doi.org/10.1037/0022-006X.74.5.920

Rique, J., & Camino, C. (2010). O perdão interpessoal em relação a variáveis psicossociais e demográficas [Interpersonal forgiveness in relation to psychosocial and demographic variables]. *Psychology: Research and Review, 23*(3), 525–532. https://doi.org/10.1590/S0102-79722010000300013

Rique Neto, J., Camino, C. P. S., Santos, W. S., & Gouveia, V. V. (2009). Analises confirmatória da escala de atitudes para o perdão – EFI [Confirmatory analysis of the Enright Forgiveness Inventory – EFI]. *Avaliação Psicológica, 8*(2), 169–178.

Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software, 17*(5), 1–25.

Rizopoulos, D. (2015). *ltm: Latent trait models under IRT*. (Version 1.0-0). http://rwiki.sciviews.org/doku.php?id=packages:cran:ltm

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores* (Psychometric Monograph No. 17) Psychometric Society. http://www.psychometrika.org/journal/online/MN17.pdf

Subkoviak, M. J., Enright, R. D., Wu, C., Gassin, E., Freedman, S., Olson, L., & Sarinopoulos, I. (1995). Measuring interpersonal forgiveness in late adolescence and middle adulthood. *Journal of Adolescence, 18*(6), 641–655. https://doi.org/10.1006/jado.1995.1045

Waltman, M. A., Russell, D. C., Coyle, C. T., Enright, R. D., Holter, A. C., & Swoboda, C. (2009). The effects of a forgiveness intervention on patients with coronary artery disease. *Psychology and Health, 24*, 11–27. https://doi.org/10.1080/08870440801975127

Worthington, E. L. Jr., & Wade, N. G. (2020). *Handbook of forgiveness* (2nd ed.). Routledge.

History
Received May 21, 2020
Revision received January 27, 2021
Accepted February 8, 2021
Published online July 5, 2021
EJPA Section / Category Clinical Psychology

Acknowledgment
We wish to thank all who participated in data collection in all of the nations included in this research.

Open Science
We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact *p* values, effect sizes, and 95% confidence or credible intervals.
Open Data: The information needed to reproduce all of the reported results is accessible in ESM 1.
Open Materials: The information needed to reproduce all of the reported methodology is accessible in ESM 2 and 3.
Preregistration of Studies and Analysis Plans: This study was not preregistered.

Funding
Authors acknowledge the financial support received from CAPES and CNPQ, Brazil, for the work developed in this article.

ORCID
Julio Rique
https://orcid.org/0000-0001-7383-5111

Julio Rique
Department of Psychology
Federal University of Paraiba
Joao Pessoa, PB
Campus I
Cidade Universitária, CCHLA, 58051-900
Brazil
julio.rique@hotmail.com