DOCUMENT RESUME

ED 430 014 TM 029 727

AUTHOR	Kim, Ahyoung; Lim, Eun-Young
TITLE	How Critical Is Back Translation in Cross-Cultural Adaptation of Attitude Measures?
PUB DATE	1999-04-00
NOTE	38p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150) -- Tests/Questionnaires (160)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Attitude Measures; *Cross Cultural Studies; *Elementary School Students; Foreign Countries; Intermediate Grades; *Test Construction; Test Format; *Translation
IDENTIFIERS	*South Korea

ABSTRACT

The purpose of this study is to compare the effectiveness of three types of practices applied in Korea in enhancing the validity and equivalency of test instruments when cross-cultural adaptation of attitude measures is necessary. The three types of practices are: (1) translation and review (translation version); (2) translation, back translation, and review (back translation version); and (3) translation, back translation, review, and empirical validation study (validation version). The focus was on the relative effectiveness of back translation applied to the construction of Korean versions of instruments. Participants were 734 fifth graders from 3 public elementary schools in Seoul (Korea). Responses on the three test versions and two other motivation scales were collected within a 3-week period at approximately 1-week intervals. Results show that the back translation version is superior to the translation version in terms of its similarity to the validation version and construct-related evidence. However, results from item-response theory analyses reveal that the quality of the translated items is similar. The nature of adapted attitude scales is discussed. Appendixes contain the Academic Failure Tolerance Scale (M. Clifford 1988, 19991) and two back translation versions. (Contains 13 tables and 33 references.) (Author/SLD)

How Critical is Back Translation in Cross-Cultural Adaptation of Attitude Measures?

Ahyoung Kim
&
Eun-Young Lim
Ewha Womans University
Seoul, Korea

Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada. April, 1999.

How critical is back translation procedure in cross-cultural adaptation of attitude measures?

Abstract

The purpose of the present study is to compare the effectiveness of three types of practices applied in Korea in enhancing the validity and equivalency of test instruments when cross-cultural adaptation of attitude measures is neccessary. The three types of practices are: (1) translation and review (Translation version); (2) translation, back translation, and review (Back translation version); (3) translation, back translation, review, and empirical validation study (Validation version). The present authors are particularly interested in the relative effectiveness of back translation as it is applied to the construction of Korean versions of instruments. Seven hundred and thirty four 5th graders from three public elementary schools in Seoul, Korea participated in this study. Reponses on the three test versions and two other motivation scales were collected within a 3 week period with approximately one week intervals during last October. Results show that the back translation version is superior to the translation version in terms of its similarity to the validation version and construct-related evidence. However, results from IRT analysis reveal that the quality of the translated items are similar. Discussions are provided in terms of the nature of adapted attitude scales.

Key words: back translation, cross-cultural test adaptation, graded response model, IRT, Korean, MULTILOG, psychological equivalence

Introduction

When one investigates certain human characteristics by adopting a theory that has been developed and tested in a foreign language and culture, replication of the findings and confirmation of applicability of the theory to his or her own culture are due procedures. These procedures also provide an expansion of the universality and generalizability of the theory. Therefore, researchers investigating cultural differences in human psychological traits, especially in the affective domain, need to have equivalent research materials including psychological testing instruments for measuring the traits in all involved cultures. Consequently, researchers should adapt the instrument written in the original researcher's language. An appropriate adaptation procedure is required to secure psychological equivalency between the original (source) and target language versions of the instrument.

The validity of psychological test adaptation has long been an issue for cross-cultural researchers (e.g., Cattell, 1970; Eysenck, & Eysenck, 1983; Geisinger, 1994; Hambleton, 1993). To the extent that the adaptation is valid, acceptance of the research findings in that culture is judged valid. Because of this reason, numerous attempts have been made all around the world to improve the equivalency and validity of cross-cultural test adaptation [e.g., Cheung (1985) in Hong Kong; Manos (1985) in Greece; Savasir & Erol (1990) in Turkey]. To the present authors' knowledge, insufficient effort has been made to improve the validity and equivalency of instruments used in Korean cross-cultural test adaptation practice.

Theory and Methods of Cross-Cultural Test Adaptation

Psychological Equivalence

Berry and Dasen (1974) have pointed out that there are three aspects of psychological equivalence that should be taken into consideration when cross-cultural adaptation is neccessary: These are functional, conceptual, and metric equivalencies. Some researchers (Butcher & Garcia, 1978; Butcher & Han, 1996) proposed scalar equivalence in addition to the three aspects.

Functional equivalence. Functional equivalence exists when certain behaviors that the instrument attempts to represent function identically in all involved cultures. For example, "when personality characteristics measured by one scale are highly related to those measured by another scale in a different culture, it can be said that these two scales, thought manifestly different, are functionally equivalent across cultures (Butcher & Han, 1996, p. 45)." Statistical analysis techniques, such as factor analysis and intercorrelation pattern analysis are applied to assess functional equivalence between scales (Butcher & Han, 1996). When the functional equivalence can be considered to be present, then securing conceptual equivalence is the next concern.

Conceptual equivalence. When there are semantic similarities between the words, conceptual equivalence is considered to be present. *Translation*, *back translation*, and small group discussion for *review* have been adopted to ensure conceptual or linguistic equivalence of source and target language versions (Brislin, 1971; Hulin, 1987). Back translation in particular has been identified as an effective procedure to secure conceptual equivalence.

Metric equivalence. Metric equivalence can be acquired when the instrument is validly adapted. Various statistical analyses have been proposed to ensure metric equivalence, such as: computation of intercorrelation among subcomponents, examination of point-biserial correlation between item responses, and the total scale score between the different language versions of the scales. Differences in item-total correlations are assumed to reflect psychometric differences introduced by the translation from the source to the target language.

Scalar Equivalence. Along with the above mentioned three types of equivalence, scalar equivalence has been proposed by some researchers (e.g., Butcher & Garcia, 1978; Butcher & Han, 1996). Scalar equivalence is said to be established when the two instruments measure certain characteristics with the same degree, intensity, or magnitude. Thus, mean score similarity is not sufficient to demonstrate scalar equivalence of two instruments. Butcher and Han (1996) illustrates that the scalar equivalence has been established when two persons who have MMPI T scores of 75 on the social subscale are socially introverted to approximately the same degree. However, scalar equivalence is the most difficult one to establish among the four types, and only indirect approaches have been provided.

Statistical Methods

Factor Analysis. The most commonly applied statistical analysis to confirm the underlying factor structures of the source and target language versions of a scale is factor analysis. If two scales are representing the same traits, the factor structure obtained from the analyses of two response sets will be similar. Commonly used methods of factor structure comparison are examination of factor congruence coefficients, factor score correlation, and maximum likelihood confirmatory factor analysis [see Butcher & Han (1996) for details].

Item Response Theory. While factor analysis techniques do not allow individual item comparisons, IRT method provides assessment of the similarity of invariant individual item characteristics across samples (Butcher & Han, 1996; Bontempo, 1993). Differences in the item characteristic curve(ICC) indicate that the two items are not equivalent. Thus, such items will produce nonequivalent scales. IRT can be used to ensure translation adequacy. Securing high-fidelity translations from source to target language is essential to ensuring metric equivalence in the two versions.

As Hulin (1987) noted, metric equivalence is determined by the equivalence of responses to two different versions. If two versions of an item elicit equal probabilities of a specified response from individuals at the same level of the trait assessed by the item, metric equivalence of the two items is supported (Hulin, 1987). On this ground, cross-cultural test adaptation researchers have acknowledged the effectiveness of IRT-based techniques in ensuring the quality and equivalence of test items between the source and target language versions (e.g., Candell & Hulin, 1986; Ellis, Becker, & Kimmel, 1993; Drasgow, 1984; Hulin, Drasgow, & Komoar, 1982). These researchers claim that the classical test theory-based item analysis techniques can not achieve psychometric equivalence between the target and source language versions because of the sample-specific nature of item difficulties and discriminations.

Since traditional IRT method presumes dichotomous response items, other response scales such as rating scale measures have often been treated as dichotomous ones, which raised serious limitations in the adoption of the IRT method to affective scales. But this problem has been solved with the development of a graded response model which can handle polytomous responses obtained from multiple choice or Likert-type items (Samejima, 1969; Tissen, 1992).

As Butcher and Han (1996) noted, it is difficult to distinguish and establish the four types of equivalence separately. Thus, it is proposed that cross-cultural test adaptation researchers should first improve an instrument by proper translation techniques, and then establish conceptual equivalence and functional equivalence by constructing nomological network or by factor analysis, followed by application of IRT or regression methods to test item/metric equivalence and scalar equivalence (Hui & Triandis, 1985 cited in Butcher & Han, 1996).

Back Translation

Back translation involves, first, the process of translating the translated target language version back to the source language by a bilingual person. The back translated version is then compared with the original version in terms of general meaning of the sentences, complexity levels, forms, semantic similarity of words, and grammatical structures. Items which don't match the original version are retranslated, back translated and compared again. Multiple iterations are recommended to produce equivalence between the two language versions. A small group of bilinguals are involved in the translation, back translation, and review discussion process for item correction. Functional and conceptual equivalence are tested and secured via psychometric procedures. In this sense, rigorous procedure of translation of the original into target language version is fundamental prior condition for achieving the validity and equivalence of the two.

Korean Adaptation Practice

For valid test adaptation, it is proposed to follow all of the above mentioned procedures through empirical research (Butcher & Han, 1996; Geisinger, 1994). Nevertheless, few Korean cross-cultural test adaptation researchers have applied the recommended procedures adequately. In Korea, it is observed that four different practices have been attempted in cross-cultural test adaptation. These practices are based on either a partial procedure or the whole procedure that has been proposed by the researchers, such as Bracken and Barona (1991), Butcher (1985), Geisinger (1994), and Hambleton and Kanjee (1993) and others. The four types of practices applied in Korea will be described below.

The adaptation procedure starts with the translation of the original scale into a new language version. Thus, the *first* and simplest way of adapting the original scale is to translate the original version into Korean and use it without any further validation. The *second* and most commonly used practice is to translate the scale, then set up a small review committee which edits or revises the translated items carefully to ensure correct understanding and content validity of the instrument. In some instances, if certain items are not appropriate in Korean culture, those are eliminated. The *third* practice is that, after first translation, back translation procedure is adopted. Items for which the original version and back translated version do not match are subjected to another translation by the first translator (this procedure is called double back translation), or sent to a review committee to be edited or revised as was mentioned in the first type of practice above, i.e., without any double back translation. The *fourth* and the most desirable practice is that, after both second and third practice procedures are completed, empirical validation study is conducted. That is, after back translation and editing and revising items, a test is assembled and administered to a sample from the target population. Item analysis and factor analysis are conducted to select good items, and the factor structure and other validity evidences are examined to ensure equivalency to the original instrument.

Purpose of the Present Study

In the present study, we are concerned with the relative effectiveness of the second and third types of practice for the following reasons: (1) The second practice is the most commonly used in Korea and some researchers (e.g., Hambleton, 1993) claimed that back translation did not significantly improve the validity of the translated version in many empirical studies; (2) nevertheless, some researchers (e.g., American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1985; Butcher, 1985) contend that back translation enhances the validity of cross-cultural test adaptation; (3) the simplest first practice is least recommended. We are going to use the fourth type as the criterion in examining the relative effectiveness of the two types.

We will judge the differential effectiveness in enhancing equivalence and validity of the two types of procedure by comparing the similarity of the translated and the back translated versions to the validated version in the following aspects: (1) a general tendency of subjects' response, (2) the total and subscales reliability coefficients, (3) patterns of item-total correlations, (4) factor structures, (5) patterns of intercorrelation among factors, (6) patterns of relationships with external variables, such as other motivation variables like general self-efficacy and locus of control that have been included in the previous studies, and (7) item parameters estimated via IRT method.

Methods

Subjects

The subjects used in the present study were 734 5th graders attending three typical public elementary schools in a middle class residential area of metropolitan Seoul, Korea. Intact classrooms were the unit of sampling. Data from 711(357 males, 354 females) students' were used in the final analysis. Data from 10 students were excluded due to the incompleteness of the responses in three repeated administrations of three versions of the scales used in this study.

Instruments

To examine the effects of test adaptation practices, this study used Margaret M. Clifford's Academic Failure Tolerance Scale (Clifford, 1988, 1991, hereafter AFT) as the original test instrument (Appendix 1). The AFT was developed as an academic motivation measure that assesses students' reactions following failure experience. The AFT consists of 27 6-point(1: strongly disagree to 6: strongly agree) Likert-type scale items with three 9-item subscales, each measuring preferred task difficulty, feelings following failure, and behavior following failure. High scores represent positive attitude following failure. Technical properties, such as validity and reliability, of the original instrument were already reported from US samples (Clifford, 1988, 1991) and the original AFT has been adapted into Korean version. The Korean version of AFT (K-AFT) scale is one of few available instruments for measuring attitude, which has applied a valid adaptation procedure which includes translation, double back translation, review, and empirical validation studies (Kim, 1993, 1994, 1997).

The results from the two validation studies for K-AFT were relatively satisfactory to conclude the equivalency to American AFT (Kim, 1994; 1997).

Reliability of the subscales, factor structures and loadings, patterns of intercorrelation among subscales, the predictability in academic achievement, developmental trend, and gender differences among subscales were all quite similar to the original version (Kim, 1994). In addition to these two studies, Item analysis via polytomous IRT technique also shows that K-AFT is a fairly good test for measuring academic failure tolerance (Seong, 1998). Upon completion of the full adaptation procedure and validation studies, the K-AFT resulted in 24 items while the original AFT had 27 items.

The ***Instrument*** used in the present study was based on Clifford's 1991 AFT scale. Excluding 3 items that were eliminated in K-AFT, the remaining 24 corresponding AFT items were translated and reviewed, composing the first set (translation version: T, hereafter). This first set items was back translated (Appendix 2). Back translated items were compared with the original English items and 10 out of 24 items didn't sufficiently converge with the original meanings. These items were then revised, back translated (Appendix 3), and revised again. These 10 items were merged with the remaining items which resulted in the second set (back translation version: BT, hereafter). Translation and back translation was done by 2 college graduates independently. Translation was done by a Korean who lived for 7 years and received B.A. degree in business in the US. Back translation was done by a Korean bilingual who lived for 15 years and received B.A. degree in English in the US. The review group consisted of 4 psychology majors in a Korean graduate school. The third set items were from K-AFT scale (validation version: V, hereafter).

Since comparison among the three procedural types was our purpose, repeated reponses to all three sets from all participants were required. Items from each version were scrambled with items of 2 other scales (Korean General Self-efficacy Scale: K-GS; Korean Locus of Control Scale: K-LC). The Korean General Self-efficacy Scale (24 Likert-type items) was developed and modified by Kim and Cha (Kim & Cha, 1996; Kim, 1997), and Korean Locus of Control Scale

(16 Likert-type items), developed by Clifford (1988), has been adapted by Kim (1996, 1997). These two scales were used as criterion variables to test concurrent and construct validity as was done in Kim's validation study (Kim, 1997).

Procedure

Subjects received three forms of test booklets, each of them consisting of 48, 40, and 24 items, respectively. To eliminate order effects of the administration sequence of the three adaptation versions, Latin-square design was employed by counterbalancing three administration sequences to each of the three groups. Each administration sequence consisted of three alternative forms which contained three versions. For effective use of test administration, items of K-GS and K-LC were included in two of the three administrations (Table 1 shows the content and order of the administerd test booklets).

<insert Table 1 about here>

Test administrations were repeated three times to intact classrooms by homeroom teachers in a manner similar to standardized testing situations. There was at least a one-week separation between the three sessions for all repeated administrations. Instructions were read aloud and explained by the teachers and sample items were answered together following teachers' request for sincere response. Average testing time was 15 to 20 minutes depending on the test booklets. As is shown in Table 1, to eliminate school effect, all three forms of the test booklets were distributed to the classes of all three schools.

Analyses

The scrambled items were sorted to restore the original scale sets, representing T, BT, V, K-GS, and K-LC. Since V can be assumed to be valid and equivalent to the original AFT, comparisons were to be made between the

1st and 3rd sets and the 2nd and 3rd sets.

Differences were examined as follows: Basic descriptive statistics, item-total correlations, and reliability indices were compared. Factor analysis was conducted and factor structures and loadings were examined and compared. Item qualities were examined using item parameters estimated from graded response model (Samejima, 1969; Tissen, 1991). For the comparison of the pertinent construct-related validity evidence, correlational analysis was conducted and the patterns of interrelationship among subscale scores, general self-efficacy scale scores, locus of control scale scores were compared. Statistical Analyses System (SAS Institute Inc., 1996) and Multilog 6.0 (Tissen, 1991) programs were used for statistical analyses.

Results and Discussion

Response Tendency

Preliminary analyses of the subjects' responses to individual items showed that the responses for each item were normally distributed and that the means and the score variabilities of the total scale and the feeling subscale (Feel), preferred task difficulty subscale (PD), and behavior subscale (Beh) of the three versions (T; BT; V) were similar. The score variabilities of all the scales were similar to the results of antecedent studies (Kim, 1994; 1996). However, while the means of Feel in the three versions were somewhat higher in the present study than in the Kim's 1996 data, the means of the Beh subscales were somewhat lower in the present study. Since the subjects of Kim's 1996 study were from 6 representative regional strata in Korea and the subjects of the present study were from one of such strata, this discrepancy can be interpreted as group difference.

Since sex differences were not our primary concern, the data was not analyzed separately. Table 2 shows basic descriptive statistics of the total and subscales of the three versions and those from Kim's 1996 data.

<insert Table 2 about here>

Correlations among the Three Versions in All Scales

Table 3 shows the correlations among the three versions in the total scale and the subscales. As can be seen in Table 3, the patterns of correlations among three versions are quite similar in all the total and the subscales. To be specific, the correlations between V and any of the other two versions are virtually the same for each scale. However, the correlations between T and BT are consistently lower than the correlation between V and any of the other versions. This reveals that the relationship between T and BT is the least among the possible correlations between any pair of the three versions. However, we can say that the three correlations between any pair of the three versions are large enough to support or extract one superordinate method factor. This suggests that the three versions can be treated as alternative measures for each other.

<insert Table 3 about here>

Reliability and Item-total Correlations

The α coefficients for internal consistency were obtained to assess the reliability of the total and subscales in the three versions. Although α coefficients of Beh in T and BT are .64 and .69 which are not very high, α coefficients of all other scales are satisfactory for attitude measures, ranging from .73 to .84. In PD and Beh, V and BT show reliability better than T. However, T shows the highest reliability in the Feel subscale.

The similarity in the patterns of item-total correlation among the three versions was examined. Table 4 shows the item-total correlations and changes of α when the given item is removed from the scale for each subscale in the three versions. For the Feel subscale, only 1 item of BT has item-total correlation lower than .30. For the PD subscale, 2 items of T have item-total correlation lower than .30. For the Beh subscale, 2 of T, 3 of BT, and 1 of V have this pattern. In summary, V has less poor items than the other two versions, but BT turned out to be no better than T in regard to the quality of items.

<insert Table 4 about here>

Factor Structures

Factor analysis was performed to compare the underlying factor structures of the three versions. As was done in the previous studies (Clifford, 1988; Kim, 1994), the common factor model (method=prinit, priors=SMC, nfactor=3 in SAS PROC FACTOR) with varimax rotation was estimated. Results are given in Tables 5, 6, and 7.

<insert Tables 5, 6, 7 about here>

In terms of the size of explained common variance, V and BT are virtually the same, ordered as PD(36%), Feel(34%), and Beh(30%, 29%). However, T shows quite a different pattern from the other two versions: Feel factor takes the largest portion(39%) of explained common variance, PD factor the least(29%), and Beh factor the medium(32%). It seems that BT is closer to V than T is.

For T, 4 items are less interpretable. For BT, 1 item originally from PD seems to be a better indicator of the Beh factor. Other than that all the other items are consistent with V. With respect to the quality of items indicating the factors, T is the worst, while BT and V perform similarly and are better than T.

Factor loadings of items on the three factors in the three versions were compared. Items are rearranged by the size of factor loadings in the validation version. Factor loadings and their ranks of corresponding items of the other two versions are also presented (Table 8). If the three versions are equivalent, the ranks of the factor loadings of the three versions should coincide. Spearman's rank-order correlation coefficients between each pair of versions for each subscale were computed. Rank-order correlation coefficients between T and V, and BT and V are .81 and .76 in the Feel factor, respectively; these coefficients are .55 and .95 in the PD factors and .86 and .92 in the Beh factor, respectively. According to these results, BT is more similar to V in their factor loading pattern than the T in the PD and Beh factors, but not in the Feel factor.

<insert Table 8 about here>

Intercorrelations between Three Versions and External Variables

It is recommended to examine the relationship between focal variables and external criterion variables in assessing the validity of the focal variables. In the present study we use K-GS and K-LC as the external variables which are expected to have a certain degree of correlation with the three subscales. The relations in each subscale and both K-GS and K-LC have been studied earlier by the first author (Kim, 1996; 1997). The correlations are given in Table 9.

<insert Table 9 about here>

In Table 9, we present the result from Kim's data as evidence of convergent validity for the validation version. The results from Kim's data and V are very similar. We then compared the similarity of T and BT to V. Regarding the Feel subscale, no version shows a significant correlation with K-LC and all the versions show significant correlation with K-GS. Judging from the size of correlation between both T and BT, and V, BT is more similar to V than T is. Regarding the PD subscale, all the versions have significant correlations with the two external variables. BT is less similar to V than T is in its correlation with K-LC. However, BT is more similar to V than T is in its correlation with K-GS.

Regarding the Beh subscale, BT is more similar to V than T is in its correlation with both K-LC and K-GS. All in all, the BT shows more similarity to V than T does, yielding additional evidence favoring for BT over T.

Item Response Theory

Since the factor analysis shows 3 distinct subscale factors as expected, we applied IRT to analyze each subscale. Items of each subscales were analyzed with Multilog program. For each subscale, items from the three versions were entered simultaneously in the model to estimate the item parameters and test information function.

Parameter estimation. Item parameters for the three versions of Feel, Beh, PD are shown in Tables 10, 11, and 12, respectively. Items were judged by the discrimination parameter (a) and location parameters of boundary characteristics curve (b_k). Tables show these parameters for the 8 items in the three versions of the 3 subscales.

<Insert Tables 10, 11, 12>

Items with high discrimination power and equally spreaded range of category boundary span are judged to be good (Baker, 1992). Baker suggested that the item discrimination parameter estimates could be judged according to the following criteria: a below .65 is low; from .65 to 1.34 is appropriate; from 1.35 to 1.69 is high; above 1.70 is very high. The attribute (attitude trait) of the person being measured by the test (θ) is usually arbitrarily placed on a z-score scale, thus in practice, ranges roughly from -3.0 to +3.0. Therefore, Items that have location parameters within this range and have approximately equal intervals between b_k's are judged to be good.

An examination of the quality of the items using item parameter estimates reveals that 9 items of T, 9 items of BT, and 5 items of V have unrealistic b_k values (below -3.0 and over +3.0) and that 4 items of T, 3 items of BT, and 1 item of V have *a* lower than .65. Overall, 4 items (#10, #12, #20, & #23) of T, 3 items (#2, #17, & #20) of BT, and 1 item (#20) of V have both low *a* and unrealistic value of b_k's. These results show that BT is slightly better than or similar to T in their item qualities, and V is better than the other two.

Test Information Function. Table 13 shows the test information function for the subscales of the three versions. The test information function values are generally similar across the attribute levels(θ) of −1.0 to 1.5 in Feel, −1.5 to 1.5 in PD, −2.0 to 2.0 in Beh, showing that the Beh subscale provides similar information over the widest range. Regarding the Feel subscale, T shows the most information and BT the least. However, V shows the best information for the PD and Beh subscales. BT shows more information than T for the PD subscale, but the reverse is observed for the Beh subscale.

< Insert Table 13 about here >

From the overall results based on the IRT analyses, we can conclude that item quality of V is definitely superior to the other two versions and BT is not particularly superior to T in its item quality.

Conclusions

The purpose of the present study is to assess the relative effectiveness of back translation procedure in the cross-cultural test adaptation, particularly in the measurement of affective characteristics. Prevalent practice of ignoring proper adaptation procedure in Korea would bring about adverse effects on the generalization of certain theories originated from different cultures. Although numerous international studies have provided accumulated evidences that back translation is an essential technique of ensuring psychological equivalence between source and target language versions (Brislin, 1970; Butcher, 1993; Thorndike, 1974), cross-culturally adapted Korean instruments rarely report such practices. In this respect, this paper attempted to emphasize the importance of back translation procedure for securing psychological equivalence and provided empirical evidences which were supportive to its goal.

The results of the present study show that the back translation version is more similar to the validation version in the pattern of intercorrelation among subscales, of factor structure, and of its relations with external variables. However, the similarity in the response tendency, item-total correlations, and the item quality are not particularly in favor of the effectiveness of back translation. This result can be understood from the fact that the complexity level of the meaning and sentences used in the AFT is very simple and clear. As Thorndike noted, "maintaining comparability under translation becomes a progressively more serious problem as the material to be translated becomes more difficult (Thorndike, 1974, p. 9)," which implies that the relative efficiency of back translation procedure may vary with the nature of the sentences used. The material used in the present study was not complex enough to reveal the problem of misunderstanding caused by inaccurate translation. The similarity of response tendency and item quality support this interpretation. The item quality assessed by IRT suggests that all three versions can be judged to be an acceptable measure of academic failure tolerance, evidencing the scalar equivalence.

However, an adoption of back translation procedure enhances construct-related validity which results in conceptual and metric equivalences. Especially, the factor similarity of BT to V is more salient than that of T to V. In addition, the more equivalent relations with the two external variables support this contention.

All in all, as was evidenced by Brislin's early work, back translation procedure can confirm the quality of translator and translation (Brislin, 1970), which leads to functional, conceptual, metric, and even scalar equivalence between the source and target language versions. With the results of the present study, we can strongly recommend the use of back translation in the cross-cultural test adaptation. It is suggested that future research should be conducted in Korea using more abstract and complex psychological instruments which are used in the assessment of personality and in clinical settings. However, it should be noted that the consistent superiority of the validation version in terms of its reliability, factor structure clarity, and item quality confirms the importance of a proper validation procedure.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). *Standards for educational and psychological testing.* Washington D. C.: American Psychological Association.

Baker, F. B. (1992). *Item Response Theory: Parameter Estimation Techniques.* New York: Marcel Dekker, Inc.

Berry, J. W., & Dasen, P. (Eds.) (1974). *Culture and cognition.* London: Methuen.

Bontempo, R. (1993). Translation fidelity of psychological scales: An item response theory analysis of an individualism-collectivism scale. *Journal of Cross-Cultural Psychology, 24.* 149-166.

Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12*, 119-132.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*, 185-216.

Butcher, J. N. (1985). Current developments in MMPI use: An international perspective. In J. N. Butcher & C. D. Spielberger(Eds.), *Advances in personality assessment,* (Vol. 4, 83-94). Hillsdale, NJ: Erlbaum.

Butcher, J. N., & Garcia, R. (1978). Cross-national application of psychological tests *Personnel and Guidance, 56,* 472-475.

Butcher, J. N., & Han, K. (1996). Methods of establishing cross-cultural equivalence. In J. N. Butcher (Eds.), *International adaptations of the MMPI-2* (pp.44-63). Minneapolis, Mn.: University of Minnesota Press.

Candell, G. L., & Hulin, C. I. (1986). Cross-language and cross-cultural comparisons in scale translation: Independent sources of information about item nonequivalence, *Journal of Cross-Cultural Psychology*, *17*, 417-440.

Cattell, R. B. (1970). The isopodic and equipotent principles for comparing factor scores across different populations. *The British Journal of Mathematical and Statistical Psychology*, *23*, 23-41.

Cheung, F. M. (1985). Cross-cultural considerations for the translation and adaptation of the Chinese MMPI in Hong Kong. In J. N. Butcher & C. D.

Spielberger, (Eds.), *Advanced in personality assessment,* (Vol 4, pp.131-158) Hillsdale, NJ: Erlbaum.

Clifford, M. M. (1988). Failure Tolerance and risk-taking in ten- to twelve-year-old students. *British Journal of Educational Psycholo*gy, *58*, 15-27.

Clifford, M. M. (1991). Risk-taking: Theoretical, empirical, and educational considerations. *Educational Psychologist, 26*, 263-297.

Drasgow, F. (1984). Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variable are the central issues. *Psychological Bulletin, 95.* 134-135.

Ellis, B. B., Becker, P., & Kimmel, H. D. (1993). An item response theory evaluation of an english version of the Trier personality inventory (TPI), *Journal of Cross-Cultural Psychology*, *24*, 133-148.

Eysenck, H. J., & Eysenck, S. B. G. (1983). Recent advances in the cross-cultural study of personality. In J. N. Butcher & C. D. Spielberger(Eds.), *Advances in personality assessment,* (Vol. 2, pp. 41-69). Hillsdale, NJ: Erlbaum.

Geisinger, K F. (1994). Cross-cultural normative assessment: Translation and adaption issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*, 304-312.

Hambleton, R. K. (1993). Technical standards for translating tests and establishing test score equivalence. In Symposium conducted at the 1001st Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Hambleton, R. K., & Kanjee, A. (1993). Enhancing the validity of cross-cultural studies: Improvements in instrument translation methods. Paper presented at the Annual Meeting of American Educational Research Association. Atlanta, Georgia.

Hulin, C. I. (1987) A psychometric theory of evaluations of item and scale translations: Fidelity across languages, *Journal of Cross-Cultural Psychology, 18.* 115-142.

Hulin, C. I., Drasgow, F., & Komoar, J. (1982). Applications of item response theory to analysis of attitude scale translations. *Journal of Applied Psychology, 17.* 417-440.

Kim, A. (1993). Failure tolerance in Korean grade school students: Development and sex differences. Paper presented at the Annual Meeting of American Educational Research Association. Atlanta, Georgia.

Kim, A. (1994). Development of Korean version of the academic failure tolerance scale. *Korean Journal of Educational Research*. *32*, 59-75. (Korean).

Kim, A. (1996). Data base '96.

Kim, A. (1997). A study on the academic failure tolerance and its correlates. *Korean Journal of Educational Psychology, 11*, 1-19. (Korean).

Kim, A., & Cha, J. (1996). General self-efficacy and its measurement. Paper presented at the Annual Meeting of Korean Industrial and Organizational Psychological Association, Seoul, Korea. (Korean).

Manos, N. (1985). Adaption of the MMPI in Greece: Translation, standardization, and cross-cultural comparison. *Advances in personality assessment,* (Vol, 4, pp.159-208). Hillsdale, NJ: Erlbaum.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* No. 17.

Savasir, I., & Erol, N. (1990). The Turkish MMPI: Translation, standardization and validation. *Advances in personality assessment* (Vol. 8, pp.49-62). Hillsdale, NJ: Erlbaum.

Seong, T. J. (1998). Item parameter estimation of the Korean Academic Failure Tolerance Scale and person's trait estimation via the grade response model. *Korean Journal of Educational Psychology, 12.* 203-218. (Korean).

Thorndike, R. L. (1974). Methodological problems in developing instruments for cross-national studies. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Chicago, Illinois, April 1975.

Tissen, D. (1991). *MULTILOG Version 6.0 user's guide.* Chicago, Illinois: Scientific Software, Inc.

Table Captions

<Table 1> Counterbalanced Content and Order of Test Adminstration Sequences

<Table 2> Means and Standard Deviations of Total and Subscale Scores of the Three Versions

<Table 3> Intercorrelations Among 3 Versions in Total Scale and Subscales

<Table 4> Item-Total Correlations of 3 Versions of 3 Subscales

<Table 5> Factor Analysis Result for T-version

<Table 6> Factor Analysis Result for BT-version

<Table 7> Factor Analysis Result for V-version

<Table 8> Rank Order of Factor Loadings of 3 Versions in Each Subscale

<Table 9> Correlations with External Variables

<Table 10 > Estimated Item Parameters of the 3 versions of FEEL

<Table 11> Estimated Item Parameters of the 3 versions of PD

<Table 12> Estimated Item Parameters of the 3 versions of Beh

<Table 13 > Test Information Functions of the Subscales

<Table 1> Counterbalanced Content and Order of Test Adminstration Sequences

group / order	Group A (7 classes)	Group B (6 classes)	Group C (6 classes)
1st administ.	**Booklet A1 (48 items)** T-version (24) + K-SG scale (24)	**Booklet B1 (24 items)** BT-version	**Booklet C1 (40 items)** V-version (24) + K-LC scale (16)
2nd administ.	**Booklet A2 (40 items)** BT-version (24) + K-LC scale (16)	**Booklet B2 (48 items)** V-version (24) + K-GS scale (24)	**Booklet C2 (24 items)** T-version
3rd administ.	**Booklet A3 (24 items)** V-version	**Booklet B3 (40 items)** T-version (24) + K-LC scale (16)	**Booklet C3 (48 items)** BT-version (24) + K-GS scale (24)

Note: All three groups included three different schools. To avoid confusion, we marked on each envelope to indicate which class should go on which day.

<Table 2> Means and Standard Deviations of Total and Subscale Scores of the Three Versions

N = 711

		version	Mean	SD
Total		T-version	3.44	.68
		BT-version	3.50	.66
		V-version	3.47	.69
		Kim data*	3.44	.73
	Feel	T-version	3.23	1.15
		BT-version	3.12	1.03
		V-version	3.35	1.11
		Kim data	2.96	1.00
	PD	T-version	3.31	.94
		BT-version	3.53	.99
		V-version	3.31	1.01
		Kim data	3.31	1.14
	Beh	T-version	3.77	.74
		BT-version	3.87	.78
		V-version	3.76	.84
		Kim data	4.06	.97

* N = 856 for Kim's data

<Table 3> Intercorrelations Among 3 Versions in Total Scale and Subscales

		1	2	3	4	5	6	7	8	9	10	11	12
TOT	1. T-vers	1.00											
	2. BT-vers	.76**	1.00										
	3. V-vers	.81**	.80**	1.00									
FEEL	4. T-vers	.71**	.50**	.54**	1.00								
	5. BT-vers	.49**	.62**	.48**	.74**	1.00							
	6. V-vers	.54**	.48**	.63**	.78**	.76**	1.00						
PD	7. T-vers	.74**	.63**	.65**	.16**	.10*	.11*	1.00					
	8. BT-vers	.58**	.77**	.64**	.12*	.10	.09	.76**	1.00				
	9. V-vers	.63**	.65**	.77**	.14**	.08	.09	.78**	.77**	1.00			
BEH	10. T-vers	.70**	.50**	.56**	.18**	.06	.10*	.50**	.44**	.50**	1.00		
	11. BT-vers	.48**	.66**	.53**	.07	.05	.03	.46**	.50**	.51**	.60**	1.00	
	12. V-vers	.47**	.51**	.64**	.02	.00	.01	.48**	.50**	.52**	.65**	.65**	1.00

* p<.01 ** p<.001 (N=711)

<Table 4> Item-Total Correlations of 3 Versions of 3 Subscales

		T-version α = .84		BT-version α = .80		V-version α = .82	
	Item No.	Item-total correlations	α changed*	Item-total correlations	α changed	Item-total correlations	α changed
FEEL	1	.663	.810	.592	.772	.685	.781
	2	.426	.839	.271	.815	.424	.817
	3	.625	.814	.563	.776	.572	.797
	4	.552	.824	.531	.781	.472	.811
	5	.660	.809	.597	.770	.584	.796
	6	.643	.812	.694	.755	.621	.790
	7	.357	.846	.312	.812	.396	.821
	8	.647	.812	.588	.772	.588	.796
		α = .78		α = .84		α = .84	
PD	9	.576	.738	.693	.808	.697	.803
	10	.158	.803	.549	.826	.472	.831
	11	.594	.732	.609	.818	.602	.815
	12	.282	.782	.377	.846	.487	.829
	13	.591	.733	.668	.810	.622	.812
	14	.607	.731	.608	.818	.569	.819
	15	.569	.737	.509	.830	.541	.823
	16	.469	.755	.578	.822	.564	.819
		α = .64		α = .69		α = .73	
BEH	17	.367	.605	.240	.690	.415	.700
	18	.471	.580	.402	.655	.526	.678
	19	.305	.622	.289	.682	.329	.719
	20	-.086	.724	.255	.694	.126	.758
	21	.545	.556	.576	.611	.620	.657
	22	.476	.578	.532	.626	.549	.675
	23	.258	.634	.313	.675	.367	.711
	24	.500	.572	.497	.635	.486	.687

<Table 5> Factor Analysis Result for T-version

Item No.	factor 1 Feel	factor 2 Beh	factor 3 PD
T1	***.755***	.032	.017
T8	***.736***	.065	.068
T5	***.723***	.021	.043
T6	***.715***	.032	-.014
T3	***.686***	-.144	.107
T4	***.598***	-.109	.120
T2	***.464***	.113	.028
T7	***.394***	.227	.389
* **T20**	.167	-.156	-.084
T21	-.064	***.733***	.140
T24	-.087	***.691***	.119
T22	-.067	***.624***	.170
T18	.058	***.608***	.159
T17	.267	***.351***	.256
T19	.140	***.337***	.134
* **T12**	-.128	.296	.239
* **T23**	.110	.267	.147
T14	-.025	.318	***.664***
T11	-.057	.349	***.611***
T9	.041	.316	***.591***
T13	-.021	.398	***.572***
T16	.175	.164	***.557***
T15	.050	.356	***.557***
* **T10**	.079	-.093	.264
eigen value	3.556	2.942	2.623
% of variance	39	32	29

* less interpretable items that shows loading value lower than .30.

<Table 6> Factor Analysis Result for BT-version

Item No.	factor 1 Feel	factor 2 Beh	factor 3 PD
B9	***.743***	-.001	.253
B13	***.726***	-.022	.256
B16	***.640***	.117	.193
B14	***.630***	.018	.205
B10	***.587***	.117	.178
B11	***.551***	-.031	.411
B15	***.442***	-.014	.345
B6	.053	***.777***	.072
B5	.025	***.669***	.049
B1	-.007	***.660***	.076
B8	.060	***.651***	.110
B3	.099	***.639***	-.149
B4	.008	***.582***	-.039
B7	.305	***.343***	.291
B2	-.050	***.321***	-.076
B21	.236	-.064	***.699***
B22	.247	-.068	***.651***
B24	.244	-.106	***.611***
B18	.185	-.101	***.488***
# **B12**	.295	-.072	***.368***
B23	.190	.104	***.350***
B19	.181	.058	***.317***
B20	.029	.264	***.311***
B17	.098	.156	***.304***
eigen value	3.218	3.053	2.740
% of variance	36	34	30

items that seem to be an indicator of other factors than originally expected.

BEST COPY AVAILABLE

28

<Table 7> Factor Analysis Result for V-version

Item No.	factor 1 Feel	factor 2 Beh	factor 3 PD
V9	***.702***	.011	.322
V16	***.662***	.104	.163
V13	***.661***	-.007	.251
V14	***.625***	-.040	.268
V11	***.596***	-.028	.283
V10	***.547***	.124	.088
V15	***.506***	.014	.322
V12	***.456***	.033	.277
V1	.013	***.781***	-.009
V6	-.042	***.714***	.035
V3	-.019	***.661***	-.154
V8	.055	***.656***	.034
V5	-.044	***.645***	-.080
V4	.086	***.498***	.007
V2	-.014	***.475***	.026
V7	.257	***.428***	.193
* **V20**	.141	.183	.049
V21	.228	-.043	***.756***
V22	.181	-.042	***.663***
V18	.277	.006	***.611***
V24	.265	-.118	***.564***
V23	.173	.038	***.403***
V17	.321	.178	***.389***
V19	.206	.068	***.304***
eigen value	3.383	3.179	2.716
% of variance	36	34	29

* uninterpretable item

BEST COPY AVAILABLE

29

<Table 8> Rank Order of Factor Loadings of 3 Versions in Each Subscale

FEEL				PD				BEH			
Item #	V-vers (rank)	BT-vers (rank)	T-vers (rank)	Item #	V-vers (rank)	BT-vers (rank)	T-vers (rank)	Item #	V-vers (rank)	BT-vers (rank)	T-vers (rank)
1	.781(1)	.660(3)	.755(1)	9	.702(1)	.743(1)	.591(3)	21	.756(1)	.699(1)	.733(1)
6	.724(2)	.777(1)	715(4)	16	.662(2)	.640(3)	.557(5)	22	.663(2)	.651(2)	.624(3)
3	.661(3)	.639(5)	.686(5)	13	.661(3)	.726(2)	.572(4)	18	.611(3)	.488(4)	.608(4)
8	.656(4)	.651(4)	.736(2)	14	.625(4)	.630(4)	.664(1)	24	.564(4)	.611(3)	691(2)
5	.645(5)	.669(2)	.732(3)	11	.596(5)	.551(6)	.611(2)	23	.404(5)	.350(5)	.267(7)
4	.498(6)	.582(6)	.598(6)	10	.547(6)	.587(5)	.264(7)	17	.389(6)	.304(7)	.351(5)
2	.475(7)	.321(8)	.464(7)	15	.506(7)	.442(7)	.557(5)	19	.305(7)	.317(6)	.337(6)
7	.428(8)	.343(7)	.394(8)	12	.456(8)	.295(8)	.239(8)	20	.049(8)	.304(7)	-.156(8)

BEST COPY AVAILABLE

<Table 9> Correlations with External Variables

N = 711

		K-LC	K-GS
Feel	T-version	-.05	.34*
	BT-version	-.08	.27*
	V-version	-.07	.29*
	Kim data#	-.06	.20*
PD	T-version	.41*	.61*
	BT-version	.38*	.58*
	V-version	.44*	.59*
	Kim data	.44*	.61*
Beh	T-version	.49*	.59*
	BT-version	.43*	.54*
	V-version	.46*	.49*
	Kim data	.45*	.53*

* p<.001

Kim's 1996 data from 856 5th graders.

<Table 10 > Estimated Item Parameters of the 3 versions of FEEL

Item	a	attitude trait levels (b_k)				
		1	2	3	4	5
T1	1.46	-1.14	-.25	.70	1.25	2.22
BT1	1.43	-1.08	-.01	1.05	1.63	2.51
V1	1.99	-1.00	-.21	.56	.96	1.62
T2	.86	-2.49	-1.10	.26	1.02	2.35
* BT2	**.54**	**-4.26**	**-2.23**	**-.10**	**1.59**	**4.29**
V2	.87	-2.45	-.97	.14	.96	2.44
T3	1.67	-1.06	-.18	.63	1.06	1.77
BT3	1.34	-.71	.30	1.33	1.80	2.69
V3	1.42	-1.23	-.19	.66	1.10	2.15
T4	1.33	-1.26	-.13	.80	1.28	2.16
BT4	1.17	-1.22	.02	1.00	1.63	2.71
V4	1.03	-1.37	-.17	.96	1.52	2.65
T5	1.63	-1.23	-.42	.28	.70	1.34
BT5	1.66	-1.25	-.46	.31	.74	1.35
V5	1.62	-1.52	-.69	.05	.51	1.28
T6	1.79	-.95	-.06	.62	.97	1.68
BT6	1.95	-.81	.06	.81	1.31	2.11
V6	1.65	-1.15	-.29	.50	.90	1.82
T7	.65	-2.27	-.67	.95	1.97	3.87
BT7	.67	-3.04	-1.22	.43	1.33	3.09
V7	.82	-2.47	-1.21	.06	.90	2.43
T8	1.61	-1.59	-.59	.28	.79	1.70
BT8	1.40	-1.52	-.61	.37	1.01	1.97
V8	1.37	-1.82	-.83	.26	.92	1.96

* poor quality item

BEST COPY AVAILABLE

<Table 11> Estimated Item Parameters of the 3 versions of PD

Item	a	attitude trait levels (b_k)				
		1	2	3	4	5
T9	1.45	-1.82	-.80	.28	1.26	2.22
BT9	1.89	-1.64	-.69	.20	1.27	2.20
V9	1.91	-1.44	-.48	.49	1.46	2.47
* T10	.30	-1.82	2.34	4.99	6.46	9.03
BT10	1.18	-1.91	-.59	.52	1.44	2.66
V10	.97	-1.55	-.18	1.03	1.94	3.45
T11	1.56	-1.85	-.97	-.16	.62	1.53
BT11	1.42	-2.16	-1.29	-.43	.70	1.78
V11	1.50	-1.95	-.88	.13	1.08	2.15
* T12	.51	-6.59	-4.80	-2.99	-.59	1.93
BT12	.70	-4.83	-3.34	-1.75	-.10	1.82
V12	.97	-3.08	-1.68	-.49	.82	2.19
T13	1.47	-1.71	-.76	-.07	.82	1.78
BT13	1.84	-1.77	-.84	-.01	.98	2.11
V13	1.73	-1.58	-.81	-.03	1.02	2.08
T14	1.72	-1.67	-.79	.12	1.02	1.88
BT14	1.53	-1.95	-1.02	.03	1.15	2.28
V14	1.59	-1.70	-.76	.21	1.35	2.24
T15	1.32	-2.07	-.92	-.01	.90	1.85
BT15	.99	-3.72	-2.25	-1.03	.36	1.68
V15	1.24	-2.58	-1.38	-.57	.43	1.62
T16	1.11	-1.17	.06	1.42	2.37	3.27
BT16	1.30	-1.35	-.39	.66	1.40	2.72
V16	1.23	-1.67	-.84	.31	1.33	2.55

* poor quality items

<Table 12> Estimated Item Parameters of the 3 versions of Beh

Item	a	attitude trait levels (b_k)				
		1	2	3	4	5
T17	.78	-4.57	-2.87	-1.30	.08	1.96
* BT17	.49	-7.20	-4.99	-2.66	-.62	2.27
V17	.93	-3.72	-2.38	-.98	.02	1.98
T18	1.15	-3.31	-2.27	-1.21	.24	1.85
BT18	1.01	-3.42	-2.27	-.97	.77	2.51
V18	1.34	-2.33	-1.47	-.56	.61	1.94
T19	.78	-2.79	-.90	.34	1.80	3.35
BT19	.80	-2.60	-.94	.29	1..71	3.31
V19	.75	-2.60	-1.10	.25	1.77	3.53
* T20	.21	-6.87	-.53	3.54	6.97	11.41
* BT20	.44	-5.07	-2.12	-.27	1.35	3.77
* V20	.39	-6.47	-2.18	.38	2.41	5.27
T21	1.88	-2.21	-1.30	-.54	.57	1.64
BT21	1.75	-2.13	-1.32	-.57	.50	1.65
V21	2.00	-1.82	-1.08	-.29	.61	1.81
T22	1.33	-2.59	-1.31	-.29	1.15	2.49
BT22	1.51	-2.47	-1.27	-.30	1.00	2.36
V22	1.64	-2.26	-1.26	-.36	.88	2.23
* T23	.62	-4.65	-2.80	-.71	1.04	2.93
BT23	.69	-4.63	-2.97	-1.10	.48	2.73
V23	.79	-3.00	-1.18	.10	1.18	2.99
T24	1.62	-2.64	-1.68	-.78	.43	1.73
BT24	1.53	-2.67	-1.63	-.60	.68	2.03
V24	1.36	-2.90	-1.87	-.99	.05	1.30

* poor quality items

<Table 13 > Test Information Functions of the Subscales

scale	version	θ -2.0	-1.5	-1.0	-.5	0	.5	1.0	1.5	2.0
FEEL	T	2.91	4.05	4.77	5.04	5.14	5.21	5.17	4.88	4.16
	BT	2.35	3.32	4.07	4.41	4.53	4.56	4.57	4.40	4.04
	V	3.01	4.02	4.65	4.87	4.95	4.97	4.93	4.64	3.95
PD	T	3.36	3.83	4.00	4.05	4.06	4.07	4.04	3.90	3.50
	BT	4.14	4.70	4.87	4.91	4.88	4.83	4.83	4.74	4.52
	V	4.02	4.76	5.01	5.06	5.05	5.02	4.99	4.94	4.70
BEH	T	3.29	3.38	3.37	3.33	3.26	3.26	3.22	3.15	2.82
	BT	3.07	3.17	3.18	3.14	3.08	3.07	3.03	2.98	2.78
	V	3.62	3.87	3.91	3.89	3.83	3.77	3.67	3.55	3.28

<Appendix 1 > Clifford's "Academic Failure Tolerance Scale"

FEEL	1. I feel terrible when I make a mistake in school.
	2. If I do poorly in my school work, I try not to let anyone know.
	3. A low mark in my school work makes me feel very sad.
	4. I worry a lot about making errors in my school work.
	5. I feel like hiding whenever I get a bad mark in school.
	6. If I make lots of mistakes in school, I feel very moody or angry.
	7. I really dislike school work on which I make mistakes.
	8. If I give a wrong answer to teacher's question, I feel terrible.
PD	9. I like to do school work that is difficult for me.
	10. I would rather work problems I can do in a hurry than those that take much time and thought.
	11. I like to try difficult assignments even if I get some wrong.
	12. School work that really makes me think is fun.
	13. I would rather study a difficult course than a very easy one.
	14. If I could chose my math problems, I would pick hard ones rather than very easy ones.
	15. It is fun to try to answer questions that are difficult or challenging.
	16. The easier school work is for me, the more I like it.
BEH	17. If I can't succeed at a new school task, I give up quickly.
	18. When I make mistake in my school work, I just keep trying and trying.
	19. If I do not understand something, I ask the teacher to explain.
	20. I would rather guess at something and get it wrong than ask a question that may sound silly.
	21. If I get a low grade in my school work, I study my errors and rework the problems I get wrong.
	22. I usually study and correct the errors I makes on school work, even if I don't have to.
	23. I don't like to set goals for my school work. I just do the work and forget about it.
	24. If I get a low score, I usually make up my mind to buckle down and study hard.

<Appendix 2 > First Back Translation Version

1. I get very upset when I make a mistake at school.
2. If I do poorly in a subject, I try not to let anyone know.
3. When I get a low mark I feel really sad.
4. I worry a lot about making mistakes at school.
5. When I get a low mark in a subject, I want to hide
6. If I make a lot of mistakes at school, I get really depressed or angry.
7. I really hate assignments in which I make mistakes.
8. If I answer a teacher's question incorrectly, I feel really bad.

9. I feel that I want to do difficult assignments.

* 10. I would do the short questions before the questions which require more time and thought.

* 11. I want to answer difficult homework questions even if I might get them wrong.

12. Assignment which make me think are enjoyable

13. I would rather study a difficult subject than a really easy one

* 14. If I could choose my own math problems, I would choose the hard ones rather than easy ones

* 15. It's fun to try to solve problems that are difficult or hard to attempt

16. When an assignment is easier I like it better.

* 17. I give up easily when I can't continue a new school assignment.

* 18. If I make a mistake in school I keep at it

19. If there is something that I don't understand, I ask the instructor to explain.

* 20. I'd rather think through something on my own than ask a stupid question.

* 21. When I get a low mark in a subject, I study my mistakes and re-do the problems I got incorrect.

22. Even when it's not necessary I usually study and correct the mistakes I've made in a subject.

* 23. I don't like to set goals for myself in my studies. I just study and try to forget.

* 24. When I get a low score I just pick myself up and study harder.

* items that show discrepancy between the original and translated versions.

<Appendix 3 > Second Back Translation Items

10. I prefer sticking to problems I can do quickly to problems which require a lot of time and thought.
11. Even if I may do it incorrectly, I want to have difficult homework.
14 .If I could only choose my own math problems I would pick tough ones rather than plain ones
15. It's fun to try to answer difficult or challenging questions.
17. New school work gets abandoned if I can't continue.
18. I just keep trying even when I make mistake in a subject.
20. Rather than risk sounding silly by asking a question, I would just think through it alone and get it wrong.
21. If I get a low score in a subject I study the mistakes I made and review the problems I got wrong.
23. I don't like setting academic goals I just study and forgot it.
24. If I get a low score I usually redirect myself and study hard

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029727

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: How critical is back translation in cross-cultural adaptation of attitude measures?

Author(s): Ahyoung Kim & Eun-Young Lim

Corporate Source: Ewha Womans University, Seoul, Korea

Publication Date: April, 1999

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[✓]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Ahyoung Kim	Printed Name/Position/Title: Ahyoung Kim / Assist. Prof.
Organization/Address: Ewha Womans Univ., Seoul, Korea	Telephone: 822-3277-2720 FAX: 822-3277-2638
	E-Mail Address: aykim@mm.ewha.ac.kr Date: 4/12/'99

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:	Ahyoung Kim
Address:	Dept. of Psychology, Ewha Womans Univ. Seodaemun-gu, Seoul, 120-750 Korea
Price:	

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:	
Address:	

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com