別紙様式３（第３条関係）

論　文　要　旨

氏　名　トッド・トーネ

論文題目（外国語の場合は、和訳を併記すること。）

Adaptation, Analysis and Critique of Four Psychometric Instruments into the Japanese Context to Measure University Students' Causal Perceptions for Success and Failure in Second Language Acquisition (SLA)

第二言語習得の成功と失敗に対する大学生の因果的認識を測定するための４つの心理測定用具の日本のコンテクストへの適用、分析及び批評

論文要旨（別様に記載すること。）

（注）１．論文要旨は、A4版とする。
２．和文の場合は、４０００字から８０００字程度、外国語の場合は、２０００語から４０００語程度とする。
３．「論文要旨」は、ＣＤ等の電子媒体（１枚）を併せて提出すること。
（氏名及びソフト名を記入したラベルを張付すること。）

Adaptation, Analysis and Critique of Four Psychometric Instruments into the Japanese Context to Measure University Students' Causal Perceptions for Success and Failure in Second Language Acquisition (SLA)

1.0 Introduction

Attribution theory (Weiner, 1979; 1985, 2010; Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971) has partially re-emerged as an important research trajectory owing to its migration into the field of applied linguistics (Gobel & Mori, 2007; Hsieh & Kang, 2010; Hsieh & Schallert, 2008; Peacock, 2010) where it is finding increasing attention. The rationale for this migration was based in part, on a need to understand students' perceptions for success and/or failure in second language acquisition (SLA) so that appropriate intervention could be provided (Banks & Woolfson, 2008; Graham, 1997; Stipek & Weisz, 1981). The theoretical components of attribution theory used in the area of SLA, and inherited from the general educational research trajectory, include the following four potential causal attributions; ability, effort, luck, and task difficulty. Weiner identified these causal attributions as being the typical attributions for achievement-related outcomes. However, the main concern that surrounds this issue is the development and use of appropriate instrumentation that can accurately measure these causal attributions and produce valid scores.

The research reported in this study contributes to the endeavor of establishing adequate, evidence-based instrumentation for the research area within SLA, by adapting the Causal Dimension Scale II (CDS II) developed by McAuley, Duncan and Russell (1992), the Critical Incident Attribution Measure (CIAM) developed by Vispoel and Austin (1995), the Sydney Attribution Scale (SAS) originally developed by Marsh, Cairns, Relich, Barnes, and Debus (1984) and revised by Marsh (1984), and the Survey of Achievement Responsibility (SOAR) developed by Ryckman and Rallo (1983) into the Japanese context. The psychometrics of scores for each of the instruments were examined following the prerequisite of appropriate attention to measurement outlined by Wilkinson and the Task Force on Statistical Inference (1999) and by following the guidelines of the International Test Commission (Hambleton, Merenda, & Spielberger, 2005). This study is conducted with a view to satisfying, specifically, the deficit with respect to the way in which some psychometric instruments have migrated across domains into the area of applied linguistics, and, more generally and positively, to contribute to a sound psychometric footing for attribution theory within applied linguists at an early stage in the emergence of this research trajectory.

2.0 Literature Review

Attribution theory has a relatively long history within educational research. The early work began in the 1950s, and there was significant evolution in theory and constructs until the 1970s when it is arguable that the typical constructs employed in theory reached something close to their final state. For a more complete review of this evolution refer to Weiner (1979; 1985). In brief summary of the final state of constructs employed in the area, four constructs comprise the foundation of attribution theory and these are locus of causality, stability, personal control and external control. These causal attribution constructs are represented in the general line of instrumentation as ability (locus of causality), effort (personal control), luck (external control), and task ease/difficulty (stability). These perceived causes, if measurable, are useful in determining future success and/or failure, and in determining the motivation level for a future task (Heider, 1958; Weiner et al., 1971).

3.0 Methodology

The data obtained from the instruments was placed in a Microsoft Office Access 2010 database. IBM/Statistical Package for the Social Sciences (SPSS) software (Version 19.0) was used to determine descriptive statistics and the reliability estimates (Cronbach's alphas) for the scores. AMOS (Version 5.0.1) was used to conduct a Confirmatory factor analysis (CFA). A CFA was conducted in addition to establishing the Cronbach's alpha because only a CFA can determine the unidimensionality of scales (Gerbing & Anderson, 1988).

The original instruments were all developed in English. Consequently, forward and back translation methods were used following the guidelines of the ITC (Hambleton et al., 2005) as the first step in adapting them for use in the Japanese population. The forward translation was performed by a near-native speaker of English and the back translation was performed by a different near-native speaker of English. Both translators had some training and experience in test construction. The back-translated version was then compared to the original English version and a few notable contradictions emerged. The two near-native translators were consulted and modifications were made to accommodate the discrepancies. The Japanese version of the instrument was then deemed suitable in terms of both language equivalency and cultural context, and ready to be empirically tested in the field.

3.1 Participants

The total number of participants in this study was 1885 Japanese university students.

This number was derived based on two important considerations: each instrument requires a minimum sample size using the 10:1 ratio recommended by Byrne (2001) and Kline (2011) for structural equation modeling (SEM), and as a counter measure against missing data that may occur in large studies of this kind. However, the missing values in three of the datasets were not systematically missing, and therefore deletion of cases was not judged to have systematically altered the properties of the sample.

The respondents who volunteered for this study were asked to complete the adapted version of the instruments and provide the following background information; age, major, academic year, and the date of administration. All responses were completely anonymous because no identifying information was collected and informed consent was given by simply filling out the questionnaire. The time required to complete the instruments was about 15 minutes.

3.1.1 Causal Dimension Scale II

The dataset for this instrument came from 213 SLA students at a university in western Japan studying within the fields of science (n = 42), engineering (n = 92), English (n = 49), literature (n = 19), and medicine (n =11). Age ranged from 18 years through 26. There were 125 males and 88 females. There were no missing values.

3.1.2 Critical Incident Attribution Measure

The dataset for this instrument came from 579 SLA students at four universities in western Japan whose major field of study included English (n = 42), welfare (n = 47), science (n = 45), education (n = 78), law (n = 83), engineering (n = 130), medicine (n = 135), and business (n = 19). However, due to missing values on some response forms, 43 records were deleted from the database and the statistical analyses were performed on the dataset for 536 participants. Age ranged from 18 years through 30 years.

3.1.3 Sydney Attribution Scale

The dataset for this instrument came from 439 SLA students at four universities in western Japan whose major field of study included science (n = 150), education (n = 73), law (n = 71), engineering (n = 77), medicine (n = 9), business (n = 34), and communication (n = 25). However, due to missing values on some response forms, 25 records were deleted from the database and the statistical analyses were performed on the dataset for 414 participants. Age ranged from 18 to 23 years and there were 226 males and 189 females.

3.1.4 Survey of Achievement Responsibility

The dataset for this instrument came from 654 SLA students studying at four different universities in western Japan in the fields of English (n=23), social welfare (n=15), science (n=31), education (n=112), law (n=29), engineering (n=258), medicine (n=85), business (n=30), communication (n=48), and Japanese (n=23). A number of missing values were identified and 90 records were deleted as a result. The statistical analysis was based on the data from the remaining 569 respondents. Age ranged from 18 to 24 years. There were 311 males and 253 females (5 participants did not indicate their gender) present in the final sample.

4.0 Results

The results for the CDS II, CIAM, SAS, and SOAR instruments are presented in three sections. The first section covers item normality (skew & kurtosis). The second section deals with Cronbach's alpha. The third section reports on the results a CFA.

The procedure for evaluating skew and kurtosis was to determine the critical ratio which is calculated by dividing the unstandardized value for skew and kurtosis by the standard error, and then comparing the value for this computation against a criterion of 3.0 (stipulated in advance).

This study adopted Hu and Bentler's (1999) recommended fit indexes. They were the root mean square error of approximation (RMSEA), the standardized root mean squared residual (SRMSR), the Tucker-Lewis index (TLI), and the comparative fit index (CFI).

4.1 Causal Dimension Scale II

4.1.1 Normality

Four items presented values that were skewed and none of the items were kurtotic at the 3.0 threshold.

Mardia's coefficient (Mardia, 1985) indicated multivariate nonnormality. The critical ratio was 11.37. This lead to the use of the Bollen-Stine bootstrap procedure to cope with this nonnormality and to assist with adjudicating model fit in the CFA.

4.1.2 Cronbach's Alpha

For all four hypothesized scales the lower-bound for the 95% confidence level fell below Nunnally and Bernstein's (1994) criterion of .70 for scale reliability. The derived value for alpha fell below the threshold on all hypothesized scales except Personal Control which produced a value for alpha of .74. However, Cortina (1993) and Green, Lissitz and Muliak (1977) point out that alpha is biased by the number of items on a

scale, with larger numbers of items producing higher alphas. Only three items comprise each subscale on the CDS II which is comparatively low, and in fact close to the minimum for measurement of latents; and therefore, there is a compelling argument for Nunnally and Bernstein's (1994) criterion of .70 to be relaxed in the interpretation of values for alpha in the case of the CDS II.

4.1.3 Confirmatory Factor Analysis

The model comprised 78 sample moments, 30 free parameters and 48 degrees of freedom meeting the criterion for overidentification.

The values derived in this study were as follows (Hu and Bentler's cutoffs are in parentheses): RMSEA .05 (< .06), SRMSR .06 (< .08), TLI .92 (> .95), and CFI .94 (> .95).

The Bollen-Stine bootstrap procedure was adopted as a further analytical tool in dealing with the multivariate nonnormality of the data. One thousand samples were extracted in the bootstrap procedure. The model fit better in 941 of these samples and worse in 59. The resulting Bollen-Stine bootstrap p value was p =.060. This result was not significant at either the .01 or .05 level. In terms of the logic of CFA, this means the model is accepted or that the model fits the data.

4.2 Critical Incident Attribution Measure

4.2.1 Normality

Ten items presented values that were skewed and 6 of the items were kurtotic at the 3.0 threshold for the failure outcome, and 9 items presented values that were skewed and 4 of the items were kurtotic at the 3.0 threshold for the success outcome.

4.2.2 Cronbach's Alpha

All of the eight hypothesized subscales for both the failure and success outcomes produced alphas above Nunnally and Bernstein's (1994) criterion of .70 for scale reliability. The lower bound of the 95% confidence interval (.68) for the strategy subscale in the failure outcome fell slightly below the threshold of .70.

4.2.3 Confirmatory Factor Analysis

The model had 300 distinct sample moments, 76 parameters, and 224 degrees of freedom. This met the criterion for overidentification for a direct test.

The values derived in this study for the hypothesized model on the failure outcome were as follows (Hu and Bentler's cutoffs are in parentheses): RMSEA .06 (< .06),

SRMSR .06 (< .08), TLI .91 (> .95), and CFI .93 (> .95). The values derived for the success outcome were as follows: RMSEA .05 (< .06), SRMSR .03 (< .08), TLI .96 (> .95), and CFI .97 (> .95). The TLI value and the CFI value for the failure outcome fell slightly below the recommended thresholds. All of the values derived for the success outcome were within the recommended thresholds.

4.3 Sydney Attribution Scale

4.3.1 Normality

Thirteen items presented values that were skewed and 3 of the items were kurtotic at the 3.0 threshold.

4.3.2 Cronbach's Alpha

Two of the three hypothesized subscales for a successful outcome (Ability and Effort) produced alphas above Nunnally and Bernstein's (1994) criterion of .70 for scale reliability. However, the External subscale produced an alpha of .67. The upper bound of the 95% confidence interval (.72) for this subscale fell slightly above the threshold of .70. On the other hand, only the Ability subscale for an outcome of failure produced alphas that were above this criterion of .70. The External subscale produced a similar alpha to that of the External subscale for the successful outcome. There was some asymmetry between the Ability construct across the success and failure outcomes—but with both values being acceptable against the threshold of .70. A notable difference occurred on the Effort subscale with acceptable alphas on the success outcome (.83) but a poor alpha on the failure outcome (.57).

4.3.3 Confirmatory Factor Analysis

The model had 171 distinct sample moments, 39 parameters, and 132 degrees of freedom. This met the criterion for overidentification for a direct test.

The values derived in this study for the hypothesized model on the success outcome were as follows (Hu and Bentler's cutoffs are in parentheses): RMSEA .09 (< .06), SRMSR .09 (< .08), TLI .84 (> .95), and CFI .87 (> .95). The values derived for the failure outcome were as follows: RMSEA .08 (< .06), SRMSR .07 (< .08), TLI .79 (> .95), and CFI .82 (> .95). All values derived for the model fell short of the recommended thresholds and could not be considered as satisfactory.

4.4 Survey of Achievement Responsibility

4.4.1 Normality

Thirty-nine items presented values that were skewed and 14 of the items were kurtotic at the 3.0 threshold.

4.4.2 Cronbach's Alpha

All four hypothesized scales for both the success and failure outcomes produced alphas above Nunnally and Bernstein's (1994) recommended value of .70 for scale reliability. However, the lower bound of the 95% confidence level for Luck in the failure table fell slightly below the threshold (.69).

4.4.3 Confirmatory Factor Analysis

The model had 528 distinct sample moments, 70 parameters, and 458 degrees of freedom. This meant that the model was overidentified and suitable for a confirmatory test.

The values obtained in this study for the success outcome were as follows with Hu and Bentler's cutoffs presented in parentheses: TLI .82 (> .95), CFI .84 (> .95), RMSEA .07 (< .06), and SRMSR .09 (< .08). The values derived in this study for the failure outcome were as follows: TLI .74 (> .95), CFI .76 (> .95), RMSEA .07 (< .06), and SRMSR .09 (< .08). All values for both the success and failure outcomes did not meet the recommended thresholds offered by Hu and Bentler (1999). This indicates that the model hypothesized by the originating authors for the instrument is problematic under this adaptation of the instrument in the present dataset.

5.0 Discussion & Conclusion

The goal of this study was two-fold. First, it was conducted with a view to satisfying the deficit with respect to the way in which some psychometric instruments have migrated across domains into the area of applied linguistics; and second, to contribute to a sound psychometric footing for attribution theory within applied linguists at an early stage in the emergence of this research trajectory. Both of these goals were achieved.

Results for two of the instruments (CDS II & CIAM) were satisfactory and the results for the other two instruments (SAS & SOAR) showed that the adaptation process is not complete. The following is a brief discussion of each of the instruments as well as the limitations encountered in this study and possible future research.

5.1 Causal Dimension Scale II

The distributions for 4 out of the 12 items were skewed and none of the 12 items were kurtotic at the threshold of 3.0. To improve the normality of the instrument the

Bollen-Stine bootstrap procedure was used and the resulting p value was p =.060. This result was not significant at either the .01 or the .05 level, which in terms of the logic of CFA, means that the model is accepted or that the model fits the data.

The reliability estimates for 3 of the 4 hypothesized scales produced alphas that fell below Nunnally and Bernstein's (1994) criterion of .70. Only the Personal Control scale produced an alpha (.74) that was above this criterion. However, as stated earlier, Cortina (1993) and Green, Lissitz and Muliak (1977) point out that alpha is biased by the number of items on a scale, with larger numbers of items producing higher alphas. Therefore, Nunnally and Bernstein's criterion of .70 may not be suitable for determining the reliability estimates for the CDS II instrument.

The values produced for the selected indexes on the CFA were all within Hu and Bentler's (1999) recommended thresholds, which means that the scores in this study fit the four-factor oblique model hypothesized by the original authors.

5.2 Critical Incident Attribution Measure

The distributions for 10 out of the 24 items for the failure outcome were skewed and 6 of the 24 items were kurtotic at the threshold of 3.0. The distributions for 9 out of the 24 items for the success outcome were skewed and 4 of the 24 items were kurtotic at the threshold of 3.0.

The reliability estimates for seven of the eight hypothesized scales for the failure outcome produced alphas that fell above Nunnally and Bernstein's (1994) criterion of .70, and all eight of the hypothesized scales for the success outcome produced alphas that fell above this criterion. Only the lower bound of the 95% confidence level for strategy on the failure outcome produced an alpha that was slightly lower (.68).

Two of the values produced for the selected indexes on the CFA for the failure outcome were on (RMSEA .06) or within (SRMSR .06) Hu and Bentler's (1999) recommended thresholds, and two were slightly outside (TLI .91; CFI .93) the thresholds. However, all the values produced on the CFA for the success outcome were within the recommended thresholds. This means that there was a reasonable fit between the scores produced in this study and the eight-factor oblique model hypothesized by the original authors.

5.3 Sydney Attribution Scale

The distributions for 13 out of the 36 items were skewed and for 3 out of the 36 items kurtotic at the threshold of 3.0. The reliability estimates were questionable with only three (success/Ability; success/Effort; failure/Ability) of the scales producing alphas

that were above Nunnally and Bernstein's (1994) criterion of .70. However, two of the subscales (success/External; failure/External) produced 95% confidence intervals where the upper bound of the interval (.72 in both cases) was above the threshold of .70.

The values produced for selected indexes on the CFA were outside Hu and Bentler's (1999) recommended thresholds, which means that the scores in this study did not fit the three-factor oblique model hypothesized by the original authors.

However, these unsatisfactory results do not mean that the adaptation of the SAS instrument should be abandoned. Adaptation is a process that sometimes involves multiple datasets as appropriate changes are made and then empirically evaluated in the population for intended use. It is not unusual that an instrument adapted into a new culture and domain will encounter problems in the first phase of adaptation. The first phase functions somewhat diagnostically in a trajectory of ongoing research.

5.4 Survey of Achievement Responsibility

The data from 39 out of the 64 items was skewed, and 12 out of the 64 items indicated kurtosis at the threshold of 3.0.

All the values produced by the CFA were outside Hu and Bentler's (1999) recommended cutoffs for the indexes which mean that the data in this study did not fit the four-factor oblique model hypothesized by the original authors. Only the reliability estimates produced results which could be considered satisfactory as these were all above Nunnally and Bernstein's (1994) criterion of .70 (except for the lower bound of the 95% confidence level for Luck on the failure outcome which was just below the threshold with a value of .69).

Even though these results are unsatisfactory, the data obtained can be used as an empirical guide for the next phase of the adaptation process.

5.5 Limitations & Future Research

The goal of this study was two-fold. First, it was conducted with a view to satisfying the deficit with respect to the way in which some psychometric instruments have migrated across domains into the area of applied linguistics; and second, to contribute to a sound psychometric footing for attribution theory within applied linguists at an early stage in the emergence of this research trajectory. Both of these goals were achieved.

References

Banks, M., & Woolfson, L. (2008). Why do students think they fail? The relationship between attributions and academic self-perceptions. *British Journal of Special Education, 35*(1), 49-56.

Byrne, B. M. (2001). *Structural equation modeling with AMOS.* London: Lawrence Erlbaum.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

Gerbing, D. W., & Anderson, J. C. (1988). An updated paradigm for scale development incorporating unidimensionality and its assessment. *Journal of Marketing Research, 15*, 186-192.

Gobel, P., & Mori, S. (2007). Success and failure in the EFL classroom: Exploring students' attributional beliefs in language learning. *Eurosla Yearbook, 7*(21), 149-169.

Graham, S. (1997). Using attribution theory to understand social and academic motivation in african american youth. *Educational Psychologist, 32*(1), 21-34.

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement, 37*, 827-838.

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment.* Mahwah New Jersey: Lawrence Erlbaum Associates.

Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley.

Hsieh, P. P., & Kang, H. (2010). Attribution and self-efficacy and their interrelationship in the Korean EFL context. *Language Learning*, 1-22. doi: 10.1111

Hsieh, P. P., & Schallert, D. L. (2008). Implications from self-efficacy and attribution theories for an understanding of undergraduates' motivation in a foreign language course. *Contemporary Educational Psychology, 33*, 513-532. doi: 10.1016

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Kline, P. (1994). *An easy guide to factor analysis.* London & New York: Routledge.

Mardia, K. V. (1985). *Mardia's test of multinormality.* New York: Wiley.

Marsh, H. W. (1984). Relations among dimensions of self-attribution, dimensions of

self-concept, and academic achievements. *Journal of Educational Psychology, 76*(6), 1291-1308.

Marsh, H. W., Cairns, L., Relich, J., Barnes, J., & Debus, R. L. (1984). The relationship between dimensions of self-attribution and dimensions of self-concept. *Journal of Educational Psychology, 76*(1), 3-32.

McAuley, E., Duncan, T. E., & Russell, D. W. (1992). Measuring causal attributions: The revised causal dimension scale (CDSII). *Personality and Social Psychology Bulletin, 18*, 566-573. doi: 10.1177

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Peacock, M. (2010). Attribution and learning English as a foreign language. *ELT Journal, 64*(2), 184-193.

Ryckman, D. B., & Rallo, P. H. (1983). Survey of achievement responsibility. Seattle: Authors.

Stipek, D. J., & Weisz, J. R. (1981). Perceived personal control and academic achievement. *Review of Educational Research, 51*(1), 101-137.

Vispoel, W. P., & Austin, J. R. (1995). Success and failure in junior high school: A critical incident approach to understanding students' attributional beliefs. *American Educational Research Journal, 32*(2), 377-412.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3-25.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573.

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist, 45*, 28-36.

Weiner, B., Frieze, I. H., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1971). *Perceiving the causes of success and failure.*

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*(8), 594-604.