JALT Journal

JALT Journal is the research journal of the Japan Association for Language Teaching (JALT). It is published semiannually, in May and November. As a nonprofit organization dedicated to promoting excellence in language learning, teaching, and research, JALT has a rich tradition of publishing relevant material in its many publications.

Links

- JALT Publications: http://jalt-publications.org
- *JALT Journal*: http://jalt-publications.org/jj
- *The Language Teacher*: http://jalt-publications.org/tlt
- *Conference Proceedings*: http://jalt-publications.org/proceedings

- JALT National: http://jalt.org
- Membership: http://jalt.org/main/membership

Attribution Theory: Measurement Foundations for an Emerging Research Area Within Applied Linguistics

Todd Tournat
Kumamoto University

This measurement study is a report on the adaptation of a measure for attribution theory, the Causal Dimension II Scale (CDS II), into the Japanese SLA context. The contribution of this study is three-fold. First, it partially addresses deployment of the instrument in the SLA domain for the first time without attention having been paid to evidence for its capacity to generate unidimensional scores in this domain. Second, it makes an initial contribution to putting use of the instrument in the Japanese context for both research and practice on an evidence-based footing. Third, it assists with resetting the initial research interest in attribution-theory constructs within SLA from their potential influence on learning in the classroom to the necessary and prior issue of their adequate measurement. Results from this study were interpreted as indicating that model fit for scores on the adapted instrument was satisfactory rather than meritorious, but not negative.

本研究は、帰属理論の心理測定用具Causal Dimension Scale II（CDS II）を日本の第二言語習得（SLA）環境に適合させる試みについて報告する。この研究では以下の3点の成果が得られた。第1に、この領域における一次元のスコアを生成するための能力に対する証拠に注意を払うことなく、本研究が日本において初めて　SLA　におけるこの測定用具を使用したことを指摘した。第2に、本研究が、証拠に基づく展開の基礎とすべく、この測定用具を研究及び実践の両方において日本の文脈で用いた初の貢献となることを述べた。第3に、本研究が、SLA　における帰属理論研究において、教室内学習への影響の可能性から適切な測定ということまで、必要かつ重要な問題へのリセットにつながることを指摘した。本研究から、CDS　IIが日本のSLA環境に十分適合することを示した。

Attribution theory, which represents an attempt to theorize the ascription of causes by individuals to their success or failure in educational settings, has been applied to a wide range of academic subjects (e.g., Grant & Dweck, 2003; McClure et al., 2010; Mueller & Dweck, 1998; Rees, Ingledew, & Hardy, 2005; Ryckman & Peckham, 1987; Yang & Montgomery, 2011). Contributions have come in the form of theoretical understandings of four hypothesized causal attributions for success and failure (ability, effort, task difficulty, and luck), which have been used to predict the outcome of an achievement-related event and to estimate the level of future motivation for a task, in the form of attempts to empirically confirm these theoretical understandings.

More recently, attribution theory has become an area of interest within SLA research. This interest is informed by the concern that outcomes in terms of proficiency in foreign languages vary, with successful acquisition not at all certain. In this context, attributions for success and failure have recently come to be seen as one potential area of theoretical explanation for such varied outcomes (Hsieh & Kang, 2010; Hsieh & Schallert, 2008).

Naturally, any emerging research trajectory requires appropriate instrumentation so that measurement for subsequent inferences related to the testing of theoretical claims can be put on a secure empirical footing. The research reported in this paper contributes to the endeavor of establishing adequate, evidence-based instrumentation for the research on attribution theory within SLA by examining the psychometric properties of scores for the Causal Dimensional Scale II (CDS II), which has recently migrated from mainstream educational research on attribution theory into the area of SLA (Hsieh & Schallert, 2008). This migration has occurred with insufficient attention to evidence-based measurement, and thus the guidance of Wilkinson (1999) and the Task Force on Statistical Inference (an influential document outlining recommended practices for research that was published by the Board of Scientific Affairs of the American Psychological Association), is pertinent here; in the absence of secure evidence-based measurement, subsequent causal inferences should be considered questionable even when appropriate inferential statistics are used to assist with these inferences.

Guidelines for test adaptation (whether this involves domain or population adaptation) offered by the International Test Commission (Hambleton, Merenda, & Spielberger, 2005) exemplify the requirements for adherence to the recommendations of Wilkinson and the Task Force on Statistical Inference (1999). The research reported in this paper was conducted with the goals of redressing, specifically but partially, the shortfall with respect to the way in which the CDS II has migrated across domains into the area of SLA and across populations (Hsieh & Kang, 2010; Hsieh & Schallert, 2008), and, more generally and positively, of contributing to a more sound psychometric footing for attribution theory within SLA at this early stage in its emergence within the domain, through the provision of either positive or negative evidence.

The paper reports on the adaptation from the domains of sports and psychology to SLA and from the North American college population (the population for which the instrument was originally developed) to the Japanese college population, using the method of Confirmatory Factor Analysis (CFA) as the primary analytical procedure. CFA is the measurement component of Structural Equation Modelling (SEM; Byrne, 2001; Isemonger, 2012) and directly tests the dimensionality of scores produced by an instrument against a measurement model (i.e., the structural validity of the scores); the measurement model is invariably implied by the scoring regime for the instrument. If the measurement model is rejected in a direct test using CFA, then the practitioner should be wary, if not wholly skeptical, about using the instrument for classroom diagnostics as the dimensional properties of scores produced will not cohere with the scoring regime for the instrument. This can lead to false or mistaken counsel to students and inappropriate interventions by the teacher. Also, the researcher should be equally skeptical about using the instrument, as causal inferences based on data derived from the instrument will be questionable.

Literature Review

Attribution theory has a long lineage within educational research. Foundational work was conducted from the 1950s through the 1970s and the reader can refer to Weiner (1985, 2010) for a useful review of these origins and the place of his work in the research record during the late 1970s and early 1980s. The reviews conducted by Weiner cover the ultimate emergence of the four principal constructs—namely, locus of causality, stability, personal control, and external control—which are the pillars of his theory of causal attribution. Weiner also cites Kelley (1971, 1983) and Heider (1958) in his review, both of whom figure prominently in the early theorization of causal attributions.

Weiner (1985) initially explicated the perceived causes for success and failure in terms of three core constructs: locus of causality, stability, and controllability. The locus of causality construct represents the dependence of causal attributions on two conditions: factors within the person (internal) and factors within the environment (external). The stability construct represents the theoretical notion that among the internal and external causes, some fluctuate and others remain relatively constant. Constant internal causes cited in attributions for success or failure include such things as ability and aptitude, and fluctuating internal causes include such things as effort and moods. The controllability construct represents causes that can be regulated by the individual, such as the amount of effort one is willing to give to a task.

A variety of findings relating to the above constructs have emerged in the general educational context with clear pedagogical implications, raising the prospects for this theoretical area within the domains of SLA and language pedagogy. For example, a number of authors (Elliot, 2005; Grant & Dweck, 2003; McClure et al., 2010; Mueller & Dweck, 1998) have provided evidence and argumentation for the identification of some learners as mastery oriented and others as performance oriented; this identification is associated with tendencies in the way causes are attributed to outcomes. Mastery-oriented learners tend to attribute outcomes to effort and are more likely to focus on learning new skills. Conversely, performance-oriented learners tend to attribute outcomes to stable causes, such as ability, and are more likely to be concerned with how their performance appears to others. Important gender-related findings have also emerged with respect to causal attributions (Andrews, 1987; Ryckman & Peckham, 1987; Stipek, 1984; Sweeney, Moreland, & Gruber, 1982).

Of course, all empirically related, inferential research of the above nature is premised on secure measurement, and within mainstream educational research, a significant part of the literature on attribution theory has been concerned with this issue. For example, Weiner's (1985) theoretical contribution was informed by previous findings in the area that used the following three mathematical techniques to analyze the underlying causal structure of responses given by participants in their research: factor analysis (Foersterling, 1980; Meyer, 1980; Meyer & Koelbi, 1982; Wimer & Kelley, 1982), multidimensional scaling (Falbo & Beck, 1979; Michela, Peplau, & Weeks, 1982; Passer, Kelley, & Michela, 1978), and correlations with a priori schemes (Stern, 1983). Weiner concluded that some of the research was questionable in that not all of the causes were adequately measured. Attention to psychometric adequacy of scores generated by a new line of instrumentation is important, and the research cited above indicates a significant concern with this issue within attribution theory in the past. However, attention to psychometric adequacy of scores is equally as important at points of migration for instrumentation either into new domains or into new populations (Hambleton et al., 2005). It is this issue that forms the rationale for the contribution of the research reported in this paper and further argumentation below will bring this into more precise focus.

One of the most important of the instruments that have drawn upon the theoretical work of Weiner is the Causal Dimension Scale (CDS) developed by Russell (1982). Russell developed the CDS as a measure of how individuals attribute causes to success or failure. Using the CDS, the respondent codes items operationalizing the causal attribution dimensions of locus of causality, stability, and control on a semantic differential scale. However, McAuley, Russell, and Gross (1983) and Russell, McAuley, and Tarico (1987) published research expressing concerns about the low internal consistency of the control dimension and its tendency to correlate highly with the locus of causality dimension. These findings were cited by McAuley, Duncan, & Russell (1992) as a rationale for their revision of the CDS. In the 1992 article, the controllability dimension of the CDS was modified to correct this perceived flaw and the CDS II was developed. In the CDS II, the control dimension was separated into the personal control and external control dimensions. The causality and stability dimensions were unchanged.

The CDS II, in the process of its revision (McAuley et al., 1992), had its scores generated from within the domains of sports and psychology. These scores were directly tested via a CFA for fit with the hypothesized model, which was the unidimensional model (also known as the common factor model). Unidimensionality refers to the property of a score indicating relatively exclusively the single construct that it is purported to indicate, and not indicating any other constructs (either constructs explicitly specified in the model for the instrument or other sources of nonrandom variance not specified in the model and perhaps unanalyzed in the error). The fact that the instrument was tested under the unidimensional model via a CFA is important because secure interpretation of derived scores is premised on this property being present in the scores. Thus, this study met a reasonably high threshold for evidence-based claims on the structural validity of scores generated by the CDS II. This high threshold has not been met, however, in subsequent migrations of the instrument into new domains and populations.

The instrument in its initial form (CDS) migrated to the domain of business, where it was used to measure causal attributions for employment success and failure (Schaufeli, 1988) without appropriate attention to the measurement properties of scores. The CDS II has also migrated to other academic domains. First, it migrated to the medical domain to measure causal attributions made by nurses following an error (Meurier & Vincint, 1998). Second, and more recently, it migrated to the domains of applied linguistics and SLA to measure the causal attributions made by university students for success and failure in learning a second language (Hsieh & Kang, 2010; Hsieh & Schallert, 2008). In the first of these studies relating to SLA (Hsieh & Schallert, 2008), the migration was one of domain only, but in the second (Hsieh & Kang, 2010) the migration was one of both domain and population because a Korean translation of the instrument was administered to a sample of 9th-grade Korean students. In both studies, the domain migration is from sports and psychology, for which there is validity evidence from a psychometric study (McAuley et al., 1992), to SLA (Hsieh & Kang, 2010; Hsieh & Schallert, 2008), for which there is no validity evidence. Cronbach's alphas were reported in Hsieh and Kang, but this index is inadequate for demonstrating unidimensionality of scores (Gerbing & Anderson, 1988), which is critical to score interpretation.

Therefore, the rationale for this study was to provide a partial redress of these weaknesses by providing data on the psychometric properties of scores for the use of the instrument in the SLA domain, but in this case, on a sample of the Japanese university student population. The data reported in this study provide evidence for the structural validity of scores in a new domain (SLA) and evidence for the structural validity of scores in a new population (Japanese university students), but the data cannot provide evidence on the psychometrics of scores with respect to the Korean school population. This discrepancy will have to be addressed within the Korean population.

Under the above rationale, the research reported in this paper covers the adaptation of the CDS II into the Japanese SLA context through an appropriate process for translation and the use of measurement methods to provide initial evidence for the scores produced. Therefore, what is addressed in this study is the absence of evidence for structurally valid scores in the domain migration of the instrument from sports and psychology to SLA; furthermore, the study makes an initial contribution to what should be a cumulative process of gathering evidence for the plausibility of the subscales comprising the CDS II under the Japanese translation and in the Japanese population.

Methodology

Instrument

The CDS II (see Appendix) uses a 9-point semantic differential scale as the response format for each item making up the total instrument. There are 12 items in total, with three items measuring each of the four subscales that represent the theoretical constructs purportedly being measured. The items are, by design, presented at random to the respondents. Items 1, 6, and 9 measure the locus of causality construct; items 3, 7, and 11 measure the stability construct; items 2, 4, and 10 measure the personal control construct; and items 5, 8, and 12 measure the external control construct.

Because the original instrument was developed in English, forward and back translations were used as the first step in adapting it into the Japanese context, in accordance with the International Test Commission Guidelines (Hambleton et al., 2005). 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 in test construction. The back-translated version was then compared to the original English version and no notable contradictions emerged. It was then decided that the Japanese version was suitable in terms of both language equivalency and cultural context.

Participants and Administration

The participants for this study were freshmen students (N = 213) at a university in western Japan from the fields of engineering, science, medicine, and literature who were studying EFL in the General Education Program. Ages ranged from 18 years through 26 years with 96% of the sample between, and including, 18 years and 21 years. There were 125 males and 88 females. In order to reduce the number of independent variables and isolate each student's causal attributions for success and failure to the same achievement-related event, all participants went through the same basic EFL course and were taught by the same teacher. Therefore, a random sample of the general student population was not possible in this study's design; it will rarely be possible given the exigencies of having a learning or achievement event common to all participants as the object of the causal attributions. Meta-analyses of this study and studies that might follow may help to address this design constraint, facilitating more confident generalizations. The nonrandom sampling approach in this study is consistent with previous studies that have reported the psychometrics of scores for the population in which the instrument was developed (e.g., McAuley et al., 1992) and with studies that have used the instrument without such psychometric reporting (Hsieh & Kang, 2010; Hsieh & Schallert, 2008). The EFL course was 15 weeks long and all classes used the same English textbook. At Week 8 of the course, all students took the same midterm examination, the result of which functioned as the reference point for making causal attributions for success and failure.

The exam was a written exam and comprised three sections: a listening section, a vocabulary section, and a grammar section. This exam was an achievement test, with all questions bound tightly to the material taught in the course. This was done so that the outcome was, as far as possible, an achievement-related event rather than a proficiency-related event.

The results of the exam were given back to students 1 week later, and they were then given an open-ended questionnaire asking them to consider and report explicitly the causes for the grade they had received. At the beginning of the following class a week later, they filled out a form asking for the following background information: age, major, academic year, and the date they filled out the questionnaire (informed consent was also included). They then filled out the Japanese version of the CDS II. The administration of the CDS II requires the researcher or practitioner to use it after explicit written reflection on an achievement-related event and when the outcome in terms of success or failure is known by the respondents (McAuley et al., 1992). In this timing of the administration of the CDS II, this study departs from those of Hsieh and Kang (2010) and Hsieh and Schallert (2008) but not from McAuley et al. (1992).

Analytical Procedure

The analytical procedure was to initially consider the data from the point of view of descriptive statistics. This involved a focus on score distribution, including both univariate normality (i.e., skew and kurtosis) and multivariate normality. Next, reliability estimates (Cronbach's alpha) were calculated, including the 95% confidence intervals for alpha. This was followed by a CFA of the hypothesized model.

The procedure for evaluating skew and kurtosis was first to determine the critical ratio, which is calculated by dividing the values in the skew and kurtosis statistic columns by the respective standard error values, and then to compare the values for this computation against an interpretive criterion of both 2.0 and 3.0, stipulated in advance by the researcher. The value of 2.0 was taken as being a strict criterion and 3.0 was taken as being a more relaxed criterion.

Mardia's coefficient (Mardia, 1985) was calculated as an indication of multivariate nonnormality. This is important because maximum likelihood estimation (MLE), which was used as the method of estimation for the CFA in this study, assumes that the data has multivariate normal properties. Mardia's coefficient gives an indication of the extent to which this assumption is satisfied.

With respect to calculating the reliability estimates (Cronbach's alpha), and consistent with Fan and Thompson's (2001) recommendations, the confidence intervals (95%) for alpha, using the central F distribution, are also reported. Nunnally and Bernstein's (1994) criterion of .70 for scale reliability was adopted, tentatively and critically, as the first point of reference for interpreting the derived values for alpha and the derived confidence intervals.

CFA was performed using AMOS 5.0.1 (Arbuckle, 2003). A CFA was conducted using the common factor model. When the common factor model is directly tested using the method of CFA (in other words, with items hypothesized to load on only one factor and with the error uncorrelated), the unidimensionality of subscales can be determined. This cannot be established using Cronbach's alpha (Gerbing & Anderson, 1988), which was the only index reported in Hsieh and Schallert (2008) and Hsieh and Kang (2010). The Bollen-Stine bootstrap procedure (Byrne, 2001) was also used in order to cope with the multivariate nonnormality of the data, and this was interpreted via the *p* value for the associated process of comparing the chi-square for multiple bootstrapped samples to the chi-square for the observed data.

Results

Descriptive Statistics

Table 1 shows the values for item mean, standard deviation, skew, and kurtosis. The results for skew and kurtosis are discussed below.

Skew & Kurtosis

The calculated values for comparison against the stipulated criteria of 2.0 and 3.0 (referred to above in the methods section) are presented in Table 2. The values marked with a single asterisk indicate skew and kurtosis values that fell on, or exceeded, the more stringent threshold of 2.0, and the values marked with double asterisks indicate skew and kurtosis values that fell on, or exceeded, the more relaxed threshold of 3.0 adopted in this study. The values that exceeded the threshold of 3.0 were not acceptable according to either criterion set by the researcher.

Table 1. Item Means, Standard Deviations (SD), Skew and Kurtosis for Scores Derived on Items Comprising the CDS II (N = 213)

Test items	*M*	*SD*	Skew		Kurtosis	
			Statistic	*SE*	Statistic	*SE*
CDS II 01	5.65	2.19	-.43	.17	-.61	.33
CDS II 02	6.20	2.15	-.64	.17	-.40	.33
CDS II 03	4.10	1.93	.51	.17	-.15	.33
CDS II 04	6.03	1.96	-.47	.17	-.43	.33
CDS II 05	3.92	1.76	.13	.17	-.49	.33
CDS II 06	6.66	1.65	-.60	.17	.03	.33
CDS II 07	4.06	2.12	.54	.17	-.42	.33
CDS II 08	4.07	1.83	.14	.17	-.50	.33
CDS II 09	6.77	1.54	-.39	.17	-.25	.33
CDS II 10	6.24	1.74	-.29	.17	-.24	.33
CDS II 11	4.86	2.24	.13	.17	-.95	.33
CDS II 12	4.12	1.79	.09	.17	-.45	.33

Table 2. Calculated Values for Skew and Kurtosis

Test items	Calculated values	
	Skewness	Kurtosis
CDS II 01	*2.6	1.8
CDS II 02	**3.8	1.2
CDS II 03	**3.0	0.5
CDS II 04	*2.8	1.3
CDS II 05	0.8	1.5
CDS II 06	**3.6	0.1
CDS II 07	**3.3	1.3
CDS II 08	0.8	1.5
CDS II 09	*2.3	0.7
CDS II 10	1.7	0.7
CDS II 11	0.7	*2.8
CDS II 12	0.5	1.4

Note. * Test item is skewed at a threshold of 2.0. ** Test item is skewed at a threshold of 3.0.

With respect to skew, three items presented values that fell on or above the 2.0 threshold, and four items presented values that fell on or above the 3.0 threshold (for a total of seven items with questionable skew). With respect to kurtosis, only one of the items fell above the 2.0 threshold and none of the items fell above the 3.0 threshold.

However, McAuley et al. (1992), the original authors of the CDS II, did not report the normality of distribution for scores generated by the instrument; therefore, it is difficult to determine whether scores on these items might have been skewed or kurtotic in the original as well. In other words, it is difficult to determine whether nonnormal properties for scores for some of the items are native to the instrument or an outcome of the adaptation undertaken in this study. Therefore, in an effort to remain faithful to the original instrument and to directly test the inherited model from the literature, all items were entered into the model specification hypothesized by McAuley et al. The subsequent analytical procedure, therefore, and the associated results reported below were sensitive to distortions as a result of this. For example, by adopting the Bollen-Stine bootstrap procedure, the researcher attempted to accommodate for nonnormality in the analysis and the uncorrected chi-square was not interpreted; this is consistent with recommendations offered by Kline (2011).

Finally, Mardia's coefficient (Mardia, 1985) indicated multivariate nonnormality. The critical ratio was 11.37, and thus the Bollen-Stine bootstrap procedure was adopted to cope with this form of nonnormality and to assist with adjudicating model fit in the CFA (for an explanation of this procedure, refer to Byrne, 2001, pp. 267-271). McAuley et al. (1992) did not address multivariate nonnormality in their study, nor did they adopt the bootstrapping procedure. Normality was also not reported by Hsieh and Schallert (2008) nor by Hsieh and Kang (2010). The nonnormality of some items enters into the interpretation of the CFA later in this paper.

Reliability Estimates

Reliability estimates (Cronbach's alpha) for scores on the four subscales hypothesized for the CDS II by the original authors (McAuley et al., 1992) are presented in Table 3. For all four hypothesized subscales, the lower bound for the 95% confidence level fell below the interpretive threshold of .70 (referred to above in the methods section). The derived value for alpha fell below the threshold on all hypothesized subscales except personal control, which produced a value for alpha of .74. However, Cortina (1993) and Green, Lissitz, and Muliak (1977) pointed out that alpha is biased by the number of items on a scale, with larger numbers of items producing higher alphas. This property of alpha, and the propensity for some research areas to neglect this when interpreting alpha, has been explained by Isemonger (2012) in a recent and approachable critical review. Only three items comprise each subscale on the CDS II, which is a comparatively low number of items for a subscale and, in fact, close to the minimum for measurement of a latent. Therefore, there is 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. This will be further addressed in the discussion section.

Table 3. Reliability Estimates, Confidence Intervals for Alpha (95%), Means and Standard Deviations for Scores on Subscales of the CDS II

Subscale	Cronbach's α		Subscale	
	Value	95% CI	*M*	*SD*
Locus of causality	.66	[.57, .73]	19.08	4.20
Stability	.66	[.57, .73]	13.02	4.87
Personal control	.74	[.68, .80]	18.47	4.77
External control	.66	[.57, .73]	12.10	4.16

Confirmatory Factor Analysis (CFA)

A CFA was conducted to directly test the researcher-hypothesized (McAuley et al., 1992) four-factor structure of the CDS II. MLE was employed to evaluate the fit of the four-factor oblique model to the data in this study, which directly simulated the model tested by McAuley et al. in the original population for which the instrument was designed. The model comprised 78 sample moments—30 free parameters and 48 degrees of freedom—meeting the criterion for overidentification.

Hu and Bentler (1999) recommended a number of cutoff values for important fit indexes used in adjudicating model fit to the data matrix. The purpose of these fit indexes and the appropriate use of cutoff values is to overcome the sensitivity of the chi-square statistic to sample size. This sensitivity can mean that trivial departures from the hypothesized model in the data lead to unwarranted model rejection or Type I error. Hu and Bentler's cutoffs were empirically derived to ensure that the probability of both Type I and Type II error was low. The following indexes were used in this study:

the root mean squared error of approximation (RMSEA), the standardized root mean squared residual (SRMSR), the Tucker-Lewis index (TLI), and the comparative fit index (CFI). These indexes were selected in advance and are consistent with recommendations by Hu and Bentler. The RMSEA was also selected because it rewards model parsimony (Byrne, 2001), which means that simpler models are rewarded above more complex models. The SRMSR represents the average of all standardized residuals derived from the fitting of the hypothesized model to the sample data. In other words, it rejects based on the difference between the observed and predicted correlations. The TLI and CFI are incremental-fit indexes that evaluate the hypothesized model against the baseline model (the independence model).

The values derived in this study for the above-cited indexes were as follows (Hu and Bentler's cutoffs are in parentheses): RMSEA .057 (< .06), SRMSR .06 (< .08), TLI .92 (> .95), and CFI .94 (>.95). In the case of the RMSEA, the lower bound of the 90% confidence interval was .03 and the upper bound was .07.

As stated above, multivariate nonnormality was a property of the scores, and thus the Bollen-Stine bootstrap procedure (Byrne, 2001) was adopted as a further analytical tool in dealing with this problem. 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 $p = .06$. This result was not significant at either the .01 or .05 level. In terms of the logic of CFA (the reverse of traditional inferential statistics), this means the model is accepted. In other words, the structure of the data and scores is not significantly different from the model being hypothesized to explain the scores.

Discussion

As mentioned above, the conventional criterion of 3.0 was decided on in advance as an acceptable threshold for skew and kurtosis, but results under a more stringent threshold of 2.0 were also inspected. The distribution of scores for some of the items on the CDS II was slightly skewed, and this property of some of the scores provides a departure point for improvement of the instrument in the future. Bollen-Stine bootstrapping (Byrne, 2001) was also adopted to overcome the limitations associated with the multivariate nonnormality of the data, and it is important to note the associated positive result from this procedure, which is discussed further below.

Nunnally and Bernstein's (1994) criterion of .70 was initially set as the threshold to evaluate the reliability estimates (Cronbach's alpha) for each of the subscales, but this criterion should be adopted critically (Isemonger, 2012). The number of items, which was low for each construct, was also considered in the application of the criterion. Also, it is important to recognize that given the shortcomings of alpha, the results for alpha in this study are subservient to the results of the CFA, which is the more powerful method.

The reliability estimates (Cronbach's alpha) for three of the four hypothesized subscales produced alphas that fell below Nunnally and Bernstein's (1994) criterion of .70. Only the personal control subscale produced an alpha (.74) that was above this criterion. However, as stated earlier, Cortina (1993) and Green et al. (1977) demonstrated that alpha is biased by the number of items on a scale, with larger numbers of items producing higher alphas. Cortina directly counseled the consideration of the number of items in a scale when interpreting the alpha value, stating that "alpha is very much a function of the number of items in a scale, and although alpha is also a function of item intercorrelation, it must be interpreted with number of items in mind" (p. 102).

The CDS II is a relatively small instrument with only three items for each of the four constructs (12 items in total). The alpha values derived in this study are also relatively consistent with McAuley et al.'s (1992) results in the original population for which the instrument was designed. Therefore, Nunnally and Bernstein's criterion of .70, as a rule of thumb, is not suitable for determining the reliability estimates for subscales on the CDS II. It is arguable that the values for alpha derived on the scores for this instrument, under the adaptation, are acceptable. One notable point is that the alpha Hsieh and Kang (2010) obtained for the stability subscale (.26) was extremely low even for instruments with only three items per construct and is inconsistent with the findings in this study and in McAuley et al. If items on this subscale are not consistent with one another, it raises the question of whether they might be consistent with items on a subscale for which they should not be consistent given the model and scoring regime for the instrument. This low value for alpha on the stability subscale could be a result of Hsieh and Kang's translation; this possibility reinforces the necessity for attention to be paid to measurement in the future with respect to the Korean version of the instrument.

Finally, it should also be noted that Cronbach's alpha was not the most crucial analytical tool used in this study; the CFA was, and Cronbach's alpha has rather severe limitations. The most important of these is that the index does not demonstrate unidimensionality of scores (Bentler, 2009; Cortina, 1993; Green et al., 1977; Green & Yang, 2009a, 2009b; Nunnally & Bernstein, 1994; Revelle & Zinbarg, 2009; Sijtsma, 2009a, 2009b). This provided the rationale for the CFA, which was the main part of the study.

Therefore, in order to avoid Type I and Type II error, two absolute fit indexes (RMSEA & SRMSR) and two incremental fit indexes (TLI & CFI) associated with CFA were used. Empirically derived cutoffs were used for interpreting the values produced for these indexes (Hu & Bentler, 1999), and this assisted with triangulating the decision on model fit, both among these indexes themselves and with the bootstrap procedure, which uses a test statistic and which is not an approximate fit index. It is important to note in interpreting these cutoffs or thresholds for the indexes selected that they are just that; that is, they are "approximate fit indexes" and not "test statistics," and interpretation occurs along a continuum (see Hu & Bentler, 1999). The sign is principally there to indicate directionality of interpretation and not as an absolute and precise threshold for accepting or rejecting, as is the case with the statistical levels associated with a test statistic.

In terms of the abovementioned combination of indexes and the bootstrap p value, the adaption process for the CDS II did not produce an instrument, and associated scores, with exemplary results, but the results were arguably satisfactory and not negative. The Bollen-Stine bootstrap procedure (Byrne, 2001) produced a nonsignificant value of .06, indicating evidence for the model. The p value for the bootstrap is based on repeated comparisons of the chi-square (for each of the 1000 bootstrap samples) to the chi-square for the observed data (81.0). In the spirit of criticism of the interpretation taken here of the p value (which was .06) as evidence for the model, it bears saying that although this interpretation is consistent with common practice and precedent in the literature, there are important critics within the area of CFA who argue that the statistical level can potentially be set to be more strict (Hayduk, 1996), that is, to thresholds above .05 (which again, it must be reiterated, is the reverse logic from that of traditional inferential statistics, where a stricter threshold would be represented by a statistical level less than .01).

The values produced for the RMSEA and SRMSR indexes on the CFA were within Hu and Bentler's (1999) recommended thresholds. Some sources offer a threshold of $< .05$ for the RMSEA (see Byrne, 2001; Kline, 2011), but it is important to emphasize that these authors also specifically identified this as the threshold for a meritorious result, and a zone of interpretation exists above .05, which is for a satisfactory result. Byrne characterizes the threshold as "$< .05$ to .08" (p. 152). In other words, below .05 is the meritorious result, and between .05 and .08 is the range for interpretation of a satisfactory result. The threshold adopted in this study was set in advance as .06, because this was recommended by Hu and Bentler in one of the most important pieces of primary literature in the area, and because their research indicates that in small samples (< 250) the RMSEA will be moderately underestimated. The values produced for the TLI and CFI were only slightly below, although the case for the TLI is more problematic. In her highly influential book, Byrne (2001) used her own data to simulate the inferential process that uses these indexes and interpreted a CFI of .94 (which is .01 short of the threshold) as "relatively well-fitting" (p. 152). A value of .94 was also obtained on this same index for scores on the CDS II reported in the current study and the same interpretation is taken. On the TLI, the departure in the wrong direction from the cutoff is only slightly more significant. Byrne characterized a value of .90 for the goodness-of-fit index (GFI)—which has the same threshold of .95 as the CFI—as "marginally adequate" (p. 152). Therefore, the value of .92 derived for the CFI in this study was interpreted as adequate. In addition, the *p* value under the Bollen-Stine bootstrap procedure indicated model fit, so there is a defensible case that this result is positive evidence for the model.

Overall, this means that the scores in this study reasonably fit the four-factor oblique model hypothesized by McAuley et al. (the original authors of the CDS II). Thus, these results also mean that researchers and practitioners who use this adaptation in the Japanese college population and the SLA domain have some evidence-based grounds to proceed with interpreting their scores on the assumption that they are indeed unidimensional and therefore interpretable. The instrument therefore holds good prospects for use within applied linguistics and within the Japanese population because there is initial evidence for the structural validity of scores generated by the instrument. Evidence for the structural validity of scores generated by an instrument is critical to being able to interpret scores for both pedagogy and research. Thus, preliminary evidence for structural validity represents a contribution to the literature. However, it should be emphasized that it is only an initial contribution. Validity evidence is typically seen as a cumulative process; further studies are required to extend the evidence. Also, the evidence for model fit, characterized in this study as satisfactory rather than meritorious (but not negative), should not preclude the possibility of further refinement of this instrument.

It is worth drawing these findings back to the more grounded issue of how research using instruments in the area of attribution theory might contribute to applied linguistics. On one level, it would seem fairly self-evident that understanding how students ascribe responsibility for success and failure will have an impact on success and failure; this means that this area is of interest by itself. However, on another level, it is important to note that the area of learner autonomy is attracting greater and greater interest within applied linguistics, and one of the concerns has been that despite this interest, it remains an undertheorized construct (Benson, 2007, 2009; Stewart & Irie, 2011). There is a clear notional relationship between autonomy and the relative behaviors of ascribing causes for success or failure as either from within oneself or from outside oneself. In this context, it is arguable that the area of causal attribution has potential for the theoretical elaboration of the construct of autonomy. However, before this elaboration can occur and be tested, the measurement of causal attributions needs the benefit of evidence, and that is where this paper makes an initial and partial contribution.

Conclusion

There were two important limitations to this study. First, the way in which the sample was selected for this study had limitations with respect to the representativeness of the specified population (Japanese students in tertiary education). However, it was explicitly argued that this is a design constraint that is difficult to avoid given that all respondents in a particular analysis need to be responding to the same achievement-related event, and that this restraint is evident in previous research as well. It was also argued that meta-analyses and secondary research of this and subsequent studies will assist with providing more secure grounds for generalization. Second, there was no access to data from the original instrument and from the original population for which the instrument was designed, limiting the potential for the study to address measurement invariance across populations. The original population was not highly specified by McAuley et al. (1992) but could be described as the English-speaking North American college population. Future research should also be focused on addressing this limitation by gathering data from different populations and examining these sets of data using the methods of measurement invariance to establish equivalency of measurement across populations. This kind of work would also significantly assist with cross-cultural research because results across the populations could be compared once equivalence of measurement was demonstrated. This would also add to the potential that attribution theory has for theoretical elaboration of the construct of autonomy, because there has been significant interest in how autonomy varies and expresses itself across cultures.

On a final note, with respect to future research, an additional and important future direction would include examining how well the CDS II instrument works at the high school level (a younger and different target population). If the results indicated limitations in this regard, additional research and development could be undertaken to further adapt the instrument for this younger population.

Todd Tournat is an adjunct faculty member of the University of Kumamoto. He has a wide range of research interests, which include psychometric testing, attribution theory, motivation, and learner autonomy.

References

Andrews, P. H. (1987). Gender differences in persuasive communication and attribution of success and failure. *Human Communication Research, 13*, 372-385. http://dx.doi.org/10.1111/j.1468-2958.1987.tb00110.x

Arbuckle, J. L. (2003). *AMOS 5.0 update to the AMOS user's guide*. Chicago, IL: Smallwaters.

Benson, P. (2007). Autonomy in language teaching and learning. *Language Teaching, 40*, 21-40. http://dx.doi.org/10.1017/S0261444806003958

Benson, P. (2009). Making sense of autonomy in language learning. In R. Pemberton, S. Toogood, & A. Barfield (Eds.), *Maintaining control: Autonomy and language learning* (pp. 13-26). Hong Kong: Hong Kong University Press.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*, 137-143. http://dx.doi.org/10.1007/s11336-008-9100-1

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

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104. http://dx.doi.org/10.1037/0021-9010.78.1.98

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52-72). New York: Guilford.

Falbo, T., & Beck, R. C. (1979). Naive psychology and the attributional model of achievement. *Journal of Personality, 47*, 185-195. http://dx.doi.org/10.1111/j.1467-6494.1979.tb00198.x

Fan, X., & Thompson, B. (2001). Confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement, 61*, 517-531. http://dx.doi.org/10.1177/0013164401614001

Foersterling, F. (1980). A multivariate analysis of perceived causes for success and failure. *Archives of Psychology, 133*, 45-52.

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.

Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology, 85*, 541-553. http://dx.doi.org/10.1037/0022-3514.85.3.541

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.

Green, S. B., & Yang, Y. (2009a). Commentary on coefficient alpha: A cautionary tale. *Psychometrika, 74*, 121-135.

Green, S. B., & Yang, Y. (2009b). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika, 74*, 155-167. http://dx.doi.org/10.1007/s11336-008-9099-3

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

Hayduk, L. A. (1996). *LISREL issues, debates and strategies*. Baltimore, MD: Johns Hopkins University Press.

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, 60*, 606-627. http://dx.doi.org/10.1111/j.1467-9922.2010.00570.x

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. http://dx.doi.org/10.1016/j.cedpsych.2008.01.003

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. http://dx.doi.org/10.1080/10705519909540118

Isemonger, I. (2012). Perceptual learning styles and lessons in psychometric weakness. *JALT Journal, 34*, 5-33.

Kelley, H. H. (1971). *Attributions in social interactions*. Morrison, NJ: General Learning Press.

Kelley, H. H. (1983). The situational origins of human tendencies. *Journal of Personality and Social Psychology Bulletin, 9*, 8-30. http://dx.doi.org/10.1177/0146167283091003

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

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

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.

McAuley, E., Russell, R., & Gross, J. B. (1983). Affective consequences of winning and losing: An attributional analysis. *Journal of Sport Psychology, 5*, 278-287.

McClure, J., Meyer, L. H., Garisch, J., Fischer, R., Weir, K. F., & Walkey, F. H. (2010). Students' attributions for their best and worst marks: Do they relate to achievement? *Contemporary Educational Psychology, 36*, 71-81.

Meurier, C. E., & Vincint, C. A. (1998). Nurses' responses to severity dependent errors: A study of the causal dimensions made by nurses following an error. *Journal of Advanced Nursing, 27*, 349-354.

Meyer, J. P. (1980). Causal attributions for success and failure: A multivariate investigation of dimensionality, formation, and consequences. *Journal of Personality and Social Psychology, 38*, 704-715. http://dx.doi.org/10.1037/0022-3514.38.5.704

Meyer, J. P., & Koelbi, S. L. M. (1982). Dimensionality of students' causal attributions for test performance. *Personality and Social Psychology Bulletin, 8*, 31-36.

Michela, J. L., Peplau, L. A., & Weeks, D. G. (1982). Perceived dimensions of attributions for loneliness. *Journal of Personality and Social Psychology, 43*, 929-936. http://dx.doi.org/10.1037/0022-3514.43.5.929

Mueller, C. M., & Dweck, C. S. (1998). Praise for intelligence can undermine children's motivation and performance. *Journal of Personality and Social Psychology, 75*, 33-52. http://dx.doi.org/10.1037/0022-3514.75.1.33

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

Passer, M. W., Kelley, H. H., & Michela, J. L. (1978). Multidimensional scaling of the causes for negative interpersonal behavior. *Journal of Personality and Social Psychology, 36*, 951-962. http://dx.doi.org/10.1037/0022-3514.36.9.951

Rees, T., Ingledew, D. K., & Hardy, L. (2005). Attribution in sport psychology: Seeking congruence between theory, research and practice. *Psychology of Sport and Exercise, 6*, 189-204. http://dx.doi.org/10.1016/j.psychsport.2003.10.008

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and glb: Comments on Sijtsma. *Psychometrika, 74*, 145-154.

Russell, D. (1982). The Causal Dimension Scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology, 42*, 1137-1145.

Russell, D., McAuley, E., & Tarico, V. (1987). Measuring causal attributions for success and failure: A comparison of methodologies for assessing causal dimensions. *Journal of Personality and Social Psychology, 52*, 1248-1257.

Ryckman, D. B., & Peckham, P. (1987). Gender differences in attributions for success and failure situations across subject areas. *Journal of Educational Research, 81*, 120-125.

Schaufeli, W. B. (1988). Perceiving the causes of unemployment: An evaluation of the causal dimensions scale in a real-life situation. *Journal of Personality and Social Psychology, 54*, 347-356.

Sijtsma, K. (2009a). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107-120. http://dx.doi.org/10.1007/s11336-008-9101-0

Sijtsma, K. (2009b). Reliability beyond theory and into practice. *Psychometrika, 74*, 169-173. http://dx.doi.org/10.1007/s11336-008-9103-y

Stern, P. (1983). *A multimethod analysis of student perceptions of causal dimensions* (Unpublished doctoral dissertation). University of California, Los Angeles.

Stewart, A., & Irie, K. (2011). Realizing autonomy: Contradictions in practice and context. In K. Irie & A. Stewart (Eds.), *Realizing autonomy: Practice and reflection in language education contexts* (pp. 1-17). Basingstoke, UK: Palgrave Macmillan.

Stipek, D. J. (1984). Sex differences in children's attributions for success and failure on math and spelling tests. *Sex Roles, 11*, 969-981.

Sweeney, P. D., Moreland, R. L., & Gruber, K. L. (1982). Gender differences in performance attributions: Students' explanation for personal success or failure. *Sex Roles, 8*, 359-373.

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

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist, 45*, 28-36. http://dx.doi.org/10.1080/00461520903433596

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604. http://dx.doi.org/10.1037/0003-066X.54.8.594

Wimer, S., & Kelley, H. H. (1982). An investigation of the dimensions of causal attribution. *Journal of Personality and Social Psychology, 43*, 1142-1162.

Yang, Y., & Montgomery, D. (2011). Behind cultural competence: The role of causal attribution in multicultural teacher education. *The Australian Journal of Teacher Education, 36*(9). http://dx.doi.org/10.14221/ajte.2011v36n9.1

Appendix

Causal Dimension Scale (CDS II)

Instructions: Think about the reason or reasons you have written above. The items below concern your impressions or opinions of this cause or causes of your performance. Circle one number for each of the following questions.

Is (are) the cause(s) something			
1	that reflects an aspect of yourself	9 8 7 6 5 4 3 2 1	that reflects an aspect of the situation
2	manageable by you	9 8 7 6 5 4 3 2 1	not manageable by you
3	permanent	9 8 7 6 5 4 3 2 1	temporary
4	you can regulate	9 8 7 6 5 4 3 2 1	you cannot regulate
5	over which others have control	9 8 7 6 5 4 3 2 1	over which others have no control
6	inside of you	9 8 7 6 5 4 3 2 1	outside of you
7	stable over time	9 8 7 6 5 4 3 2 1	variable over time
8	under the power of other people	9 8 7 6 5 4 3 2 1	not under the power of other people
9	about you	9 8 7 6 5 4 3 2 1	about others
19	over which you have power	9 8 7 6 5 4 3 2 1	over which you have no power
11	unchangeable	9 8 7 6 5 4 3 2 1	changeable
12	other people can regulate	9 8 7 6 5 4 3 2 1	other people cannot regulate