Population and Domain Adaptation of the Critical Incident Attribution Measure (CIAM)

Todd Tournat

Abstract

This measurement study reports on the adaptation of the Critical Incident Attribution Measure (CIAM) from the domains of psychology, sport and education into the Japanese second language acquisition (SLA) domain at the tertiary level. The procedure was based on the guidelines of the International Test Commission (ITC; Hambleton, Merenda, & Spielberger, 2005) and the recommendations of Wilkinson and the Task Force on Statistical Inference (1999). This involved doing forward (into Japanese) and back (into English) translation using two near-native speakers of English with some training in psychometric testing to check for any cultural themes that may not be relevant in the target population. The analytical procedure involved looking at the univariate and multivariate normality of score distribution, determining the reliability estimates, and performing a confirmatory factor analysis (CFA) using structural equation modeling (SEM). The participants for this study were 579 SLA students at four universities in western Japan whose major field of study included English, welfare, science, education, law, engineering, medicine, and business. The results were satisfactory indicating that the CIAM is arguably ready for further testing in the field.

Keywords: Critical Incident Attribution Measure (CIAM), validity, population study, domain adaptation, Japanese version, Confirmatory Factor Analysis (CFA)

1.0 Introduction

Attribution theory (Heider, 1958; Kelley, 1971; Kelley, 1983; Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971) attempts to explain the causal relationships between motivation and achievement (McClure, Meyer, Garisch, Fischer, Weir, & Walkey, 2010; Meyer, McClure, Walkey, Weir, & Mckenzie, 2009; Weiner, 1972, 1979). The research that went into the emergence of attribution theory occurred under the scholarship of several researchers from the 1950s to the 1970s. The question that fueled this research trajectory was the question as to why certain events occur and what are the causes that create both positive and negative outcomes

(Foersterling, 1980; Weiner, 1976, 1985; Weiner, et al., 1971). These researchers proceeded from the theoretical contention that if the underlying properties of causes could be found, quantitative causal comparisons could be made (Weiner, 1985; Weiner, et al., 1971).

Recently, attribution theory has migrated from its original domains of psychology, sport and education to second language acquisition (SLA) in the field of applied linguistics (Gobel & Mori, 2007; Hsieh & Kang, 2010; Hsieh & Schallert, 2008; Mori, Gobel, Thepsiri, & Pojanapunya, 2010; Thang, Gobel, Nor, & Suppiah, 2011). However, the migration process did not follow the recommendations of the International Test Commission (ITC) guidelines (Hambleton, Merenda, & Spielberger, 2005) and/or the recommendations of Wilkinson and the Task Force on Statistical Inference (1999). The guidelines developed by the ITC are important for the adaptation of psychometric instrumentation for use in a different domain and population from which the instrument was originally intended, and the guidelines by Wilkinson and the Task Force on Statistical Inference advocate the proper testing of an instrument before using it in the field. Neglecting the above mentioned guidelines is a serious problem for educators because all subsequent inferences made in the absence of secure evidence-based instrumentation are put in question. Therefore the aim of this study was two-fold; first, to adapt the CIAM into the Japanese SLA context and, second, to put it on a secure measurement footing at its initial entry point with respect to use in the Japanese university population.

2.0 Literature Review

The early work for the eventual configuration of attribution theory began with Heider (1958; as cited in Weiner, 1985) who first stated that causes could be placed under two categories; factors within the person and factors within the environment. However, Rotter (1966; as cited in Weiner, 1985) later classified individuals into 'internals' and 'externals' and the subsequent research formed the basis for the internal-external bipolar dimension, also known as the locus and control dimension (Collins, Martin, Ashmore, & Ross, 1974; Weiner, 1985). Weiner (1985) later relabeled 'locus and control' to 'locus of causality' because 'locus and control' was often confused with 'locus of control.'

In 1971, Weiner and his colleagues argued that a second dimension was needed based on the following rationale; some internal and external causes fluctuate while others remain constant. In terms of internal causes, they viewed ability and aptitude as being constant, and effort and moods as being variable. This theoretical distinction between constant and variable causes found psychometric expression as the stability dimension (Weiner et al., 1971).

Rosenbaum (1972), identified a third dimension as being controllability. He reasoned that mood, fatigue, and effort are all internal and unstable causes and that only effort is controllable. In other words, an individual can decide how much effort they want to give to a task but they are unable to consciously alter their mood or level of fatigue (Weiner, 1985).

Subsequent to the above theoretical developments, Weiner et al. (1971) combined the locus of causality, stability, and controllability dimensions to establish the theoretical framework for attribution theory and hypothesized that the main causal attributes that influence achievement outcomes are ability (internal, stable and uncontrollable), effort (internal, unstable and controllable), task difficulty (external, stable and uncontrollable), and luck (external, unstable and uncontrollable).

However, whenever a new theoretical framework emerges in any field, and this includes applied linguistics, the issue of instrumentation immediately becomes important. In order to measure something, secure empirically tested instrumentation is essential. In response to this need, a number of instruments were developed in the domains of psychology, sport and education that incorporated Weiner et al.'s (1971) four main casual attributes (ability, effort, task difficulty, and luck): the Survey of Achievement Responsibility (SOAR) developed by Ryckman and Rallo (1983); the Sydney Attribution Scale (SAS) developed by Marsh (1984); the Causal Dimension Scale II (CDS II) developed by McAuley, Duncan and Russell (1992); and the instrument used in this study, the Critical Incident Attribution Measure (CIAM) developed by Vispoel and Austin (1995). Marsh (1984) did not include luck as one of the main causal attributes and Vispoel and Austin (1995) included strategy, interest, teacher influence, and family influence for a total of eight causal attributes.

Research questions within measurement research do not follow the typical formulation characteristic of inferential research. The research question is typically related to the correspondence between the model hypothesized by the author/s of an instrument and the scores it actually generates-and the model hypothesized by the author/s should be the unidimensional factor model if scores are to be interpreted unambiguously by the practitioner. Thus, for the purpose of this study the research question is the following:

RQ1: Does the adapted version of the CIAM generate scores that correspond with the unidimensional factor model hypothesized by the original authors?

3.0 Method

The method for this study is reported in terms of the participants for the study, the instrument itself, and the analytical procedure.

3.1 Participants

The participants for this study were SLA students (N = 579) 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/cases were deleted from the database and the statistical analyses were performed on the dataset for 536 participants. The missing values were not systematically missing, and therefore deletion of cases where missing values occurred was not judged to have systematically altered the properties of the sample. Age ranged from 18 years through 30 years with 96% of the sample between, and including, 18 years and 22 years. There were 326 males and 210 females and participation in the study was voluntary. Administration took about 15 minutes to complete.

3.2 Instrumentation

Original Instrument

Vispoel and Austin (1995) developed the CIAM as a way to measure causal attributes in real-life academic settings rather than in hypothetical scenarios or in contrived laboratory tasks. The original CIAM comprises 64 subscales covering four different school subject areas (math, English, general music, and physical education), under two potential outcomes (success or failure), and eight causal dimensions (ability, effort, task difficulty, luck, strategy, interest, teacher influence, and family influence). The subject areas, outcomes and dimensions, respectively, produce 64 subscales under the formula 4 x 2 x 8 = 64 (Vispoel & Austin, 1995). The original instrument comprises 24 real-life experiences with randomly ordered possible causes for a total of 192 items. For each of these subject areas there is a "part one" and a "part two." Part one comprises 24 items that measure failure and part two comprises 24 items that measure success. The items that measure failure are identical to the items that measure success with the only difference being that the items are worded negatively. For example, Item 1 on the form measuring failure reads, "I was unlucky" and on the form measuring success reads, "I was lucky." On both forms Items 3, 11, and 19 measure Ability, Items 4, 12, and 20 measure Effort, Items 2, 10, and 18 measure Task Difficulty, Items 1, 9, and 17 measure Luck, Items 8, 16, and 24 measure Strategy, Items 5, 13, and 21 measure Interest, Items 7, 15, and 23 measure Teacher Influence, and Items 6, 14, and 22 measure Family Influence. To help the respondent recall a particular time that they were unsuccessful and successful, the authors provide a list of four activities the respondent can choose from. In addition to these choices, the authors provide a blank space for the respondents to write down an incident that may not have been included in the above mentioned list of activities. For English, the activities listed were; writing an essay, giving a speech or oral report, reading and understanding a story, and taking a test. Once the respondents select the activity they rate the items on a 6-point Likert scale (Strongly Agree = 6 through Strongly Disagree = 1). The coefficient alphas for the 64 attribution subscales using scores in the 1995 study (4 subject areas x 2 outcomes x 8 attributions) ranged from .79 to .96 with a median of .89 (Vispoel & Austin, 1995).

Adapted Version

The adaptation of the CIAM, reported in this study, is an abbreviated version of the original instrument because only one subject area or domain is represented with this being the English oral communication experience of the students surveyed. The original version (Vispoel & Austin, 1995), as stated above, included the domains of math, English, general music and physical education. The adaptation reported in this study restricts the domain to English oral communication. This restriction provides an appropriate level of specificity for interpretive purposes.

In this study, each item was responded to on a Likert scale incorporating 6 points of discrimination with the following semantic anchors; 6 = Strongly Agree; 5 = Agree, 4 = Slightly Agree, 3 = Slightly Disagree, 2 = Disagree, and 1 = Strongly Disagree. The adapted version, tested in this study, comprises 16 subscales representing one subject area (English oral communication), two outcomes (success or failure), and eight causal dimensions (ability, effort, task difficulty, luck, strategy, interest, teacher influence, and family influence). The formula 1 x 2 x 8 = 16 expresses the subscale structure in simple form. The adapted version of the CIAM is divided into two parts. The first part represents the failure outcome and the second part represents the success outcome. Both outcomes are measured with 24 statements, which are randomly ordered, and which sum to a total of 48 items. This means that, with respect to the subscales, each is measured by 3 of the 24 items. Table 1 shows the subscales and items for both the failure and success outcomes.

Table 1
Subscales and Items Indicating the Subscales for Success and Failure Outcomes

Subscale	Failure Items (Part I)	Success Items (Part II)
Ability	3, 11, 19	3, 11, 19
Effort	4, 12, 20	4, 12, 20
Task Difficulty	2, 10, 18	2, 10, 18
Luck	1, 9, 17	1, 9, 17
Strategy	8, 16, 24	8, 16, 24
Interest	5, 13, 21	5, 13, 21
Teacher Influence	7, 15, 23	7, 15, 23
Family Influence	6, 14, 22	6, 14, 22

Vispoel and Austin's (1995) original instrument was developed in English. The purpose of this study was to adapt the instrument into Japanese and into the SLA context. Thus, methods recommended by the ITC (Hambleton, et al., 2005) involving forward and back translation procedures were followed. These procedures constituted the initial step in adapting the CIAM into the Japanese SLA context.

A near-native speaker of English did the forward translation into Japanese and a different near-native speaker of English did the back translation into English. Both of these near-native speakers of English were doctoral students who had some training in test construction. The back-translated version and the original English version were compared and no inconsistencies were identified. The author then proceeded to administer the Japanese version of the instrument in the Japanese SLA context.

3.3 Analytical Procedure

The data obtained from the CIAM 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 (Arbuckle, 2003) was used to conduct a CFA for the unidimensional model hypothesized for the instrument. The rationale for conducting a CFA in addition to calculating the Cronbach's alphas was that a CFA is particularly powerful as an analytical tool for determining the unidimensionality of scales (Anderson & Gerbing, 1988).

4.0 Results

Results for the CIAM are presented in the following four sections; descriptive statistics, normality for each item, the reliability of scales using the confidence intervals (95%) for Cronbach's alpha, and the results from a CFA.

4.1 Descriptive Statistics

Table 2 and Table 3 show the values for the scale mean, standard deviation, skew, and kurtosis for each of the test items for both the failure and success outcomes.

Table 2
Scale Mean, Standard Deviation (SD), Skew, and Kurtosis on the failure outcome for the CIAM

Test Items	N	Mean	SD	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
CIAM 01	536	2.29	1.260	1.051	.106	.594	.211
CIAM 02	536	3.80	1.103	-.349	.106	-.172	.211
CIAM 03	536	4.77	1.101	-.944	.106	.759	.211
CIAM 04	536	3.26	1.282	.033	.106	-.665	.211
CIAM 05	536	3.64	1.272	-.123	.106	-.650	.211
CIAM 06	536	3.14	1.633	.191	.106	-1.233	.211
CIAM 07	536	2.34	1.263	.963	.106	.371	.211

CIAM 08	536	3.22	1.121	-.146	.106	-.475	.211
CIAM 09	536	2.32	1.228	.932	.106	.439	.211
CIAM 10	536	3.17	1.130	.100	.106	-.309	.211
CIAM 11	536	3.71	1.274	-.119	.106	-.497	.211
CIAM 12	536	4.15	1.195	-.465	.106	-.008	.211
CIAM 13	536	3.38	1.234	.052	.106	-.448	.211
CIAM 14	536	3.08	1.455	.259	.106	-.844	.211
CIAM 15	536	2.42	1.116	.839	.106	.539	.211
CIAM 16	536	3.44	1.111	-.168	.106	-.320	.211
CIAM 17	536	2.38	1.259	.865	.106	.288	.211
CIAM 18	536	3.68	1.083	-.254	.106	-.157	.211
CIAM 19	536	4.16	1.220	-.432	.106	-.271	.211
CIAM 20	536	3.67	1.276	-.073	.106	-.520	.211
CIAM 21	536	3.69	1.282	-.167	.106	-.563	.211
CIAM 22	536	3.10	1.505	.247	.106	-.966	.211
CIAM 23	536	2.46	1.097	.667	.106	.312	.211
CIAM 24	536	3.92	1.159	-.302	.106	-.196	.211
Valid N	536						

Table 3
Scale Mean, Standard Deviation (SD), Skew, and Kurtosis on the success outcome for the CIAM

Test Items	N	Mean	SD	Skewness		Kurtosis	
				Statistic	Std. Error	Statistic	Std. Error
CIAM 01	536	3.57	1.390	-.208	.106	-.711	.211
CIAM 02	536	4.07	1.042	-.370	.106	.221	.211
CIAM 03	536	2.65	1.059	.387	.106	.287	.211
CIAM 04	536	4.27	1.045	-.450	.106	.259	.211
CIAM 05	536	3.79	1.230	-.223	.106	-.407	.211
CIAM 06	536	2.53	1.156	.637	.106	.160	.211
CIAM 07	536	3.74	1.196	-.314	.106	-.296	.211
CIAM 08	536	3.17	.975	-.050	.106	.289	.211
CIAM 09	536	3.34	1.398	-.037	.106	-.827	.211

CIAM 10	536	4.21	.953	-.381	.106	.649	.211
CIAM 11	536	2.63	1.049	.393	.106	.127	.211
CIAM 12	536	3.93	1.179	-.288	.106	-.263	.211
CIAM 13	536	3.76	1.180	-.231	.106	-.237	.211
CIAM 14	536	2.57	1.115	.539	.106	.217	.211
CIAM 15	536	3.36	1.275	.079	.106	-.546	.211
CIAM 16	536	3.10	.984	-.003	.106	.061	.211
CIAM 17	536	3.32	1.461	-.028	.106	-.903	.211
CIAM 18	536	3.94	1.064	-.310	.106	.264	.211
CIAM 19	536	2.65	1.035	.283	.106	.039	.211
CIAM 20	536	4.09	1.191	-.484	.106	-.109	.211
CIAM 21	536	3.74	1.235	-.143	.106	-.390	.211
CIAM 22	536	2.55	1.094	.520	.106	.178	.211
CIAM 23	536	3.40	1.161	-.121	.106	-.328	.211
CIAM 24	536	3.11	1.022	-.011	.106	.020	.211
Valid N	536						

The skew and kurtosis values in Tables 2 and 3 are used in the section below for an assessment of the normality of score distribution.

4.2 Normality

The analytical procedure for evaluating skew and kurtosis was to first determine the critical ratio which is calculated by dividing the value in the skew and kurtosis statistic columns by the respective standard error. The critical ratio was then compared against a predetermined criterion of both 2.0 and 3.0. These calculated values are available for inspection in Table 4 and Table 5. Table 4 represents the failure outcome, and Table 5 represents the success outcome. The values marked with a single asterisk indicate skewness and kurtosis values that fell on or exceeded the threshold of 2.0, and the double asterisk indicate skewness and kurtosis values that fell on or exceeded the threshold of 3.0. The values that exceeded the threshold of 3.0 are not acceptable according to the criterion set by the author.

Table 4
Critical Ratios for Skew and Kurtosis on the failure outcome for the CIAM

Test Items	Skewness	Kurtosis
	Critical Ratio	Critical Ratio
CIAM 01	**9.9	*2.8
CIAM 02	**3.3	.82
CIAM 03	**8.9	**3.6
CIAM 04	.31	**3.2
CIAM 05	1.2	**3.1
CIAM 06	1.8	**5.8
CIAM 07	**9.1	1.8
CIAM 08	1.4	*2.3
CIAM 09	**8.8	*2.1
CIAM 10	.94	1.5
CIAM 11	1.1	*2.4
CIAM 12	**4.4	.04
CIAM 13	.50	*2.1
CIAM 14	*2.4	**4.0
CIAM 15	**7.9	*2.6
CIAM 16	1.6	1.5
CIAM 17	**8.2	1.4
CIAM 18	*2.4	.74
CIAM 19	**4.1	1.3
CIAM 20	.70	*2.5
CIAM 21	1.6	*2.7
CIAM 22	*2.3	**4.5
CIAM 23	**6.3	1.5
CIAM 24	*2.8	.93

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

Table 5
Critical Ratios for Skew and Kurtosis on the success outcome for the CIAM

Test Items	Skewness	Kurtosis
	Critical Ratio	Critical Ratio
CIAM 01	*2.0	**3.4
CIAM 02	**3.5	1.1
CIAM 03	**3.7	1.4
CIAM 04	**4.2	1.2
CIAM 05	2.1	1.9
CIAM 06	**6.0	.50
CIAM 07	*2.9	1.4
CIAM 08	.47	1.4
CIAM 09	.35	**3.9
CIAM 10	**3.6	**3.1
CIAM 11	**3.7	.60
CIAM 12	*2.7	1.2
CIAM 13	*2.2	1.1
CIAM 14	**5.1	1.0
CIAM 15	.75	*2.6
CIAM 16	.03	.30
CIAM 17	.30	**4.3
CIAM 18	*2.9	1.3
CIAM 19	*2.7	.20
CIAM 20	**4.6	.52
CIAM 21	1.4	1.8
CIAM 22	**4.9	.84
CIAM 23	1.1	1.6
CIAM 24	.10	.09

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

After calculating the critical values for skewness for the failure outcome and then comparing them against the criteria selected, 4 items presented values that fell above the 2.0 threshold, and 10 items presented values that fell above the 3.0 threshold. After calculating the critical values for kurtosis for the failure outcome and then comparing them against the criteria selected, 8 items presented values that fell above the 2.0 threshold, and the values for 6 items fell above the 3.0 threshold.

The number of skew values for the success outcome was similar to the number of skew values found in the failure outcome. Six items (4 items in the failure outcome) presented values that fell above the 2.0 threshold and 9 items (10 items in the failure outcome) presented values which fell above the 3.0 threshold; for a total of 15 items which were skewed. On the other hand, kurtosis for the success outcome was better than kurtosis for the failure outcome. Only 1 item (8 items in the failure outcome) presented a value which fell above the 2.0 threshold and only 4 items (6 items in the failure outcome) fell above the 3.0 threshold; for a total of 5 items with a kurtotic distribution.

The original author of the revised CIAM instrument (Vispoel & Austin, 1995) did not report the normality of scores for items making up the revised instrument, and therefore it is not possible to determine whether the departures from normality in items making up this adapted version mirror the revised version offered by Vispoel and Austin, or whether these departures have been introduced in the adaptation.

4.3 Cronbach's Alpha

Nunnally and Bernstein's (1994) criterion of .70 for scale reliability was used in this study. Table 6 and Table 7 present the results for alpha, as well as the confidence intervals (95%) for alpha.

Table 6
Reliability Estimates, Confidence Intervals for Alpha (95%), Scale Means, and Scale Standard Deviations for Failure Outcome

Scale	Cronbach's alpha	95% Confidence Intervals for Cronbach's alpha		Scale Mean	SD for Scale
		Lower Bound	Upper Bound		
Ability	.76	.73	.80	12.64	2.97

Effort	.81	.77	.83	11.08	3.18
Task Difficulty	.76	.72	.79	10.65	2.73
Luck	.94	.93	.95	6.99	3.54
Strategy	.73	.68	.76	10.58	2.73
Interest	.85	.82	.87	10.71	3.32
Teacher Influence	.84	.81	.86	7.22	3.03
Family Influence	.90	.89	.92	9.32	4.21

Table 7
Reliability Estimates, Confidence Intervals for Alpha (95%), Scale Means, and Scale Standard Deviations for Success Outcome

Scale	Cronbach's alpha	95% Confidence Intervals for Cronbach's alpha		Scale Mean	SD for Scale
		Lower Bound	Upper Bound		
Ability	.89	.87	.90	7.93	2.84
Effort	.90	.89	.92	12.29	3.13
Task Difficulty	.83	.81	.86	12.21	2.65
Luck	.92	.91	.94	10.23	3.96
Strategy	.87	.85	.89	9.37	2.65
Interest	.91	.89	.92	11.29	3.35
Teacher Influence	.84	.82	.87	10.50	3.17
Family Influence	.90	.89	.92	7.65	3.08

All of the eight hypothesized subscales for both the failure and success outcomes produced alphas above the threshold of .70. The lower bound of the 95% confidence interval (.68) for the Strategy subscale in the failure outcome fell slightly below the threshold of .70, but the majority of the subscales produced satisfactory alphas; i.e. they were well above the threshold.

4.4 Confirmatory Factor Analysis

A CFA was conducted to directly test the author-hypothesized (Vispoel & Austin, 1995), eight-factor structure for both the success and failure outcomes of the CIAM in an adaptation of the instrument to the Japanese population. The CFA only needed to be conducted once because there is only one domain for the adaptation reported in this study. The model had 300 distinct sample moments, 76 parameters and 224 degrees of freedom, which met the criterion for overidentification for a direct test. The unidimensional model was tested meaning that error terms were not permitted to correlate and items were only permitted to indicate the factor they were hypothesized to measure. Factors were permitted to correlate – in other words they were oblique – and this is consistent with procedures in the social sciences where it is unlikely that orthogonal constructs will be found (Kline, 1994).

The chi-square value for the failure outcome was 729.49 ($p < .01$) and the chi-square value for the success outcome was 483.04 ($p < .01$). In the logic of CFA, this means that the model should be rejected for both outcome conditions. This is because the logic is the reverse of traditional inferential statistics, and the significant result implies that the model departs from the dimensionality of the data. However, when CFA models are directly tested, fit indexes are typically used in addition to the chi-square test (see Hu & Bentler, 1999). This is because the chi-square test becomes extremely sensitive as the sample size increases leading to over-rejection of good-fitting models or Type I errors, and the use of fit indexes has arisen as a way to compensate for this problem with the chi-square and its sensitivity. The most important thing about fit indexes is that the researcher has to decide on a criterion in advance to determine whether the value obtained for an index is acceptable or not. To assist in this task, cutoff values have been recommended and the most widely used are those recommended by Hu and Bentler (1999), because they were empirically derived using data simulations to minimize both Type I and Type II errors. The following indexes were used in this study: the RMSEA, the SRMSR, the TLI, and the CFI. It is also critical to point out (see Hu & Bentler again) that the criteria adopted in advance for interpreting the values obtained for each index are not to be used absolutely – i.e. if the value obtained falls just short of a particular criterion this does not automatically mean that you reject it, as you would do with a test statistic (like the chi-square test statistic). This is because the criterion is being applied for an index which requires interpretation along a continuum, and is not the same as the interpretation procedure for a test statistic. Obviously, if the departure of the value derived for the index is substantial a negative interpretation will be taken, but the criterion needs to be interpreted with judgment, and in the context of the results for the other indexes under the principle of triangulation.

The values derived in this study for the hypothesized model on the failure outcome were as follows (Hu and Bentler's cutoffs/criteria are in parentheses): RMSEA .07 ($< .06$), SRMSR .06 ($< .08$), TLI .91 ($> .95$), and CFI .93 ($> .95$). The values derived for the same model on the success outcome were as follows: RMSEA .05 ($< .06$), SRMSR .03 ($< .08$), TLI .96 ($> .95$), and CFI .97 ($> .95$). On the failure outcome, the RMSEA value fell slightly above the threshold and the TLI and CFI values fell slightly below the recommended thresholds. All of the values derived for the success outcome met the required thresholds.

5.0 Discussion

The research question outlined at the beginning of this paper, to determine whether or not the adapted version of the CIAM could generate scores that correspond with the unidimensional factor model hypothesized by the originating authors, was answered in the affirmative but with a cautionary note. The affirmative answer to the research question is based on the generally positive results for the values derived for the indexes used in the study (the RMSEA, the SRMSR, the TLI, and the CFI) – i.e. very close to meeting the thresholds on the failure outcome, and meeting them on the success outcome. The cautionary note to this interpretation is because the chi-square was significant, and although it is accepted within the literature that this statistic tends to over reject models because of its sensitivity in large datasets, it would still be preferable to derive a chi-square that supports the model. Nonetheless, it is almost always the case in practice that the chi-square result is not satisfactory, and this is why the indexes are primarily used. Additionally, the cautionary note is also based upon the distributional departures from normality on some of the items making up the instrument and difficulties with multivariate normality as well.

Turning first to the issue of distributional normality, the distributions of scores for 10 out of the 24 items for the failure outcome were skewed, and for 6 out of the 24 items they were kurtotic (assuming a threshold of 3.0). On the success outcome, the distributions for scores on 9 out of the 24 items were skewed and 4 out of the 24 items were kurtotic (again assuming the threshold of 3.0). This is the most important area for potential revision of the instrument. If items are skewed or kurtotic the information yield from the item is effectively reduced. The outcomes on the indexes reported were nonetheless acceptable – arguably excluding the RMSEA, CFI and TLI values for the failure outcome (there is, however, further discussion on the interpretation of results for the values returned for these three indexes below).

The reliability estimates (Cronbach's alpha) 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 interval for Strategy on the failure outcome produced an alpha that was slightly lower (.68). While overall, this is a welcome result, it should be noted that CFA is a more powerful method than Cronbach's alpha, particularly with regard to the issue of unidimensionality, and these results for alpha and their interpretation should be seen as secondary to the results for the CFA.

As reported in the results, one of the values produced for the selected indexes on the CFA for the failure outcome was within (SRMSR = .06) Hu and Bentler's (1999) recommended thresholds, and three indexes produced results which were slightly outside (RMSEA = .07; TLI = .91; CFI = .93) the respective thresholds. All of the values produced on the CFA for the success outcome were within the recommended thresholds.

With respect to the three cases where the value returned was above or below the threshold set in advance (informed by Hu and Bentler, 1999), some further argumentation is required in interpreting them. As stated above, when interpreting in terms of these cut-off values it is important to keep in mind that they are not "test statistics" they are "approximate fit indexes." A test statistic is generally used to reject or accept a hypothesis using significance levels of 95% (alpha level of 0.05) or 99% (alpha level of 0.01), and an approximate fit index does not follow the absolute forms of interpretation associated with such test statistics. Fit indexes are interpreted along a continuum (see Hu & Bentler, 1999) and the greater than and less than signs indicate the direction of interpretation. Thus, it is possible, and in fact rational, that if the departure to the wrong side of the sign is by a trivial amount, the research will interpret the result as still satisfactory. For example Byrne (2001), in her highly influential book, and also using Hu and Bentler's (1999) criteria, interprets a CFI of .94 (which is just short of the criterion of .95) as "relatively well-fitting," and a value of .90 for the GFI (which is more significantly short of .95) as "marginally adequate" (p. 152). Byrne also identifies a threshold range to distinguish between meritorious results and satisfactory results. For example, she uses the following threshold as a guide to help interpret her results "<.05 to .08" (p. 152) for the RMSEA. Values below .05 are considered meritorious results and values between .05 and .08 are considered satisfactory results.

Hu and Bentler's (1999) recommended thresholds were adopted for this study because the results of their detailed research in this area has been the standard when choosing approximate fit indexes for assisting in the interpretation of results, and their paper on the issue is among the most frequently cited within the field as a whole. As mentioned above, only three of the values were outside Hu and Bentler's recommended thresholds (RMSEA = .07; TLI = .91; CFI = .93) – notably all on the failure outcome rather than the success outcome. In the case of the RMSEA the value of .07 is not outside of the range which Byrne adopts for a satisfactory, rather than a meritorious or well-fitting outcome, and thus the interpretation taken here is consistent with Byrne when this value is interpreted as satisfactory but not well-fitting. In the case of the CFI, the result is only slightly less than what Byrne interprets as relatively well-fitting, and in the case of the TLI the interpretation taken in this study is similar to Byrne's interpretation of a slightly lower value for the GFI, which is that the result is marginally adequate.

Thus, and including all the results which were within the threshold in this triangulated approach of using multiple indexes, an interpretation is taken that there was a reasonable fit between the scores produced in this study and the eight-factor oblique model implicitly hypothesized by the original authors in the scoring regime for the instrument; but the fit is slightly better in the case of the success outcome than in the case of the failure outcome where there were minor negative departures from the thresholds set in advance for three of the indexes.

Follow-up research for this instrument should mainly focus on improving the non-normal distributions (skew and kurtosis) of some of the items. This could be done by examining the items for problems of wording and recalibrating the expressions to shift the response distributions to normality. In some of the extreme cases, items could be discarded and replaced with new ones. In addition, the Likert scale could potentially be extended from six response points to a more refined scale to make it more sensitive to the respondent's disposition, and as a result, less likely to produce scores with a skewed or kurtotic distribution. Once these adjustments are made, the instrument could then be empirically assessed in a new dataset on a different sample of Japanese university students and the items could then be reexamined to see if the non-normality problem has been overcome – or at least significantly improved. If the problem is substantially reduced, CFA could then again be adopted to see if the improvements in distribution of scores impacts the overall model fit and plausibility.

The limitations encountered in this study are in two areas. The first area of limitation inherits from the field itself and the instrument which was adapted in this study, and this is that a CFA was not conducted by the original authors on the instrument. This means that it is difficult to determine whether misfit in the model is an inherent feature of the scores generated by the instrument or a feature following the changes made in the adaptation process. One of the reasons for the deficit of CFAs may be that the instrument was developed in the early 1990s and conducting a CFA was not a widely practiced statistical procedure at that time. Cronbach's alpha was the dominant index and many believed that it was not only a good indicator of reliability, but also a good indicator of unidimensionality. However, this belief has been thoroughly refuted (Gerbing & Anderson, 1988). To date, CFA is the most powerful way to test the unidimensionality of scores derived on an instrument (within classical test theory) and it is a crucial statistical procedure for test construction and adaptation.

The second area of limitation concerns the issue of validity. On reviewing the large body of modern scientific research and current views on the notion of validity, it is apparent that validity is taken as a cumulative process. In other words, each new study builds upon the work done in previous studies, and the new findings assist by adding empirical evidence for the capacity of an instrument to generate valid scores in what is a cumulative research process. The work done in this study was one such step and therefore should not be viewed as an empirically sufficient study for final conclusions to be drawn. It represents a contribution to an ongoing endeavor.

Limitations related to this second area involve two additional sub-issues. Firstly, the way in which the sample was selected for this study (Japanese students in tertiary education). Secondly, the data obtained from the original instrument for the original English speaking population was not available. The first limitation was brought on by time and monetary constraints which made it difficult to get a broad-scale random sample. This limitation is not at all uncommon in the literature and real-world research practice. However, and as a result of these time and monetary constraints, the data had to be derived from a sample of convenience (Japanese students from universities where the author of this paper worked). The second limitation of not having access to data obtained from the original instrument in the English population from which it was created, negated any chance of doing a measurement invariance study to test if the adaptation of the instrument in this study measured equivalently with the original instrument in the original population for which it was first designed.

Future research trajectories should be focused on addressing these limitations in the following two ways. First, there should be an attempt to gather data from two populations, one from the original population for which the instrument was originally intended and the other from the Japanese population. The scores available in these two sets of data should then be examined using the methods of measurement invariance to establish equivalency of measurement across populations. Second, the CIAM should be tested on other samples from the Japanese university student population to assist with the accumulation of evidence for the instrument's capacity to generate structurally valid scores in the target population (Japanese students in tertiary education). Such studies, if accumulated, would also assist with sample bias, by providing multiple samples for meta-analyses where the goal would be to average-out sample bias (which is an inevitable consequence of samples of convenience).

6.0 Conclusion

The goal of this study was two-fold. First, it was conducted in an effort to satisfy 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, and particularly in Japan. Both of these goals were achieved.

Results for the CIAM were satisfactory, although cautioned upon, and as stated above, clearly indicate that further empirical research is both promising and warranted. There are two empirical conclusions from this study. The first is that the CIAM, in this Japanese adaptation, is capable of producing scores whose dimensionality plausibly fits the model specified by the original authors. The second is that these constructs related to causal attribution are measureable.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423.

Arbuckle, J. L. (2003). *Amos 5.0 Update to the Amos User's Guide*. Chicago IL: Smallwaters Corporation.

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

Collins, B. E., Martin, J. C., Ashmore, R. D., & Ross, L. (1974). Some dimensions of the internal-external metaphor in theories of personality. *Journal of Personality and Social Psychology, 29*, 381-391.

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.

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.

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, 60*(3), 606-627.

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.

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.

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.

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

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.

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.

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*.

Meyer, L. H., McClure, J., Walkey, F., Weir, K. F., & Mckenzie, L. (2009). Secondary student motivation orientations and standards-based achievement outcomes. *British Journal of Educational Psychology*, *79*, 273-293.

Mori, S., Gobel, P., Thepsiri, K., & Pojanapunya, P. (2010). Attributions for performance: A comparative study of Japanese and Thai university students. *JALT Journal, 32*(1), 5-28.

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

Rosenbaum, R. M. (1972). *A dimensional analysis of the perceived causes of success and failure. Unpublished doctoral dissertation*. University of California. Los Angeles.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monograph, 80*, 1-28.

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

Thang, S. M., Gobel, P., Nor, N. F. M., & Suppiah, V. L. (2011). Students' attributions for success and failure in the learning of English as a second language: A comparison of undergraduates from six public universities in Malaysia. *Pertanika Journal of Social Sciences and Humanities, 19*(2), 459-474.

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. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research, 42*(2), 203-215.

Weiner, B. (1976). An attributional approach for educational psychology. *Review of Research in Education, 4*(1), 179-209.

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., Frieze, I. H., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. (1971). *Perceiving the causes of success and failure*.: Morristown, NJ: General Learning Press.

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

Appendix

Critical Incident Attribution Measure

Part I

Directions: Think about your past experiences in high school English oral communication classes. Try to remember a time in which you did particularly POORLY on an activity that was important to you. The activity you are thinking of might be listed below. If so, circle the letter preceding the activity. If the activity is not listed below, please circle the letter preceding "other" and describe the activity in the space provided. Be sure to circle only one letter.

A. Participating in a group task done in English
B. Giving a speech or oral report in English
C. Answering a question by the teacher in English
D. Taking an oral communication test
E. Participating in a skit or dialogue done in English
F. Other ____________________________(please specify)

There may have been many different reasons why you did poorly on the activity you just circled. The following statements are possible reasons why you might have done poorly. Read each statement carefully and circle the number to indicate the extent to which you agree or disagree with each statement. Be sure to respond to all of the statements.

I DID POORLY ON THE ACTIVITY BECAUSE:	Strongly Agree	Agree	Slightly Agree	Slightly Disagree	Disagree	Strongly Disagree
1. I was unlucky.	1	2	3	4	5	6
2. The activity was difficult.	1	2	3	4	5	6
3. I have weak skills in English as a foreign language.	1	2	3	4	5	6
4. I didn't try hard.	1	2	3	4	5	6
5. I disliked the activity.	1	2	3	4	5	6
6. My parents have weak skills in English as a foreign language.	1	2	3	4	5	6
7. I disliked the teacher.	1	2	3	4	5	6
8. I used the wrong study or practice methods.	1	2	3	4	5	6
9. I had bad luck.	1	2	3	4	5	6
10. The activity was complicated.	1	2	3	4	5	6
11. I'm not talented in English as a foreign language.	1	2	3	4	5	6

12. I made a weak effort.	1	2	3	4	5	6
13. I wasn't interested in the activity.	1	2	3	4	5	6
14. Talent in English as a foreign language doesn't run in my family.	1	2	3	4	5	6
15. I didn't get along with the teacher.	1	2	3	4	5	6
16. I used ineffective learning or training strategies.	1	2	3	4	5	6
17. The odds worked against me.	1	2	3	4	5	6
18. The activity wasn't easy.	1	2	3	4	5	6
19. I don't have natural ability in English as a foreign language.	1	2	3	4	5	6
20. I didn't try to do my best.	1	2	3	4	5	6
21. I didn't find the activity enjoyable.	1	2	3	4	5	6
22. My parents aren't talented in English as a foreign language.	1	2	3	4	5	6
23. The teachers didn't understand me.	1	2	3	4	5	6
24. I didn't know the best ways to study or practice.	1	2	3	4	5	6

Part II

Directions: Think about your past experiences in high school English oral communication classes. Try to remember a time in which you did particularly WELL on an activity that was important to you. The activity you are thinking of might be listed below. If so, circle the letter preceding the activity. If the activity is not listed below, please circle the letter preceding "other" and describe the activity in the space provided. Be sure to circle only one letter.

A. Participating in a group task done in English
B. Giving a speech or oral report in English
C. Answering a question by the teacher in English
D. Taking an oral communication test
E. Participating in a skit or dialogue done in English
F. Other (please specify)

There may have been many different reasons why you did well on the activity you just circled. The following statements are possible reasons why you might have done well. Read each statement carefully and circle the number to indicate the extent to which you agree or disagree with each statement. Be sure to respond to all of the statements.

I DID WELL ON THE ACTIVITY BECAUSE:	Strongly Agree	Agree	Slightly Agree	Slightly Disagree	Disagree	Strongly Disagree
1. I was lucky.	1	2	3	4	5	6
2. The activity was easy.	1	2	3	4	5	6
3. I have strong skills in English as a foreign language.	1	2	3	4	5	6
4. I tried hard.	1	2	3	4	5	6
5. I liked the activity.	1	2	3	4	5	6
6. My parents have strong skills in English as a foreign language.	1	2	3	4	5	6
7. I liked the teacher.	1	2	3	4	5	6
8. I used the right study or practice methods.	1	2	3	4	5	6
9. I had good luck.	1	2	3	4	5	6
10. The activity was not complicated.	1	2	3	4	5	6
11. I'm talented in English as a foreign language.	1	2	3	4	5	6
12. I made a strong effort.	1	2	3	4	5	6
13. I was interested in the activity.	1	2	3	4	5	6
14. Talent in English as a foreign language runs in my family.	1	2	3	4	5	6
15. I got along with the teacher.	1	2	3	4	5	6
16. I used effective learning or training strategies.	1	2	3	4	5	6
17. The odds worked for me.	1	2	3	4	5	6
18. The activity was easy.	1	2	3	4	5	6
19. I have natural ability in English as a foreign language.	1	2	3	4	5	6
20. I tried to do my best.	1	2	3	4	5	6
21. I found the activity enjoyable.	1	2	3	4	5	6
22. My parents are talented in English as a foreign language.	1	2	3	4	5	6
23. The teachers understood me.	1	2	3	4	5	6
24. I knew the best ways to study or practice.	1	2	3	4	5	6