

(Abstract)

## **Background**

This study reports on the statistical analysis of the structural properties of scores produced by the Kambara (1982, 1987) locus of control scale (K-LoC). The K-LoC, a psychometric instrument, which is both statistically and qualitatively analyzed in this paper, fits into a larger research theme of attempts at measuring learner autonomy with respect to second language English students in Japan. The scale has two versions: the short form, K-LoC18 (18 items, 1982) and the long form, K-LoC43 (43 items, 1987).

Psychometric instruments are used to measure psychological traits and aptitudes as an aid to researchers and practitioners in a variety of fields, including second language acquisition and language education. While statistical factor analysis and other statistical methods can be used to assess the reliability and validity of scores generated by such instruments when are being created, adapted and improved, it can also be valuable to gain a qualitative perspective, such as can be had through focus groups (FG).

In order to frame the K-LoC scale in the wider research context of attempts at measuring learner autonomy, this paper begins with a literature review which elaborates on the complex nature of the learner autonomy construct. This is followed by a survey of attempts by researchers at creating a unified model of learner autonomy. The focus of the literature review is then turned towards learner autonomy in the Japanese second language learning context. This is followed by detailing the deficit in the literature with regard to the development of plausible learner autonomy instrumentation. The notional relationship between learner autonomy and various related psychological constructs is laid out in order to give a background on alternative approaches to measuring learner autonomy directly, an endeavor which has yet to bear results.

Based on a search of the literature, the locus of control construct was found to be a widely used construct notionally related to learner autonomy, and in the Japanese ESL context, the K-LoC scale was found to be both prevalent and yet at the same time lacking a crucial vetting of the model through confirmatory factor analysis (CFA), having only exploratory factor analyses (EFA) conducted thus far. Thus the reason for the K-LoC scale being chosen as the subject of study for this paper is outlined, followed by a background on LoC, both in general and with regards to second language acquisition. Finally, the use of the K-LoC scale in the Japanese research context is elaborated upon. The literature review led to the five research questions (RQs) of this study.

## **Research Questions**

RQ 1: Does a two-factor (I-LoC and E-LoC) measurement model for the K-LoC18 (Kambara, 1982), consistent with original theory and conception, provide plausible fit in a direct and a priori test using CFA as the method?

RQ 2: Does a two-factor (I-LoC and E-LoC) measurement model for the K-LoC43 (Kambara, 1987), consistent with original theory and conception, provide plausible fit in a direct and a priori test using CFA as the method?

RQ 3: Does a three-factor (Factor 1, Factor 2 and Factor 3) measurement model for the K-LoC43 (Kambara, 1987), consistent with Kambara's (1987) a posteriori EFA analysis, provide plausible fit in a direct and a priori test using CFA as the method?

RQ 4: Does a three-factor (Effort, Contingency and Environment) measurement model for the K-LoC43 (Hosaka, 2007), consistent with Hosaka's (2007) a posteriori EFA analysis, provide plausible fit in a direct and a priori test using CFA as the method?

RQ 5: What changes should be made to improve the instrument in general and inform future adaptation for tertiary education based on focus group analysis?

## **Methodology**

This research is conducted under a mixed methods sequential explanatory/exploratory design, with the quantitative data collection and analysis (3.1) being followed by qualitative data collection and analysis (3.2) with the dual aims of assisting in explaining and interpreting the results of the quantitative study as well as exploring possible ways to improve and adapt the instrument to the tertiary level. It should be noted however that the results from the CFAs have special status as direct and a priori tests of models hypothesized for the dimensionality of scores making up the K-LoC18 and K-LoC43.

## **Quantitative Methodology**

The quantitative methodology is reported in terms of the instrument under study, the participants, the data collection procedure, and the analytical procedure. The analytical procedure is further sub-divided into three phases, namely, descriptive statistics, the *a priori* statistical analysis (i.e. CFAs of all models and associated Cronbach's alphas for factors in each model), and finally, the *a posteriori* statistical analysis (i.e. the EFAs conducted on both versions, the K-LoC18 and the K-LoC43).

A data set (N = 1125) was collected from Japanese high school students, and four models were tested a priori; two were the originally hypothesized models for the short and long forms, and two were based on models derived through EFAs and which are present in the literature. The purpose of this study was twofold, and as follows: 1) to evaluate the plausibility of these models for an instrument which has been widely used in the Japanese research context thus far, and in the absence of such a priori testing; and 2) to establish whether this instrument could be a viable proxy for measuring learner autonomy in Japan, given that learner autonomy is notionally related to locus of control, that it is an area of vital interest to language teachers and researchers, and that failure to measure it adequately remains a perpetual obstacle to progress. The second purpose is fundamentally premised upon the first. If a suitable measurement model, consistent with theory, can be identified, and if scores from the instrument can be shown to have structural properties in-line with this model, then the instrument can be extended to the domain of language teaching.

A series of EFAs were also conducted on scores generated by the instrument (both versions, i.e. the K-LoC18 and K-LoC43) to gain further insights into their properties. The purpose of the EFA tests, which as a method is subordinate to the CFA, was to further explore the data set in order to provide additional explanatory value to the results of the CFA, as well as to the outcomes of the FG discussions.

Both the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity indicated that the data matrix was suitable for EFA to be conducted. The first EFA for each of the two instruments was based on an evaluation of the original un-rotated extractions, taking into consideration the eigenvalue greater than one rule (Kaiser, 1960) and the scree plots (Cattell, 1966). The second was based on the two factors which inform the original conception for the instrument; i.e. Internal and External. The purpose of this second EFA was to scrutinize for items which may not be expressing the latent construct they were

originally purported to measure.

## **Qualitative Methodology**

Five mixed-gender FGs consisting of first-year university students majoring in the sciences were created with each FG ranging in size from four to six members each, with a total of 27 participating students (N = 27). Members were randomly assigned to a FG. The author was also a participatory member in the discussions, and all sessions were audio recorded, transcribed and translated by the author.

With regard to the rationale for choosing university students as opposed to high school students, there were several reasons for this decision. The first was the matter of practicality, as university students are far more accessible for such extended length discussions. It was felt that it would be a significant incursion on the scheduling of participating high schools to ask for FG sessions with students who had already given time to answer the K-LoC43. Second, the instrument was originally not restricted to the high school level, but was modified by expansion (into the K-LoC43) to include questions relating more specifically to high school students' lives. It is reasonable to assume that first year university students who have only just graduated from high school, should have a fairly similar perspective, or perhaps even better insights into their high school days, having just completed their high school lives. Also, the FGs were not being tested/surveyed via the instrument for their levels of I-LoC and E-LoC but were rather evaluating the instrument, especially the content and language used therein. Obviously, these participants' evaluation of the instrument does not constitute expert criticism; they are after all not psychometricians. However, much of their evaluation involves the exercise of critical faculties which would not be as developed among high school students, especially the more junior high school participants. Finally, it was a secondary aim to be able to gain insights on changes that might make the instrument more appropriate for the tertiary level in Japan, and thus the university student FGs were the ideal choice for this study.

## **Results**

The results of this study indicated that all models represented unsatisfactory fit to the dimensionality of the data. Thus the answers to RQs 1 through 4 were all negative. This has implications for past and future research, which relies upon structurally valid measurement, and also has implications for the practitioner. It also means that the instrument, as it stands in unrevised form, is not suitable as a proxy measure for autonomy, despite its significant presence in the literature.

The EFAs conducted on both of the instruments produced interesting and useful data. The first notable result was that the two-factor extraction on the K-LoC18 showed loadings which were in accordance with the original conception of the instrument. This is a positive result for the instrument from one point of view, and had a more powerful CFA not been conducted in this study (unlike previous studies where EFA is the only analysis brought to bear on the data), this positive result could be taken as the definitive one. However, in light of the fact that the CFA showed poor model fit, and the fact that these first two components explain only 32% of the variance, such positive views must be tempered by the analytical point that approximately 68% of the variance remains unexplained. The EFAs conducted for both the short and long form suggest that increasing the number of items from 18 to 43 does not actually represent a step forward, but rather a reduction in the clarity of the scale. This is an unfortunate situation given that the K-LoC18 is rarely if ever used in the Japanese research context, whereas the K-LoC43 is quite frequently used. It may be that researchers are erroneously laboring under the false premise that more items are better, and that the revised instrument inevitably improves

upon its predecessor. Such problems have occurred with other such long LoC instruments such as the Reid-Ware Three-Factor Internal-External Scale, as detailed in the literature review, wherein the longer, 45-item version, was also found to produce better results when shortened. The lengthening of an instrument becomes even less beneficial when it is done by adding further items which have significant overlap in content, as the extra items are not broadening the operational expression of the construct being measured, but rather are simply introducing operational redundancy

With regards to the qualitative results which emerged from the Focus Group (FG) study conducted in order to gain more insight into how students experienced and responded to questions comprising the K-LoC43 instrument. The results of this study also showed that the use of FG yielded a wide variety of useful insights into how students reacted to individual items; insights which would not be obtainable by a purely statistical approach to the data.

Based on the FG feedback from the students, the changes that should be made to the instrument, and the answer to RQ 5 is elaborated next. First, it is highly recommended that the instrument be shortened, preferably with five or six items per construct. Second, it is necessary to eliminate repetitive items. Third, instrument modifications should be made to avoid the unclear and age inappropriate items. Fourth, it is recommended to make the Likert scale more refined by adding further points of discrimination to it, and reverse the ordering of the scale to conform with the what the students perceive to be the standard ordering. Fifth, and finally, it is advisable to adopt FG suggestions for Japanese language changes where appropriate or necessary.

## **Conclusion**

It is the author's opinion, based on the evidence, that at a very minimum, given 1) the multitude of criticisms of the length of the K-LoC43 2) it having failed the CFA test for model plausibility 3) the shorter K-LoC18 two-factor EFA showed much better factor loadings for the underlying LoC constructs on an item per item basis, that the K-LoC18 be used in place of the K-LoC43. This should be until such time as a revised Kambara LoC scale which is even shorter, less redundant, and more refined in terms of Likert scale discrimination points can be tested and confirmed with a priori data from the Japanese population. The K-LoC18 also has the benefit of being more of a general LoC instrument and thus useable in the university context with no explicit item alterations required. In terms of a future revision of the K-LoC43, a useful starting point would be to make a ten to twelve item scale using the 5 or 6 items (taking into consideration redundancy) from both I-LoC and E-LoC, which had the highest factor loadings.