Measuring Learner Autonomy: A Confirmatory Factor Analysis of Scores on an EFL Learner Autonomy Questionnaire

by
Kayoko Horai

Abstract

Four different models of the EFL Learner Autonomy Questionnaire, which was originally developed by Shimo (2008), are examined for plausibility in this study. While the importance of learner autonomy in language education has been frequently discussed, there has been little quantitative research because of the lack of suitable instrumentation (Apple, 2011; Benson, 2001; Macaskill & Taylor, 2010; Shimo, 2008). The instrument examined in the present study is a self-report scale comprising 18 items originally scored under a two-factor model (Shimo, 2008). In subsequent research, Shimo (2009) tested four other models emphasizing two of these four as preferable. One was a two-factor solution with 16 items and was advanced in the pre-survey phase of the study, and the other was a three-factor solution with 16 items advanced in the post-survey phase of the study. Apple (2011) questioned these models and conducted research suggesting a two-factor solution with 13 items. In this study, these four hypothesized models emerging out of the literature from the respective authors were examined under a Confirmatory Factor Analysis. Data was collected from 388 university students in western Japan in 2012. Results indicated that none of the models were plausible for scores in the a priori test reported in this study.

Key Words:　confirmatory factor analysis, learner autonomy, measurment scale

1. Introduction

The term 'learner autonomy' has been used over the last four decades in the field of applied linguistics. The initial introduction of the term in language education is known to have a connection to the emergence of a self-access center as part of the Council of Europe's Modern Languages Project in 1971 at the University of Nancy in France. Since then, many scholars and educators have advocated the importance of fostering learner autonomy as an important goal in the field of education (Benson, 2001; Benson & Voller, 1997; Little 1991), and this advocacy continues. The concept of learner autonomy in Japan has gained an increasing presence particularly over the past decade. This has been strongly influenced by accelerating globalization and rapid technological advancements. The Ministry of Education, Culture, Sports, Science and Technology (MEXT, 2003) has announced a policy outline called the "Action Plan to Cultivate Japanese with English Abilities" which cites the development of communicative competence in English among Japanese people as

＊Advisor/Lecturer, Sojo University

an important educational goal. This policy includes a focus on autonomy to foster the notion of life-long learning among Japanese. This policy shift and educational reform has changed English learning environments in public schools dramatically. It has emphasized that the focus of teaching should shift to communicative competence from the long-established grammar translation approach in Japan. This reform has drawn dichotomous attention focusing on the positive and negative impacts for learners and teachers. Shifting to Communicative Language Teaching (CLT) may not be as easy to implement as MEXT might have imagined without good preparation for the transition. One of the reasons for difficulty in the transition and implementation is the gap between the idea and the confronted reality. Most teachers who will have to undertake the changes have learned in a traditional way (i.e. teacher-centered, grammar-translation approaches) and often successfully, and therefore some of them may not sit comfortably with CLT methods. Put another way, they may feel that if they only focus on CLT, they will not be able to fully support their students in passing entrance exams which remain in place and which are a critical hurdle for students. Thus there is a dilemma because the two different goals (communicative competence and success in entrance exams) require pedagogies with a different emphasis.

CLT emphasizes authenticity of language input, and it involves interaction among students and their teacher. It implies a heavier workload of preparation for teachers who have been using traditional teaching styles. To deal with natural conversation as a teaching tool requires a broad knowledge base from teachers and confidence in their own communicative skills which are on display in far less controlled circumstances than in a teacher-fronted grammar class. Also, grammar-focused activities, pattern practice and translation may be easier to make lesson plans for because there is more explicit structure, answers are not so ambiguous, and objectives are easier to define.

From the learner point of view, while many learners seem to enjoy learning through a communicative approach, they, at the same time, still need to perform well on their tests and prepare for entrance exams to enroll in tertiary institutions. Thus young Japanese students have to focus on improving two different facets of English-language performance; one is communicative competence and the other is knowledge of English as a tested language. These two facets are not easy to reconcile and for the student the most immediate and tangible consequences for failure relate to the latter facet.

Having said that the most immediate and tangible consequences for failure relate to underperforming in English as a tested language, English has become the world's lingua franca, and it is necessary for Japanese to be able to use English as a communicative tool with other nations. Thus there are less immediate, and probably less tangible (but real and important), consequences to failure with regard to communicative competence. In this context of pedagogical dilemma, learner autonomy emerges as an important aspect in reconciling the two learning focuses the student needs to maintain for overall success, because it very plausibly plays a vital role in negotiating diverse learning purposes. It also plays a role in helping the learner to progress in contexts for learning other than those formulaically offered by classroom education, thus leveraging what the teacher and formal-education classroom has to offer. Therefore, successfully fostering learner autonomy is not only an idealistic goal in English education, but also a necessary component for a successful national strategy for English education in Japan.

The number of articles and books published about learner autonomy within the EFL literature has increased, but the evidence offered by sound quantitative studies showing learning outcomes is insufficient. A dearth of empirical research has

been cited by Apple (2011). He states that "there is still an overall lack of generalizable empirical evidence that autonomy leads to increased foreign language proficiency" (p. 194).

The present study explored the structural validity of scores generated by an existing instrument, the EFL Learner Autonomy Questionnaire (LAQ), developed by Shimo (2008). This instrument comprised 18 items in the original version and a two-factor solution was advanced by the author in the original study of 2008. However, a subsequent study by Shimo in 2009 examined four alternative models with two of these being explored in pre-survey data and two in post-survey data. These models were also based on a reduced set of items (16 items from the earlier 18 items). Apple (2011) conducted research aiming to understand whether a two or three-factor solution best fitted the dimensionality of scores generated by the instrument. Apple's (2011) research supported the two factor solution, but the number of items was further reduced to 13 from 18. These various models emerging from previous literature suggest that questions still remain concerning the structural validity of the scores it generates. This study addressed this deficit, incrementally, by testing the models which have emerged in the literature thus far in a direct, a priori test using a new dataset.

2. Learner Autonomy

The difficulty of defining learner autonomy is well known. One of the commonly cited definitions of learner autonomy in language learning was offered by Holec (1981) who referred to it as "the ability to take charge of one's own learning" (p. 3). In all definitions, however, one will detect the notion that autonomy refers to some sense of personal agency which can potentially find expression in different domains such as one's own personal life, work life or educational life. In language education, learner autonomy often refers to behaviors which one would presume to come from a person with a strong sense of personal agency such as setting goals, making and carrying out plans, choosing materials, reflecting on methods of learning, using strategies, actively asking questions of others, and so on.

The emerging and developing interest in autonomy in education was initially influenced by human rights movements in Europe in the 1960s (Gremmo & Riley, 1995). Gremmo and Riley also cite "shifts in educational philosophy, reactions against behaviorism, linguistic pragmatism, wider access to education, increased internationalism, the commercialization of language provision and easier availability of educational technology" as important drivers of the gathering interest in the notion of autonomy (p. 151). Benson (2009) has illustrated how the concept of autonomy in language learning was originally imported from outside of the field in studies related to areas such as moral and political philosophy, and he points to this as the reason why it is such a complex and multifaceted concept.

The characteristics of autonomy may be influenced by various factors, including a socio-cultural factor. And while it is difficult to frame and illustrate autonomy thoroughly, there is no doubt that it is an important aspect of the individual's ability to function adaptively and effectively within the social milieu. The autonomous intervention of a single individual in his or her social space influences society and the actions of the other. The importance of autonomy for membership of society was emphasized by Benson (2001) who stated that "when learners succeed in developing autonomy, they not only become better language learners but they also develop into more responsible and critical members of the communities in which they live" (p. 1).

But besides these factors related to the individual in social space, the important issue from a pedagogical point of view is that autonomy plausibly supports behaviors conducive to achieving an educational goal. Most learning

requires action and endurance. Second language learning occurs over a long period of time and the leaner's motivation may decrease over this period for several reasons, but most importantly because progress is incremental and often not tangible. Learners are not aware of their own responsibility in this process, have no freedom with respect to their own learning, are forced by others to study, or otherwise have diminished personal agency. In this context it is important to recognize that autonomy is variable among individuals, and it is not stable even in the same person over time. It is arguably also something subject to intervention if a teacher is sensitive to it and wants to act to change it in a student or group of students. Therefore, it is important for educators to recognize learners' autonomy levels to support and foster its growth for better learning. This is not only important with respect to meeting the immediate learning goal, but also with respect to life-long learning.

2.1 Importance of learner autonomy in EFL in Japan

Japan's EFL context offers limited opportunities for exposure to English in authentic contexts. To increase the opportunities for using English, each learner has to make an effort beyond the classroom which means each learner has to be more responsible for his or her own learning. This effort requires a high degree of autonomy and requires that the learner take charge of much of their own learning process. Moreover, as previously stated, educational reform in English education has resulted in their being two demands upon the learner-knowledge of English as a tested language and English communicative competence. Both are significant for learners and it is essential for learners to control their learning beyond the classroom to achieve success, especially when either or both of them are receiving insufficient attention for improvement in formal education. Thus, learner autonomy is an essential component of dealing with the challenges which Japan faces in the EFL context.

2.2 Instruments for Measuring Learner Autonomy

If learner autonomy is, therefore, an important area for research and pedagogical intervention in the Japanese EFL context, then its presence and the levels of its presence need to be able to be measured. Teachers need instruments which generate reliable scores so that they can identify students whose levels of autonomy put them at risk of failure, and researchers need the same instruments to place research concerning autonomy on a secure measurement footing.

A little over 10 years ago, Benson (2001) stated that "[f]or the purpose of research and the evaluation of practice, it would indeed be convenient if we had a reliable method of measuring degrees of autonomy" (p. 51). Apple (2011) has more recently reported the remaining deficit with respect to good instrumentation in the area of learner autonomy in a study examining scores for a more recently developed instrument in the area of learner autonomy (Shimo, 2008). Shimo developed the questionnaire to measure learner autonomy in the EFL domain and concluded with a two-factor measurement model for 18 items. The author conducted subsequent research in 2009 with the same instrument. In this study, the author examined the components emerging from scores on the instrument over time. The research questions asked by Shimo (2009) were: 1) "[w]ill the components change over time?" and 2) "[i]f so, what will explain the changes?" (p. 36). Arguably, the research questions were not well-conceived because instrumentation presumes a certain level of stability with respect to dimensionality over time and seeks to measure changing levels on these dimensions. The point is that it is the levels measured of a construct which are presumed to change over time, and not the construct itself. If constructs change over time, then longitudinal measurement is not possible because the tool being

used for measurement is unstable. Shimo's research questions seem to imply an assumption that the instrument is inherently unstable overtime, and her findings imply such instability given that a different factor structure was extracted in the pre-survey and the post survey (a two-component solution changed to a three-component solution).

Given the research record with respect to Shimo's instrument, the competing models for its scores, and the absence of a Confirmatory Factor Analysis (CFA) to directly test models which have so far emerged, research reported in this study examines all four models of the LAQ which have emerged in Shimo (2008, 2009) and Apple (2011) under a CFA.

2.3 Models of the LAQ

In the section below, the models which have emerged, (these models are labeled for testing in this study), and the rationale of each model, as reported by the respective authors, is outlined. The models are labeled as follows for convenience: Model 1 is a two-factor solution with 18 items which emerged in Shimo (2008); Model 2 is a two-factor solution with 13 items which emerged in Apple (2011); and Models 3A and 3B represent two models which emerged in Shimo (2009). Model 3A is a two-factor solution with 16 items from the pre-survey conducted in Shimo (2009), and Model 3B is a three-factor solution with the same 16 items from the post-survey conducted in the same study. Shimo (2009) also examined two additional models, but these models were not ultimately advanced in her study and they are not tested in this study.

2.3.1 Model 1

The phases for developing the LAQ took the following steps as reported in Shimo (2008). Items for the questionnaire were sought in light of five hypothesized dimensions of learner autonomy which were: the capacity to "set learning goals," "plan one's own learning," "choose learning materials," "reflect upon one's own learning," and "evaluate one's progress and growth" (p. 158). However, these five hypothesized dimensions were re-categorized into four hypothesized dimensions, because some were argued to be closely related to each other and re-naming was as follows: "a) setting goals and planning learning processes, b) evaluation of learning ways and tools, c) evaluation of one's own language abilities, and d) creating learning opportunities" (p.158). Following this, the instrument that was now hypothesized to have 4 factors with 17 items was examined in a pilot survey with 54 university students. A Principle Component Analysis (PCA) was conducted with the data set, and a three-factor solution was then obtained with these factors being labeled as follows: "orientation for reflecting on learning processes," "orientation for enhancing learning opportunities," and "orientation for reflecting on language abilities" (p.159). In the revised questionnaire after the pilot study, 3 items were eliminated and 4 new items were created for Factor 3 leading to an 18-item instrument. Finally, the revised model was examined again in the main study with a sample of 106 university students (non-English majors). The author reported that two components were extracted from the scores using a PCA, thus contradicting the initial solution of three factors in the pilot study. Two components were then extracted under a Varimax rotation which accounted for 46% of the variance. Although the reasons for selecting two components are under-reported in the study, the explanation was that:

> [t]wo of the hypothesized three factors in the pilot study involved reflecting acts (i.e., reflecting on learning processes and reflecting on language abilities), but delineating between the two is difficult because learners use the language to learn as well as learn the language to use and language learning and actual abilities which are reflected in the use of the language often overlap each other. (p. 165)

In summary, at the final stage of the initial development of the instrument in Shimo (2008),

the author concluded that scores for the instrument model were best explained by a two-factor solution with 18 items comprising the total instrument. The two factors were labeled as "orientation towards improving learning environment" and "orientation towards reflective learning" (p.165). Items which loaded on the first factor were Items 1, 16, 8, 11, 10, 5, 14, and 4. Items which loaded on the second factor were Items 12, 13, 7, 15, 6, 18, 2, 3, 17, and 9.

2.3.2 Model 2

Shimo (2009) conducted a second study (reported on below) using the LAQ and concluded with two different models, arguing that the structure of scores had changed over time, namely, between the period of the pre-survey and the period of the post-survey. One of these models was a two-factor solution (pre-survey) and the other was a three-factor solution (post-survey). Apple's (2011) research rested on the claim that "whether the questionnaire contains two or three factors remains open to examination" (p. 196).

He examined the validity of scores generated by the instrument, and also studied their relationship with English-proficiency levels measured on the TOEIC. He employed Rasch measurement analysis, parallel analysis (which was reported as a separate analysis but is actually an analysis to determine the number of factors to extract in an Exploratory Factor Analysis [EFA]), EFA, and multiple regression analysis. The data was collected from 204 engineering students from a Japanese technical college. After data collection, 19 participants were removed from the data set, because they showed inappropriate responses which the author reported as "incomplete, illegible or set response patterns (i.e., answering all questions with the same response) that would adversely affect the correlational matrixes in later analyses" (p. 196). In the Rasch analysis, 5 (Items 3, 4, 9, 14, and 17) of the 18 items were removed prior to the EFA, because of misfit. The data for the revised 13 items were factor analyzed into two components. This two-factor solution with 13 items is labeled as Model 2 for testing purposes in this study. The first factor for the model comprised Items 8, 1, 5, 16, 11, 18, and 15 and the second factor comprised Items 12, 6, 7, 13, 2, and 10.

2.3.3 Model 3A and 3B

Shimo (2009) conducted a subsequent study with the LAQ after the initial launch of the instrument in 2008. The stated purpose of the study was to investigate the types of components of learner autonomy in language learning, and to study whether the components change over time (Shimo, 2009).

Data collection was conducted in the fall semester of 2007 in the first phase (pre-survey) with 159 university students participating. The second phase (post-survey) was conducted in the 2008 spring semester and 151 students participated. All data (pre-survey and post-survey) was then bundled and examined for suitability for factor analysis. The Kaiser-Meyer-Oklin value was calculated and the Barlett's Test of Sphericity was conducted. This approach of bundling is arguably unorthodox as responses from participants occurred twice in the overall data set and are therefore not independent. In a PCA, 2 items (Items 9 and 17) were eliminated for the subsequent analysis of both data sets. For further details on the rationale for removing these two items see Shimo (2009, p. 38). This procedure of eliminating two items from both data sets is questionable. It is not explicit in the article, but it implied these two data sets were bundled at the time of elimination, which undermines the rationale for the study which was to examine change in dimensionality over time. It would not make sense to bundle two sets of scores with different dimensionality.

In subsequent analyses four models emerged. Two models emerged for the pre-survey scores, and two models emerged for the post-survey scores. In the case of each set of scores one model was

advanced by the author and one was discarded or not supported by the author. The models advanced by the author are covered below.

2.3.3.1 Model 3A

The 16 items in the pre-survey were subjected to a PCA, and the author decided on a two-factor solution including all 16 items entered into the PCA. This model is labeled as Model 3A in this study. Prior to the PCA, the Kaiser-Meyer-Oklin value was calculated (.88) and Barlett's Test of Sphericity was conducted and was reported to have reached statistical significance. The eigenvalue-greater-than-one rule indicated 4 factors, but the two-factor solution was chosen based on interpretation of a screeplot, which is how Model 3A emerged with two factors. The two component solution was then analyzed under a Varimax rotation. Loadings under this rotation were reported by the author. Items 8, 1, 11, 16, 5, 10, 4, 2, and 14, loaded on Factor 1 and Items 12, 7, 13, 3, 15, 18 and 6 loaded on Factor 2. These two factors were labeled as "orientation towards enhancing learning opportunities and planning learning processes" and "orientation towards reflecting on learning processes and language abilities" (p. 43).

2.3.3.2 Abandoned Model in Pre-Survey (Shimo, 2009)

An additional three-factor solution with 16 items was examined for scores in the pre-survey which was to be compared with the two-factor model (labeled as Model 3A in this study and covered immediately above). The rational for testing three components was not made explicit. The author stated that "PCA revealed the presence of four components with eigenvalues exceeding 1" (p. 38), but the four-factor solution was abandoned after inspecting and interpreting the screeplot. Also, instead of extracting the four-factor solution suggested by the abandoned criterion of the eigenvalue-greater-than-one rule, the author extracted three components. Another statement related to this issue was that the "[t]he three-component solution was also conducted in order to compare the results" (p. 40). But this does not serve as the rationale for three factors rather than four. The reason for this comparison was not clear and the reader has to make some presumptions for why the three-factor solution was extracted, and these presumptions relate to the first article by the author (Shimo, 2008) with respect to this instrument and the appearance of a three-factor solution in that article. She hypothesized three factors as a result of the pilot study, but concluded on two factors after the main study (Shimo, 2008). The 2009 study implies the initial three-factor solution appearing in the first study at one point in what was reported in that study (the pilot-study point), even though this was not the final solution settled upon in the main study, when Shimo states that "[t]he revised questionnaire consisted of 18 items based on the three re-hypothesized components" (p. 36). Thus, in the study of Shimo (2009), neither the eigenvalue-greater-than-one rule nor inspection of the screeplot indicated three components, but the author did extract a three-factor solution anyway.

This three-factor solution presented with loadings as follows: Factor 1 loaded with Items 8, 5, 1, 11, 4, and 2; Factor 2 with Items 12, 3, 7, 13, 6, 15 and 18; and Factor 3 with Items 14, 16, and 10.

In the present study, this model was not examined because the author (Shimo, 2009) did not advance this model as the more plausible of the two examined on the pre-survey data (i.e. the two-factor solution which is Model 3A in this study, and this three-factor solution).

2.3.3.3 Model 3B

Model 3B comprises 16 items under a three-factor solution and was derived on the post-survey data reported in Shimo (2009). The Kaiser-Meyer-Oklin value was calculated (.89) and Barlett's test of Sphericity was conducted and statistical

significance was reported as having been reached. A PCA was run and three components were extracted according to the eigenvalue-greater-than-one rule. Inspection of a screeplot was also reported to reveal a break after the third component, and it was decided to retain three components. Varimax rotation was then performed for further inspection. Factor 1 loaded with Items 7, 18, 15, 13, 2, 10, and 14. Factor 2 loaded with Items 1, 11, 8, 16, and 5. Factor 3 loaded with Items 3, 12, 6, and 4. These factors were labeled as "orientation towards reflecting on and planning learning processes," "orientation towards enhancing learning opportunities," and "orientation towards reflecting on language abilities" (p. 44).

2.3.3.4 Abandoned Model in Post-Survey (Shimo, 2009)

An additional two factor solution with 16 items was extracted in the post-survey set of scores in Shimo (2009) to compare with Model 3B. There were ambiguities in the rationale for the alternative factor solution that were similar to the ambiguities in the rationale for the similar comparative extraction on the pre-survey data. First, the reason for comparing a two-factor solution was not made explicit. The author had reported that the PCA suggested three components under the eigenvalue-greater-than-one rule and this was supported by inspection of the screeplot. There is no explicit explanation of the rationale for extracting a two-factor solution for comparison against this three-factor solution. The two-factor solution loaded as follows: Factor 1 loaded with Items 12, 13, 15, 18, 6, 14, 2, 3, 10, 7, and 4; and Factor 2 with Items 1, 11, 8, 16, and 5. Shimo abandoned this model as the less plausible of the two for the post-survey data, and thus this model is not tested via CFA in this study.

3. Research Question

In this study, four different models advanced in previous studies for the LAQ (Apple, 2011; Shimo, 2008, 2009) and discussed above (Models 1, 2, and, 3A and 3B) are directly tested a priori on a new dataset collected for the purpose. The research question for this study is therefore as follows:

RQ: Are any of the previous models for the LAQ, advanced in previous literature, plausible in an a priori test using CFA as the method?

Answering this question directly will be informative with respect to the potential for the instrument to be used for classroom diagnostics or research using scoring regimes aligned with models advanced in the literature so far.

4. Method

Permission to use the instrument (Shimo, 2008; 2009) was obtained from the author before data collection. The method is outlined below in terms of procedure, instrument, data set and data analysis.

4.1 Procedure

The data was collected in 2012 from three different universities in western Japan. The survey was conducted in May, 2012, which was approximately a month after the new educational year starts (April). All collected data was scanned visually to identify any questionable cases such as incompletion and false responses (e.g. marking all the items with the same responses). Nine cases were removed from the data set in this process. All remaining data was input into a Microsoft Access database and then imported into IBM/Statistical Package for the Social Sciences (SPSS). Following this, a second phase of data scanning was conducted which involved looking for missing responses (i.e. one or two responses were omitted). There were 14 cases where missing values occurred. The omissions were inspected across cases for systematic patterns, and none were found. In other words, omissions were observed to be at

random. Thus these 14 cases were eliminated from the dataset without the threat of biasing it. Therefore, 23 surveys were eliminated in total out of the 411 initial participants leaving a total sample of N=388. Descriptive statistics were conducted to review the normality of distribution for scores on the items. Finally, a CFA was conducted to directly test the models (Models 1, 2, and, 3A and 3B).

4.2 Instruments

The LAQ (Shimo, 2008; 2009) was used in this study. A cover letter indicating the procedure for informed consent, and some questions seeking biographical information were printed in Japanese. The instructions for informed consent told the participants that completing the questionnaire and providing personal information such as major, grade, age and gender was an indication of voluntary consent, and not completing it was an indication of no consent. This was done so that the participants did not have to offer consent by signing the document with their name, thus surrendering anonymity. The cover letter stated that their identity would remain anonymous and that cooperation was completely voluntary.

4.3 Data Set

There were no particular problems reported by the instructors who participated by conducting the survey in their classes. The cover letter in Japanese included a request to respond faithfully to both questionnaire items and all questions concerning biographical information.

There were 338 participants from one university (National University Corporation), 43 from a second university (Prefectural University Corporation) and 7 from a third university (Private University). Majors varied: Literature (142), Engineering (93), Science (56), Education (45), Law (6), Medicine (19), Pharmacy (1), Economics (3), Social Welfare (22), and Commerce (1).

The gender distribution of the participants was almost equal with 192 male students and 196 female students. Their ages ranged from 18 to 25 years old. There were 153 first-year students, 118 second-year students, 112 third-year students and 5 fourth-year students. Almost all students were Japanese (377), but there were 5 Korean students, 5 Chinese students, and 1 Vietnamese student.

4.4 Data Analysis

The normality of distribution of scores for each item was assessed by considering skew and kurtosis. CFA was then conducted using AMOS (5.1) which is structural equation modeling (SEM) software.

5. Results

The results section consists of two sub-sections, namely, Descriptive Statistics for Items on the LAQ (5.1) and Confirmatory Factor Analysis for Items on the LAQ (5.2). For Section 5.2, each model tested is presented in its own section.

5.1 Descriptive Statistics for Items on the LAQ

The values for skewness and kurtosis provide an indication of the normality of distribution for scores on each item. Table 1 presents the descriptive statistics of the 18 items making up the LAQ, as well as information on skewness and kurtosis. The values for skewness and kurtosis (which are represented in Table 1) were divided by the value for the standard error (also represented) resulting in a critical ratio which was then compared against a criterion or cutoff decided in advance. There are two commonly used cutoff points. A value of 2 for the cutoff is widely used and sometimes a more relaxed cutoff value of 3 is used. In this study, items are calculated against both thresholds/cutoffs with items exceeding the value of 2 being marked in Table 1 with a single asterisk and items exceeding the value of 3 being marked with a double asterisk.

With respect to skewness, Items 1, 3, 4, 5, 6, 9, 11, and 12 exceeded the value of 3, and these are

Table 1. Descriptive statistics.

	Mean		Std. Deviation	Skewness		Kurtosis	
	Statistic	Std. Error	Statistic	Statistic	Std. Error	Statistic	Std. Error
Item_1	2.43	.064	1.263	**.787	.124	-.002	.247
Item_2	3.20	.062	1.215	.141	.124	-.457	.247
Item_3	4.00	.060	1.177	**-.468	.124	-.074	.247
Item_4	4.39	.059	1.168	**-.688	.124	.290	.247
Item_5	2.43	.067	1.322	**.819	.124	-.083	.247
Item_6	4.38	.060	1.185	**-.820	.124	.482	.247
Item_7	2.99	.060	1.183	.245	.124	-.191	.247
Item_8	3.30	.071	1.395	.149	.124	**-.753	.247
Item_9	4.44	.054	1.068	**-.752	.124	**.847	.247
Item_10	3.40	.065	1.277	.057	.124	*-.626	.247
Item_11	2.29	.064	1.261	**.996	.124	.487	.247
Item_12	4.09	.059	1.160	**-.471	.124	.025	.247
Item_13	3.65	.060	1.179	-.195	.124	-.356	.247
Item_14	3.59	.071	1.398	-.064	.124	**-.853	.247
Item_15	3.43	.065	1.285	-.054	.124	*-.591	.247
Item_16	3.28	.073	1.444	.213	.124	**-.809	.247
Item_17	3.68	.065	1.275	-.238	.124	-.392	.247
Item_18	3.38	.066	1.310	.026	.124	*-.587	.247

Note: For skewness and kurtosis, items which failed to meet the stricter cutoff/threshold of less than absolute-value 2 are marked with an asterisk, and those which failed to meet the more relaxed threshold of less than absolute-value 3 are marked with two asterisks.

marked with a double asterisk. Items 2, 7, 8, 10, 13, 14, 15, 16, 17 and 18 met the stricter criterion/cutoff of 2 (10 items out of 18 items, or 56 % of the items).

With respect to kurtosis, Items 10, 15, and 18 did not meet the stricter criterion of 2, and these items are marked with a single asterisk. Items 8, 9, 14, and 16 exceeded the criterion of 3, the more relaxed criterion, and these items are marked with a double asterisk. The results indicate that 11 items met the stricter criterion/cutoff of 2 (61 % of the items). Non-normal distributions for some items are a problem for the instrument and are important areas for revision of the instrument. For the purposes of this study, these items were left intact so that previous models advanced in the literature could be directly tested in exactly the way they have appeared in the literature and on the items which have underpinned these models.

5.2. CFA of the EFL Learner Autonomy Questionnaire

All models advanced in the previous literature and tested in this study had their respective model fit judged using four indexes which were the Tucker-Lewis index (TLI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA) and the standardized root mean squared residual (SRMSR). The selection of four indexes with different rationales underlying their conception helps to triangulate the decision as to whether the model is satisfactory or not. The values produced for these indexes are typically interpreted in terms of a criterion or threshold/cutoff; the most widely used and recommended being those criteria offered by Hu and Bentler (1999). The cutoffs recommended by Hu and Bentler for the four indexes selected for this study are as follows: TLI, >.95; CFI, >.95; RMSEA, <.06; SRMSR, <.08. These cutoffs are used for judging model fit for all models below.

5.2.1 Model 1

The 18-item, two-factor model emerging from Shimo (2008) is represented in Figure 1. The model has 171 distinct sample moments, 37 distinct parameters for estimation and 134 degrees of freedom. The model was overidentified.

The results for fit indexes were as follows (cutoffs in parentheses): TLI, .84 (>.95); CFI, .86 (>.95); RMSEA, .09 (<.06); SRMSR, .07 (<.08). Only the SRMSR produced a satisfactory value in terms of the associated cutoff. The values for the other indexes were not satisfactory.

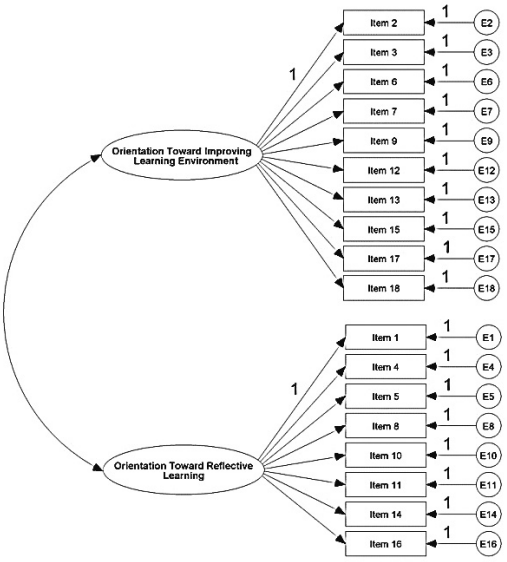


Figure 1. Confirmatory factor analysis model of the two factors underlying the 18-item version of the instrument as hypothesized in Shimo (2008).

5.2.2 Model 2

The 13-item, two-factor model emerging from Apple (2011) is represented in Figure 2. The model has 91 distinct sample moments, 27 distinct parameters for estimation and 64 degrees of freedom. The model was overidentified.

The results for fit indexes were as follows (cutoffs in parentheses): TLI, .82 (>.95); CFI, .85 (>.95); RMSEA, .12 (<.06); SRMSR, .08 (<.08). Only the SRMSR produced a marginally satisfactory value in terms of the associated cutoff. The values for the other indexes were not satisfactory.

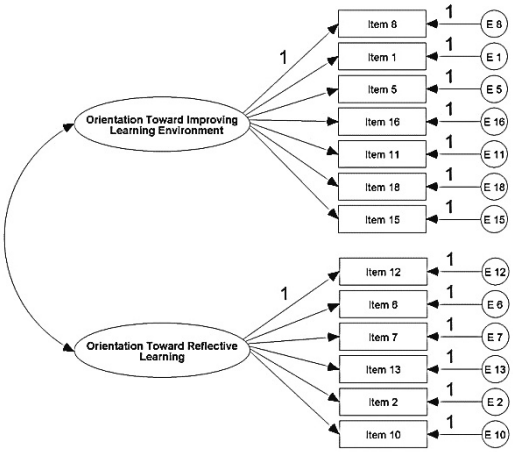


Figure 2. Confirmatory factor analysis model of the two factors underlying the 13-item version of the instrument as hypothesized in Apple (2011).

5.2.3 Model 3A

The 16-item, two-factor model emerging from a pre-test in Shimo (2009) is represented in Figure 3. The model has 136 distinct sample moments, 33 distinct parameters for estimation and 103 degrees of freedom. The model was overidentified.

The results for fit indexes were as follows (cutoffs in parentheses): TLI, .82 (>.95); CFI, .85 (>.95); RMSEA, .11 (<.06); SRMSR, .07 (<.08). Only the SRMSR produced a satisfactory value in terms of the associated cutoff. The values for the other indexes were not satisfactory.

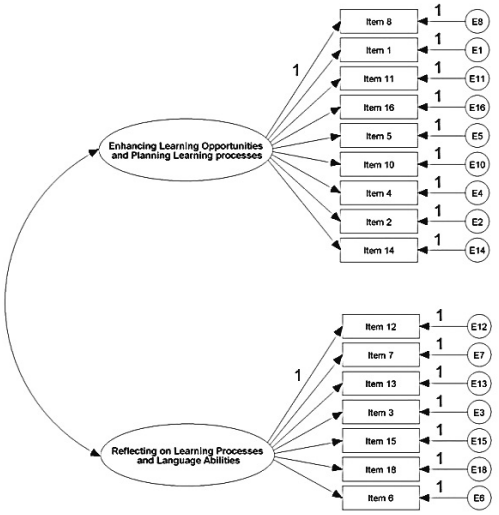


Figure 3. Confirmatory factor analysis model of the two factors underlying the 16-item version of the instrument as hypothesized in Shimo (2009).

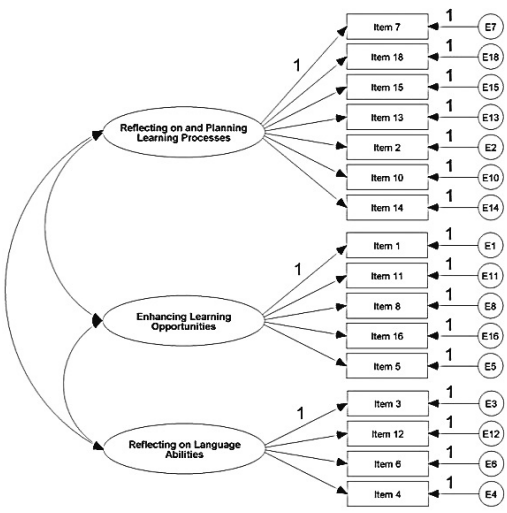


Figure 4. Confirmatory factor analysis model of the three factors underlying the 16-item version of the instrument as hypothesized in Shimo (2009).

5.2.4 Model 3B

The 16-item, three-factor model emerging from a post-test in Shimo (2009) is represented in Figure 4. The model has 136 distinct sample moments, 35 distinct parameters for estimation and 101 degrees of freedom. The model was overidentified.

The results for fit indexes were as follows (cutoffs in parentheses): TLI, .90 (>.95); CFI, .92 (>.95); RMSEA, .08 (<.06); SRMSR, .06 (<.08). Only the SRMSR produced a satisfactory value in terms of the associated cutoff. The values for the other indexes were not satisfactory.

Discussion and Conclusion

In this study, scores from the LAQ (Shimo, 2008) were examined under a range of models which emerged in Shimo (2008) and subsequent literature (Apple, 2011; Shimo, 2009). The purpose of the study was to conduct a direct and a priori test of whether any of the previous models hypothesized in past literature would fit scores from a new dataset collected for this purpose. The following discussion is presented in terms of the results reported above for each model followed by an overall set of comments about how the findings in this study should inform future lines of research with respect to this instrument.

With respect to Model 1 (Shimo, 2008), which was the model first advanced for the instrument, the results overall indicate that the model should be rejected. Both the TLI and CFI produced similar results which were, in both cases, significantly below the cutoff recommended by Hu and Bentler (1999). In addition, the RMSEA, an index which rewards for model parsimony, produced a value which was significantly higher than the threshold

adopted in this study. The only index which produced a satisfactory result was the SRMSR which is simply an expression of the size of residuals.

Model 2 was a two-factor model with 13 items advanced for the instrument by Apple (2011). The results overall indicate that the model should be rejected. Both the TLI and CFI produced similar results which were, in both cases, significantly below the cutoff, and the values for both indexes were even slightly lower than for Model 1. In addition, the RMSEA produced a value which was significantly beyond the threshold adopted in this study and which was informed by Hu and Bentler (1999). The SRMSR produced a value equal to the cutoff.

Model 3A was advanced for the instrument in the pre-survey phase of Shimo (2009) and was a two-factor solution with 16 items. The results overall indicate that the model should be rejected. Both the TLI and CFI produced values which were the same as for Model 2 and which were significantly below the cutoff adopted in this study. The RMSEA produced a value close to Model 2. The only index which produced a satisfactory result was the SRMSR.

Finally, Model 3B (Shimo, 2009) was a three-factor model with 16 items advanced in the post-survey phase of the study. The results overall indicate that the model should be rejected. However, compared with the other three models, the values produced by the various indexes were better, even though most of the values did not meet the stipulated threshold. For example, both the TLI and CFI produced slightly lower values than the cutoffs, which indicated the model did not fit well, but the values were much closer to the thresholds than for the other models tested in this study. The RMSEA indicated a similar result. The value for the index was higher than the cutoff, but was nonetheless better than for other models. The SRMSR produced a satisfactory result of .06 which was also the best result for all models. This implies that the three-factor solution may have more potential than two-factor solution for further research.

Recommendations for future research include engaging with the non-normal distributions of scores in a revision of the instrument, investigating more complex models if items remain roughly the same, or revisiting the items completely under a clear rationale for constructs which are stipulated in advance.

With respect to model complexity, it is important to note that the three-factor model performed better than any of the two-factor models. This indicates that there is probably more dimensionality in the scores than can be adequately represented in a two-factor model. It is important to note, however, that more complex models will often fit better than simpler models as parsimony is sacrificed. Therefore, the route of moving towards a more complex model for the instrument should be undertaken critically. It is important to note that the theoretical conception of the instrument, initially, involved more than two dimensions, and Shimo (2008, 2009) seems to have collapsed some of these categories with the explanation that they are associated. This may have resulted in a multi-dimensional construct from the conception, and perhaps the alternative route of completely revising the items under a construct-framework, which is clearer and more explicitly articulated, would be a more fruitful long-term approach to continuing research with respect to this instrument.

References

Apple, M. (2011). Autonomy as a predictor of English proficiency. *OnCUE Journal*, *4* (3), 191-216.

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.

Benson, P. (2001). *Teaching and researching autonomy*

in language learning. London: Pearson Education.

Benson, P., & Voller, P. (1997). Introduction: Autonomy and independence in language learning. In P. Benson, & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 1-12). London: Longman.

Gremmo, M. J., & Riley, P. (1995). Autonomy, self-direction and self access in language teaching and learning: The history of an idea. *System*, *23* (2), 151-164.

Holec, H. (1981). *Autonomy in foreign language learning*. Oxford: Pergamon.

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.

Little, D. (1991). *Learner autonomy 1: Definitions, issues and problems*. Dublin: Authentik.

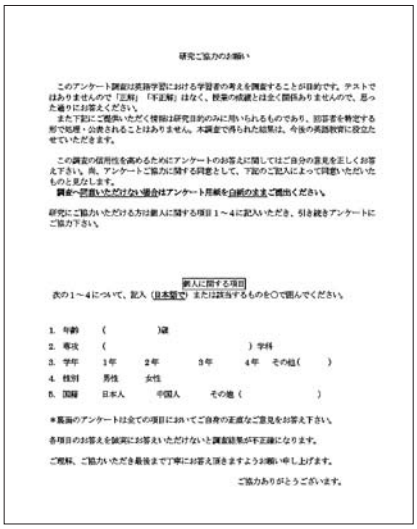
Macaskill, A., & Taylor, E. (2010). The development of a brief measure of learner autonomy in university students. *Studies in Higher Education*, *35* (3), 351-359.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2003). Retrieved October 10, 2010, from Regarding the Establishment of an Action Plan to Cultivate “Japanese with English Abilities” (MEXT): http://www.mext.go.jp/english/topics/03072801.htm

Shimo, E. (2009). Dynamics of learner autonomy in language learning: Components and changes. *JACET Kansai Journal*, *11*, 33-49.

Shimo, E. (2008). Learner autonomy and English language proficiency: An exploration among non-English majors at a Japanese university. *Kinki University Department of Language Education Bulletin*, *8* (2), 153-178 .

Appendix



研究ご協力のお願い

このアンケート調査は英語学習における学習者の考えを調査することが目的です。テストではありませんので「正解」「不正解」はなく、授業の成績とは全く関係ありませんので、思った通りにお答えください。
また下記にご提供いただく情報は研究目的のみに用いられるものであり、回答者を特定する形で処理・公表されることはありません。本調査で得られた結果は、今後の英語教育に役立たせていただきます。

この調査の信用性を高めるためにアンケートのお答えに関してはご自分の意見を正しくお答え下さい。尚、アンケートご協力に関する同意として、下記のご記入によって同意いただいたものと見なします。
調査へ協力いただけない場合はアンケート用紙を白紙のままご提出ください。

研究にご協力いただける方は個人に関する項目１～４に記入いただき、引き続きアンケートにご協力下さい。

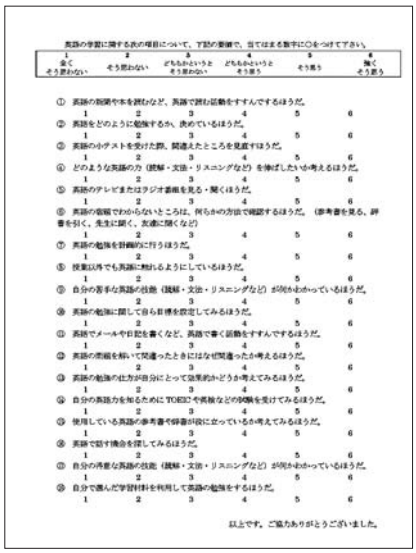
個人に関する項目

次の１～４について、記入（日本語で）または該当するものを○で囲んでください。

1. 年齢　（　　）歳
2. 専攻　（　　）学科
3. 学年　1年　2年　3年　4年　その他（　　）
4. 性別　男性　女性
5. 国籍　日本人　中国人　その他（　　）

＊英語のアンケートは全ての項目において自身の正直なご意見をお答え下さい。
各項目のお答えを論理にお答えいただけないと調査結果が不正確になります。
ご理解、ご協力いただき最後まで丁寧にお答え頂きますようお願い申し上げます。

ご協力ありがとうございます。



英語の学習に関する次の項目について、下記の基準で、当てはまる数字に○をつけて下さい。

1	2	3	4	5	6
全く そう思わない	そう思わない	どちらかというと そう思わない	どちらかというと そう思う	そう思う	強く そう思う

① 英語の雑誌や本を読むなど、英語で読む活動をすすんでするほうだ。
1 2 3 4 5 6
② 英語をどのように勉強するか、決めているほうだ。
1 2 3 4 5 6
③ 英語の小テストを受けた際、間違えたところを見直すほうだ。
1 2 3 4 5 6
④ どのような英語の力（読解・文法・リスニングなど）を伸ばしたいか考えるほうだ。
1 2 3 4 5 6
⑤ 英語のテレビまたはラジオ番組を見る・聞くほうだ。
1 2 3 4 5 6
⑥ 英語の意味でわからないところは、何らかの方法で確認するほうだ。（参考書を見る、辞書を引く、先生に聞く、友達に聞くなど）
1 2 3 4 5 6
⑦ 英語の勉強を計画的に行うほうだ。
1 2 3 4 5 6
⑧ 授業以外でも英語に触れるようにしているほうだ。
1 2 3 4 5 6
⑨ 自分の苦手な英語の技能（読解・文法・リスニングなど）が何かわかっているほうだ。
1 2 3 4 5 6
⑩ 英語の勉強に関して自ら目標を設定してみるほうだ。
1 2 3 4 5 6
⑪ 英語でメールや日記を書くなど、英語で書く活動をすすんでするほうだ。
1 2 3 4 5 6
⑫ 英語の意味を調べて間違ったときにはなぜ間違ったか考えるほうだ。
1 2 3 4 5 6
⑬ 英語の勉強の仕方が自分にとって効果的かどうか考えてみるほうだ。
1 2 3 4 5 6
⑭ 自分の英語力を知るために TOEIC や英検などの試験を受けてみるほうだ。
1 2 3 4 5 6
⑮ 使用している英語の参考書や辞書が何に役立っているか考えてみるほうだ。
1 2 3 4 5 6
⑯ 英語で話す機会を探してみるほうだ。
1 2 3 4 5 6
⑰ 自分の得意な英語の技能（読解・文法・リスニングなど）が何かわかっているほうだ。
1 2 3 4 5 6
⑱ 自分で選んだ学習材料を利用して英語の勉強をするほうだ。
1 2 3 4 5 6

以上です。ご協力ありがとうございました。