

Confirmatory Factor Analyses of Hypothesized Measurement Models for Three Learner Autonomy Instruments Intended for the Japanese College Population

Student ID # 119—G9204

Kayoko HORAI

Summary

1.0 Introduction

Interest in *autonomy* in applied linguistics has been increasing. Nonetheless, there is a shortage of quantitative research (Apple, 2011) and a dearth of instruments for measuring autonomous learning (Apple, 2011; Benson, 2001; Macaskill & Taylor, 2010).

The rationale for the research reported in this thesis is to adapt and test instrumentation for the measuring of learner autonomy within the Japanese population. The main argument supporting the necessity of this line of research is that any emerging research field needs good instrumentation to establish a secure basis for measurement. On top of that, the necessity for secure instrumentation to measure autonomy in Japan is amplified given issues related to educational policy. The Ministry of Education, Culture, Sports, Science and Technology (MEXT) has emphasized the need to improve communicative competence in English in response to internationalization (MEXT, 2003). This is a clear and compelling goal which is driven by globalization and rapid technological advancements. This educational policy also includes a focus on autonomy to foster lifelong learning among Japanese. This means that developing the field of learner autonomy is important in Japan to meet this goal as well. However, the learning goals for English learners in Japanese schools are mostly associated with passing entrance exams for their targeted university, or as Nakata (2011) described it, they are in an “entrance-exam-driven educational culture” (p. 900). These demands require learners to make a herculean effort to achieve two very dissimilar goals under the current Japanese

educational system: namely, passing exams that mostly assess a technical knowledge of English and trying to achieve communicative competence in English. In this challenging context, learner autonomy emerges as an important aspect of integrating these dichotomous focuses to maintain a level of overall success. If practitioners are to assist with this, and if researchers are to guide practitioners in how to assist with this, then learner autonomy needs to be measured and this measurement needs to be sound and supported by evidence.

Therefore, I attempt to contribute to English learning in Japan by conducting research which helps to provide evidence for the suitability or non-suitability of instrumentation for measuring learner autonomy in the Japanese context. This study involves two research agendas. One is to test instrumentation which is available in Japanese and currently advocated, but which still lacks empirical evidence for this advocacy. The other is to adapt existing instrumentation into the field of English language learning and the Japanese context.

2.0 Literature Review

In the literature review, I discuss the concept of autonomy in language education from different perspectives. First, I review the theoretical background of autonomy in language education. Early evidence of the emergence of the autonomy concept and its adoption within language learning appears in the early 1970s when the Council of Europe's Modern Language Project started its promotion of independent learning. However, autonomy was not a dominant issue in the 1970s, although, the concept was indicated implicitly and educators, researchers and promoters of language learning programs focused on individual learning, learning strategies, and independent learning by using self-access learning centers that are heavily related to autonomous learning behaviors. The concept of autonomy started to be discussed more actively and explicitly in applied linguistics in the 1990s. However, the notion became more theoretically complicated because it became a challenging issue to explain autonomy in language education considering the sociality of language learning and the use of language. In the 2000s, autonomy theory became mainstream in applied linguistics and this resulted in the

boosting of the number of publications related to the issue (e.g. Aoki & Nakata, 2011; Benson, 2007; Pemberton, Toogood, & Barfield, 2009). In the late 2000s, there was increased critique about the incompleteness of theorizing about the concept of autonomy in language education and the challenge of reconstructing our theoretical understanding has been referred to (Benson, 2007, 2009; Esch, 2009; Little, 2007; Stewart & Irie, 2011).

For the purpose of articulating the concept of autonomy, this study goes back to the origin of the concept in philosophy. The prime origin for the notion is from the seventeenth-century work of Immanuel Kant, who was the progenitor of the notion and who claimed that it played a significant role in moral thinking based on human dignity. His notion of autonomy and enlightened belief in human rationality had a strong influence on developments in Western culture. However his notion of autonomy was both too rigorous and too restricted to moral philosophy to adapt into modern educational theories (Schmenk, 2005). The identified notions of the original concept of autonomy, therefore, were not transmitted into language education; however, the fundamental concept is clearly rooted therein.

In the literature review, I then focus on educational psychology; particularly focusing on educational paradigms (i.e. behaviorism, cognitive psychology, constructivism, and humanistic psychology) as well as methodologies in language education (audiolingualism, the grammar translation method, and Communicative Language Teaching or CLT). I argue with respect to how each notion relates to the concept of autonomy. With regard to behaviorism, there appeared to be no explicit concern with autonomy and the paradigm is essentially neglectful of this human capacity. Cognitive psychology, on the other hand, emphasizes information processing and does not have an explicit goal of explaining autonomy but is much more consistent with the notion of autonomy and conscious control than behaviorism. Most importantly, engages with the notion of autonomy quite well and, in fact, the notion is central to what constructivists are trying to explain. Finally, humanistic psychology in education focuses on the whole person through lifelong learning and development throughout life. The idea of humanistic psychology in education led to a focus on individuality in learning and shifted education to a learner-centered curriculum. The field of humanistic psychology resulted in valuing a

stronger sense of personal agency and learner autonomy through individual responsibility for one's own learning with the aim of achieving lifelong learning. Narrowing the focus from educational psychology to the dominant methodologies within language teaching approaches, I reviewed three methodologies. First, audiolingulism is highly based on behaviorism, and the method neglects autonomy even though the behaviors it promotes (rote learning and pattern practice) may require some autonomy, in the sense of personal drive, to follow through to completion. Nonetheless, the method of learning by pattern practice is a passive learning approach and clearly does not make much pedagogical accommodation for notions of learner autonomy. The grammar translation method is less explicitly dominated by behaviorism than audiolingulism; although there is still an assumption that systematic pattern practice rather than free production is important. Therefore, it is similar to audiolingualism in that it does not explicitly engage with the notion of autonomy. CLT emphasizes autonomy in many aspects such as the importance placed on the need for free production and interactions in class which requires the autonomy required in natural human interaction as well as the need to be positive and creative. In other words, learners need to be autonomous to participate actively in a CLT class.

The idea of autonomy in language learning should not be one of the individual learning in self-driven isolation because language is after all used in a social context to communicate with others. With regard to language learning, sociocultural theory has gained particular appeal. The insight of sociocultural theory, advocated by Vygotsky, views learning as a process of social mediation and thus learning is not as an activity typically pursued in social isolation (at least as part of early childhood development). Explaining autonomy in terms of social context is challenging and remains under-theorized, though the issue has started to receive more attention over the past decade.

I then turn to the Japanese context to discuss more specifically Japanese educational policy and the notion of *yutori* education which is connected to the theoretical notion of learner autonomy. The educational reform underlying *yutori* education was aimed at a 'zest for living.' The idea of the policy did not explicitly suggest autonomy as important

at the launch in the 1990s though the curriculum underlying *yutori* education required autonomy implicitly by suggesting the need for opportunities for free-productive activities in school. Finnish education was highly regarded at the end of the twentieth century and the insights of this system were applied into educational reform in Japan. While the Finnish performed well in the Program for International Student Assessment (PISA) under their system, Japan dropped in the rankings after the educational reform underlying *yutori* education, and based on the Finnish system, were implemented. It could be argued that the different outcomes for the two nations from the policy emphasizing autonomy could be the level of prerequisite autonomy for the policy to work.

Having discussed the situation in Japan specifically, I then returned to the notion of autonomy in language learning but more explicitly. Autonomy theory became part of the mainstream of applied linguistics research in the 2000s; though the concept is still not resolved and remains challenging for theorists in the field. Diversified study of the concept has caused miss-conceptualizing in pedagogy and stimulated critiques about the concept of autonomy in language education. The ambiguity of the autonomy concept has resulted in calls for the necessity of reconstruction of the concept in the late 2000s (Benson, 2007; Little, 2007; Stewart and Irie, 2012). It is within this lack of clarity with respect to the concept that I draw attention to what has not been achieved so far which is to locate the concept clearly within a social learning context.

Finally, I articulate my particular interest and the contribution represented in this study for research in the area of autonomy and language learning, indicating a focus on measuring autonomy. I reviewed the search I undertook within the literature for instruments with the potential for adaptation into the Japanese context. I conducted preliminary research for autonomy instruments both in and outside of applied linguistics. However there are few instruments adaptable into language education and into the Japanese context. Of those surveyed, I finally settled on three available instruments for either adaptation or further investigation. One is Chang's Learner Autonomy Questionnaire (which I abbreviate as CLAQ to differentiate it from the LAQ mentioned immediately hereafter) developed by Chang in 2007, another is the Learner Autonomy

Questionnaire (LAQ) developed by Shimo in 2008 and the final instrument is the Autonomous Learning Scale (ALS) developed by Macaskill and Taylor in 2010.

To summarize, the current situation with respect to the study of autonomy is that the concept of autonomy in language education is under-theorized in terms of explaining autonomy in social space and language learning is more than independent or private learning; it requires a notion of autonomy which copes with social space. The central issue with respect to this thesis is the current instruments within the literature. This thesis deals with three of these instruments and all of them seem to reflect this theoretical incompleteness. All three instruments are underpinned by a theoretical model which is very close to the independent-learning and personal-motivation versions of the notion of autonomy. Nonetheless, they still measure an important, if incomplete, aspect of autonomy. The research questions for this study, therefore, concern whether the hypothesized models for each instrument dealt with in this thesis explain the dimensionality of data derived in the Japanese context.

3.0 Research questions

The research questions for this study concern whether the hypothesized models for three instruments (the CLAQ, the LAQ and the ALS) dealt with in this thesis explain the dimensionality of data derived in the Japanese context.

For the CLAQ (Chang, 2007) there are three hypothesized models. The research questions concern whether the hypothesized models (Model 1A, 1B and 2) for the CLAQ are plausible or not in a direct test of the instrument for scores generated in the Japanese context with an adapted Japanese version using Confirmatory Factor Analysis (CFA) as the method (one research question per model, respectively). With regard to the LAQ (Shimo, 2008), there are four hypothesized models advanced for it in the literature. Model 1 emerged in Shimo (2008), Model 2, emerged in Apple (2011), Model 3A emerged in a pre-survey in Shimo (2009), and Model 3B emerged in a post-survey in Shimo (2009). The research questions concern whether the hypothesized models (Model 1, 2, 3A, and 3B) for the LAQ are plausible or not in a direct test of the instrument for scores generated in the Japanese context using CFA as the method (one research question

per model, respectively). Finally, the ALS has one hypothesized model (Model 1) which was advanced by the authors of the instrument. The associate research question in this thesis asks whether this model is plausible in a direct test of the adapted instrument (Japanese version) for scores generated in the Japanese context using CFA as the method.

4.0 Methodology

In this study, a quantitative approach is used to analyze the psychometric properties of scores generated with the three instruments indicated above. Research was conducted according to three main stages or phases. In the preliminary or first stage potential instruments were searched for both in and outside of applied linguistics. In the second stage, each survey was prepared and data was collected. In the final stage, the collected data was analyzed. The three potential instruments chosen for this study indicated reasonable prospects for both domain adaptation and population adaptation on initial inspection, and had the most psychometric information available when compared with other instruments-although this is not to suggest that this information was completely adequate. The three instruments were directly tested, under a variety of models, in this study. The number of models tested for each instrument represents all models advanced previously (latent models) in the literature.

To conduct surveys, I requested of the author/authors of each instrument, permission to translate and/or use the respective instruments in the Japanese population. Regarding the LAQ, the original version was developed in Japanese; therefore, I only requested to use it. As for the other two instruments, I translated them from English into Japanese and asked two English-Japanese bilinguals with professional knowledge of academic English and Second Language Acquisition (SLA) to back-translate them and after the back-translation, some revisions were carefully made.

The data for each instrument was collected at different times and from different groups of participants. No participants responded to more than one instrument. The data for the CLAQ was collected in 2009, for the LAQ in 2011, and finally for the ALS in 2012. The participants in the case of each data set were $N = 373$ for the CLAQ, $N = 388$ for the LAQ, and $N = 1,068$ for the ALS. The details of each data set (i.e. gender,

student's year, and majors are reported respectively in the thesis. The collected data for each instrument was entered into a Microsoft Office Access 2010 database and then imported into IBM/Statistical Package for the Social Sciences (SPSS) software (Version 17.0). Descriptive statistics were conducted for the purpose of reviewing the normality of distribution of the scores by considering skew and kurtosis on the items for each adapted instrument. In addition, reliability (Cronbach's alpha) for subscales making up each instrument was estimated. Finally, CFAs were conducted and the scores were examined under the various models hypothesized using AMOS (Version 5.0.1), the measurement branch of Structural Equation Modeling (SEM).

5.0 Results

The analytical method includes the normality of score distribution, reliability estimation using the value of Cronbach's alpha, and testing for unidimensionality of sub-scales with a CFA. With respect to the CLAQ result, the values of the critical ratios for the skewness of the items on the Autonomy Beliefs (A.Bel) scale were not good with only 40% of items meeting the strict criterion (2.0), while kurtosis values were better with 70% of items meeting the strict criterion. With respect to skewness on the Autonomy Behaviors (A.Beh) scale, 80% of items met the strict threshold while for kurtosis 70% met the strict criterion. The reliability estimates for the scores on subscales of the CLAQ (Cronbach's alpha) were reasonably high, A.Bel (.87) and A.Beh (.84), when interpreted in terms of the criterion of .7 offered by Nunnally and Bernstein (1994). The results from the CFA, using the four indexes (TLI, CFI, RMSEA, and SRMSR) suggested by Hu and Bentler (1999) indicated that all models were not plausible (Models 1A, 1B and 2) and that they should be rejected as dissatisfactory explanations of the dimensionality of scores in the new dataset reported in this study. Only the SRMSR (one out of four indexes) produced values which could be considered either satisfactory or close to satisfactory, and this index provides an indication of the size of residuals.

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). With respect to the skewness for the items on the LAQ, 56% met the strict criterion (2.0), and

with respect to kurtosis 61% of values met the same strict criterion. The reliability estimates for the scores on subscales of the LAQ on Model 1 with two factors were satisfactory (Cronbach's alpha for Factor 1 was .87 and for Factor 2 was .87) when compared with the criterion of .70 offered by Nunnally and Bernstein (1994). Model 2 with two subscales indicated .87 (Factor 1) and .83 (Factor 2). Model 3A, also with two subscales, indicated .89 (Factor 1) and .83 (Factor 2) and Model 3B, a three-factor model, was .86 (Factor 1), .87 (Factor 2) and .77 (Factor 3). The results from the CFA using the four indexes indicate that none of the models were plausible (Models 1, 2, 3A, and 3B) and should be rejected. Only the SRMSR produced values which were either satisfactory or close to satisfactory.

With regard to the ALS and skewness for scores on the items making up the scale, 25% met the strict criterion, while with respect to kurtosis 58% met the strict criterion. Non-normality of distributions for scores on some items is a problem for the instrument in the case of this sample. The reliability estimates for scores on the ALS (Cronbach's alpha) were satisfactory with respect to Factor 1 (.76), however Factor 2 (.33) was significantly below the criterion of .70 offered by Nunnally and Bernstein (1994). The cause of the negative result in the case of Factor 2 in this study are difficult to specify with sufficient determinacy and further research with modified items would have to be undertaken in an effort to make inferences in this regard. The study conducted by Macaskill and Taylor (2010) reported that both alpha values were satisfactory which indicates doubt for the adaptation of this particular subscale. The results from the CFA, using the four indexes, indicated that the model was not plausible and should be rejected. Among the four indexes, only the SRMSR produced a satisfactory value.

6.0 Discussion and Conclusion

The main purpose of this study was to contribute to the development of sound psychometric instruments for measuring autonomy in English learning for the Japanese population. This study was conducted to examine whether any of the previous models (for the three instruments dealt with in this thesis) hypothesized in past literature would fit scores from a new dataset collected for this purpose. The results of this study indicated

that none of the models in the three instruments (the CLAQ, the LAQ and the ALS) were plausible in the Japanese context. This means that the issue of improving autonomy instrumentation in applied linguistics remains pertinent, and this study fulfills a negative though critical function of identifying weakness in the field with respect to current instrumentation.

With respect to the non-normality of distributions of scores on some items for all instruments, this is a problem which requires attention with respect to the modification of these instruments and is also an issue which requires more adequate reporting from the originating authors of these instruments. It is not possible to directly compare the normality of scores in this study with previous studies because none of the original studies reported these details, and this is an obstacle to further improvement of instrumentation for applied linguistics with respect to autonomy and with respect to these specific instruments.

With respect to the alpha values for all subscales making up two of the instruments (the CLAQ and the LAQ), these values met the criterion of .70. However, with respect to the ALS, one subscale did not meet the criterion. The alpha value indicates the reliability of a scale; however it the value should be considered critically because there is bias caused by the number of items on a scale. Furthermore, alpha does not indicate unidimensionality, and the more critical part of this study relates to the results of the CFAs (which do indicate unidimensionality) and which are summarized immediately below.

In terms of explaining the unsatisfactory results for all the models using CFA and with respect to the three instruments considered in this research, it is difficult to determine whether the cause of the problem is with the original design of the instrument, or with the adaptation into the Japanese context. Nonetheless, the scores derived in this study from a sample of the Japanese university-student population indicate that the models implied by the researches are not plausible in this population. The question now arises (for future research) as to whether they were ever plausible because previous researchers did not conduct CFAs (a more powerful psychometric tool). Recommendations for future research include engaging with the normality of

distributions of scores in the original data generated by the original authors of these instruments. Furthermore, in the absence of CFAs having been conducted in the original population for which the instrument was originally advocated as useful, the original authors should return to their data under a CFA. This will clarify the origin of the weakness as being in either the original conception of the instruments or in these adaptations.

Overall, this may appear to be a disappointing result as it is negative in the nature of conclusions which can be drawn for current instrumentation in the field, but this is nonetheless an important contribution. An empirical contribution is important whichever way the outcome falls because it facilitates inferences on the basis of evidence. In this case, the negative result based on empirical findings indicates that the field of autonomy has a pressing need for continued work on providing adequate instrumentation for the measurement of the construct, and the contribution of this study comes in the form of a negative contribution by illuminating the weakness of current instrumentation and cautioning against reliance on this instrumentation.

Finally, it should be stated that this study is attended by the limitations which frequently attend studies unassisted by the financial resources needed for a very wide-scale, representative sample of the intended population. Conclusions reached in this study require the tentativeness required of all studies where the sample could have been larger and more representative. Further and cumulative studies conducted along the lines of this study, but with other samples from the intended populations (college samples whether in Japan or other contexts), will assist with a gradual solution to this problem.