Measuring learner autonomy: A pilot study

Fumiko Murase
Macquarie University

This paper presents part of a project which aims at developing a new instrument to quantitatively measure the degree of learner autonomy with a focus on Japanese university students. On the basis of an operationalisation of the construct within this particular context, a pilot version of the instrument was developed and administered to 90 first-year Japanese students enrolled at a Japanese university. Following a brief description of the theoretical background of this study, major findings of the pilot study are presented, with a focus on the process of revising the pilot-tested instrument.

　本稿では、日本人大学生の学習者オートノミーを測定する新しい尺度の開発を目的とした研究について述べる。筆者は学習者オートノミーという概念の操作化を行った上で尺度の試作版を作成し、ある日本の大学に在籍する日本人の大学１年生９０人を対象に予備調査を実施した。本稿は、まずこの研究の理論的背景を説明した上で、予備調査で使用された尺度を改良していく過程に焦点を当てながら、予備調査の結果について報告するものである。

F ostering students' autonomy as learners of English, or helping students develop their autonomy, has been a goal for teachers in Japan (e.g., Aoki & Smith, 1999). Toward that goal, much effort has been made by teachers both in and outside the classroom. However, how do they know if their approach has been effective? In other words, how do they know if or how much their students have developed their autonomy?

This paper reports the pilot study for an ongoing research project, whose main objective is to develop a quantitative instrument that would enable teachers to measure their students' autonomy as language learners. The target participants are Japanese university students studying English at university level in Japan. Although learner autonomy may be measured both qualitatively and quantitatively, the aim of this project is

to develop a quantitative measure, as there is no such tool available in this context.

Background

In developing any sort of quantitative measure, it is crucial to establish a strong conceptual framework which underpins the measure by operationalising the construct to be measured. Operationalisation of learner autonomy is, however, a difficult task as there has been no consensus about the construct definitions and the construct has been widely considered as "multidimensional" (e.g., Benson, 2001, p. 13).

For the purpose of this research, the construct has been operationalised on the basis of an interdisciplinary literature review, which covers publications in philosophy, psychology, general education, and applied linguistics that are relevant to learner autonomy in language learning. First of all, I attempted to reconceptualise the construct by including as many aspects of learner autonomy as possible, so that I could reflect its multidimensionality and measure it from different perspectives. Having reviewed several possible ways of categorising the multidimensional construct (e.g., Benson, 1997; Holliday, 2003; Oxford, 2003; Schwienhorst, 2003), it seems that four main categories within the construct can be identified: *technical*, *psychological*, *political-philosophical*, and *socio-cultural* autonomy. Each of these categories can be divided into subcategories. For instance, technical autonomy, which means learners' ability or skills to take control of their own learning, is divided into two subcategories, *behavioural autonomy* and *situational autonomy*.

The four categories are considered to be interrelated to each other. For example, the learners may show their *ability* to use learning strategies simply because they are told to do so by their teacher, but I would not recognise this as autonomous learning. In order for the learners to become genuinely autonomous, therefore, it seems crucial that technical autonomy, which involves the learners' *ability* to use learning strategies, is supported by their psychological autonomy, their internal "capacity" (e.g., Little, 1991) to control their own learning: Technical autonomy and psychological autonomy should be positively correlated to each other.

Pilot study

Based on this operationalisation, the pilot version of the instrument was developed and tested. The main purpose of this pilot study was to test the reliability and other technical aspects of the questionnaire, before revision for its full-scale administration. This study consisted of two phases: (1) the questionnaire administration to Japanese university students, and (2) an interview with a teacher of the student participants. This paper focuses on the former, the questionnaire administration.

Methods
Context of study

As mentioned earlier, this project is aimed at the context of EFL classrooms at university level in Japan. Having this target context in mind, I conducted the pilot study by visiting EFL classrooms at a university in Japan and administering

the questionnaire to 90 Japanese students. All the students were English majors and enrolled in an English reading class, a required course for the first-year students in their department (though there were some students who were repeating the course). The 90-minute reading class met once a week over 15 weeks and was taught by a native-speaker teacher. There were three native-speaker teachers involved in this course. According to one of the teachers, although he had the concept of learner autonomy in mind, he had not adopted any explicit approaches specifically to develop learner autonomy in his class. Therefore, it can be said that this study was independent of their teaching and did not intend to measure the effectiveness of any specific learner autonomy training as such.

Table 1 shows demographic information about the participants. Please note that there were 10 students who did not answer the questions about their demographic information.

Materials

The material used was the pilot version of the instrument, *Measuring Instrument for Language Learner Autonomy* (or MILLA), which I developed specifically for this research. The MILLA was a paper-based questionnaire and written in Japanese, the L1 of the target participants. (The Japanese version was translated from a version which was developed in English for research and publication purposes; the questionnaire items presented in this paper are taken from the English version.) The pilot version consisted of three sections, and it took approximately 20 minutes for the participants to complete the whole questionnaire.

Table 1: Demographic information on the participants (N=90)

First language (L1)	
Japanese	80
Other language	0
Age	
18 years old	64
19 years old	15
20 years old	1
Gender	
Male	12
Female	68
Year of study	
1st year	79
2nd year	1
Experience of living overseas	
Yes	8
No	72

The first section, Section 1, consisted of 143 questions which were designed to measure the students' autonomy in terms of the four different categories. These *categories* were, however, not stated on the questionnaire form: There was no heading showing the category each question belonged to. Therefore, the respondents did not know which aspect of learner autonomy was being measured by each question. All the questions were on five-point Likert scale and there were two different sets of responses: (A) Never-Rarely-Sometimes-Often-Always, or (B) Strongly disagree-

Disagree-Neither agree nor disagree-Agree-Strongly agree, depending on the type of question. Basically, type (A) was for the questions asking what the students *do* for English learning (e.g., *I set goals for the day before I start studying English* (Q3)) while (B) was for those asking what they *think* or *believe* about their English learning (e.g., *Improvement in English depends on my own efforts* (Q52)). In addition to these two response sets, the "Not applicable (N/A)" option was added in order to discover the questions which were not applicable and to differentiate them from the neutral position (i.e., "Neither agree nor disagree").

The second section contained six questions designed to gain written feedback about Section 1. The questions included the following: (1) how long it took to complete Section 1, (2) if the instructions/ layout were clear, and (3) if there were any statements that were not clear or applicable to them (if any, they were asked to describe them). Since I did not have access to my participants for follow-up interviews or focus groups, it proved to be good to have this section. Admittedly, students' voices may be limited when they are asked for written feedback by open-ended questions. Indeed, not so many students wrote comments; however, I did obtain some very useful comments from participants in this section.

In the last section, there were six questions asking about their demographic information as shown in Table 1.

Data analysis

The data collected were analysed statistically by *item analysis* and *factor analysis* with the use of the students' written comments as a source of useful reference.

Item analysis

According to Dörnyei (2007), after a questionnaire is developed and pilot-tested, the next step is to conduct what is called *item analysis*. In this process, researchers should investigate the following:

- "missing responses and possible signs that instructions were not understood correctly"

- "the range of the response elicited by each item"

- "the internal consistency of multi-item scales" (Dörnyei, 2007, pp.99-100)

I analysed my data by investigating these three issues.

Missing responses

Firstly, I investigated the questions which had missing responses. There were some missing responses found, though not very frequent. I assumed that one possible reason for them could be a problem with the layout of the questionnaire. Therefore, I looked for the students' responses to the question: *Was the layout of the questionnaire clear?* in Section 2. There were three students who claimed that the layout was not clear. The following are some of the students' comments about the layout. (The students' comments, originally written in Japanese, were translated into English by the author.)

As it was closely printed, I marked in the wrong line several times.

As I am left-handed, it was not very easy to mark in the right box.

These comments show that what may look like a trivial issue, such as the layout of the questionnaire, can influence the responses. Therefore, the physical appearance of the questionnaire should be refined in the revising process in order to gain as many valid responses as possible.

"Not applicable" option

Many of the items responded to with N/A were those asking about *technical autonomy*, which is related to what the students *do* in learning English. For example, there were 13 students who responded with N/A to the question: *I study English at an English conversation school* (Q28). The problem here was that there were 52 people who chose "Never": If a student was not studying English at an English conversation school, he/she could have chosen either "N/A" or "Never." These examples showed the importance of carefully examining the match between each question and the response set.

Clarity of the instructions

There were seven students who responded with "No" to the question: *Were the instructions clear?* When I looked at their comments in Section 2, it seemed that there were three issues related to the clarity of the instructions: (1) N/A option, (2) the questions, and (3) the match between the questions and the descriptors of the response sets. The following are some examples of the students' comments about each issue:

> *I don't know what the difference between "Strongly disagree" and "Not applicable" is.*

> *I don't think "Not applicable" is necessary.*

> *Does "Not applicable" mean that I have never thought about it?*

These three comments about the N/A option indicate that the option has confused the students.

The second issue was about the questions themselves. For example, a student gave the following comment:

> *[The questions were] not very clear because similar questions were repeated.*

It was my intention that similar questions appeared in the questionnaire. For example, some questions in the category of *technical autonomy* were intentionally matched to questions in the category of *psychological autonomy*. For example, the question: *I set long-term goals in learning English* (Q1) was matched to the question: *Every student ought to set long-term goals in learning English* (Q63). It was intended to investigate the correlation between the students' technical and psychological autonomy. This attempt is important because, ideally, an autonomous learner is one who has the ability to learn autonomously (i.e., technical autonomy) and whose autonomous learning behaviour is based on his/her autonomous learning capacity (i.e., psychological autonomy). However, the appearance of similar questions seemed to confuse the students.

Thirdly, the match between the questions and the descriptors in response sets was another important issue as seen in the following comment:

> *As there were some questions that were not matched well to the descriptors, I had difficulties in responding to those questions.*

As previously mentioned, this match should be carefully examined.

Descriptive statistics

Secondly, I investigated the range of the responses. As statistical tests depend on a reasonable range of responses, it is not desirable to have extreme responses, which show *ceiling effect* or *floor effect*. The ceiling effect refers to the case where many of the respondents choose the highest score (i.e., "Always" or "Strongly agree" in this study) whereas the floor effect refers to the case where many of the respondents choose the lowest score (i.e., "Never" or "Strongly disagree"). If these effects occur, it means that the item is not *discriminative*, and it is preferable to cut out such items in the revising process.

In order to investigate floor and ceiling effects, the descriptive statistics were computed using SPSS and the frequency of the responses was examined. As a result, there were items which had these effects, particularly in the items asking about the following issues:

- Keeping records of learning (e.g., *I keep records of what I learned from my English study* (Q19): *floor effect*)

- Motivation towards English learning (e.g., *I want to communicate with people from foreign countries by using English* (Q39): *ceiling effect*)

- Responsibility for success/failure (e.g., *Improvement in English depends on my own efforts* (Q52): *ceiling effect*)

- Priority of English learning (e.g., *I give a higher priority to studying English than studying other academic subjects* (Q60): *ceiling effect*)

A possible reason for the floor effect here is the students' lack of instruction or experience in that particular learning strategy (i.e. keeping records of learning). Regarding the ceiling effect, many students chose the highest score for these items probably because all the participants were English majors and generally highly motivated to study English. Having considered these possible reasons for the extreme responses, most items exhibiting floor and ceiling effects were removed, but some were retained as they should give insights into the population as a whole.

Internal consistency

Thirdly, I examined the internal consistency of the questionnaire. That is, I investigated if each item on the questionnaire correlated with the other items and with the total score. For that purpose, the *item-total statistics* for the 143 items in Section 1 was computed using SPSS. Table 2 shows part of the result set (only a subset is presented here due to the limited space).

These results show, for example, that the correlation between Q1 and the whole section (i.e., 143 items) is .501 and that the overall reliability of the questionnaire will become .949 if we delete Q1 from the questionnaire. If the deletion of any particular item would dramatically increase the reliability, that item would better be deleted in the revising process. In this study, however, no item was deleted simply depending on the result of these item-total statistics

Table 2: Item-total statistics

Item	Corrected item-total correlation	Cronbach Alpha if item deleted
1 I set long-term goals in learning English	.501	.949
2 I make long-term plans for studying English	.489	.949
3 I set goals for the day before I start studying English	.242	.950
Cronbach Alpha for the 143 items = .950		

since the Cronbach Alpha for the 143 items was .950 and this would be considered fairly high for a questionnaire.

Factor analysis

After the item analysis, a factor analysis was conducted to investigate the construct validity: I investigated if each item in the questionnaire and the whole questionnaire was actually measuring what I intended to measure. If there are any items that are not measuring what they were designed to measure, those items need to be deleted.

The factor analysis was performed by using SPSS to examine how the responses to the items clustered with each other and if the factors extracted by the analysis corresponded with my hypothesised categories and subcategories. In this respect, my analysis was *confirmatory*, rather than exploratory.

First of all, 16 items were excluded from the factor analysis, as they were found to be problematic as a result of the item analysis. With the 127 items left, 26 items asking about what

the students *do* in English learning and 101 items asking what they *think* or *believe* about English learning were factor analysed separately. The analysis was repeated several times by each time removing the items with low factor loadings (lower than .300) and the ones which did not cluster together with the other items that were designed to measure the same (sub-)category. A total of 30 items were removed throughout this process, and the number of items in Section 1 was reduced in the end to 97. These 97 items seemed to fall into 13 main factors and the extracted factors are shown in Table 3:

Table 3: Extracted factors

Factor	Label	Category
1	Ability to use learning strategies during/after learning	TEC
2	Ability to use learning strategies before learning	TEC
3	Ability to keep records of learning	TEC
4	Knowledge about metacognitive strategies	PSY
5	Knowledge about one's own learning and affective strategies	PSY
6	(a) Extrinsic motivation	PSY
	(b) Teacher's role as authority	P-P
	(c) Teacher-dependence	S-C
7	Learning with other learners	S-C
8	Intrinsic motivation	PSY
9	Views about learning in different cultures	S-C
10	Individual autonomy	P-P
11	Positive freedom	P-P
12	One's view of group autonomy	P-P
13	One's awareness of policy	P-P

Note: TEC=technical; PSY=psychological; P-P=political-philosophical; S-C=socio-cultural.

Although I am aware that these factors may contain problematic areas (e.g., Factor 6 overlaps three different categories), I assume that these factors generally corresponded to my hypothesised subcategories.

Implications for further research

This study was not free from limitations. After describing the two main limitations this pilot study experienced, sample size and sample composition, this section discusses two main issues that I learned from this study: (1) issues related to sampling, and (2) problems to be dealt with in revising the questionnaire.

Firstly, I should admit the sample size (N=90) was not large enough for conducting factor analysis. Hence, I did not thoroughly rely on factor analysis, although the result gave a very useful reference for the revising process. For example, the items in Section 1 were reduced down to 97 items through the factor analysis, but 16 items were added to it at the later stage of the revising process: Some were once deleted but put back, while others were newly written. This manipulation was done partly because I wanted to see how those "problematic" items would work with a larger sample in the main study and also because I wanted to make sure that each factor contain at least five items in it: Some factors had only two or three items after factor analysis. Regarding the sample *composition*, the participants were all English-majors and this might have influenced the responses (e.g., extreme responses) as mentioned in an earlier section. If the data had been collected also from non-English major students, the results might have been different. Having considered these sampling issues, the main study will need

to have a larger number of participants and more variety in the sample composition (e.g., major of study).

The second issue I learned was about the justifications for the revising process. As a result of item analysis and factor analysis, the pilot-tested MILLA has been revised by reducing the number of items and then adding 16 items to it as described in the previous section. During this process, some items were moved to a different category or a different section. For example, the question: *If I ask my teacher for help in learning English, I know how I want him/her to help me* (Q120) was originally in a subcategory of socio-cultural autonomy but later moved to a category of psychological autonomy, as it clustered well together with other items in the latter category. In addition, 10 items designed to measure technical autonomy were moved to Section 2 and turned into questions asking about out-of-class learning experiences. This revision was done as those items had problems with the match between the question and the response set.

Once I finalised all the items to be included in the final version, those items were refined with a focus on the wording by incorporating the feedback from my fellow students and some researchers. This process involved two main issues: (1) rewording of the questions and (2) improving the quality of translation. The result of factor analysis indicated that the wording may have affected some groups of questions with the same phrasing clustering together. In order to reduce the influence of the wording, those questions were reworded. For example, there were 18 questions beginning with *Every student ought to* and some of these were replaced with *A good learner of English* or *All students ought to*. Regarding the translation issue, I tried

to make the translation of the MILLA as good as possible. Although the questionnaire administered to the students was written in Japanese, I first produced the English version and then translated it to Japanese myself. Therefore, even though I am a native-speaker of Japanese, some questions in Japanese appeared awkward and needed to be refined. I incorporated the feedback from two Japanese-English bilingual speakers for the final version.

Having gone through all these revising processes, the final version of the MILLA was developed. The final version basically follows the same format as its pilot version: it is paper-based and written in Japanese. However, it has only two sections: (1) 113 items on a five-point Likert scale for measuring learner autonomy, and (2) six main questions for getting demographic information. This version does not have the feedback section that the pilot version had, although a space for free comments is provided. The N/A option was not included in this version as it turned out to be confusing to the respondents.

Conclusion

This paper has described the pilot study of a project which aimed at developing a new measure for learner autonomy. Although this was not a full-scale study, it has provided great insights for further research: It showed that pilot-testing is a necessity in developing measurement instruments. It is hoped that this pilot-study experience will contribute greatly to the process of developing better and more efficient instruments for further research.

I would like to conclude this paper with a teacher's voice. This teacher, one of the three native-speaker teachers involved in the teaching of the 90 student participants, was asked if/how he thought a quantitative measure of learner autonomy could be useful or helpful. Here is his comment:

> I think it [a quantitative measure of learner autonomy] will help me as a teacher…it will show me where the gaps are in my class – what they know, what they don't know about learner autonomy…their behaviour, their awareness, their everything.…I would use it as an evaluative tool to show me…what I need to do and who I need to help…I think also…this helps remind the teacher what autonomy is…and also it becomes a tool for discussion in the classroom. So, after the students have answered the questions, then you can get them to go through and discuss the questions together.… I think it could become a good motivating tool but also a tool to raise students' awareness.

Although this is not specifically about the MILLA, his comments coincide with the starting point of this research project, which was to develop a valid and reliable measure of learner autonomy in a Japanese EFL context. It is hoped that this and further research will contribute to better teaching and learning in the future by devising an effective tool for both teachers and learners.

Fumiko Murase is currently a PhD candidate at Macquarie University in Sydney, Australia. Her main research interests include learner autonomy, learning strategies, and learning styles. <fumiko.murase@ling.mq.edu.au>

References

Aoki, N., & Smith, R. (1999). Learner autonomy in cultural context: The case of Japan. In S. Cotterall & D. Crabbe (Eds.), *Learner autonomy in language learning: Defining the field and effecting change* (pp. 19-27). Frankfurt am Main: Peter Lang.

Benson, P. (1997). The philosophy and politics of learner autonomy. In P. Benson & P. Voller (Eds.), *Autonomy and independence in language learning* (pp. 18-34). London: Longman.

Benson, P. (2001). Teaching and researching autonomy in language learning. Harlow: Pearson Education.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Holliday, A. (2003). Social autonomy: Addressing the dangers of culturism in TESOL. In D. Palfreyman & R. C. Smith (Eds.), *Learner autonomy across cultures: Language education perspectives* (pp. 110-126). Basingstoke: Palgrave Macmillan.

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

Oxford, R. L. (2003). Toward a more systematic model of L2 learner autonomy. In D. Palfreyman & R. C. Smith (Eds.), *Learner autonomy across cultures: Language education perspectives* (pp. 75-91). Basingstoke: Palgrave Macmillan.

Schwienhorst, K. (2003). Neither here nor there? Learner autonomy and intercultural factors in CALL environments. In D. Palfreyman & R. C. Smith (Eds.), *Learner autonomy across cultures: Language education perspectives* (pp. 164-179). Basingstoke: Palgrave Macmillan.