Foreign Language Learning Mode Preferences of
Japanese University Students

Kunihiro KUSANAGI
Kazumichi ENOKIDA
Shusaku KIDA
Tatsuya SAKAUE
Fuyuko TAKITA
Keiso TATSUKAWA
Koji UENISHI
Goro YAMAMOTO
Lisa YOSHIIKAWA
Hiroshima University

Abstract

The present study aims to develop a practical and statistically sound inventory that helps in examining Japanese university students’ foreign language learning mode preferences (LMPs). LMP, like numerous other learners’ traits such as learning styles, beliefs, and motivation, is one of the behavioral and meta-cognitive variables that potentially moderate the variance of the learning outcome under a specific learning mode. The present study conducted a questionnaire survey, sampling university students who were learning English as a foreign language (N = 1,003) in a Japanese national university and developed a Foreign Language Learning Mode Preference Inventory (ver. 1), whose items (K = 12) were all written in Japanese. The inventory yielded three subscales with relatively higher reliability coefficient for each: (a) cooperative learning preference (α = .91), (b) e-learning preference (α = .90), and (c) face-to-face instruction preference (α = .70). The factorial validity of the inventory was tested using structure equation modeling. The results showed that the theoretical measurement model fit well to the observations. Furthermore, multiple sample analyses supported the measurement invariance with the constraints of equal loadings and equal intercepts across genders. Some practical applications of this inventory, specifically in the context of foreign language teaching in Japanese higher education, were discussed.
1. Background

1.1 Diversified Learning Modes in Higher Education

Nowadays, most foreign language learning curricula, programs, and classes in Japanese higher education consist of various learning modes or modalities, such as (a) pair and group works, (b) e-learning, and (c) face-to-face instruction. To give a concrete example, a compulsory English language learning program, which Hiroshima University’s Institute for Foreign Language Education and Research supplies, is systematically designed to cross multiple learning modes. A common student in this cross-mode program thus experiences face-to-face instruction and some communicative language activities in class and is required to engage in an e-learning program for basic vocabulary in another online course for the same academic term.

Such a recent trend in diversification of learning modes in higher education is not simply a matter of everyday teaching practice. Rather, it has been complexly affected by various academic, social, and political movements, such as (a) a cooperative learning perspective in higher education research (e.g., Johnson & Johnson, 2008); (b) striking developments in information and communication technology (Zhao, 2003); (c) recent educational policies in Japan and other countries (e.g., active learning; Prince, 2004 for review); and possibly, (d) second language acquisition (SLA) research. These movements may interact complexly in educational decision-making at any level, and consequently, such interactions may form cross-mode language learning curricula, programs, and classes. This diversification trend will continue for a while.

Blended learning and other similar terms may have captured the situation in an appropriate way. Blended learning research focuses on the learning outcome of the combinations of some learning modes rather than that of a single mode. As is easily expected, a couple of meta-analyses reported the relatively higher efficiency of blended learning (e.g., the combination of e-learning and face-to-face instruction) in comparison to single-mode learning, such as purely e-learning (e.g., Means, Toyama, Murphy, & Baki, 2013; Means, Toyama, Murphy, Bakia, & Jones, 2009). However, it should be noted that the results are far from decisive, and researchers should consider the inevitable effects of some moderator variables.

1.2 Learning Mode Preferences as a Moderator of Learning Outcomes

One can easily claim that there is no optimal combination for all the students and can also assume the existence of some cognitive, psychological, and behavioral variables that statistically moderate one’s learning outcome, as the classical aptitudes-treatment interaction framework in educational psychology formulates (Snow, 1991). For example, one who favors a certain treatment will greatly benefit from the treatment and will not benefit from another treatment. Simply put, this also goes for the combinations of these treatments.

Regardless of its historical success, there is no doubt that understanding and controlling some learners’ information concerning aptitudes, and other inter-learner variables that moderate
the learning outcome, is potentially beneficial for any type of educational decision-making, such as in (a) student placement, (b) classroom size setting, (c) curriculum design, (d) teaching material selection, and (e) optimization of everyday teaching practice for each student. Such applications are exactly what the present study views as a goal.

In psychology, numerous researchers have paid greater attention to learning styles (e.g., Dunn, 1984; Dunn & Dunn, 1978; Kolb, 2015). Learning style research is basically theory-oriented and is based on a cognitive architecture of learning in general, and thus it sometimes fails to explain the direct causality between learning styles and learning outcome in practice (see also criticisms to learning style research, such as Lynn, 1990). Various theoretical learning style models have emerged, such as Kolb’s model, the basis of which is experiential learning (Kolb, 2015) as well as related learning style inventories, such as Kolb’s Learning Style Inventory (Kolb, 2007). However, these models and inventories were not specifically designed to directly examine the moderator variables regarding the currently major learning modes, and they do not fit the diversification trend in higher education, specifically, in the context of foreign language teaching.

Likewise, SLA research has voluminous literature on learner variables such as motivation, beliefs, and strategies (Winke, 2007, for overview). It cannot be denied that such extensive research on learner variables has historically brought us enormous scientific insights. However, most of the constructs in SLA research have not been fine-tuned to measure the moderator variable of the learning outcome, and some of them lack statistical appropriateness in their measurement.

In this study, we take a different angle from learning style research and SLA research. Namely, we attach importance to more practical aspects. Unlike learning style research, which often focuses on genetic differences of cognitive architecture, we look at the closer causality behind learning. We assume that a behavioral and meta-cognitive trait, learning mode preference (LMP), exists and that this LMP is one of the causes of individuals’ learning engagement, and thus LMP moderates the learning outcome. We statistically model LMP as a latent variable under the reflective model (cf., Edwards & Bagozzi, 2000; Borsboom, Mellenbergh, & Heerden, 2003). This assumption is guided by a perspective on students’ own decision-making about learning activities rather than genetic individual differences in cognition or epistemological differences of one’s mental state such as integrative motivation vs. instrumental motivation. In the LMP framework, it can be said that one who shows higher corporate learning preference (see detail in the next section) can be meta-cognitively aware that he or she favors cooperative learning and initiatively decides how much he or she should be engaged in learning under the cooperative learning mode than in another mode. This type of simple causality can be more easily statistically modeled and behaviorally tested. For instance, compare the three causality scenarios here: (a) “those who have visual and audio learning styles tend to benefit more in e-learning,” (b) “those who have higher instrument motivation tend to benefit more in e-learning,” and (c) “those who prefer e-learning tend to benefit more in e-learning.”
Although we take a practical approach based on decision-making and discuss some applications of this approach, we never ignore the literature on learning styles and other learner variables. As shown in the path diagram in Figure 1, the present study presumes that the effect of learning styles and other possible factors is behind the latent variable, LMP.

![Figure 1](image)

**Figure 1.** A schematic path diagram representing the basic concept of LMP, where (a) LMP is measured as a reflective model, (b) learning styles and other variables are one of the possible causes of LMP, and (c) LMP potentially explains the variance on the outcome measure, via the decision-making function, which is not manifestly given.

### 1.3 The Measurement Model and Inventory Design

Based on the conceptualization of LMP above, the present study attempts to develop a practical and statistically sound inventory that measures LMP. Now, LMP can be more formally defined as “a behavioral and meta-cognitive trait of learners, which predicts one’s engagement, and outcome of learning under a specific learning mode.” The present study assigns three currently major learning modes in Japanese higher education: (a) cooperative learning preference (CLP), (b) e-learning preference (ELP), and (c) face-to-face instruction preference (FTFIP). It should be assumed that the estimated values of the subscales are not orthogonal, and the present study does not consider any higher order structure of the subscales. These statistical conditions come down to a simple three-factor model. Moreover, regarding the practicality of the inventory, we designed a shorter one; we set the target item number of the final model at 12, which comes to 4 items for each subscale. The exact measurement model can be seen in Figure 2.

CLP refers to the preference for cooperative learning. Cooperative learning perspectives have long debates on its definition (e.g., Johnson & Johnson, 2008; Kagan, 1989). Taking a relatively broader view, the present study expects that cooperative learning typically occurs in pair and group works only when (a) the activities are mutually beneficial to the members; (b) the members share their goal, efforts, thoughts, and information; and (c) the goal cannot be accomplished without cooperation (Johnson & Johnson, 2008). Therefore, the term cooperative learning takes a narrower domain than pair and group works; not all pair and group works are
Although we take a practical approach based on decision-making and discuss some applications of this approach, we never ignore the literature on learning styles and other learner variables. As shown in the path diagram in Figure 1, the present study presumes that the effect of learning styles and other possible factors is behind the latent variable, LMP.

Figure 1.
A schematic path diagram representing the basic concept of LMP, where (a) LMP is measured as a reflective model, (b) learning styles and other variables are one of the possible causes of LMP, and (c) LMP potentially explains the variance on the outcome measure, via the decision-making function, which is not manifestly given.

1.3 The Measurement Model and Inventory Design

Based on the conceptualization of LMP above, the present study attempts to develop a practical and statistically sound inventory that measures LMP. Now, LMP can be more formally defined as “a behavioral and meta-cognitive trait of learners, which predicts one’s engagement, and outcome of learning under a specific learning mode.” The present study assigns three currently major learning modes in Japanese higher education: (a) cooperative learning preference (CLP), (b) e-learning preference (ELP), and (c) face-to-face instruction preference (FTFIP). It should be assumed that the estimated values of the subscales are not orthogonal, and the present study does not consider any higher order structure of the subscales. These statistical conditions come down to a simple three-factor model. Moreover, regarding the practicality of the inventory, we designed a shorter one; we set the target item number of the final model at 12, which comes to 4 items for each subscale. The exact measurement model can be seen in Figure 2.

CLP refers to the preference for cooperative learning. Cooperative learning perspectives have long debates on its definition (e.g., Johnson & Johnson, 2008; Kagan, 1989). Taking a relatively broader view, the present study expects that cooperative learning typically occurs in pair and group works only when (a) the activities are mutually beneficial to the members; (b) the members share their goal, efforts, thoughts, and information; and (c) the goal cannot be accomplished without cooperation (Johnson & Johnson, 2008). Therefore, the term cooperative learning takes a narrower domain than pair and group works; not all pair and group works are cooperative. Cooperative learning perspectives may have some similar characteristics as task-based language learning and the active learning concept. Some meta-analyses in higher education research have supported the effectiveness of cooperative learning (e.g., Johnson, Johnson, & Stanne, 2000; Johnson, Maruyama, Johnson, Nilson, & Skon, 1981).

Figure 2. The measurement model of the present study.

ELP is related to e-learning, web-based training, digital learning devices, multimedia, and social networking. Nowadays, this learning mode is applied to numerous areas such as occupational training, language learning, and other school subjects. Some meta-analyses have reported its efficacies (e.g., Bernard et al., 2009; Boghikian-Whitby & Mortagy, 2016, Means et al., 2009), but some studies have also reported the equivalent effect of e-learning with face-to-face instruction (e.g., Neuhauser, 2010).

FTFIP is about so-called traditional teaching methods, such as grammar translation, language exercises, lectures, and other common classroom activities that are typically featured in one-directional communication between a teacher and students. Face-to-face instruction is frequently contrasted with e-learning, especially in distance education research (e.g., Neuhauser, 2010; Johnson, Argon, Shaik, & Palma-Rivas, 2000). Some studies showed a higher preference for this on average than for e-learning (e.g., Beard & Happer, 2002; Kishore, Tabrizi, Ozan, Aziz, & Wuensch, 2009). However, this does not imply that the same preferences apply in the context of Japanese higher education.

1.4 Other Related Surveys

An applied linguistics researcher, Fushino, conducted a questionnaire survey relating to learner beliefs about second language group work (Fushino, 2010). The scale that Fushino developed is multi-factorial, unlike the present study, which put all its information into a single factor relating to cooperative learning. Furthermore, Fushino’s scale is about learner beliefs, which in the present study we treat as one of the indirect factors behind LMP.
In Japanese foreign language teaching research, Kawaguchi and Kusanagi (2016) developed a scale that examines computer-assisted language learning (CALL) attitudes. Kawaguchi and Kusanagi’s scale also yields multiple factors. As Fushino (2010) and Kawaguchi and Kusanagi (2016) had addressed conceptually similar constructs with the present study, a future study should examine the correlational evidence between the present study and these studies.

2. The Preliminary Study

2.1 Purpose

In order to develop a statistically sound inventory, the present study followed a two-step procedure. First, we conducted a preliminary study, whose main purpose was to select the questionnaire items before the main study.

2.2 Participants

In total, 94 university students who learned English as a foreign language participated in the preliminary study. All of the participants were sampled from one national university.

Some of the demographic differences were (a) academic years (second-year students and above in the preliminary study and mostly freshmen in the main study); (b) gender distribution; and (c) English language proficiency. However, the present study presumed that the effects of the demographic differences were minuscule. In fact, the main study exactly reproduced the targeted factorial structure.

2.3 Materials

The first author clarified the initial item pool \( K = 32 \). The item writer picked up sets of keyword lists for each subscale widely from the literature (e.g., “pair and group” for CLP) and built up sentences that contained the keywords. Then, the item pool was checked by other authors through a consultation. All the items were written in Japanese and had meta-cognitive and self-reflective features. The present study administered the items through a seven-point Likert scale examining the degree of agreement with each sentence.

2.4 Procedure and Results

The participants answered a questionnaire that consisted of the 32 items using pen and paper. It took almost 15 minutes on average for them to complete the survey. Then, a research assistant input the responses into computer readable media by hand with some times of validations. All the answers were valid without missing values.

We first checked the distribution of each item. Through this procedure, 10 items that violated the normal distribution were excluded. The remaining items \( k = 22 \) were submitted to
classical multidimensional scaling in order to visually select the items, confirming the covariance structure. We excluded 8 items through this process.

Finally, by repeatedly simulating exploratory and confirmatory factor analyses, 2 items were excluded by referring to the simulated factor loads and goodness of fit indices. Then, the remaining 12 items were submitted to confirmatory factor analysis with a theoretical measurement model. The results showed relatively fine goodness of fit, CFI = .959, TLI = .930, RMSEA = .068, SRMR = .061 and acceptable reliability coefficient values for each item, CLP \((k = 4, \alpha = .86)\), ELP \((k = 4, \alpha = .83)\), FTFIP \((k = 4, \alpha = .70)\). The 12 items were selected for the main study (see the items in the Appendix).

3. The Main Study

3.1 Purpose

The purpose of the main study was to establish the inventory. The present study performed structure equation modeling with the measurement model and multiple sample analyses to test the measurement invariance across genders.

3.2 Participants

In total, 909 university students who learned English as a foreign language participated in the main study. All the participants were sampled from the same national university, and they were all freshmen with a couple of exceptions.

Their departments were variously ranged: medicine (19.58%), dental (8.91%), pharmaceutical (5.28%), law (14.85%), education (11.55%), engineering (26.51%), biology (10.89%), and science (2.42%). The gender ratios were as follows: men (57.21%), women (42.13%), and transgender (0.66%). The responses of the transgendered participants were not submitted to multiple sample analyses. However, this does not imply that they were not of interest; this decision was based on purely statistical considerations. No participant attended both studies.

Most of the participants (91.31%) reported their scores in the Test of English for International Communication (TOEIC), which they had taken during an academic semester. The distribution is shown in Figure 3. The data can be modeled by fitting the normal distribution at \(\mu = 522.58, \sigma = 101.33\), with the goodness of fit at AIC = 10025, or Gamma distribution at \(\alpha = 26.27, \beta = 0.05\), with the goodness of fit at AIC = 10015.

3.3 Materials

The 12 items selected in the preliminary study were adopted. Each subscale had 4 items. Namely, items 1 to 4 referred to CLP, 4 to 8 to ELP, and 9 to 12 to FTFIP. The present study administered the items through a Likert scale, like in the preliminary study.
3.4 Procedure

The participants answered the items on an online form. There were no missing values. All the item responses were visualized using multiple histograms. Then, the moment statistics and variance and covariance matrices were reported. Next, we fit the measurement model to the observations using structure equation modeling. The standardized coefficients and goodness of fit under the models were reported. We also calculated reliability coefficients for each subscale. Then, multiple sample analyses were conducted. All the procedures were performed on R, statistical software (R Core Team, 2016) with the packages psych (Revelle, 2016), lavaan (Rosseel, 2012) and semTools (semTools contributors, 2016).

3.5 Results

The empirical distribution of all the item responses are shown in Figure 4, and based on them, we can confirm the normality of each variable. Table 1 shows the moment statistics, and Table 2 shows the variance and covariance matrices.

![Figure 3](image-url)  
*Figure 3. The empirical distribution of self-reported TOEIC scores (n = 830).*

![Figure 4](image-url)  
*Figure 4. Multiple histograms showing the distribution of the responses.*
3.4 Procedure
The participants answered the items on an online form. There were no missing values. All the item responses were visualized using multiple histograms. Then, the moment statistics and variance and covariance matrices were reported. Next, we fit the measurement model to the observations using structure equation modeling. The standardized coefficients and goodness of fit under the models were reported. We also calculated reliability coefficients for each subscale. Then, multiple sample analyses were conducted. All the procedures were performed on R, statistical software (R Core Team, 2016) with the packages psych (Revelle, 2016), lavaan (Rosseel, 2012) and semTools (semTools contributors, 2016).

3.5 Results
The empirical distribution of all the item responses are shown in Figure 4, and based on them, we can confirm the normality of each variable. Table 1 shows the moment statistics, and Table 2 shows the variance and covariance matrices.

Table 1.
The Moment Statistics of the Item Responses

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 1</td>
<td>4.18</td>
<td>1.69</td>
<td>-0.15</td>
<td>-0.79</td>
</tr>
<tr>
<td>Item 2</td>
<td>3.93</td>
<td>1.54</td>
<td>0.00</td>
<td>-0.54</td>
</tr>
<tr>
<td>Item 3</td>
<td>4.19</td>
<td>1.52</td>
<td>-0.12</td>
<td>-0.56</td>
</tr>
<tr>
<td>Item 4</td>
<td>4.01</td>
<td>1.56</td>
<td>-0.04</td>
<td>-0.65</td>
</tr>
<tr>
<td>Item 5</td>
<td>4.37</td>
<td>1.44</td>
<td>-0.36</td>
<td>-0.19</td>
</tr>
<tr>
<td>Item 6</td>
<td>4.12</td>
<td>1.43</td>
<td>-0.13</td>
<td>-0.38</td>
</tr>
<tr>
<td>Item 7</td>
<td>4.09</td>
<td>1.47</td>
<td>-0.16</td>
<td>-0.46</td>
</tr>
<tr>
<td>Item 8</td>
<td>4.17</td>
<td>1.54</td>
<td>-0.17</td>
<td>-0.52</td>
</tr>
<tr>
<td>Item 9</td>
<td>4.32</td>
<td>1.38</td>
<td>-0.32</td>
<td>-0.22</td>
</tr>
<tr>
<td>Item 10</td>
<td>4.57</td>
<td>1.35</td>
<td>-0.38</td>
<td>0.05</td>
</tr>
<tr>
<td>Item 11</td>
<td>4.99</td>
<td>1.27</td>
<td>-0.57</td>
<td>0.51</td>
</tr>
<tr>
<td>Item 12</td>
<td>4.30</td>
<td>1.35</td>
<td>-0.28</td>
<td>-0.14</td>
</tr>
</tbody>
</table>

Table 2.
The Variance and Covariance Matrices of the Item Responses

<table>
<thead>
<tr>
<th></th>
<th>I1</th>
<th>I2</th>
<th>I3</th>
<th>I4</th>
<th>I5</th>
<th>I6</th>
<th>I7</th>
<th>I8</th>
<th>I9</th>
<th>I10</th>
<th>I11</th>
<th>I12</th>
</tr>
</thead>
<tbody>
<tr>
<td>I1</td>
<td>2.85</td>
<td>1.89</td>
<td>1.57</td>
<td>1.83</td>
<td>0.22</td>
<td>0.22</td>
<td>0.26</td>
<td>0.31</td>
<td>-0.07</td>
<td>0.18</td>
<td>0.36</td>
<td>0.32</td>
</tr>
<tr>
<td>I2</td>
<td>1.89</td>
<td>2.37</td>
<td>1.62</td>
<td>1.89</td>
<td>0.20</td>
<td>0.34</td>
<td>0.34</td>
<td>0.33</td>
<td>-0.04</td>
<td>0.13</td>
<td>0.31</td>
<td>0.28</td>
</tr>
<tr>
<td>I3</td>
<td>1.57</td>
<td>1.62</td>
<td>2.32</td>
<td>1.84</td>
<td>0.16</td>
<td>0.19</td>
<td>0.19</td>
<td>0.30</td>
<td>0.09</td>
<td>0.32</td>
<td>0.39</td>
<td>0.56</td>
</tr>
<tr>
<td>I4</td>
<td>1.83</td>
<td>1.89</td>
<td>1.84</td>
<td>2.44</td>
<td>0.23</td>
<td>0.25</td>
<td>0.24</td>
<td>0.31</td>
<td>0.01</td>
<td>0.15</td>
<td>0.35</td>
<td>0.40</td>
</tr>
<tr>
<td>I5</td>
<td>0.22</td>
<td>0.20</td>
<td>0.16</td>
<td>0.23</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
<td>0.25</td>
<td>0.16</td>
<td>0.20</td>
<td>0.28</td>
<td>0.33</td>
</tr>
<tr>
<td>I6</td>
<td>0.22</td>
<td>0.34</td>
<td>0.19</td>
<td>0.25</td>
<td>1.40</td>
<td>2.06</td>
<td>1.70</td>
<td>1.53</td>
<td>0.16</td>
<td>0.20</td>
<td>0.28</td>
<td>0.33</td>
</tr>
<tr>
<td>I7</td>
<td>0.26</td>
<td>0.34</td>
<td>0.19</td>
<td>0.24</td>
<td>1.41</td>
<td>1.70</td>
<td>2.16</td>
<td>1.59</td>
<td>0.11</td>
<td>0.23</td>
<td>0.30</td>
<td>0.35</td>
</tr>
<tr>
<td>I8</td>
<td>0.31</td>
<td>0.33</td>
<td>0.30</td>
<td>0.31</td>
<td>1.51</td>
<td>1.53</td>
<td>1.59</td>
<td>2.38</td>
<td>-0.03</td>
<td>0.17</td>
<td>0.26</td>
<td>0.28</td>
</tr>
<tr>
<td>I9</td>
<td>-0.07</td>
<td>-0.04</td>
<td>0.09</td>
<td>0.01</td>
<td>-0.10</td>
<td>0.16</td>
<td>0.11</td>
<td>-0.03</td>
<td>1.91</td>
<td>1.02</td>
<td>0.70</td>
<td>0.83</td>
</tr>
<tr>
<td>I10</td>
<td>0.18</td>
<td>0.13</td>
<td>0.32</td>
<td>0.15</td>
<td>0.12</td>
<td>0.20</td>
<td>0.23</td>
<td>0.17</td>
<td>1.02</td>
<td>1.82</td>
<td>0.95</td>
<td>0.83</td>
</tr>
<tr>
<td>I11</td>
<td>0.36</td>
<td>0.31</td>
<td>0.39</td>
<td>0.35</td>
<td>0.25</td>
<td>0.28</td>
<td>0.30</td>
<td>0.26</td>
<td>0.70</td>
<td>0.95</td>
<td>1.61</td>
<td>0.63</td>
</tr>
<tr>
<td>I12</td>
<td>0.32</td>
<td>0.28</td>
<td>0.56</td>
<td>0.40</td>
<td>0.28</td>
<td>0.33</td>
<td>0.35</td>
<td>0.28</td>
<td>0.83</td>
<td>0.83</td>
<td>0.63</td>
<td>1.81</td>
</tr>
</tbody>
</table>

Note. “I” represents Item. The bold figures signify the variance and covariance values among the items in the same subscale.

In order to examine the factorial validity, we conducted structure equation modeling. The estimation method was maximum likelihood estimation. Standard errors were calculated using nonparametric bootstrapping at B = 1,000. The results showed a sufficient goodness of fit, CFI
= .959, TLI = .946, RMSEA = .074, with 90% CI [.066, .082], SRMR = .053, AIC = 33382, BIC = 33512. The estimated standardized coefficients can be seen in the path diagram in Figure 5. All the coefficients were statistically significant at $\alpha = .05$. The covariance among the latent variables showed lower values. This means that this inventory is practically informative.

In order to attain more practicality, the present study also examined a statistically simpler model, setting equality constraints on all of the factor loading values. This alternative model was submitted to structure equation modeling. The results showed a slightly degraded, but practically acceptable, goodness of fit, CFI = .947, TLI = .942, RMSEA = .077, with 90% CI [.069, .084], SRMR = .062, AIC = 33442, BIC = 33528. This statistical evidence can justify the use of summated scale scores instead of calculating factor scores. Furthermore, reliability coefficients for the subscales are summarized in Table 3. All the subscales signify sufficient reliability coefficients.

Table 3. Reliability Coefficients and their 95% Boundaries

<table>
<thead>
<tr>
<th>k</th>
<th>$\alpha$</th>
<th>95% Lower Boundary</th>
<th>95% Upper Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLP</td>
<td>4</td>
<td>.91</td>
<td>.90</td>
</tr>
<tr>
<td>ELP</td>
<td>4</td>
<td>.90</td>
<td>.89</td>
</tr>
<tr>
<td>FTFIP</td>
<td>4</td>
<td>.77</td>
<td>.75</td>
</tr>
<tr>
<td>All</td>
<td>12</td>
<td>.80</td>
<td>.80</td>
</tr>
</tbody>
</table>

In the path diagram, all the numbers are significant at $\alpha = 0.05$. The correlation between CLP and ELP is 0.17. The correlation between ELP and FTFIP is 0.17. The correlation between CLP and FTFIP is 0.15.

Figure 5. Path diagram showing the estimated parameters ($n = 909$).
Using a series of multiple sample analyses, the present study tested four types of measurement invariance with their respective sets of equality constraints, as described below: (a) configural invariance; (b) equal loadings, or so-called metric invariance; (c) equal intercepts or scalar invariance; and (d) equal mean structure, or full uniqueness measurement invariance (see Milfont & Fischer, 2015, for a methodological introduction). Only when scalar invariance or full uniqueness measurement invariance is satisfied, the users of this inventory become capable of comparing the factor scores of the subscales. The results of multiple sample analyses are shown in Table 4. In the table, all of the models show similar values of goodness of fit indices, but the present study found that the scalar invariance was satisfied. This means that users can apply the same measurement model across genders and compare their scores.

Table 4.
Summary of Multiple Sample Analyses (n = 903)

<table>
<thead>
<tr>
<th>Models</th>
<th>df</th>
<th>$\chi^2$</th>
<th>CFI</th>
<th>RMSEA</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Configural</td>
<td>101</td>
<td>385.68</td>
<td>.953</td>
<td>.078</td>
<td>33097</td>
<td>33472</td>
</tr>
<tr>
<td>(b) Metric</td>
<td>111</td>
<td>398.66</td>
<td>.952</td>
<td>.076</td>
<td>33092</td>
<td>33424</td>
</tr>
<tr>
<td>(c) Scalar</td>
<td>120</td>
<td>417.83</td>
<td>.951</td>
<td>.074</td>
<td>33093</td>
<td>33382</td>
</tr>
<tr>
<td>(d) Full Uniqueness</td>
<td>123</td>
<td>441.82</td>
<td>.947</td>
<td>.076</td>
<td>33111</td>
<td>33385</td>
</tr>
</tbody>
</table>

4. Discussion

The empirical evidence reported so far is sufficient to establish a new, practical, and statistically sound inventory to measure LMS. We call this inventory a Foreign Language Learning Mode Preference Inventory (FLLMPI) Ver. 1. This inventory has some desirable statistical features: (a) it has higher reliability, (b) the factorial structure fits well to observations and is analytically steady, (c) the use of the summed scale scores is practically sufficient, (d) scalar invariance among genders is realized, (e) the inventory consists of a small number of items, and (f) the covariance among the subscales shows lower values.

Although this new inventory is statistically sound and easy to handle in practice, there remain many conceptual and technical issues to be discussed. First, the participants in the present study were sampled from a single university. It is sure that the sample was fairly large and spanned across a variety of departments. However, this sampling will never be a guarantee of generalization to other types of university students. It is necessary to conduct a survey that investigates the reproducibility of the factorial structure.

Additionally, most of the validity evidence has not been addressed yet. For instance, future studies should examine the direct causality between LMP and learners’ decision-making. Likewise, correlational evidence with some other variables such as learning styles (e.g., Kolb,
2015), beliefs (Fushino, 2010), and attitudes (Kawaguchi & Kusanagi, 2016) should be accumulated to strengthen the LMS framework. The framework itself should be more theoretically sophisticated, despite its practical orientation.

Regarding measurement invariance, scalar invariance among genders is only one of the numerous types of measurement invariance. Invariance across other groups, such as school levels, departments, academic year, and teachers, remained to be untested. Additionally, time-series stability of LMS can be an interest in future studies.

5. Conclusion and Pedagogical Significance

Although there are some issues that remain unsolved as discussed above, the FLLMPI developed in the present study will be a useful tool for teaching practitioners, specifically those in higher education.

For instance, FLLMPI can be used as prior information for classroom placement. Currently, most classroom placements in practice are administrated using proficiency data, which are quite likely filled with larger errors. Through the use of FLLMPI, the teachers can refer to the additional information. Curriculum development and course design are also a promising area for the application of FLLMPI. Since learning modes have been diversified recently, as discussed in the Background section, FLLMPI can be an important resource to learn the needs of students. Thus, it can also be applied to prior classroom size determination. These applications will enable teaching practitioners to design more efficient learning environments. Of course, from a student point of view, it is informative to know one’s own LMP, since such meta-cognitive activity may promote learner autonomy.

However, the most important point is that we should never miss the ethical issues behind the labeling and categorization of students by an arbitrary trait. Every single student has his or her own personality, aptitudes, needs, thoughts, and feelings. Researchers and teachers should not forget that LMS, or any other learner variable, is only a very small piece of students’ information. Treating students as statistical values greatly diverges from the purpose of education.

The present study attached users’ practical guidelines for FLLMPI, which was written in Japanese, to the Appendix section. Additionally, URLs to the other supplementary data, including (a) the initial item pool, (b) R code, (c) other statistical data, and (d) a simulated data set that is parametrically equal to the data of the present study, are available. The present study fully agrees with the open science movement and reproducible research.
References


### Appendices

#### Appendix A.

Supplementary materials are available online at

https://sites.google.com/site/kusanagikuni/home/
Appendix B.

尺度使用の手引

尺度名: (和文) 外国語学習形態選好性尺度 ver. 1
(英文) Foreign Language Learning Mode Preference Inventory (FLLMPI)

適用母集団: 日本の大学生のみを想定している。この母集団以外に適用した場合の因子構造の再現は保証しない。また、日本の大学に在籍する学習者を対象としているため、学習者の国籍は基本的に問わない。

測定概念: 外国語の学習形態に関する選好性を測定する。三因子からなり、因子CLPは協同学習に対する選好性を、ELPはe-ラーニングに関する選好性を、FTFIPは、対面形式の学習に関する選好性を測定する。これら三因子は非常に弱い相関をもつ。

信頼性と妥当性: 各因子にそれぞれ高い信頼性係数（およそ.70から.90程度）が見られ、十分な内的一貫性があるといえる。ただし、再検査信頼性の検討はなされていない。測定モデルが高い適合度を示したという点において一種の妥当性の証拠がある。

測定不変性: 男女間において、配置不変性、因子負荷量と切片の不変性が成立しているため、男女間で同一の採点方法を適用し、男女間の因子得点差について検討することができる。

運用と採点方法: 7件法のリッカート尺度とし、各項目について「以下の文章について、自分にどれだけよくあてはまることでください」、またはこれに類する指示文で運用すること。簡易的に各因子の合計尺度得点を求めて因子得点の代用をすることは、実務上十分である統計的証拠がある。項目の変更をしないかぎり、質問紙自体を自作することは構わない。

尺度使用の許諾: 項目を改変、翻訳する場合を除き、出典を示すことで使用を許可する。使用に関する第一著者への連絡は特段要しない。

項目の改変および翻訳: 項目の改変や翻訳を行う際、以下の連絡先にその旨を伝えること。また、本尺度の項目を別の尺度の項目と合成し、新しい尺度として発表することについても、連絡を要する。他言語への翻訳に関する諸手続きについては、本研究の責任者が協力を約束する。

出典と連絡先: 本研究を引用すること。以下が本尺度に関する照会先である。
草薙邦広（広島大学外国語教育研究センター）
Email: kusanagi@hiroshima-u.ac.jp

The English version is available online.
Appendix C.

Foreign Language Learning Mode Preference Inventory

Cooperative Learning Preference (CLP)

<table>
<thead>
<tr>
<th></th>
<th>CLP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>英語を学習するとき、ペアで会話したり、意見を交換したりする活動を好む。</td>
</tr>
<tr>
<td>2</td>
<td>CLP</td>
<td>英語を学習するとき、グループで課題に取り組む形式の授業を好む。</td>
</tr>
<tr>
<td>3</td>
<td>CLP</td>
<td>英語を学習するとき、クラスメートと一緒に課題に取り組む形式が効果的だと思う。</td>
</tr>
<tr>
<td>4</td>
<td>CLP</td>
<td>英語を学習するとき、グループで力を合わせて課題に取り組むことが重要だと思う。</td>
</tr>
</tbody>
</table>

e-Learning Preference (ELP)

<table>
<thead>
<tr>
<th></th>
<th>ELP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>ELP</td>
<td>英語を学習するとき、コンピュータを使う勉強の仕方が自分に合っていると思う。</td>
</tr>
<tr>
<td>6</td>
<td>ELP</td>
<td>英語を学習するとき、コンピュータ上で問題を解く活動が効果的だと思う。</td>
</tr>
<tr>
<td>7</td>
<td>ELP</td>
<td>英語を学習するとき、コンピュータ上で行う演習を好む。</td>
</tr>
<tr>
<td>8</td>
<td>ELP</td>
<td>英語を学習するとき、自分のコンピュータやタブレット・スマートフォンを使う勉強法を好む。</td>
</tr>
</tbody>
</table>

Face-to-Face Instruction Preference (FTFIP)

<table>
<thead>
<tr>
<th></th>
<th>FTFIP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>FTFIP</td>
<td>英語を学習するとき、黒板や資料を使った講義形式の授業を好む。</td>
</tr>
<tr>
<td>10</td>
<td>FTFIP</td>
<td>英語を学習するとき、より深い知識が得られる講義形式の授業のほうが効果的だと思う。</td>
</tr>
<tr>
<td>11</td>
<td>FTFIP</td>
<td>英語を学習するとき、先生が英語の知識を学生に伝えることが重要だと思う。</td>
</tr>
<tr>
<td>12</td>
<td>FTFIP</td>
<td>英語を学習するとき、講義形式の授業が自分に合っていると思う。</td>
</tr>
</tbody>
</table>