研究ノート

# Validating and Analyzing a Survey on the Perceived Usefulness of Moodle for English Teachers

## Moodleの有用性に関する意識調査のための
## 英語教員向けアンケート項目の分析

**Peter PARISE**

本稿の目的は、国際言語文化アカデミアの事業の1つである「英語教育アドヴァンスト研修」（高等学校英語中核教員研修）で実施した、e-ラーニングのプラットフォーム「Moodle」の有用性の意識を調査したアンケートの妥当性を分析することである。「有用性の意識」とは、このシステムが課題遂行に役立ったという受講者の確信のことを言う。Messick (1995)の6つの「内容妥当性」を妥当性の指標として使い、Raschモデルによる分析を行った。分析の結果、このアンケートの信頼性は高いものの、いくつかの質問項目がRaschモデルに適合せず、内容妥当性を確保できない可能性があることがわかった。そこで、この問題に対処するため、質問項目の再構築の方法についてさらに考察した。

## INTRODUCTION

The intention of this study is to determine the validity of a survey instrument for measuring the opinions of the Advanced Program participants regarding their use of Moodle, a website used as part of the program. The aim of the website, http://kanagawa-ilcs-training.com /moodle, is to aid Japanese teachers of English in public high schools in Kanagawa to perform action research to enhance their skills in terms of teaching methods and language ability.

The Advanced Program is a year long program managed by the Teacher Training Division of the Kanagawa Prefectural Institute of Language and Culture Studies to guide Japanese high school teachers to engage in action research. In this process, they are to find an issue that impacts their classes, develop a treatment for it, administer that treatment, and then report on the findings in a presentation in English and a report in Japanese. This program entails two observations in June and November in which the trainers from the Academia visit the respective schools of the participants to take note of their needs in terms of pedagogy or language teaching. In between these observations, the teachers must attend mandatory sessions to improve their teaching and language skills as well as hone their action research.

To support this process of teacher training, the Advanced Program makes use of Moodle ("About Moodle," 2014; Dougiamas & Taylor, 2003), a type of open-source learning management system (LMS) that helps teachers or educational institutions to build and manage a website in

order to facilitate instruction. Because of its open nature, the software is freely available, and users can customize their Moodle website. As a result, this has promoted the development of a vibrant community of teachers and programmers devoted to maintaining and updating the platform. For the Advanced program, Moodle serves a variety of purposes. One purpose is to review the material covered in the face-to-face sessions. This is facilitated using the Moodle quiz function in order to encourage the participants to review the recently learned material as a way to further their understanding, but also to inform the trainers if the information offered in the lectures was understood or not (Parise, 2016). The quiz also serves to keep the participants mindful of the process of action research by offering tasks that encourage them to submit research questions, report about a treatment, or to practice statistics relevant to their projects. The third purpose is to be a repository of materials produced by both the teacher trainers and their peers during the sessions. The idea behind this is to encourage sharing best teaching practices in terms of materials development, but also for reviewing the content of the seminars. The final purpose is to practice vital skills needed to improve their language ability such as their ability to read, write and listen to English. The website is written almost entirely in English, and this offers the participant a rich environment to interact in English in contrast with what is available at their school where most transactions are facilitated in Japanese. This study can be seen as a hybrid of computer aided language learning (CALL) and e-learning in which technology is used in a twofold manner: as a means to enhance teaching methods while at the same time improve L2 language skills vital for the teachers' professionalism.

An opinion survey was developed to obtain the *perceived usefulness* of Moodle by the participants based on their experience of the website after one year. The inquiry is to determine if the participants felt that the website served their purposes. The main goal of this article is to determine, based on the statistical analysis of two pilot administrations of this survey, if the instrument can validly measure such opinions. In short, does the survey measure what it intends to measure, or are there discrepancies in the results that can mislead our interpretations of the e-learning aspect of this course?

The second intention is to introduce a framework and statistical method to evaluate surveys. Evaluating a survey starts by posing questions in the form of a *validity argument*. To do this, different aspects of validity must be considered. For this study, the ideas presented by Messick (1995) will guide the analysis to determine if the survey needs to be adjusted in terms of wording and design in order to be fully confident in the survey instrument's results. This study will use the Rasch model as a means to evaluate the survey by providing evidence for supporting our validity argument.

# LITERATURE REVIEW

## The Construct of Perceived Usefulness

One of the main considerations when developing an opinion survey is to determine the construct of interest. For this study, we are interested in how the participants feel after interacting with the Moodle website and how their perceptions of it influenced how they have used it during the Advanced program. A fitting construct, which is quite relevant for research in technology, is perceived usefulness. This construct was identified from data from a host of studies in a variety of technology-based sectors such as programming, information reports and telecommunications and is defined as the belief that the application helps the participant perform their job (Davis, 1989, pp. 320-322). Results of the scales developed for the construct of perceived usefulness yielded high Cronbach alpha scores for two studies conducted, $\alpha = .97$ for study 1 and $\alpha = .98$ for study 2, revealing a high degree of reliability (Davis, 1989). Identifying this construct has led to its adoption in the investigation of the usefulness of websites via regression modeling. The Technology Adoption Model (TAM) is one such regression model that has perceived usefulness as a factor to predict the adoption of an application or system by its users.

## Perceived Usefulness and Moodle

The main purpose of Moodle has been to provide a space for learning in a way that promotes social constructionist principles (Dougiamas & Taylor, 2003). In terms of perceived usefulness, this construct is used along with other factors to build regression models for evaluating the users' response to Moodle. Sánchez and Hueros, (2010) used the TAM model to evaluate Moodle for students of business and education in a university in Spain. Perception of usefulness was a factor in this model and the results showed that it was positively associated with attitude towards Moodle (pp. 1637- 1638). Damnjanovic, Jednak, and Mijatovic, (2015) investigated the perspective of students using Moodle from higher learning institutions in Serbia, Lithuania, and Bosnia and Herzegovina utilizing an Information Systems (IS) Success model (DeLone & McLean, 1992; 2003) which takes perceived usefulness as a factor. In contrast with the TAM, this model emphasizes system quality, which then feeds into perceived usefulness. The results showed that while the perception of usefulness was influenced by the system and information quality of Moodle, it was significantly associated with the intention to use Moodle in the future (p. 508). These two studies show that perceived usefulness influences the users' attitude toward Moodle as well as encourage longer-term use of the website. The construct of perceived usefulness is embedded in educational uses of technology, but the issue is applying it to a unique environment that blends both language education and job training such as the Advanced Program. This article's contribution is to explore how the construct of perceived usefulness can be represented in this

context and not only focus on concepts that "helps the user perform their job" as it was originally conceived but to expand that to include another facet of the job of a language teacher such as increasing language skills, in a survey.

**Construct Validity**

In order to determine if the survey measures perceived usefulness as discussed above, the *construct validity* needs to be evaluated. There are two types of threats to construct validity, one is *construct underrepresentation*, meaning that the assessment is too narrow to adequately represent the construct, or *construct irrelevant variance*, which means that the measurements are too broad and that other constructs create "noise" in the data (Messick, 1995, p. 9). An additional concept related to the latter threat is *construct irrelevant difficulty*, in other words, the question is too difficult to endorse. If no one endorses an item or it is rarely endorsed, it may affect the interpretation since the item cannot clearly discriminate differing opinions. *Construct irrelevant easiness* is the opposite yet equally problematic. If a question item is too easy to endorse and because of this, the question may fail to show any real differences in the participants since they all answer in the same fashion. These ideas are considered when evaluating this survey and will be addressed later in this article. Construct validity is conceived as six aspects: *content, substantive, structural, generalizable, external,* and *consequential.* These aspects are a way to organize evidence to support the validity of an instrument (Messick, 1995, Wolfe & Smith Jr, 2006). As mentioned in the introduction, this article will utilize the Rasch model to obtain data related to these aspects.

**The Rasch Model**

Rasch is a type of regression model utilized for developing tests and surveys in research for education and medicine. It is a variation of Item Response Theory, which is the calculation of the probability of the respondents' abilities in relation to the difficulty of the question item. The features of this model are that the items and respondents are conceived on a linear scale. The best metaphor to describe this is that of a ruler or a thermometer. The construct and its measurement on a population is based on degrees of difficulty of the question items on this linear scale. In order to achieve this linearity, the *unidimensionality* of the construct must be maintained. In other words, the construct of perceived usefulness needs to be isolated and other conflicting constructs must not interfere in the measurement and in so doing, avoid the threat of construct underrepresentation. Another aspect is *item fit* to the Rasch model. This serves as a form of quality control for the survey as a way to ensure that the items measure the proposed construct (Bond & Fox, 2010, pp. 35-36). A related concept is the *item difficulty.* A linear scale can allow the plotting of easy and difficult items, and based on the responses of the participants to those items the

behavior of the participants is obtained. In addition, item difficulty measures guard against the threat of both construct irrelevant difficulty and ease because if the items are too hard or easy to endorse, then we lose the ability to interpret the data of the survey.

**Rasch and Contruct Validity**

Rasch analysis allows for addressing Messick's six aspects of construct validity. For this study, three of these aspects will be observed to determine the validity of the survey instrument: structural, content, and generalizable. The structural aspect is the degree of which the scoring structure matches the construct. One way to do this is to use correlations to determine the unidimensionality of the construct (Wolfe & Smith Jr, 2006, p.213). To what degree do the items of this survey truly measure perceived usefulness or are there other constructs at work that may impact the interpretation of this survey? The Rasch model can determine this with measures derived from factor analysis in order to confirm its unidimentionality. In terms of the content aspect, which reflects the "relevance and representativeness" of the items and their technical quality (Wolfe & Smith Jr, 2006, p. 205). One way is to assess the wording of the items in comparison with other instruments that measure the same construct. Another approach is to look at the technical quality of the items by using mean fit square measures (Green, 2013; Wolfe & Smith Jr, 2006) and the category function measures (Bond & Fox, 2010) of the Rasch model. Finally, the generalizable aspect, which is the determination of how the score represents the group or groups can be determined via an item-person map (Bond & Fox, 2010; Green, 2013; Wolfe & Smith Jr, 2006) of the Rasch model. These three aspects along with the relevant measures allow us to determine the validity of the survey instrument.

**Research Questions**

Based on the literature reviewed above, this study puts forth the following research questions:

1. To what degree does this survey represent the construct of perceived usefulness when inquiring to the Advanced Program participants about the use of Moodle, with the structural, content, and generalizable aspects of construct validity via the Rasch model?

2. What sort of insights can we gather about the survey items based on the results of this analysis?

3. What actions can be taken based on these insights to improve this survey?

## METHOD

**Participants**

        The participants for this study are two cohorts of the Advanced program: the 2015 and 2016 cohorts. The former consists of 25 in-service teachers and the latter 15 in-service teachers combined to make 40 respondents. All teachers have been serving in high schools throughout the Kanagawa prefecture at a minimum of 5 years.

**Advanced Program and Moodle**

        As mentioned in the introduction, the Advanced Program involves monthly sessions between the June and November observations. The program in relation to Moodle can be described a *blended online course* where the face-to-face meetings are supported with technology or a combination of "face-to-face instruction with computer-mediated instruction." (Bonk & Graham, 2006, 8-10). The face to face sessions consist of 9 monthly meetings where the participants attend workshops at the Academia. The other half is online and used to follow-up on the sessions and functions concurrently with the face-to-face sessions as shown in Figure 1.

| May | Day 1 | Goal-setting/Survey 1 | moodle |
|-----|-------|------------------------|--------|
| June | | 1st class observation (on site) | |
| July | Day 2 | Problem-identification | moodle |
| Aug | Day 3-6 | Intensive summer workshops | moodle |
| Oct | Day 7 | Rationale for Teaching Activities | moodle |
| Nov | Day 8 | Recording & Reporting AR 2nd class observation (on-site) | moodle |
| Jan | Day 9 | AR Presentations/Survey 2 | moodle |
| Feb | | Action Research Reports | |

*Figure 1.* The Advanced Program schedule with accompanying days supported by Moodle.

**Survey Instrument**

        The main instrument for this study is a 16-item paper based survey administered on the final day of the Advanced Program following the presentations in English. The first 14 items consist of Likert scaled questions which ask the participant to endorse relating to their use of Moodle. The final two items are open response items. One item is a sentence completion prompt: "The Moodle would be improved if_____." Here the respondent needs to write what they think would complete the sentence and at the same

time offer their views on how the Moodle would be improved for other cohorts in the future. The final item is a fill-in section that requires the participant to write what they feel or think about the Moodle that was not addressed in the previous items. For this paper, the focus will be on the 14 initial Likert scale items to look how these reflect the construct of perceived usefulness discussed in the literature review. Table 1 lists all the Likert scaled items. The entire survey is in Appendix.

Table 1 *Question Items for the Survey*

| Question Item |
| --- |
| 1. The Moodle website was easy to use. |
| 2. I visited the Moodle often during this program. |
| 3. I was able to access the Moodle at my school. |
| 4. I was able to access the Moodle at home. |
| 5. The Moodle helped me plan for my action research project. |
| 6. The Moodle helped me improve my teaching style. |
| 7. The quizzes on Moodle helped me review the material from the sessions at the Academia. |
| 8. I was able to practice writing in English on the Moodle. |
| 9. I was able to practice listening to English on the Moodle. |
| 10. I was able to practice reading English on the Moodle. |
| 11. I think the Moodle website is an important part of the Advanced program. |
| 12. I feel I am part of a "teacher community" through the Moodle. |
| 13. I felt that the Moodle took time away from my teaching work. |
| 14. Overall, I think the Moodle site was helpful. |

**Item Construction**

In order to address the needs of the Advanced program participants related to their work as teachers and trainees the items were written to serve various purposes related to perceived usefulness. Items 6, 12, 13 are related to their work as a teacher and ask about how the Moodle influenced their teaching. Items 8, 9, 10 are related to the development of their language ability in English. These essentially ask about the usefulness of the site for language development, in particular reading, writing and listening. Items 1, 2, 3, 4, 5, 7 are related to their work as a participant in the Advanced program, in particular in the Moodle's support of their action research projects. Item 11, 14 relate to their general impression of the site. A 6-point Likert scale is chosen for this survey, which ranges from 1, "strongly disagree" to 6, "strongly agree". Six is an optimal scale for statistics as well as ideal for the elimination of a neutral selection thereby increasing the precision of the survey (Nemoto & Beglar, 2014).

**Analysis Tools**

      Rasch Analysis will be conducted to look at the structural, content and generalizable aspects of the validity of the survey as discussed in the literature review. Winsteps 3.8.1 (Linacre, 2013) will be used for a majority of this analysis. In order to provide additional support to the results of the dimensionality of the measures, a separate Principal Components Analysis will also be conducted using R (R Core Team, 2016).

<div align="center">

**RESULTS**

</div>

Table 2 *Rasch Analysis Summary*

```
_____
|PARTICIPANT      40 INPUT       40 MEASURED         INFIT        OUTFIT     |
|         TOTAL       COUNT    MEASURE  REALSE     IMNSQ   ZSTD  OMNSQ    ZSTD|
|MEAN      57.5        13.8       .61      .31      1.01   -.1   1.02     -.2|
|S.D.      11.6          .5       .82      .06       .61   1.5    .73     1.6|
|REAL RMSE    .31 TRUE SD     .75  SEPARATION  2.40  PARTIC RELIABILITY   .85|


_____
|ITEM        14 INPUT        14 MEASURED         INFIT        OUTFIT     |
|         TOTAL       COUNT    MEASURE  REALSE     IMNSQ   ZSTD  OMNSQ    ZSTD|
|MEAN     164.1        39.6       .00      .18       .99   -.3   1.02     -.3|
|S.D.      29.1          .6       .71      .03       .58   2.2    .73     2.4|
|REAL RMSE    .18 TRUE SD     .69  SEPARATION  3.84  ITEM   RELIABILITY   .94|

_____
```

      The results reveal that the standardized residuals of the Rasch analysis had a mean of .04 and a standard deviation of 1.01. Both of these show that the data conforms to the model. In addition, in Table 2 the participant reliability is .85 that shows the possibility of obtaining similar results from a different population, though this reading may be interpreted as not entirely representing the participants. (Green, 2013, p. 159). The item reliability in Table 2 is .94, indicating that we can have a high degree of confidence that the survey items will yield the same score should it be administered to a similar population. This measure is as the same as Cronbach Alpha in terms of reliability.

Table 3 *The Principal Component Analysis of the Model*

| Standardized Residual Variance in eigen values | Observed | |
|---|---|---|
| Total raw variance in observations | 31.4 | 100.0% |
| Raw variance explained by measures | 17.4 | 55.4% |
| Raw variance explained by persons | 6.9 | 21.9% |
| Raw Variance explained by items | 10.5 | 33.5% |
| Raw unexplained variance (total) | 14.0 | 44.6% |
| Unexplained variance in 1st contrast | 2.9 | 9.3% |

Table 4 *Contrasts from Principal Component Analysis*

```
-----------------------------------------------------
|CON-  |       |             INFIT OUTFIT| ENTRY         |
| TRAST|LOADING|MEASURE  MNSQ   MNSQ |NUMBER ITEM    |
|------+-------+------------------+--------------|
|  1 1 |  .83  |   1.54 2.53 3.25 |A   13 ITEM-13 |
|  1 1 |  .57  |    .31  .93 1.02 |B    2 ITEM-2  |
|  1 1 |  .36  |    .90  .90  .93 |C    9 ITEM-9  |
|  1 2 |  .22  |    .56  .69  .67 |D    8 ITEM-8  |
|  1 2 |  .13  |    .56 2.16 2.10 |E    3 ITEM-3  |
|  1 2 |  .00  |    .24  .65  .61 |F   10 ITEM-10 |
|------+-------+------------------+--------------|
|  1 3 | -.77  |   -.67  .62  .58 |a   14 ITEM-14 |
|  1 3 | -.64  |   -.62  .75  .69 |b   11 ITEM-11 |
|  1 3 | -.61  |   -.31  .74  .75 |c   12 ITEM-12 |
|  1 3 | -.39  |   -.88  .79  .78 |d    7 ITEM-7  |
|  1 3 | -.31  |  -1.17 1.11  .92 |e    4 ITEM-4  |
|  1 2 | -.19  |   -.10  .48  .47 |f    6 ITEM-6  |
|  1 2 | -.18  |   -.28  .50  .53 |g    5 ITEM-5  |
|  1 2 | -.08  |   -.07  .98  .98 |G    1 ITEM-1  |
|------+-------+------------------+--------------|
```

**Structural Aspect**

**Dimensionality Measures**

Regarding the structural aspect of the validity of this survey, the Winsteps software performs a Principle Component Analysis of the items and produces data of the standardized residual variance as shown in Table 3. The total raw variance is explained by the Rasch model accounts for 31.4 eigen values with 55.4% explained by the measures, and 33.5% explained by the items and 21.9% explained by the actions of the participants. In terms of unidimensionality, one approach is to check the unexplained variance in the first contrast, which means that any

variance that is not explained by the observations may imply that there is a second factor at work. The eigen value is 2.9 which means almost three items in this survey are behaving as a secondary factor. According to Linacre, a second dimension needs to be the strength of three items, so these measures are on the cusp of having two dimensions (Linacre, 2014, p. 496). Another way to confirm if there is a secondary dimension is to look at the contrasts between the loadings of eigen values for the items to determine the dimensionality. Anything greater than ±.4 in terms of eigen value loading is regarded as another dimension. In Table 4, Items 13 and 2 may be a factor as well as Item 14, 11, and 12 in the lower half of the table.

Figure 2 provides a graphical confirmation of this by showing how Item 13 and 2 are clearly misaligned with the rest of the items and may be its own factor. Utilizing R (R Core Team, 2016), a Principle Components Analysis was conducted and the output in Figure 3 shows that Item 13 and 2 are clearly their own factor as F3 and the Items 14, 11 and 12 clearly are part of the first factor F1.

Item 9 seems like an ambivalent factor. While it does seem to be its own construct, its eigen values also load positively for the first factor. This ambivalence is related to the question of "listening" practice afforded by the site. For the 2015 cohort the Moodle had more listening tasks compared to 2016, and this discrepancy may be the reason for the ambivalent readings.
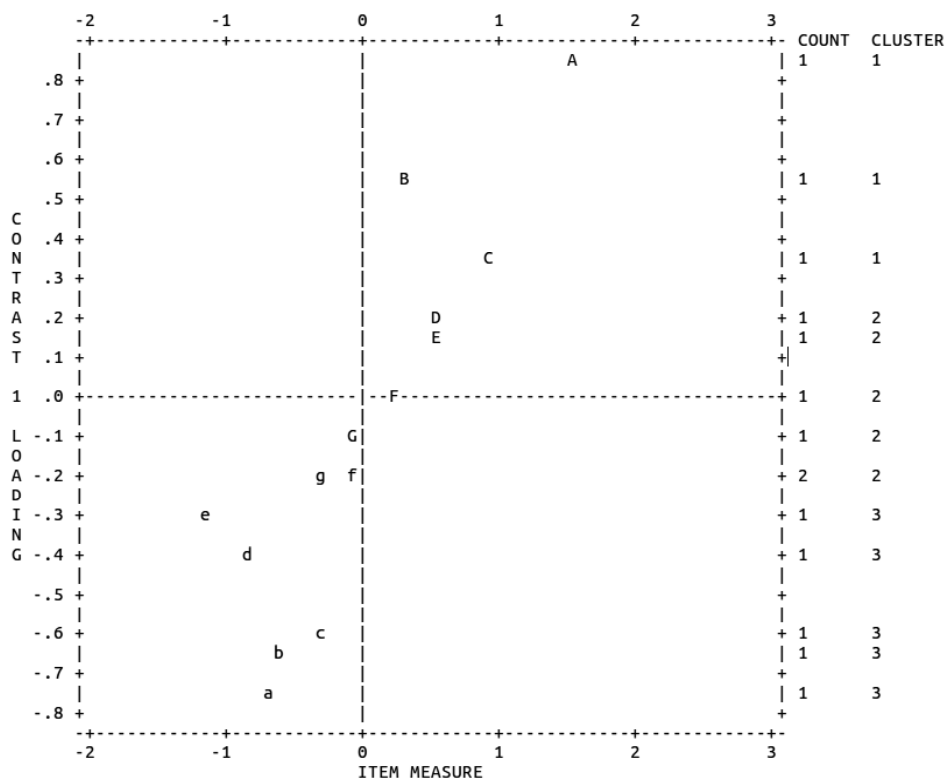


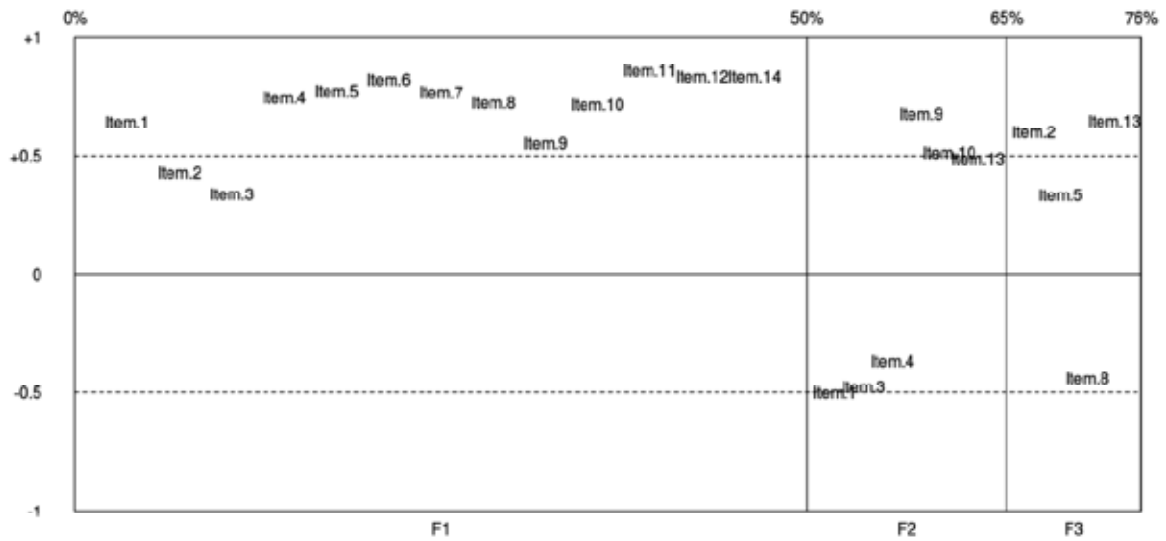*Figure 2*. Standardized Residual Plot for items.

*Figure 3.* Factor loadings of the items based on Principal Components Analysis via R.

Table 5 *Misfit Order of the Items*

```
-------------------------------------------------------------------------------
|ENTRY   TOTAL   TOTAL            MODEL|   INFIT  |  OUTFIT  |PTMEASURE-A|EXACT MATCH|      |
|NUMBER  SCORE   COUNT   MEASURE  S.E. |MNSQ  ZSTD|MNSQ ZSTD|CORR.   EXP.| OBS%   EXP%| ITEM |
|-----------------------------------------+----------+----------+-----------+-----------+------|
|   13      95      38     1.54    .16|2.53  4.9|3.25  6.0|A-.15   .58| 10.5  41.6|  13  |
|    3     139      39      .56    .15|2.16  4.3|2.10  4.0|B .42   .64| 23.1  34.0|   3  |
|    4     202      39    -1.17    .20|1.11   .5| .92  -.2|C .73   .60| 51.3  49.0|   4  |
|    2     155      40      .31    .15| .93  -.3|1.02   .2|D .46   .65| 35.0  34.8|   2  |
|    1     171      40     -.07    .16| .98   .0| .98   .0|E .57   .65| 52.5  38.2|   1  |
|    9     128      40      .90    .15| .90  -.4| .93  -.2|F .62   .63| 52.5  34.9|   9  |
|    7     199      40     -.88    .19| .79  -.8| .78  -.9|G .73   .62| 40.0  41.9|   7  |
|   11     191      40     -.62    .18| .75 -1.1| .69 -1.4|g .82   .63| 45.0  40.1|  11  |
|   12     180      40     -.31    .16| .74 -1.2| .75 -1.1|f .79   .65| 42.5  38.8|  12  |
|    8     141      39      .56    .15| .69 -1.6| .67 -1.7|e .73   .64| 43.6  34.5|   8  |
|   10     158      40      .24    .15| .65 -1.8| .61 -2.0|d .76   .65| 47.5  35.3|  10  |
|   14     188      39     -.67    .18| .62 -1.7| .58 -2.0|c .85   .64| 51.3  40.9|  14  |
|    5     179      40     -.28    .16| .50 -2.6| .53 -2.5|b .76   .65| 52.5  38.8|   5  |
|    6     172      40     -.10    .16| .48 -2.9| .47 -2.9|a .84   .65| 60.0  38.6|   6  |
|-----------------------------------------+----------+----------+-----------+-----------+------|
| MEAN   164.1    39.6      .00    .16| .99  -.3|1.02  -.3|           | 43.4  38.7|      |
| S.D.    29.1      .6      .71    .02| .58  2.2| .73  2.4|           | 12.7   3.9|      |
-------------------------------------------------------------------------------
```

**Content Aspect**

**Mean Fit Square Measures**

In addition to the dimensionality, mean fit square measures (MNSQ) allows us to evaluate the fit of the items to the Rasch model. By determining this degree of fit, we can find items that may diverge from the construct and interfere with the interpretation of the survey. Both infit (INFIT) and outfit (OUTFIT) show the degree of variation of the measures. Mean square measurements higher than 1.5 are considered unproductive for measurement due to the

unpredictability of the responses. (Green, 2013, p. 169). In Table 5, Item 13 and 3 exemplify this with measures higher than the benchmark 1.5. The Z-statistic (ZSTD) shows the degree of misfit as well with any measure over 2.0. Both Item 13 and 3 are clearly over this with 6.0 and 4.0 for outfit and 4.9 and 4.3 for infit respectively.

In addition, any mean square measurements with less than .50 also show that some items are unproductive by being too "predictable." In other words, if the items are at or below that threshold, nothing interesting can be learned about the participants. This is definitely the case with Items 5, which is at the cusp with a measure of .50 and Item 6 at .48. The z statistic for both infit and outfit is below the 1.5 benchmark, signaling that these are problematic. Further analysis and discussion of these items will be addressed in later sections of this article.
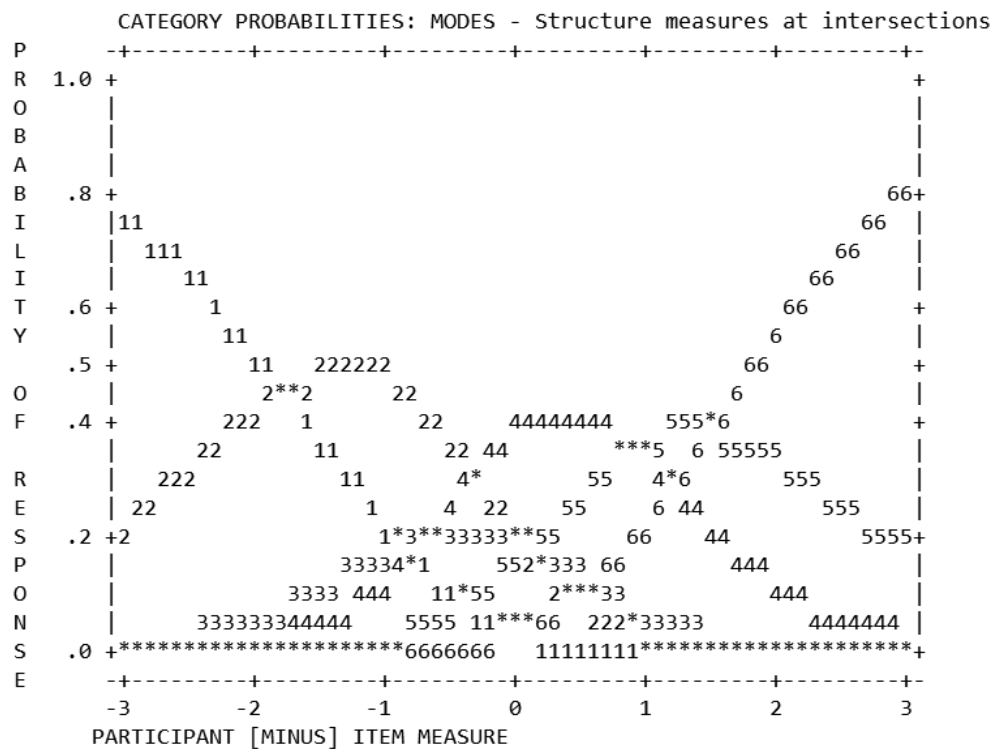
```
         CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P        -+---------+---------+---------+---------+---------+---------+-
R   1.0 +                                                              +
O        |                                                             |
B        |                                                             |
A        |                                                             |
B    .8 +                                                           66+
I        |11                                                     66 |
L        |   111                                               66  |
I        |      11                                            66   |
T    .6 +        1                                          66     +
Y        |          11                                    6        |
     .5 +            11    222222                         66        +
O        |             2**2        22                    6         |
F    .4 +            222   1          22      44444444    555*6     +
         |         22        11         22 44        ***5  6 55555  |
R        |       222           11        4*       55    4*6       555 |
E        | 22                    1      4  22     55     6 44       555 |
S    .2 +2                   1*3**33333**55      66     44          5555+
P        |                   33334*1      552*333 66         444       |
O        |               3333 444    11*55    2***33          444     |
N        |          333333344444    5555 11***66  222*33333       4444444 |
S    .0 +*******************6666666    11111111*******************+
E        -+---------+---------+---------+---------+---------+---------+-
         -3        -2        -1         0         1         2         3
         PARTICIPANT [MINUS] ITEM MEASURE
```

*Figure 4.* Category Function Measures shown as distributions.

**Category Function Measures**

A secondary consideration when considering the content aspect of validity is the accuracy of the Likert scale itself. As noted earlier, a selection of six categories ranging from 1 being "strongly disagree" to 6 being "strongly agree" is said to be the most accurate for statistical purposes. The output of Winsteps also enables the evaluation of each category and determine if a selection of six categories is feasible for gathering the opinions of the Advanced Program participants. Figure 4 shows the distribution of each category with category 1 and 6 being skewed

distributions and each with a probability of .8 or 80% of being selected as an answer to the question items. The middle categories however are much lower in frequency and are lower in terms of distribution. This is evident with category 3, "somewhat disagree" with a .2 or 20% chance of being selected. In terms of distribution, it is practically subsumed by the category 2 "disagree" and category 4 "agree". It may be feasible to collapse these categories and join both 2 and 3 together along with 4 and 5 to create a four-category Likert scale to more accurately reflect the choices of the participants (Bond & Fox, 2010, pp. 223-225).

## Generalizable Aspect

### Person-Item Map and Item Difficulty

Addressing the generalizable aspect of the validity of the survey is done with Figure 5, a participant-item map to look at how this population responded to the items and the difficulty of those items as well.

The Person-Item Map measures out in logits, the degree of probability of the skill of the participants (X) and the difficulty of the items (ITEM-). The left side of the map is the distribution of the participants, on the right the items themselves laid out in a linear scale. The higher the item is on the scale, the greater the difficulty of endorsing the item i.e. "agreement" and the higher the participant, the more likely they have endorsed some or all of the items in the survey.

The mean of the participants is set 1/2 a logit above the mean of the items revealing that most of these participants have endorsed about half of the items. The distribution of the participants divides into two peaks, both reflecting how some items were difficult to endorse. One peak centers on the mean itself and is associated with Items 3 "I was able to access Moodle at my school" and Item 8 "I was able to practice writing in English on the Moodle." Item 3 reflects the hardship certain participants experience when trying to access Moodle, which may be due to the ICT infrastructure of the participant's school. Item 8 though is much more subtle, reflecting whether the Moodle adequately provides writing practice according to the participants' conception of "writing" or that the wording of the question is too ambiguous. The second peak settles beneath Item 13 "I felt that the Moodle took time away from my teaching work." Only a small number of participants, approximately 3 teachers endorsed this item, but they do not reflect the majority opinion. In the tail end of the distribution, going into negative logits we find five participants below the mean showing that even Item 4 "I was able to access the Moodle at home" was difficult to endorse for two participants.
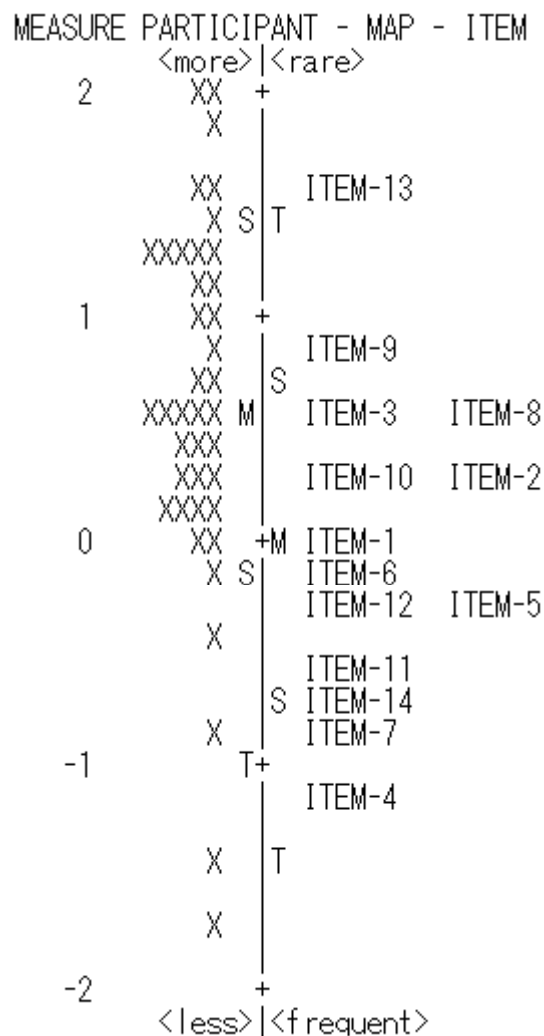
```
        MEASURE  PARTICIPANT - MAP - ITEM
                  <more>|<rare>
         2       XX   +
                  X    |
                       |
                 XX    |    ITEM-13
                  X  S|T
               XXXXX   |
                 XX    |
         1       XX   +
                  X    |    ITEM-9
                 XX   |S
               XXXXX  M|    ITEM-3    ITEM-8
                XXX    |
                XXX    |    ITEM-10   ITEM-2
               XXXX    |
         0       XX   +M  ITEM-1
                  X  S|    ITEM-6
                       |    ITEM-12   ITEM-5
                  X    |
                       |    ITEM-11
                      |S   ITEM-14
                  X   |    ITEM-7
        -1            T+
                       |    ITEM-4
                       |
                  X   |T
                       |
                  X    |
                       |
        -2            +
                  <less>|<frequent>
```

*Figure 5.* Person - Item Map of the Moodle Survey

In terms of the items, most are shown fitting the Rasch model as demarked by the "T" on the map. All the items fit except Item 13, revealing that this question may be problematic for the population. This item is difficult to endorse, and so may need to be adjusted to better fit the participants and the model. This issue will be discussed in later in this article. In terms of the participants fitting the Rasch model, as marked with the "T" two participants fall outside this demarcation, indicating that they are unique in their responses to this survey.

### DISCUSSION

**First Research Question**

The structural, content, and generalizable aspects of validity provide a robust data set for evaluating the survey and answering the initial research questions. Beginning with the first

research question, to what degree does this survey represent the construct of perceived usefulness when inquiring to the Advanced Program participants about the use of Moodle, through the structural, content, and generalizable aspects of construct validity via the Rasch model?

**Structural Aspect**

Starting with the structural aspect, the dimensionality measures of the Rasch model as well as an additional Principle Components Analysis reveal that most of the items load on one factor as shown in Figures 2 and 3 in the results. However, as shown with Tables 3 and 4 there is the presence of an additional factor, which are represented with Item 13 "I felt that the Moodle took time away from my teaching work." and Item 2 "I visited the Moodle often during this program." What these two items seem to imply is the relationship between the notion of "time", with keywords such as "took time away" and "often." In reference to the construct of perceived usefulness, referring to time in an item seems to run counter with the other items of the survey. One solution is to integrate these into the main construct by skillfully aligning "time" in relation to "usefulness." This means that these two items may be reworded to achieve this integration.

**Content Aspect**

With the content aspect of validity, the mean fit square measures of the Rasch model provided hints of what sort of items can be problematic with our survey. One interesting discovery is that the items that under and overfit the model contradict the dimensionality measures. One highly misfitting item is 13 "I felt that the Moodle took time away from my teaching work." While this is paired with Item 2 in the dimensionality results, the item that misfits is Item 3 "I was able to access Moodle at my school." These clearly exemplify the construct irrelevance in this survey, since both of these are over the 1.5 threshold and are too unpredictable to be of use in the survey. One question regarding these results is about Item 2 from the previous section. The solution to this is that misfitting items be addressed first and then the issue of dimensionality comes afterward as a secondary consideration (Bond & Fox, 2010; pp. 252-253 referencing Linacre, 1998). This means that since Item 2 fits the model, it can remain, but Item 3 and 13 need to be adjusted.

The opposite problem are items that underfit the model, in other words questions that too predictable to be of use in the survey and therefore create construct irrelevant easiness in the survey. These are Items 5 "The Moodle helped me plan for my action research project" and 6. "The Moodle helped me improve my teaching style." One possibility for these to be overly predictable is these items are stating the purpose of the Moodle and so to endorse these is simply endorsing the obvious. These are very general questions, but do not provide details about *how* the Moodle helped in terms of action research and teaching. One solution to this is to focus on specific aspects of the action research process for example in developing a research question, or thinking

of a treatment. In addition, specific questions about "teaching style" may be necessary too, in terms of what type of activities regarding the four skills of listening, reading, writing and speaking are helped by interacting with the content of the Moodle.

With the category choices of the questions, based on the data obtained from the Winsteps output regarding the six Likert scale choices, for the next iteration of the survey it may be best to reduce the selection to four. The case may be in light of the underfitting items is that the questions are too general to warrant a fine-grained response, but if the questions are more specific, then six responses may be adequate. Further piloting is warranted here.

**Generalizability Aspect**

The final aspect generalizability is shown based on the Person-Item Map data. This population was able to endorse most the items and were able to fit the Rasch model save for a few outliers. The difficulty of the items is evenly spread over the distribution with a few exceptions as discussed above. In addition, with the high reliability measures for both the items and the participants as shown in Table 2 there is a strong likelihood that if another cohort were to take this survey the results would be replicated. We must also note that should any revisions as well any additional items be added, that the measures spread out along a linear continuum as demonstrated in Figure 5. If not the items will fail to mark levels of difficulty.

**Second Research Question**

The second question: What sort of insights can we gather about the survey items based on the results of this analysis? The aspects of validity reveal that while the survey questions are generally reliable in terms of the difficulty of the items, the structural and content aspects hint that the survey must be adjusted. The most blatant change must come to Item 13, since this clearly runs counter to the Rasch model and overfits it. Should this item be adjusted then the dimensionality may be reduced to one factor alone. Regarding Items 3, 5, and 6, as discussed above, these need adjustment as well. Item 3 reflects an issue that really has more to do with the infrastructure of schools, but not the Moodle itself. Items 5 and 6 are rather obvious statements of fact that endorsement ends up being quite predictable, and so specificity is warranted here.

**Third Research Question**

The final research question is: What actions can be taken based on these insights to improve this survey? As discussed above, a rewording of the overfitting and underfitting items is the first priority. With the misfitting Item 13, "I felt that the Moodle took time away from my teaching work," if this item were to be reworded, how can "time" be combined with perceived usefulness? Time management is one vital issue for teachers in Japanese high schools that impacts

their ability to adopt more progressive teaching approaches. Because of the variety of duties they need to perform, adopting new ways of teaching English become very difficult for teachers (Sakui, 2004; Sato & Kleinsasser, 2004). The same can be said of action research. The usefulness of the Advanced Program is to support this adoption and facilitate a sense of a "teacher community" in the process. Rather than phrasing this item in a negative way as "taking away" time, using neutral words that imply integration with the teachers' school schedule may be the best strategy here. Item 13 can rewritten as "Using Moodle fit with my current teaching schedule" to both integrate the usefulness of the site in consideration of their schedule.

Other misfitting items such as Item 3 for example focus on access to the Moodle rather than the qualities of the Moodle itself. In fact, Items 2, 3 and 4 focus exclusively on this, which begs the question whether "access" should be considered synonymous with "usefulness?" One argument would be that access is an issue that is external to the workings of the website. However, access is vital to usage because trainees are expected to log in to perform the quiz or access materials for their action research, which incidentally is undertaken at their respective school. Should access be limited at school, the fallback is to access at home which relates to Item 4 "I was able to access Moodle from home." This item has no trouble fitting the Rasch model yet is phrased similarly to Item 3 and deals with access as well. One consistency for the top four misfitting items found in Table 5 is that all the questions of access are "I" statements, which seems to put the onus on the participant and this may explain the issues of misfit to the model. Issues of access, especially for a school with a compromised ICT infrastructure is not the responsibility of the respondent, so rewriting these items to emphasize the situation may be a better strategy. Item 3 can be rewritten as "The Moodle was accessible at my current school" and Item 4 "The Moodle was accessible at home." These items might have improved fit to the Rasch model as well as make the survey unidimensional in following pilot administrations because of these changes.

The strategy for replacing underfitting items 5 and 6 may need to be addressed with more items that are specific to the action research process and the enhancement of teaching methods. In Table 6 are some suggestions for more specific questions. As discussed above regarding the category function, if the items are more fine-grained, then the Likert scale selection should also reflect this as well. For the next iteration of the survey, the six-point selection will be maintained.

Table 6 *Potential Question Items for the Revised Survey*

| Question Item |
| --- |
| The Moodle helped me decide my research questions. |
| The Moodle helped me devise a treatment for action research. |
| The Moodle helped me practice statistics to analyze my data. |
| The Moodle provided materials on how to teach grammar. |
| The Moodle provided materials on how to teach vocabulary. |
| The Moodle provided materials on how to teach speaking. |
| The Moodle provided materials on how to teach reading. |
| The Moodle provided materials on how to teach listening. |
| The Moodle provided materials on how to teach writing. |
| The Moodle provided materials on how to teach discussion and debate. |
| The Moodle provided materials on how to give a presentation in English. |
| The Moodle provided useful feedback for my answers to the quizzes. |

## CONCLUSION

In the process of engaging in a validity argument and utilizing three of Messick's six aspects of content validity, this survey for evaluating the participants' views about the Moodle obtained some vital data that can lead to its improvement. The Rasch model is one robust way to evaluate the items of the survey and the data from the analysis was readily used in tandem with Messick's six aspects. By adopting the changes proposed in the discussion section, this survey can be more reliable in determining how useful this website is when assisting teachers in their participation of the Advanced Program.

One of the main motivations behind this study is to promote different constructs when evaluating e-learning. Care must be taken when designing survey instruments. This means that construct identification, validity, and reliability are vital for insuring that the instrument reflects the views of the participants. The focus for most e-learning research has been on the construct of engagement, which can be easily verified by the participants' presence on the website via log files, or completion of online tasks. It is equally important to consider other constructs. For example, the participants' experience, such as their impressions of the site, the ease of use, adoption of the site, in other words whether they choose to use the site or not are relevant constructs to consider. Having both data sets, the participants' subjective impressions of the site, such as perceived usefulness for this study, and log data that reveals their actual activity on the website can offer a richer view of how participants interact with Moodle.

Finally, it is hoped that this article serves as an introduction of the Rasch model and can encourage researchers to think about construct validity when building a survey. Creating a survey is not just the formulation of question items with the faith that they actually measure what they intend to measure. Using a validity argument framed with Messick's six aspects, researchers can

have evidence that the survey measures the target construct, which can lend greater confidence in the results.

**REFERENCES**

About Moodle. (2014, December 2). Retrieved December 2, 2014, from https://docs.moole.org /28/en/About_Moodle

Bond, T. G., & Fox, C. M. (2010). *Applying the Rasch model: Fundamental measurement in the human sciences ; [free Rasch software CD enclosed]* (2. ed., reprinted). New York: Routledge.

Bonk, C. J., & Graham, C. R. (2006). *The handbook of blended learning: Global perspectives, local designs*. San Francisco: Pfeiffer.

Damnjanovic, V., Jednak, S., & Mijatovic, I. (2015). Factors affecting the effectiveness and use of Moodle: Students' perception. *Interactive Learning Environments*, *23*(4), 496–514. https://doi.org/10.1080/10494820.2013.789062

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95.

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean model of information systems success: a ten-year update. *Journal of Management Information Systems*, *19*(4), 9–30.

Dougiamas, M., & Taylor, P. (2003). Moodle: Using learning communities to create an open source course management system. In *Proceedings of the EDMEDIA 2003 Conference*. Honolulu, HI.

Green, R. (2013). *Statistical analyses for language testers*. Houndmills, Basingstoke, Hampshire ; New York, NY: Palgrave Macmillan.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*, 266–283.

Linacre, J. M. (2013). Winsteps 3.80. 1 [computer software]. *Chicago, IL: Winsteps. Com.*

Linacre, J. M. (2014). *Winsteps Program Manual 3.81.0.*

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741.

Nemoto, T., & Beglar, D. (2014). Likert-scale questionnaires. In *JALT 2013 Conference Proceedings*. JALT.

Parise, P. (2016). Increasing L2 language skills with the Moodle quiz. *Bulletin of Kanagawa Prefectural Institute of Language and Culture Studies*, *5*, 97–112.

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Sakui, K. (2004). Wearing two pairs of shoes: Language teaching in Japan. *ELT Journal*, *58*(2), 155–163.

Sánchez, R. A., & Hueros, A. D. (2010). Motivational factors that influence the acceptance of Moodle using TAM. *Computers in Human Behavior*, *26*(6), 1632–1640.

Sato, K., & Kleinsasser, R. C. (2004). Beliefs, practices, and interactions of teachers in a Japanese high school English department. *Teaching and Teacher Education*, *20*(8), 797–816. https://doi.org/doi:10.1016/j.tate.2004.09.004

Wolfe, E., & Smith Jr, E. (2006). Instrument development tools and activities for measure validation using Rasch models: Part II–validation activities. *Journal of Applied Measurement*, *8*(2), 204–234.

**APPENDIX**

Advanced Program Survey on the ˙moodle website.

Nickname (

)

Please circle the number for each statement.

1 – Strongly disagree        2 – Disagree        3 – Somewhat disagree

4 – Somewhat agree        5 – Agree        6 – Strongly agree

The Moodle website was easy to use.

    1   2   3   4   5   6

I visited the Moodle often during this program.

    1   2   3   4   5   6

I was able to access the Moodle at my school.

    1   2   3   4   5   6

I was able to access the Moodle at home.

    1   2   3   4   5   6

The Moodle helped me plan for my action research project.

    1   2   3   4   5   6

The Moodle helped me improve my teaching style.

    1   2   3   4   5   6

The quizzes on Moodle helped me review the material from the sessions at the Academia.

    1   2   3   4   5   6

I was able to practice writing in English on the Moodle.

    1   2   3   4   5   6

I was able to practice listening to English on the Moodle.

    1   2   3   4   5   6

I was able to practice reading English on the Moodle.

    1   2   3   4   5   6

I think the Moodle website is an important part of the Advanced program.

    1   2   3   4   5   6

I feel I am part of a "teacher community" through the Moodle.

    1   2   3   4   5   6

I felt that the Moodle took time away from my teaching work.

    1   2   3   4   5   6

裏面もお願いいたします。

Advanced Program Survey on the website.

Overall, I think the Moodle site was helpful.

    1    2    3    4    5    6

I think the Moodle can be improved if:

Please freely write your opinion here:

Thank you very much. ☺

ご協力ありがとうございました。