

Article

Monitoring Engagement in the Foreign Language Classroom: Learners' Perspectives

Bradford J. Lee ^{1,*}, Hayo Reinders ² and Euan Bonner ³¹ Organization for Fundamental Education, Fukui University of Technology, Fukui 910-8505, Japan² School of Liberal Arts, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand³ Centre for Learning and Teaching Innovation, Kanda University of International Studies, Chiba 261-0014, Japan

* Correspondence: bradford-lee@fukui-ut.ac.jp

Abstract: Over the last three years, we have engaged in the development of a web-based mobile application called Classmoto that uses the Experience Sampling Method (ESM) to measure cognitive, behavioral, and emotional engagement in near real-time, with minimal disruption to the teaching and learning experience. The current study implemented Classmoto in an intact university language class in a Japanese university for an entire semester. We focused on learners' experiences of using the app for two purposes: (1) to determine its face validity, and (2) to identify any constraints on and benefits from its practical application in an authentic pedagogic context, from the learners' perspective. The results show that the instrument was able to measure the three sub-domains of engagement, as designed. In addition, participants praised the ESM for (a) giving them an opportunity for self-reflection, and (b) enabling the instructor to react to the students' feedback instantaneously, with no negative feedback reported.

Keywords: L2 engagement; experience sampling; language learning



Citation: Lee, Bradford J., Hayo Reinders, and Euan Bonner. 2024. Monitoring Engagement in the Foreign Language Classroom: Learners' Perspectives. *Languages* 9: 53. <https://doi.org/10.3390/languages9020053>

Academic Editor: Abdessatar Mahfoudhi

Received: 8 November 2023

Revised: 23 January 2024

Accepted: 23 January 2024

Published: 31 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Engagement is being increasingly recognized for its major importance in education. Recent studies have confirmed that higher levels of engagement are positively correlated with academic achievement in general (see meta-analyses by [Lei et al. 2018](#); [Tao et al. 2022](#)) and L2 acquisition specifically ([Hiver et al. 2021](#)). They have also shown that engagement levels are highly dynamic, situated, and are subject to intervention ([Symonds et al. 2021b](#)). Specifically in the area of language learning, a recent study by [Teravainen-Goff \(2022\)](#) found that even motivated learners could experience disengagement as a result of 'classroom activities and tasks, the challenging nature of language learning, low self-efficacy and confidence, competing priorities, influence of peers and teacher, and teaching style' (n.p.) This means that if teachers are able to monitor (changes in) engagement levels, they may be able to respond and alter aspects of the instructional process. For this to be feasible, teachers will need to have access to an instrument that can measure this, without undue disruption to the instructional process. In this article, we report the results of a recent study on the implementation of Classmoto in an intact language class in a Japanese classroom.

2. Literature Review

Measuring L2 Engagement

[Zhou et al. \(2021\)](#) offer a useful overview of the ways in which L2 engagement can be measured. Broadly speaking, researchers choose between:

- observations from the learner's or the observer's perspective;
- measures of engagement in-the-moment or after the event;
- measuring engagement narrowly (at a particular time and in a particular place) or broadly (for example engagement levels across a semester);

- measuring engagement directly (e.g., using physiological measures such as heart rate trackers) or indirectly (e.g., using a learner diary).

Together, these options provide a range of possible ways of approaching data collection. Common research methods in our field include observations, self-report through stimulated recall, the use of diaries, and questionnaires. However, current research suffers from a number of shortcomings, including that many measures are either impractical (e.g., collecting neurophysiological data, such as heart rate), retrospective (leading to recency bias; [Beal and Weiss 2003](#)), difficult to interpret, or intrusive (for example, asking learners to stop what they are doing to write a reflection) (see also [Reinders et al. 2023](#) for more discussion). One primary concern is that most studies do not account for the dynamic nature of engagement. As [Oga-Baldwin \(2019\)](#) points out: '[...] engagement, and its instrumentation, needs to refer to specific classrooms and events. [...] This means that measuring engagement as a series of trends (i.e., 'I enjoy learning new things in class', 'I try very hard in class', etc.) may not be the ideal measurement format (p. 7). Another widely-held concern is that 'empirical research to date has examined this construct in the context of single tasks and under laboratory conditions' ([Sulis 2022](#), p. 1) and there has been a growing call for research investigations beyond isolated tasks and in authentic language classrooms ([Hiver et al. 2021](#)). The Experience Sampling Method (ESM), also known as Ecological Momentary Assessment (EMA), offers one possible avenue to achieving this, as it involves collecting real-time, ecologically valid data by repeatedly sampling individuals' experiences, behaviors, and contextual factors in their natural environment ([Larson and Csikszentmihalyi 2014](#)). However, as shown by a recent scoping review ([Symonds et al. 2021a](#)), such studies which look at engagement in-the-moment are still very limited.

For this reason, we developed an app for use by learners and their teachers in class, to gather near real-time learner engagement data (called Classmoto: [Bonner et al. 2022](#)). In a recent study ([Reinders et al. 2023](#)), we presented validity and reliability measures of our ESM application, showing that it successfully measures cognitive, behavioral, and emotional engagement. However, we did not elicit learner feedback at that stage. In the current study, we implemented Classmoto with an intact class for a period of 12 weeks to obtain qualitative data on the learners' experiences.

Qualitative learner data is useful in L2 engagement research for three broad purposes: for self-reporting of engagement levels, to validate other measures, and to identify the impact of engagement monitoring on learners. Self-reporting is widely used in engagement research, and we have reviewed its benefits and drawbacks before ([Reinders et al. 2023](#)). As for validation, participation data (which can be used as a proxy for engagement) could be shared with learners who are asked to comment on its relationship with their perceived engagement at different times. This has been shown to be particularly important, not only for (generally beneficial) data triangulation, but also because previous studies have shown that observations and other learner-external measures are not as robust as they may seem. For example, [Mercer et al. \(2021\)](#) found that students were able to 'trick' observers into believing they were engaged, when in fact they were not. The authors note,

This [...] has important implications for practice and how teachers interpret learner behaviors, which may outwardly resemble engagement but may in fact be complete disengagement or acts of compliance as students enact the diligent learner role. Such behavior also threatens the validity of research approaches, which may rely strongly on observational data as a measure of engagement or at least behavioral engagement. (p. 143).

As for the impact of engagement monitoring on students, no research that we are aware of has occurred in an educational setting. We believe this is an important gap in the literature because if engagement monitoring is assumed to be beneficial for pedagogic purposes, then it is important that learners also perceive it to be so in practice. In the next part of this paper, we therefore describe how we obtained learners' perceptions of the implementation of our engagement monitoring application.

3. Materials and Methods

3.1. Research Questions

1. What is the face validity of Classmoto as a measure of classroom L2 engagement?
2. What are its constraints and benefits, as perceived by learners?

3.2. Regarding Classmoto

Classmoto was developed by the research team as a web-based application that provides teachers and researchers with a method of collecting students' levels of self-reported engagement. It does this by asking learners to respond to Likert-scale engagement-measuring statements in both English and in the students' native language throughout their lessons. Teachers simply select appropriate times via the app to prompt students' smart devices or laptop computers to ask for engagement feedback via three Likert-scale sliders (see Figure 1). As a web-app, it is designed to work on any internet-connected device without requiring installation, which avoids issues of teacher and student device compatibility, and is typically run on a teacher's laptop and students' smartphones. The app aims to be as non-intrusive as possible, with new opportunities to provide feedback appearing on device screens without student intervention or the need to refresh the page. The app has also been designed to be scalable, enabling the measurement of engagement in-the-moment with any number of students, lessons, and courses.

The figure shows a web-based interface titled "Provide Feedback" in red text. It contains three Likert-scale sliders, each with a blue bar and a white slider knob. The first slider is for the statement "I am exerting a great deal of mental effort right now. (今、私は精神的に大きな力を発揮しています。)" and the knob is positioned towards the "Agree" end. The second slider is for "I am participating actively in class activities right now. (今、クラスの活動に積極的に参加しています。)" and the knob is positioned towards the "Disagree" end. The third slider is for "I am feeling positive about this class right now. (今、このクラスに対して前向きな気持ちになっています。)" and the knob is positioned towards the "Disagree" end. Each slider has "Disagree" on the left and "Agree" on the right. At the bottom right, there is a grey "Submit" button.

Figure 1. The student interface for Classmoto, showing the three Likert-scale sliders used to respond to the engagement-measuring statements.

The teacher's command center page of the application displays an average of each feedback statement for the activated feedback session (known in this project as a module), as well as individualized color-coded indicators reflecting each student's engagement levels. Each student's response is converted into a color corresponding to the level of agreement with the engagement statements, with higher indicated levels of engagement corresponding with green, mid-ranges in yellow, and lower scores with red (see Figure 2).

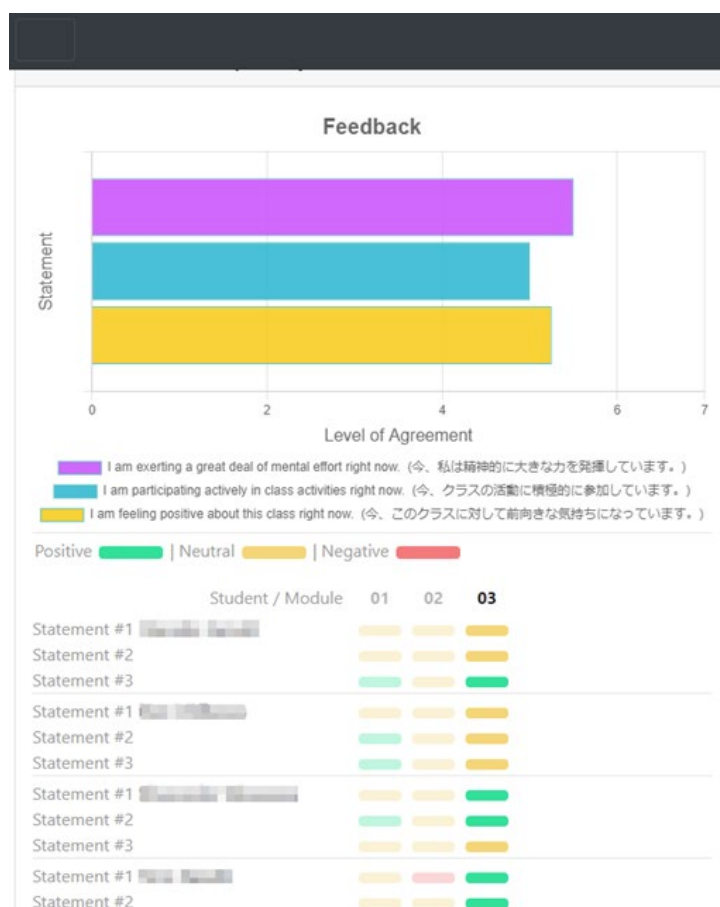


Figure 2. The teacher's command center where they can see the overall engagement levels of the class for each statement at the top and individual responses below.

Teachers are also able to review past reported engagement levels for current or previous lessons for both the entire class (see Figure 3) and individual students (see Figure 4).

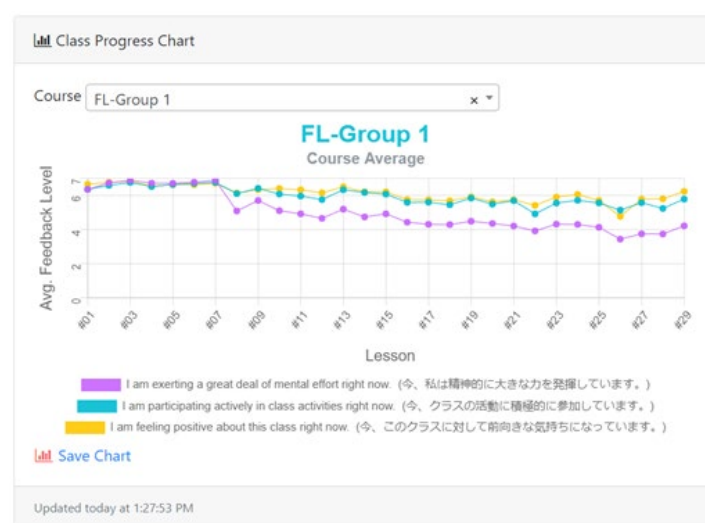


Figure 3. A class progress chart, showing an example class's overall engagement levels throughout the semester.



Figure 4. An individual student's progress chart example, showing their overall engagement levels throughout the semester.

Classmoto differs from other applications that were initially investigated for potential use in the project, such as Project MyCap (projectmycap.org/, 8 November 2023), LifeData (lifedatcorp.com/, 8 November 2023), Ethica (ethicadata.com/, 8 November 2023), and Expiwell (expiwell.com/, 8 November 2023), as none of them met all the desired needs of the project such as real-time data collection, easy creation of student accounts, teacher control over when to request engagement reports, a user-friendly interface, and easy downloading of results, while also allowing for potential scaling of the number of participants without incurring significant license costs over an extended period of time. Additionally, developing the application internally provided the researchers with complete control over the application features, enabling significant changes to be made without the need to change applications mid-project.

Precursor research to this point has involved piloting rounds whereby the instrument has been refined by validating the question prompts (Reinders et al. 2023), gathering teacher and student feedback on the use of the app and timing of data collection (Bonner et al. 2022), and general polishing of the code and interface, along with the methods of data analysis. Now that those preliminary stages have been completed with large samples (i.e., several thousand data points), the current study sought to gather more in-depth, personal reflections on the use and utility of Classmoto from a smaller, more manageable sample. A single class was therefore asked to use the program to measure their engagement for a full semester, after which the researcher sat down with each student individually for a face-to-face interview to talk about their experiences.

The researcher, an L1 English speaker, also holds the highest level, N1, in the JLPT (Japanese Language Proficiency Test). Students were therefore given the choice to speak in either English or Japanese, and could code-switch as necessary in order to fully facilitate the communication of their nuanced ideas and feelings. These sessions were recorded on video and later coded by the lead researcher. Five questions served as the basis for the interviews, which sought information relating to participants' understanding of, and impressions regarding, both the Classmoto prompts and the investigation of student engagement in general (see Appendix A). In addition to the prepared questions, the interviewer would pose off the cuff follow-up questions, depending on what issues the participants raised (e.g., asking for further information, examples, clarification, etc.). The researcher then displayed each students' individual response history and asked them to discuss, to the best of their recollection, various features of their dataset, such as any noticeable peaks or dips or consistently high (or low) responses to a particular item.

3.3. Participants

The current study was conducted at a small, private university located in western Japan. It is of note that this university, with predominantly engineering disciplines, does

not have a major in English language. All students are, however, required to take two English classes per semester to satisfy their foreign language requirements for graduation—one which focuses on oral communicative skills (i.e., listening and speaking), and one which drills explicit knowledge (i.e., grammar and vocabulary). The area in which the university is located has a very low number of L1 English speakers, making the language learning conditions very much one of EFL (English as a foreign language), as the participants' exposure to and use of English is generally restricted to these two weekly English courses.

Upon matriculation, and at the start of each academic year thereafter, each student takes a TOEIC® (Test of English for International Communication) Bridge Test for the purposes of class placement. Of note, this test only covers receptive abilities (i.e., listening and reading) while proficiency in the productive skills of speaking and writing are noticeably absent. This is unfortunate, as communicative competence is a primary focus of the curriculum, as explained previously, though necessary due to the large numbers of students being assessed simultaneously.

The students are divided into four class levels, approximately equivalent to the CEFR (Common European Framework of Reference for Languages) levels of: A1 and below, A2, A2 to B1, and B2 and above. The class selected for the current study was a B2-and-above listening skills class of first year students. With IRB (institutional review board) approval, students from this class were informed of the purpose of the research project and were invited to participate. Signed informed consent forms were obtained from all 15 students in the class indicating their acknowledgement that participation was voluntary, and could be withdrawn at any time without penalty. Students who participated in both quantitative data collection and individual interviews with the researcher were given JPY 500 (approx. USD 3.69) Amazon gift cards as honoraria.

As far as the makeup of the participants, the group consisted of 10 Japanese citizens and five foreign nationals—all male. The Japanese students followed the traditional matriculation process, meaning that they were between 18 and 19 years old. The foreign students were considered to be non-traditional, as they typically had between 2 and 3 years of intensive Japanese-language education after high school in order to prepare them to enter into a Japanese university, resulting in a mean age of $M = 20.8$. At the end of the semester, when final grades were tabulated, 13 of the participants ended the semester with the highest letter grade (i.e., between 90 and 100%), while two of the participants received the second-highest grade (i.e., between 80 and 89%).

3.4. Procedures

The first week of the experiment involved obtaining signed informed consent forms from the participants and creating their Classmote accounts. Students were given unique PIN codes with which to login to their accounts and given training on how to login and use the interface. During our precursor research, particularly the pilot study in which we validated the instrument (Reinders et al. 2023), it was observed that students tended to respond at-or-near ceiling for the first few instances of data collection before 'settling in' to the routine. We speculate that this is either due to students just entering the highest marks to test out the interface, or because they as yet, have no frame of reference with which to compare their current state of engagement. Regardless of the reason, based on these results, the instructor engaged in a training period of two weeks where the data was collected twice each week, which was not included in the final data analysis. During this training period, the three prompts, though displayed in both English and Japanese, were first explicitly explained by the instructor, including several examples of how they differed in terms of cognitive, behavioral, and emotional aspects.

Each weekly lesson followed a predictable pattern, following the textbook for the course. The class started with a warm-up period which familiarized the students with the topic of the week and relevant lexical items. These vocabulary items were introduced in a rhythmical/kinesthetic method designed to increase students' perceptive sensitivity to English suprasegmental features (i.e., prosody and stress). Students first listened to a

model of the words being spoken, after which they clapped their hands in a corresponding rhythm. Once the rhythm has been mastered, the students then listen and repeat the words themselves, ideally with accurate reproduction of target-like prosody (see [Lee 2020](#) for a further description of this methodology). It should be noted that as this methodology requires coordinated movement, it is relatively easy for the instructor to visually observe whether students are actively engaged or not, as a student who does not clap their hands, or one who is significantly out of sync with the rest of the class is quite obvious. Following this perception/production drill, the students spent a few minutes doing the drill in their textbooks which asked them to look up the meaning of the words and connect them to their Japanese translations. Once the answers were checked as a group, the students were instructed to access Classmotto and enter their responses (Entry #1).

The next portion of the lesson was an introduction to the grammar point of the week. The grammar point was introduced in both Japanese (via the textbook) and English (via the instructor), followed by textbook drills involving manipulation of the grammar pattern into various forms (e.g., declarative, interrogative, negative, etc.) and Japanese-to-English and English-to-Japanese translations. Students were then asked to enter their current states of engagement into Classmotto at the end of these drills (Entry #2). Grammar topics covered over the semester included: progressive tense, passive tense, perfect tense, helping verbs, infinitives, and relative adverbs.

The next portion of the class was focused on listening comprehension. Each unit has several model conversations which utilize the vocabulary and grammar patterns previously introduced, in support of the theme of the week. First, students listened to the conversation and answered true-or-false questions. This was followed by a dictation drill where students listened to a conversation and transcribed what they heard via a cloze activity. After these drills, students were instructed to again respond to the Classmotto prompts (Entry #3).

Finally, the class engaged in several rounds of timed free conversation. The instructor presented a topic for conversation, based on the theme of the week (e.g., summertime activities, club activities, experience overseas). Students would pair up and engage in conversation with their partner on the topic for five minutes. When the bell sounded, marking the end of the five-minute period, students had one minute to bring their conversation to a close and find a new partner. A bell would then sound to mark the start of a new five-minute round. Typically, there would be five rounds, resulting in an approximately 30 min activity. At the end of the fifth round, students completed the final data collection for the week (Entry #4).

At each instance of data entry, the instructor would monitor the students' responses via the command center. There was no time limit given for data entry, and class would not begin again until all students had entered their response scores. Once all of the feedback was in, the instructor would comment on the results to let the students know that their responses were being noted (e.g., 'Oh, some students are yellow in emotive engagement—I guess some of you didn't care for that activity so much? OK, let's change things up for next time!')

After the first two weeks of paperwork, introductions to the app, the concept of the three subtypes of engagement (i.e., cognitive, behavioral, emotive), and training on how to login and enter data accurately, official data collection began in week three. Weeks three through six had regular class sessions with data collection (four times each class). Week seven was the class's midterm test and so no data was collected. Data collection resumed in week eight and concluded in week nine.

Private interviews (approx. 15 min each) were conducted by the lead researcher with each participant on weeks ten and eleven (see Section 3.2—Instruments).

4. Results

The responses that were collected during the interviews were coded to allow for quantitative reporting. Qualitative analyses of the responses, which provide more details as to

the nuanced feelings of the participants, are provided after the tabular reporting. Table 1, below, shows the descriptive statistics for the first question of the interview, asking the participants if they felt that they understood the Classmoto app prompts. In other words, we were interested in first measuring Classmoto's construct validity by ascertaining whether the app was measuring what it was intended to.

Table 1. Codified responses to Interview Q1.

	Yes	Maybe	No
(Q1) Did you understand the three prompts in the Classmoto app clearly?	13 (86.7%)	2 (13.3%)	0

The results from Table 1 show that 86.7% of the participants reported feeling confident that they knew what sub-constructs of engagement were being probed by the app, while two participants (13.3%) reported feeling less than 100% confident about what was being asked.

In order to ascertain whether the participants' self-reported understanding of the prompts was accurate or not, Question Two followed up by asking students to explain, in their own words, what question they were responding to. Students were shown the Classmoto interface via the instructor's laptop, similar to what they had experienced during the six weeks of data collection, to ensure accurate recollection of the prompts. The students' responses were coded as either correct, ambiguous, or incorrect, as reported in Table 2 below.

Table 2. Codified responses to Interview Q2.

(Q2) In Your Own Words, Please Describe What You Think the Prompts Were Asking You.			
	Correct	Ambiguous	Incorrect
Prompt One (cognitive)	11 (73.3%)	2 (13.3%)	2 (13.3%)
Prompt Two (behavioral)	12 (80%)	3 (20%)	0
Prompt Three (emotive)	14 (93.3%)	0	1 (6.7%)

4.1. Prompt One

As shown in Figure 1, Prompt One (cognitive engagement) stated: I am exerting a great deal of mental effort now or 今、私は精神的に大きな力をはっきりしています in Japanese. However, when asked to summarize this statement in the participants' own words, two responses were ambiguous and two were incorrect. At issue is the distinction between effort and understanding. For example, accurate summaries of the prompt included the following:

- Am I trying to understand the class or not?
- This is asking my level of concentration.
- Am I focused on the lesson?

The common theme in these correct paraphrases is that they mention voluntary, directed effort on the part of the students (i.e., engagement). Additionally, in each case, an increase in the variable (i.e., trying to understand, level of concentration, or focusing on the lesson) would have correctly resulted in a higher response in the app. However, two responses were coded as being ambiguous as they referred to the lesson content:

- Is the lesson easy or difficult?

In this case, the responses focus on the qualities of the lesson, and not necessarily on their own engagement. While at first difficulty and cognitive effort may seem closely

related, note that it is entirely possible for a lesson to be easy and yet still be cognitively engaging. Likewise, a difficult lesson does not always guarantee increased engagement. In fact, the relationship between difficulty and effort in L2 listening has been described as an inverted U-shape (Zekveld and Kramer 2014) where effort increases with difficulty up to a point, after which it drops off as students reach their breaking point. Therefore, we strongly caution against linking cognitive engagement with content difficulty as we define it as sustained actions or effort on the part of the learner (e.g., Oga-Baldwin and Nakata 2017; Skinner et al. 2009a, 2009b).

However, despite the fact that this definition was slightly inaccurate, these two responses were coded as ambiguous (rather than incorrect) as it is highly likely that it did not make a difference in the data they entered in to the app for two reasons. First, as the students were at a level of CEFR B2 (see Section 3.3), the grammatical content of the course was not particularly challenging for them (see Section 3.4). Second, all students received a final score of 80% or above, with most of them scoring 90% or above (Section 3.3), affirming they understood the course content at an acceptable level. Therefore, it is highly unlikely that any units were so difficult as to cause any students to disengage, meaning that, in this case, their cognitive effort generally tracked with the content difficulty.

Finally, two responses were coded as being incorrect. These definitions referenced the students' level of comprehension of the course content:

- How much did I understand the lesson?

This sort of response is problematic for the theoretical reasons discussed previously with relation to content difficulty, but also for the practical reason that, based on this definition, the participants risk actually recording inverted scores to those that Prompt One intended. Under this assumption, a student who understood the lesson perfectly would enter the maximum score in Classmote. While this may coincide with cognitive engagement (i.e., the student thought hard about the lesson and was therefore able to understand everything), another possibility is that the content was far too easy and required no effort to comprehend—in which case the participant should enter a low score for this prompt. Therefore, given the risk to the validity of the data collected, this summary of Prompt One should be rejected as being incorrect.

4.2. Prompt Two

As shown in Table 2, $n = 12$ (80%) participants provided correct paraphrases of Prompt Two (behavioral engagement): I am participating actively in class activities right now or 今、クラスの活動に積極的に参加しています in Japanese.

Correct paraphrases needed to include language related to physical activity in order to be considered correct, such as the following:

- Am I participating in pair work or group work?
- Am I active in class or not?

These summaries accurately reflect the notion that behavioral engagement is the physical manifestation of a learner's internal motivation. In other words, this prompt is not questioning the students' desire to learn, but rather their participation/concentration level on classroom activities. A student who is highly motivated to learn may still be distracted by their smartphone, resulting in lower behavioral engagement for a particular activity. Likewise, a student who does not particularly care for the subject matter and shows low motivation might still actively participate in an activity such as a song or a game.

However, three responses to this prompt were somewhat ambiguous, such as:

- Do I try to improve my English through the class?

While responses such as the above reference trying (i.e., effort), and therefore accurately reflect some manner of engagement, they do not specifically reference physical activity. If it can be assumed that the student meant that trying to improve means dutifully participating in classroom activities, then this response would be correct, though it is pos-

sible to imagine other ways of trying to improve which are not evidenced by physical behavior, such as cognitive engagement.

4.3. Prompt Three

Finally, Table 2 reported that $n = 14$ (93.3%) participants correctly paraphrased Prompt Three (emotive engagement): I am feeling positive about this class right now or 今、このクラスに対して前向きな気持ちになっております in Japanese. Only one participant reported:

- I am not exactly sure about this one

The other 14 participants summarized the prompt using language related to their feelings about the suitability of the class and/or manner of instruction, such as:

- I am comfortable that the teaching method is good or not
- If I feel good or no confidence in this class
- Do I think this class can be useful for my future?

4.4. Response Times

Next, the interviewer asked students to estimate how long they spent answering the Classmoto prompts each time they were asked to input data (in minutes or seconds). The responses are shown below, in Figure 5. All participants reported spending less than 30 s per instance of data collection (i.e., less than 10 s per each of the three prompts).

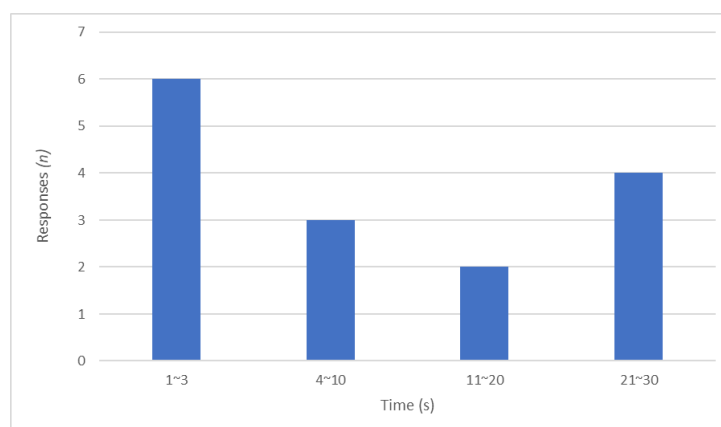


Figure 5. Responses to Interview Item 3—time spent inputting data into the Classmoto app.

4.5. Reflections on Using the App

The next two interview items asked students to think about and report any positive or negative feedback they had regarding the app, and/or its implementation in the class. As shown in Table 3, all of the participants reported positive feedback. Even when specifically asked to comment on any drawback or negative impressions, no participants chose to do so.

Table 3. Codified responses to Interview Q4 and Q5.

	Yes	No
(Q4) Did you find anything positive about using Classmoto?	15	0
(Q5) Did you feel there were any negative aspects of Classmoto?	0	15

Some excerpts are provided below for illustration:

I think it's great to get in the habit of self-reflection, an opportunity to ask myself how much I am truly understanding in-the-moment. Making time for such confirmation each lesson is, I think, very important. Not only is it ok to pause during a lesson, I think we NEED to pause and make that time.

I was able to respond at the end of different parts of the lecture, for example, vocabulary, or grammar, or listening, so I was able to analyze which parts were easy and which parts I wasn't so good at. So it was a good thing for me to think about my learning. The frequent stopping to reflect didn't interrupt my learning, I think.

Well, it was much more refined than the usual, end-of-semester type of questionnaires that just ask for a general reaction. With this, we could give our response to each individual activity, so it's much more useful, I think. I mean, everything just gets garbled into one if we only do it one time.

Responding to the survey didn't interfere with my concentration, etc. I mean, we were changing topics in the class anyway, like from vocabulary study to grammar, so my mind had to reset anyway.

4.6. Individual Discussion of Results

Finally, the interviewer displayed each students' data charts, which showed their self-reported response scores throughout the course of the semester (e.g., Figure 4). While giving the student the opportunity to see their graphs, which showed trendlines and fluctuations over time, the instructor pointed out various points of interest and sought the students' commentary for insights into what was happening in the students' minds at the time. It should be noted that this opportunity was taken to discuss the discrepancies regarding the meanings of Prompts One and Three (Table 2) with the participants. For instance, one interaction with Participant 4 went as follows:

Interviewer: *So here, there's a slight dip in your responses to Q1. Was there a reason for that?*

Student: *I think that lesson, I had trouble understanding the nuance between English my native language, because that nuance does not exist between [student's native language] and Japanese.*

Interviewer: *In that case, wouldn't your ratings go UP, as you had to concentrate harder to understand the lesson? But you input a lower score?*

Student: *I gave a lower score because I think I understood less. Oh, did I make a mistake?*

Interviewer: *Yes, but don't worry about it. But didn't you understand when I explained about the questions and how to answer at the beginning of the semester?*

Student: *Yes, I did. I think I answered correct [SIC] at first, but then one time I forgot and answered wrong thing [SIC] and then everything after that was wrong thing [SIC].*

Other interactions illustrate the way Classmoto allowed the instructor to analyze the needs of the students and make real time adjustments to the course content and pedagogy:

Interviewer: *Looking at your responses to Q2 and Q3, at first your engagement seemed a bit low, but then it rose up and stayed high?*

Student: *In the beginning of the semester, I was very worried about if I'd be able to keep up in class.*

Interviewer: *Ah, but as you got used to the class ...*

Student: *Yeah, once I got used to you and the class.*

Interviewer: *Your ratings for Q3 are high, but not at the max. So I take it to mean that you feel the class had value for you, but some things could be better?*

Student: *Yeah, we do a lot of listening and answering questions from the book, right? But I don't want to improve my writing, I want to improve my speaking, so that's where*

I think the class could be better for me. Then you made free-talk conversation warm-up activities for us and I was very happy and felt more positive after that.

5. Discussion

Our goal in the creation of Classmoto was to design a tool that would allow teachers and researchers to follow the dynamic changes in student engagement over a variety of sub-domains across any type of classroom activity. In order to achieve this, we aimed to design the instrument to be as minimally disruptive as possible, so as not to take students out of their current state of mind and/or interrupt the class to the point where learning outcomes could be negatively impacted. One way in which we sought to accomplish this was through our use of a single item per engagement sub-domain (i.e., cognitive, behavioral, and emotional). Note that theoretically, this methodology has support from other cognitive researchers who have used single or reduced item scales in their studies of motivation and/or engagement (e.g., Gogol et al. 2014; Martin 2009), not to mention our own validation of the instrument which included content validity indexing, principal component analysis, and analysis of content validity (Reinders et al. 2023). However, what had heretofore been missing was an assessment of the impact that using the app (or the ESM methodology itself) had on learners in an authentic environment. The current study sought to address this issue by collecting participants' feedback on both the technical aspects of using the app (e.g., comprehensibility, time required to operate the app) and emotive feedback (i.e., the impact that frequent data input had on their levels of concentration, or personalized discussions of their thought processes).

5.1. Research Question 1

In response to RQ1, which focused on Classmoto's face validity, the participants' responses to Q1 and Q2 of the interview offer strong evidence that the app was largely successful in measuring the three sub-domains of cognitive, behavioral, and emotive engagement (see Tables 1 and 2). What was elucidated, however, was that a small number of students reported slight misconceptions as to the nature of the prompts. This could have been the result of a number of factors, some of which may have been limitations in the current study design and not reflective of the app itself. Q2 of the interview asked students to describe what they thought the prompts were asking them, *in their own words* (see Appendix A). As a result of being asked to summarize the prompts, some participants might not have wanted to use the same language that was contained in the original questions, therefore using language such as *difficulty* to paraphrase Prompt One, which was asking about *mental effort*, for example.

However, it was clear that at least one participant began, at some point, to answer Prompt One incorrectly (see Section 4.6), despite the 2-week training session provided at the beginning of the semester. This result was highly insightful, as it suggests that administrators should either (a) more explicitly detail the difference between *mental effort* and *course difficulty* (Prompt One) and *activity* and *trying* (Prompt Two) during the training session, or (b) offer refresher training sessions periodically throughout the data collection period to ensure students' focus does not start to waver. For instance, in the current study, there was a gap in data collection between weeks six and eight (see Section 3.4). Perhaps it would have been prudent to refresh the students' memories at this point, as the routine had been interrupted due to the course's midterm exam.

5.2. Research Question 2

In answer to RQ2, which asked what benefits or constraints participants reported with regard to using Classmoto, the results were overwhelmingly positive. While some participants ($n = 4$, 26.7%) reported taking a maximum of 21–30 s to input their responses (Figure 5), the participants universally rejected the notion that data collection negatively impacted their cognitive flow or concentration in class. Quite the contrary, the current

population reported feeling very positive about the fine-tuned nature of the ESM, which asked students to report at different times throughout the lesson.

Two main themes that emerged in the responses were, (a) students valued having time set aside during their studies to pause and self-reflect because they discovered things about themselves (e.g., their strengths and weaknesses), and (b) by sharing their level of engagement with the instructor in real-time, the instructor could then better tailor the lesson to their needs in response. Essentially, the students saw that their feedback was being given immediate consideration by the instructor and that through it, they could (somewhat) influence the course pedagogy and/or curriculum. This feeling of learner agency is, of course, linked to positive learning outcomes (e.g., [Larsen-Freeman et al. 2021](#)).

6. Limitations and Future Directions

One possible limitation, which was alluded to earlier in the Discussion section, is that of asking the participants to paraphrase the three Classmoto prompts (Interview Q2—see Appendix A). In addition to the issue discussed previously (i.e., inappropriately replacing *effort* with *difficulty*), lexical knowledge might have also limited the participants' ability to address this question fully. Despite the participants being given the choice whether to speak in English or Japanese (or both), five of the participants, whose L1 was neither, may have had difficulty capturing the nuance of either the prompt, or their responses. The prompts themselves were only presented in the app in English and Japanese. Future research, or versions of the app, might look to include more languages to more accurately obtain ratings from non-L1-Japanese participants.

With regard to the finding that the participants of the current study appreciated taking frequent pauses during the lessons for self-reflection, it is important to point out that the students were all highly proficient L2-English speakers (CEFR B2 and above) who ended up receiving high grades in the class (Section 3.3). While the participants' engagement level results are not presented in this paper, as our goal here was to investigate user perspectives regarding the *use* of Classmoto/ESM and not their engagement, it can be said impressionistically that their engagement was generally high throughout the semester, some evidence of which was presented in Section 4.6. It would therefore be beneficial to survey groups with varying levels of proficiency to examine if there is a quantifiable link between topic proficiency, engagement, and the value placed on self-reflection.

However, the second major finding (i.e., that students could feel a sense of agency by having their feedback impact course pedagogy and/or curriculum) can most likely be generalized to any student population. While Classmoto was originally designed with the L2 classroom in mind, we see no limits in its use in other contexts, both course subject or learner context (e.g., age, level, proficiency, etc.). It is our hope that future studies from context other than tertiary-level L2-language populations can help add to the validity of our tool for general use.

7. Conclusions

The current study investigated the face validity of, and user feedback to, our newly developed application for measuring learner engagement, Classmoto. It was found that the app was successful in eliciting distinct measures of cognitive, behavioral, and emotive engagement from the students. This result not only adds further support to theoretical models of learner engagement which reason that these subdomains of engagement are related but independent ([Reinders et al. 2023](#)), but also that they are able to be measured by our tool via a single item scale (e.g., [Gogol et al. 2014](#); [Martin 2009](#)). The participants' generally positive reaction to the use of the app in authentic classroom settings is also a novel result which provides evidence that the ESM methodology, when combined with single item scales, is an effective research methodology to measure engagement with limited disruption of classroom activities. Lastly, the data gathered by the tool were not only shown to be valuable to the teacher ([Reinders et al. 2023](#)), but also to the participants' themselves, in terms of increased feelings of learner agency and self-reflection.

Author Contributions: Conceptualization, B.J.L.; methodology, B.J.L. and H.R.; software, E.B.; validation, B.J.L., H.R. and E.B.; formal analysis, B.J.L.; investigation, B.J.L.; resources, B.J.L. and E.B.; data curation, E.B.; writing—original draft preparation, B.J.L., H.R. and E.B.; writing—review and editing, B.J.L., H.R. and E.B.; visualization, E.B.; supervision, H.R.; project administration, B.J.L.; funding acquisition, E.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI Grant Number JP20K00839 as part of a Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) Grants-in-Aid for Scientific Research.

Institutional Review Board Statement: The study was conducted in accordance with guidelines of Fukui University of Technology, and approved by the Ethics Committee of Fukui University of Technology (人-2022-01, 11 April 2022).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data for the current study is unavailable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Interview Questions

1. Did you understand the three prompts in the Classmoto app clearly?
2. In your own words, please describe what you think the prompts were asking you.
3. How long did it take you to respond each time?
4. Did you find anything positive about using Classmoto?
5. Did you find any negative aspects of using Classmoto?

References

- Beal, Daniel, and Howard M. Weiss. 2003. Methods of Ecological Momentary Assessment in Organizational Research. *Organizational Research Methods* 6: 440–64. [\[CrossRef\]](#)
- Bonner, Euan, Kevin Garvey, Matthew Miner, Sam Godin, and Hayo Reinders. 2022. Measuring real-time learner engagement in the Japanese EFL classroom. *Innovation in Language Learning and Teaching* 17: 254–64. [\[CrossRef\]](#)
- Gogol, Katarzyna, Martin Brunner, Thomas Goetz, Romain Martin, Sonja Ugen, Ulrich Keller, Antoine Fischbach, and Franzis Preckel. 2014. "My Questionnaire is Too Long!" The assessments of motivational-affective constructs with three-item and single-item measures. *Contemporary Educational Psychology* 39: 188–205. Available online: https://kops.uni-konstanz.de/bitstream/handle/123456789/29722/Gogol_0-262761.pdf (accessed on 15 April 2023). [\[CrossRef\]](#)
- Hiver, Phil, Ali H. Al-Hoorie, Joseph P. Vitta, and Janice Wu. 2021. Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research* 28: 201–30. [\[CrossRef\]](#)
- Larsen-Freeman, Diane, Paul Driver, Xuesong Gao, and Sarah Mercer. 2021. Learner Agency: Maximizing Learner Potential [PDF]. Available online: <http://www.oup.com/elt/expert> (accessed on 8 May 2023).
- Larson, Reed, and Mihaly Csikszentmihalyi. 2014. The experience sampling method. In *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Berlin and Heidelberg: Springer, pp. 21–34. [\[CrossRef\]](#)
- Lee, Bradford J. 2020. Enhancing listening comprehension through kinesthetic rhythm training. *RELJ Journal* 53: 567–81. [\[CrossRef\]](#)
- Lei, Hao, Yunhuo Cui, and Wenye Zhou. 2018. Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal* 46: 517–28. [\[CrossRef\]](#)
- Martin, Andrew J. 2009. Motivation and engagement across the academic life span: A developmental construct validity study of elementary school, high school, and university/college students. *Educational and Psychological Measurement* 69: 794–824. [\[CrossRef\]](#)
- Mercer, Sarah, Kyle R. Talbot, and Isobel K. H. Wang. 2021. Fake or real engagement—Looks can be deceiving. In *Student Engagement in the Language Classroom*. Edited by Paul Hiver, Ali Al-Hoorie and Sarah Mercer. Bristol and Blue Ridge Summit: Multilingual Matters, pp. 143–62. [\[CrossRef\]](#)
- Oga-Baldwin, W. L. Quint. 2019. Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System* 86: 102–28. [\[CrossRef\]](#)
- Oga-Baldwin, W. L. Quint, and Yoshiyuki Nakata. 2017. Engagement, gender, and motivation: A predictive model for Japanese young language learners. *System* 65: 151–63. [\[CrossRef\]](#)
- Reinders, Hayo, Bradford J. Lee, and Euan Bonner. 2023. Tracking learner engagement in the L2 classroom with experience sampling. *Research Methods in Applied Linguistics* 2: 100052. [\[CrossRef\]](#)

- Skinner, Ellen A., Thomas A. Kindermann, and Carrie J. Furrer. 2009a. A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement* 69: 493–525. [[CrossRef](#)]
- Skinner, Ellen A., Thomas A. Kindermann, James P. Connell, and James G. Wellborn. 2009b. Engagement and disaffection as organizational constructs in the dynamics of motivational development. In *Educational Psychology Handbook Series. Handbook of Motivation at School*. Edited by Kathryn R. Wenzel and Allan Wigfield. London: Routledge, pp. 223–45.
- Sulis, Giulia. 2022. Engagement in the foreign language classroom: Micro and macro perspectives. *System* 110: 102902. [[CrossRef](#)]
- Symonds, Jennifer E., Avi Kaplan, Katja Upadyaya, Katarilna Salmela-Aro, Benjamin M. Torsney, Ellen Skinner, and Jacquelynne S. Eccles. 2021a. Momentary Engagement as a Complex Dynamic System. *PsyArXiv*. [[CrossRef](#)]
- Symonds, Jennifer E., James B. Schreiber, and Benjamin M. Torsney. 2021b. Silver linings and storm clouds: Divergent profiles of student momentary engagement emerge in response to the same task. *Journal of Educational Psychology* 113: 1192. [[CrossRef](#)]
- Tao, Yang, Yu Meng, Zhenya Gao, and Xiangdong Yang. 2022. Perceived teacher support, student engagement, and academic achievement: A meta-analysis. *Educational Psychology* 42: 401–20. [[CrossRef](#)]
- Teravainen-Goff, Anne. 2022. Why motivated learners might not engage in language learning: An exploratory interview study of language learners and teachers. *Language Teaching Research*, 13621688221135399. [[CrossRef](#)]
- Zekveld, Adriana A., and Sophia E. Kramer. 2014. Cognitive processing load across a wide range of listening conditions: Insights from pupillometry. *Psychophysiology* 51: 277–84. [[CrossRef](#)] [[PubMed](#)]
- Zhou, Shiyao, Phil Hiver, and Ali H. Al-Hoorie. 2021. Measuring L2 engagement: A review of issues and applications. In *Student Engagement in the Language Classroom*. Edited by Phil Hiver, Ali Al-Hoorie and Sarah Mercer. Bristol and Blue Ridge Summit: Multilingual Matters, pp. 75–98. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.