OPEN ACCESS

Vocabulary Learning and Instruction
ISSN 2981-9954
Volume 14, Number 1 (2025)
https://doi.org/10.29140/vli.v14n1.2161

Castledown

A Comparison of Flow State Markers Experienced Across AR, Game-based, and Analog Deliberate Vocabulary Study Activities

Adam Dabrowski
Rikkyo University

Ayako Yokogawa
Meiji University

Abstract

Flow is described as a state in which people become so involved or engrossed in an activity that nothing else seems to matter (Csikszentmihalyi, 2009). This state of consciousness seems to occur when a person is involved in a task and seemingly unable to stop. Flow states are marked by (a) a perceived balance of skills and challenge, (b) opportunities for intense concentration, (c) clear task goals, (d) feedback that one is succeeding at the task, (e) a sense of control, (f) a lack of self-consciousness, and (g) the perception that time passes more quickly (Egbert, 2003). The Japanese Flow State Scale (JFSS) is an instrument which was created specifically to measure flow states experienced during deliberate vocabulary study and is a working component of the first author's Doctor of Philosophy research project, which focuses on the deliberate study of vocabulary with augmented reality (AR) and physical word cards. Analyses with mixed effects models indicated that statistically significant differences in markers of states of flow elicited with the JFSS of 179 L1 Japanese participants based on four deliberate vocabulary study activities (AR, word card study, Quizlet live, and intensive reading) appear to exist.

Copyright: © 2025 Adam Dabrowski & Ayako Yokogawa. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Background

Csikszentmihalyi (1988) described *flow*, a theorized state of psychological concentrated engagement, as, “the state in which people are so involved in an activity that nothing else seems to matter; the experience itself is so enjoyable that people will do it even at great cost, for the sheer sake of doing it…” (Csikszentmihalyi, 2009, p. 4). Flow states are related to motivational task engagement and can be considered a “heightened level of motivated task engagement” (Dörnyei, 2019, p. 58). When a person enters a flow state, a positive experience in which a person’s skills meet the a challenge at hand, several psychological aspects interact to create a state of consciousness commonly marked by (a) a perceived balance of skills and challenge, (b) opportunities for intense concentration, (c) clear task goals, (d) feedback that one is succeeding at the task, (e) a sense of control, (f) a lack of self-consciousness, and (g) the perception that time passes more quickly (Egbert, 2003).

Surveys and interviews have been developed and used as instruments to measure flow. One of the most widely used instruments, The Flow State Scale, was devised for use in a variety of fields, such as sports medicine, education, work performance, and creative and performing arts to name a few, and boasts a great amount of validation evidence (Jackson & Marsh, 1996). Other flow state scales have been developed and tested, though not an exhaustive list, they include Jackson and Ecklund’s (2002) Dispositional Flow Scale (DFS); Rheinberg, Vollmeyer, and Engeser’s (2003) Flow Short Scale (FSS-2); Bakker’s (2008) Work-Related Flow Inventory (WOLF); and Jackson, Martin, and Eklund’s (2008) Short Dispositional Flow Scale (SDFS).

One notable issue with many of the scales above is that they are proprietary and not easily accessed by working researchers. Another issue relevant mainly to researchers in Japan is that to our knowledge at the time of writing there are no open access Japanese language instruments for measuring flow. To address these two issues, we drafted the Japanese Flow State Scale (JFSS), an instrument consisting of a set of 26 items targeting the nine subcomponents of flow, rated with a six-point Likert scale. Validation evidence was collected by subjecting 627 JFSS responses to Rasch analyses with the software Winsteps version 5.2.3 (Linacre, 2022). Please see our article currently in preparation which presents the creation of the scale and the validation evidence we gathered, which demonstrates good person and item fit, reasonable dimensionality as assessed with a PCA, excellent person and item reliability, and concludes that the JFSS is capable of measuring flow states in similar contexts (Dabrowski & Yokogawa, in preparation). See the Appendix for an OSF link to the instrument.

Though flow has been examined across various contexts in the field of SLA (see Aubrey, 2023; Egbert, 2003; Wang & Huang, 2022), the research of flow states as experienced across a variety of vocabulary study methods offers new lines of inquiry for vocabulary researchers. The main research question guiding this study asks how flow states are experienced and differ across four methods of deliberate vocabulary study, two digital and two analog: study within an AR smartphone application, study with a game-based flashcard application, study with physical word cards, and study with an intensive reading (involving identification, dictionary consultation, and note taking of unknown vocabulary items).

Method and Analyses

The dataset in this study comprises 627 JFSS survey responses submitted by 179 L1 Japanese-speaking first- through third-year university students (94 females and 85 males) recruited via a multi-site convenience sampling strategy from three universities in the Kanto region of Japan, which will be referred to as University X, University Y, and University Z. University X is a private music university, University Y is a private liberal arts school, and University Z is a public science and engineering university. 39, 30 and 110 participants were recruited from these sites respectively. The academic majors of the participants included musical composition, musical performance, music therapy, economics, engineering, and internet computer technology. The L2 English proficiencies of the participants ranged from CEFR A2 to B2.

Participants completed the JFSS within a Google Form (Google, n.d.) immediately after each of four deliberate English language vocabulary study activities which lasted ten minutes each. These rounds of deliberate study and immediate JFSS submissions were conducted on four different days to ease cognitive burden. Before each round of study, the participants were informed that they would be tested on some of the words they encountered after the activity (See Table 1). The deliberate vocabulary study activities included (1) the use of an AR smartphone application, (2) the use of Quizlet Live (Quizlet, n.d.), a class-wide vocabulary battle game, (3) physical two sided word cards, and (4) an intensive reading activity which required participants to identify unknown words, seek their meanings, and write their meanings as a note. During treatment one, *ARTourOcean* was used as an AR application (Gakken Education Publishing Co., n.d.). With this app the participants viewed the classroom in pairs through their smartphone camera and encountered AR elements which were virtual fish accompanied with English and Japanese depictions of their respective names. When participants encountered an AR fish, they wrote the L1 and L2 names of the fish on a piece of paper. Participants encountered a total of ten fish in this way. During treatment two, Quizlet Live, a game-based study component of the popular flashcard platform Quizlet, was used. Students scanned a QR code with their smartphones to join a classwide game in which they raced each other to answer written receptive form- and meaning-recognition questions. In the case of this study, the participants played against each other individually (rather than on teams) and a live progress board was displayed via projector. The participants studied a total of ten words in this way and the words studied were taken from the first sublist of Coxhead's (2000) AWL. During treatment three, participants used two-sided physical word cards to study a set of ten words relating to the parts of an academic paper. During treatment four, participants read an English adaptation of Plato's cave allegory (1943) printed on paper. Participants were instructed to identify unknown words by circling them and make a written note about their meaning after consulting an electronic dictionary. Following the administration of the JFSS, the participants were given a faux test in the form of a Google Form, which each consisted of English language written receptive meaning-recognition items (these post tests existed only as an artifact intended to motivate the participants and did not span all items studied by the participants, as such they were not graded nor were they included in the current dataset).

Table 1 *Diagram of Procedures*

Day	Study Activity (10 mins)		Collection of JFSS Data
1	Word cards	→	JFSS administered
2	AR study	→	JFSS administered
3	Quizlet Live	→	JFSS administered
4	Intensive reading	→	JFSS administered

To address the research question, mixed effects models (MEM), most appropriate for this dataset which included incomplete responses for all participants (i.e., not all participants were present on all days of treatment), were fit to the JFSS person-measure logits derived from Winsteps with the *lme4* package (Bates et al., 2015) for R (R Core Team, 2024). A maximum to minimal approach was used, and alpha was set to $p < 0.05$; in our modeling we used best practices outlined for MEMs by Meteyard and Davies (2020). Five permutations of random effects structures were fit and assessed with AIC, BIC, and log likelihood parameters. Initially, a random effects structure which included random slopes for *site* per *person* as a random effect variable and *site* as an intercept-only random effect variable was selected, but due to model convergence issues, an intercepts-only random effects structure including *site* and *person* was selected. After testing within a full fixed effects structure, this model failed to converge, and *site* was removed as a random effect. Although such intercepts-only models can be prone to Type I errors (Barr et al., 2013), the final parsimonious model resulted in the best fit, was employed, and did not exhibit singularity or convergence issues. The MEM fixed effects comprised two categorical variables, *tx* was a four-category variable representing which treatment a participant answered the JFSS immediately after, and *site* was a three-category variable representing which university the participant was recruited from. Post-hoc comparisons of these variables of interest were computed with the *emmeans* package (Lenth, 2021), which includes a Tukey adjustment to account for potential familywise errors encountered when computing multiple p values without adjustment (see Appendix for an OSF link to the dataset and R scribe which describes the model fitting procedures, results, and assumption checks).

Results, Discussion, and Limitations

The MEM random effects indicated that the variance of participants, $SD = 0.30$, explained 9% of the variance in the model. The effect size for the model, conditional $R^2 = 0.23$, indicated that it explained approximately 23% of the variance in the data, with the fixed effects accounting for 13%, $R^2 = 0.13$. The fixed effects for the model indicated that there was a significant difference between *tx* (treatment types) and *site* (universities from which the participants were recruited). The reference variable for *tx*, AR, indicated significant differences between all other treatment types ($p < 0.001$ [Quizlet], $p < 0.001$ [reading activity], and $p < 0.001$ [word cards]). The reference variable for *site*, *x* (University X), indicated an insignificant difference when compared with University Y (p = 0.37), but did demonstrate a significant difference from University Z (p < 0.001).

p values were generated with the *lmerTest* package (Kuznetsova et al., 2017) and post-hoc comparisons of *tx* and *site* were generated with the *emmeans* package (Lenth, 2021). These post-hoc comparisons can be referenced in Tables 3 and 4 respectively.

Table 2 *Mixed Effects Model Coefficients*

					Random Effects
	Fixed Effects				Person
	Coefficient	*SE*	*t*	*p*	*SD*
(Intercept)	1.32	0.1	12.499	< 0.001*	0.3
tx q (quizlet)	–0.5	0.09	–5.347	< 0.001*	
tx r (reading)	–1.06	0.1	–9.725	< 0.001*	
tx w (word cards)	–0.32	0.09	–3.442	< 0.001*	
site y	–0.11	0.13	–0.881	0.37	
site z	–0.31	0.1	–3.039	< 0.001*	

Model formula: measure ~ tx + site + (1|person).
Note. The treatment reference variable is set at AR, the site reference variable is set at site x.

Table 3 *Post-hoc Comparisons of Treatments*

Comparison	Coefficient	*SE*	*t*	*p*
AR vs. quizlet	0.503	0.0942	5.339	<0.0001*
AR vs. reading	1.06	0.1092	9.713	<0.0001*
AR vs. word cards	0.325	0.0947	3.436	0.0036*
Quizlet vs. reading	0.557	0.1087	5.125	<0.0001*
Quizlet vs. word cards	–0.178	0.0941	–1.889	0.2343
Reading vs. word cards	–0.735	0.1091	–6.739	<0.0001*

Table 4 *Post-hoc Comparisons of Sites*

Comparison	Coefficient	*SE*	*t*	*p*
site x vs. site y	0.117	0.133	0.881	0.6533
site x vs. site z	0.316	0.104	3.038	0.0078*
site y vs. site z	0.198	0.115	1.721	0.2006

Note. site x = University X, site y = University Y, site z = University Z.

Regarding the post-hoc comparisons in Table 3, all treatments except for Quizlet Live versus word cards appear to display statistical significance and differ from each other in terms of flow state markers measured with the JFSS. Regarding the post-hoc comparisons in Table 4, only University X and Z appear to display a statistically significant difference from each other (p = 0.0078). The wide variation of JFSS results in this dataset is likely a result of the flow states experienced by these participants based on their personalized experience with each study method. The indifference between the Quizlet Live activity and the word card activity might only be reasonably explained by the differences in flow state markers as elicited from the participants—perhaps Quizlet Live was just as engaging as word cards to the participants in this dataset. The statistically significant differences between University X and University Z in JFSS scores might be explained by differences in the participants' majors or areas of focus: perhaps participants from University Z who are focused on publishing in English (as they are science majors) while University X students are more so focused on music and musical performance, might differ in their perceptions of language study. Such a difference might explain this statistically significant difference between these institutions reported by the model. Figures 1 through 4 might help better visualize the differences in JFSS scores.

Some limitations of the current study are necessary to mention. First the data represented here can only reflect this context. Another limitation of the current study is the use of convenience sampling—while the current sample represents a variety of intact classes, and therefore realistic pedagogical environments—random sampling is ideal. It should also be noted that because these activities were conducted in intact

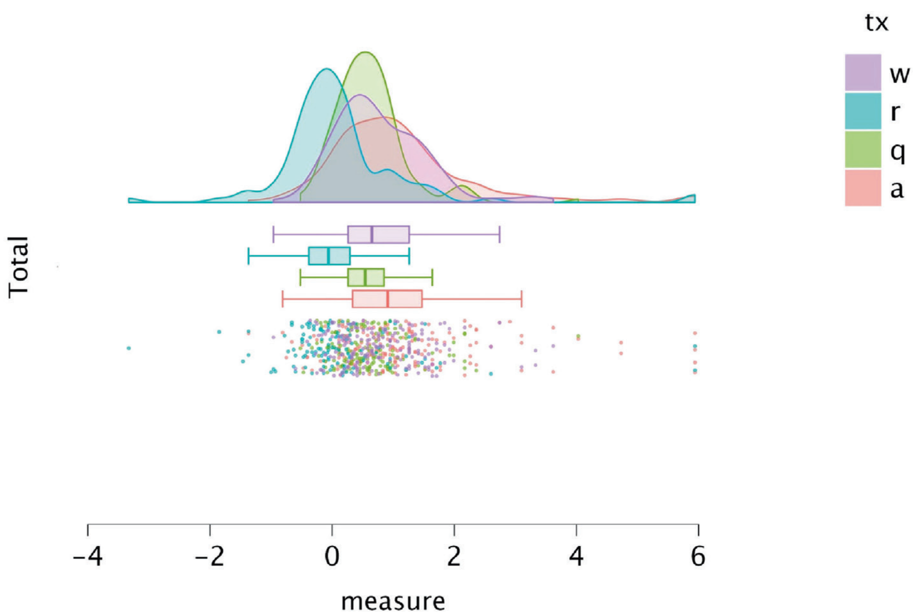

Figure 1 *JFSS Treatment Distributions.*

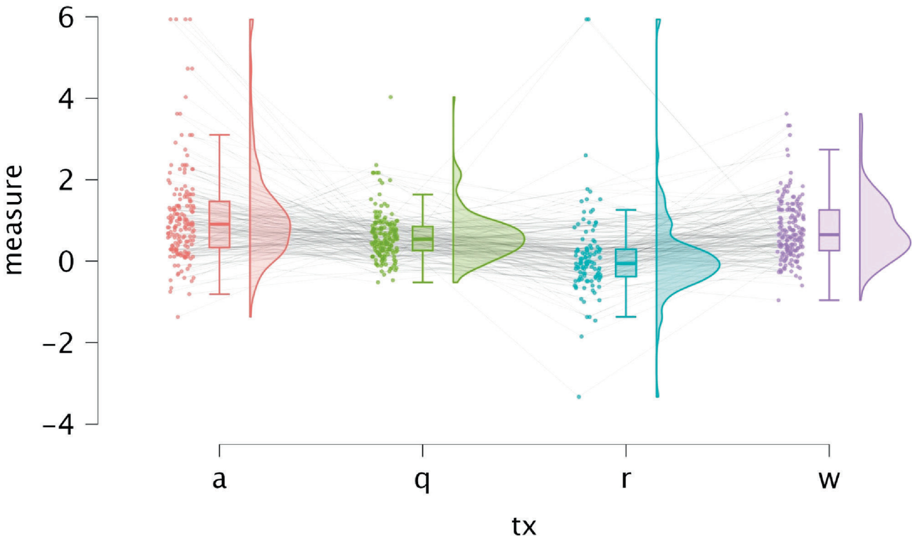

Figure 2 *Participant JFSS Changes Per Treatment.*

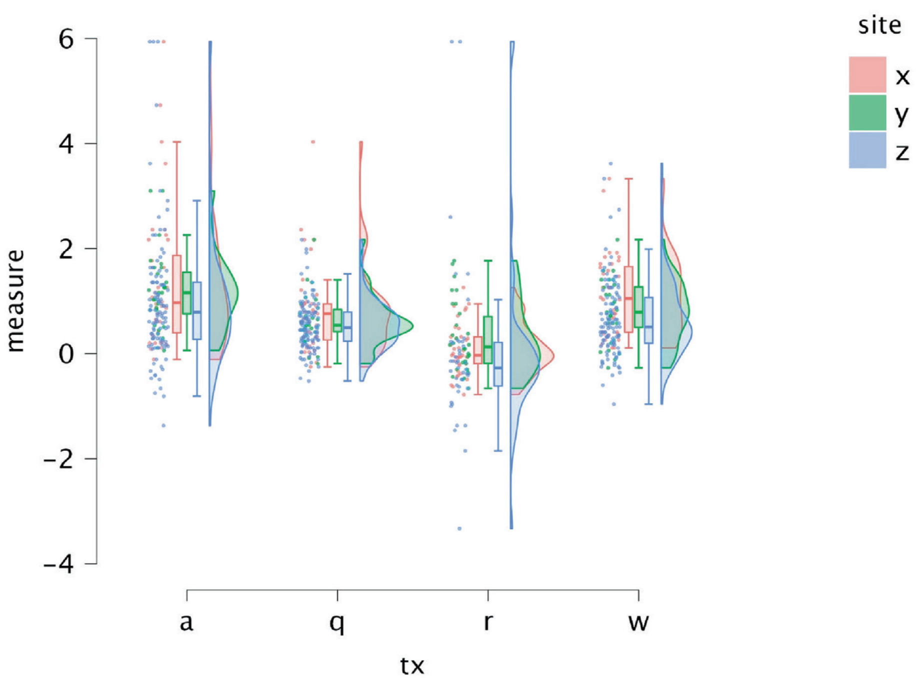

Figure 3 *JFSS Distributions Per Site Per Treatment.*

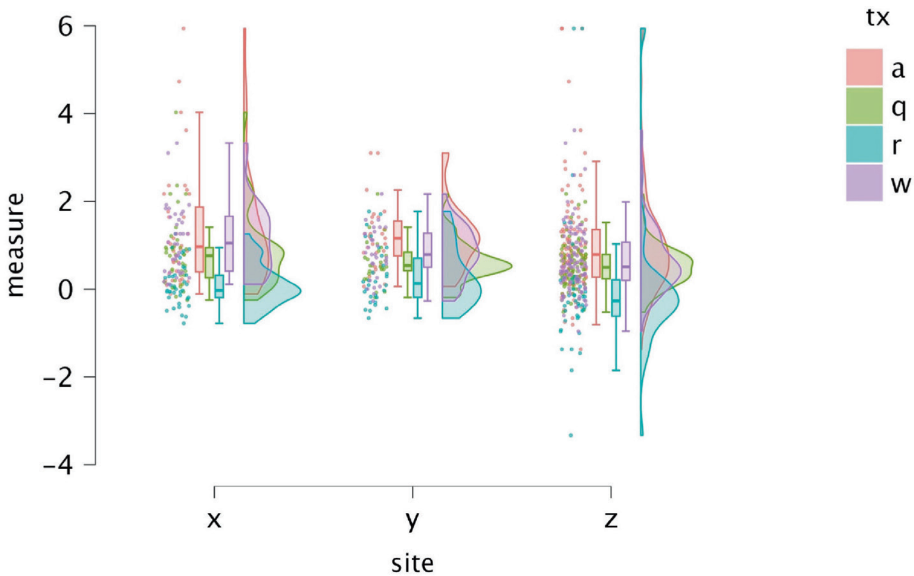

Figure 4 *JFSS Distributions Per Treatment Per Site.*

classes, a variety of moderating variables such as classroom culture, upcoming tests, and other factors associated with school life might have been present during data collection. While research was conducted on four separate days to help reduce the cognitive burden placed on the participants, this choice might also have introduced a possible confound. If replicated, the use of an orthogonal square or counterbalance would help to counteract this possibility. Finally, while this study examined perceived changes in flow based on vocabulary study method, it did not measure the effect flow had on vocabulary learning itself and this presents an opportunity for further research on this topic.

Conclusion

The statistical findings of this study indicate that participants experienced differing levels of flow state markers as measured by the JFSS depending on the deliberate vocabulary activities they participated in. Though the current study investigated two digital and two analog methods of deliberate vocabulary study, many more methods and contexts of deliberate vocabulary study remain unexamined in terms of flow states and warrant investigation.

References

Aubrey, S. (2017). Measuring flow in the EFL classroom: Learners' perceptions of challenges and skills. *Language Teaching Research*, 21(6), 686–708. https://doi.org/10.1002/tesq.387

Bakker, A. B. (2008). The Work-Related Flow Inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior, 72*(3), 400–414. https://doi.org/10.1016/j.jvb.2007.11.007

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 15–35). Cambridge University Press. https://doi.org/10.1017/CBO9780511621956.002

Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience*. Harper Collins.

Dörnyei, Z. (2019). Task motivation. In Z. Wen & M. Javad (Eds.), *Researching L2 task performance and pedagogy: In honour of Peter Skehan* (pp. 53–66). Benjamins. https://doi.org/10.1075/tblt.13.04dor

Egbert, J. (2003). A study of flow theory in the foreign language classroom. *Modern Language Journal*, *87*(4), 499–518. https://doi.org/10.1111/1540-4781.00204

Gakken Education Publishing Co. (n.d.). *ARTourOcean* [Mobile application software].

Google. (n.d.). *Google Forms*. Google LLC. https://www.google.com/forms/

Jackson, S. A., & Eklund, R. C. (2002). Assessing flow in physical activity: The Flow State Scale-2 and Dispositional Flow Scale-2. *Journal of Sport & Exercise Psychology, 24*(2), 133–150. https://doi.org/10.1123/jsep.24.2.133

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: The Flow State Scale. *Journal of Sport and Exercise Psychology*, *18*(1), 17–35.

Jackson, S. A., Martin, A. J., & Eklund, R. C. (2008). Assessing flow states in physical activity: The Flow State Scale-2. *Journal of Sport & Exercise Psychology, 30*(2), 164–186.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. V. (2021). *Emmeans: Estimated marginal means, aka least square means*. R package version 1.7.0. https://CRAN.R-project.org/package=emmeans

Linacre, J. M. (2022). *Winsteps®* (Version 5.2.3) [Computer Software]. Winsteps.com. Retrieved from https://www.winsteps.com/

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed effects models in psychological science. *Journal of Memory and Language*, *112*(104092). https://doi.org/10.1016/j.jml.2020.104092

Plato. (1943). *Plato's the republic*. Books, Inc.

Quizlet. (n.d.). *Quizlet Live*. Quizlet Inc. https://quizlet.com/live

R Core Team. (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). The Flow Short Scale (FSS). In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept* (pp. 261–279). Hogrefe.

Wang, X., & Huang, C. (2022). Conceptualization and measurement of flow in a Chinese blended English as a foreign language learning context. *Frontiers in Psychology*, *13*, (839267). https://doi.org/10.3389/fpsyg.2022.839267

Appendix

Please see the link below to view:

1) Supplementary Materials A: The Japanese Flow State Scale
2) Supplementary Materials B: Dataset CSV file
3) Supplementary Materials C: R script

https://osf.io/xcdhp/?view_only=d389563294ee471b944a6a3a2b435d64