

## Research Article

# Tracking the Longitudinal Change of Flow Experience in an EFL Conversation Course

**Yuya Arai**, *Chiba University, Japan; Waseda University, Japan*

 <https://orcid.org/0000-0001-7922-6616>

yuyaarai1@gmail.com

**Ryuki Matsuura**, *Carnegie Mellon University, USA; Waseda University, Japan*

 <https://orcid.org/0000-0002-9429-6257>

rmatsuur@andrew.cmu.edu

**Masaki Eguchi**, *Waseda University, Japan*

 <https://orcid.org/0000-0002-0791-4935>

meguchi@aoni.waseda.jp

**Shungo Suzuki**, *Nagoya University, Japan; Waseda University, Japan*

 <https://orcid.org/0000-0002-6327-3298>

suzuki.shungo.r5@f.mail.nagoya-u.ac.jp

---

## ABSTRACT

Second language researchers have recently recognized the importance of flow experience as an optimal state for learning, but still little is known about the necessary conditions for flow and its longitudinal change and impact on language learning outcomes. This study tracked the changes of 329 Japanese university students' flow experience and oral proficiency in a semester-long English as a foreign language (EFL) conversation course. A flow questionnaire was administered four times during the semester, while their oral proficiency was assessed by an AI-based interactional speaking test at the beginning and end of the semester. A series of longitudinal analyses using multilevel modeling showed a U-shaped trajectory of flow experience. Furthermore, it was found that flow experience was predicted from the learner-perceived balance between task challenge and skills, concurring with Csikszentmihalyi's (1975/2000, 1990) flow theory. In addition, flow experience contributed to improved oral proficiency scores, providing further positive evidence for the theory.

*Keywords:* flow, longitudinal research design, L2 oral proficiency development, multilevel modeling, AI-based speaking test

---

## INTRODUCTION

Flow has drawn much attention in second language (L2) research (e.g., Czimmermann & Piniel, 2016; Egbert, 2003; Jia et al., 2024; Liu & Song, 2021). Originally proposed in the research field of positive psychology (Csikszentmihalyi, 1975/2000), flow can be defined as a subjective experience “in which an individual is completely immersed in an activity without reflective self-consciousness but with a deep sense of control” (Engeser & Schiepe-Tiska, 2012, p. 1).

Examining flow in L2 learning contexts could deepen the understanding of language learning in many ways. For example, flow is regarded as “a balanced mixture” (Dörnyei & Ushioda, 2011, p. 95) of cognition, motivation, and emotion. Given that language learning cannot be fully explained only by its cognitive aspect but also needs the perspectives of motivation and emotion (Swain, 2013), flow could help view and investigate language learning through the lens of the trilogy of mind (Waninge, 2015). Furthermore, as discussed later, flow has been conceptualized as being related to skill development (Csikszentmihalyi, 1975/2000, 1990), which has also been supported by empirical studies in L2 settings (e.g., Amini et al., 2016; Engeser et al., 2005). Flow could thus contribute to the understanding of language learning processes and outcomes.

Because flow can dynamically change over time (e.g., through a language course; see Ceja & Navarro, 2012; Jia et al., 2024), it is difficult to fully capture the impact of flow on skill development without longitudinal data. However, most L2 research on flow has employed a cross-sectional research design (Jia et al., 2024), suggesting the necessity of understanding (1) how language learners’ flow experience changes through a language course; (2) what explains the change; and (3) how the change explains the development of language proficiency. To fill this research gap, this study longitudinally tracked flow experienced by Japanese university students learning English as a foreign language (EFL) throughout a semester-long conversation course. Applying Csikszentmihalyi’s (1975/2000, 1990) flow theory, the present study also aimed to examine what affected the change of their flow experience and how it contributed to the development of their L2 oral proficiency.

## LITERATURE REVIEW

### Csikszentmihalyi’s (1975/2000, 1990) Flow Theory

As Norsworthy et al. (2021) suggested, flow is a complex phenomenon, resulting in various definitions, operationalizations, and measurements employed in previous studies. Among them, the present study examined flow based primarily on Csikszentmihalyi’s (1975/2000, 1990) flow theory given its influence and empirical support for the theory (e.g., Jackson & Eklund, 2002; Kawabata & Mallett, 2011). His flow theory proposed the following characteristics of flow experience (Csikszentmihalyi, 1990; Nakamura & Csikszentmihalyi, 2009): intense concentration on the task at hand; the merging of action and awareness leading to absorption; a sense of time distortion; having a sense of control over the task; and intrinsic motivation.

Furthermore, some antecedents and consequences of flow experience have been suggested (e.g., Barthelmäs & Keller, 2021; Liu & Song, 2021; Shin, 2006). First, the following three flow antecedents are proposed: clear goals, immediate feedback, and challenge-skills (C-S) balance (Nakamura & Csikszentmihalyi, 2009). Among them, C-S balance has been seen as “a central element of flow theory” (Csikszentmihalyi, 1975/2000, p. xvi) and can subsume the other two antecedents (Barthelmäs & Keller, 2021). C-S balance means that people will experience flow when their perceptions of task challenge and their perceived competence of relevant skills strike a balance at the maximum level. As for flow consequences, flow is hypothesized to lead to skill development. Because the C-S balance is an important flow antecedent, people will not experience flow when task challenges exceed their skills ( $C > S$ ) and vice versa ( $C < S$ ). Once they stop experiencing flow, they will become eager to enter flow again because flow is an intrinsically motivating experience. In order to maximize the instances of flow experiences through a language course, people need to maintain the C-S balance, meaning that they will improve their relevant skills as tasks become more difficult. Compared to other relevant constructs of individual differences (e.g., enjoyment and engagement), Csikszentmihalyi’s (1975/2000, 1990) flow theory has an advantage in explicitly modeling the causal relationship between flow and skill development. Examining how language learners experience flow in the

classroom would thus deepen the understanding of their language learning process.

### Necessity of Longitudinal Flow Research in L2 Contexts

As noted earlier, flow has recently drawn much attention in L2 research (see Kashimura & Nakamura, 2024), but Jia et al. (2024) pointed out that most studies have employed a cross-sectional research design by administering a one-shot flow questionnaire or a pre- and post-test experimental research design by measuring flow only at the beginning and end of an intervention. To extend flow research, it is essential to track the longitudinal change and impact of flow on language learning (cf. Lazarus, 2003) given the time-varying nature of flow (Ceja & Navarro, 2012) and the developmental cycle of C-S balance and flow leading to skill development over time (Csikszentmihalyi, 1975/2000, 1990).

Among a limited number of L2 studies examining the longitudinal change of flow experience, Kirchhoff (2013) administered a flow questionnaire 14 times during an EFL extensive reading program to understand how flow experience changes in relation to the amount of reading. Another example of research on the change of flow experience is Arai's (2025) study, which served as a precursor to this study. In a semester-long EFL conversation course, Arai (2025) qualitatively examined nine Japanese EFL undergraduates' flow experience trajectories in order to understand the consequences of course placement accuracy. Jia et al. (2024) also conducted an eight-week diary study to investigate the change of flow and self-efficacy.

Those studies raise the following four important issues that should be addressed in future L2 flow research. First, Kirchhoff's (2013) study was descriptive and only used group means of the flow questionnaire scores over time to discuss flow experience in the classroom. However, various factors can be related to flow experience, such as interest and achievement motives, as they can moderate the relationship between C-S balance and flow experience (e.g., Baumann et al., 2016; Bricteux et al., 2017; Engeser & Rheinberg, 2008). Therefore, in order to better understand flow experience, more studies are needed to model the flow mechanism with both individual- and group-level information.

A second issue is whether flow is a discrete (i.e., flow or non-flow) or continuous construct (i.e., spectrum of the intensity of flow), which is still debatable in flow research (Abuhamdeh, 2020; Norsworthy et al., 2021). The participants in Kirchhoff's (2013) study were categorized into either flow or non-flow status with the group mean score as the cut-score. Even if she deliberately operationalized flow as a discrete construct, it is worth discussing whether the threshold of flow versus non-flow should have been the average score of the flow questionnaire. In contrast, Arai (2025) and Jia et al. (2024) operationalized flow as a continuous construct. While many researchers have operationalized flow as a continuous construct and employed a Likert-scale questionnaire (Norsworthy et al., 2021), treating flow as a continuous construct may question its "discriminant validity over pre-existing constructs in surrounding fields" (Abuhamdeh, 2020, p. 5), such as intrinsic motivation. However, this way of operationalization assumes the intensity of flow and therefore is conceptually consistent with Csikszentmihalyi's (1975/2000) idea of *microflow* (i.e., less-intense flow experience; e.g., humming, doodling, watching TV, etc.). Csikszentmihalyi also pointed out that "flow exists on a continuum from extremely low to extremely high complexity" (p. 141). Meanwhile, a third option is that flow is "a yes-no continuous phenomenon" (Peifer & Engeser, 2021, p. 425) and should be measured both categorically and continuously, although little is known on the exact threshold of the continuous nature of the flow experience. Further discussions of this issue are thus required in L2 flow research, where researchers need to justify why they treat flow as a discrete or continuous construct.

Third, the difference between perceived and objective C-S balances has not drawn much attention in previous studies. The difference can be important given Csikszentmihalyi's (1975) suggestion that "whether a person is going to be in flow or not does not depend entirely on the objective nature of the challenges present or on the objective level of skills ... [but] on one's *perception* of what the challenges and skills are" (p. 56, emphasis original). However, this idea needs empirical support in language learning research. While Kirchhoff (2013) and Jia et al. (2024) did not examine C-S balance as an important antecedent of flow experience, Arai (2025) qualitatively explored the relationship between flow and the following three factors

concerning the C-S balance: learner-perceived C-S balance, teacher-perceived C-S balance (operationalized as teacher judgments of course placement accuracy), and objective C-S balance (operationalized as learners' course levels based on the placement test results). The results of this study showed that learner-perceived C-S balance was closely linked to the learners' flow experience, in line with Csikszentmihalyi's (1975) assumption. In light of the qualitative nature of the study, further longitudinal research should quantitatively model the relationships between flow experience and three types of C-S balance: objective, learner-perceived, and teacher-perceived C-S balance.

Finally, it should be noted that Kirchhoff (2013), Arai (2025), and Jia et al. (2024) did not measure learning outcomes to test the effects of flow experience on skill development. Although some flow studies examined learning outcomes (e.g., Amini et al., 2016; Engeser et al., 2005), they did not examine the longitudinal change and impact of flow on language learning outcomes.

Taken together, L2 flow research has lacked empirical evidence of the applicability of Csikszentmihalyi's (1975/2000, 1990) flow theory to L2 learning from a longitudinal perspective. In addition, previous research examining the dynamic change of flow experience raised the four issues to be addressed: (1) the necessity of statistically modeling both individual- and group-level information on flow experience; (2) flow as a discrete or continuous construct; (3) the need of empirical evidence for the difference between *perceived* and *objective* C-S balances; and (4) the importance of examining the longitudinal relationship between flow and learning outcomes.

In order to address (1), this study employed multilevel regression models to track Japanese EFL university students' flow experience in a semester-long conversation course. Multilevel modeling has the potential to examine both intra- and inter-individual changes in flow experience (Hiver & Al-Hoorie, 2020). With regard to (2), this study operationalized flow as a continuous construct and measured its intensity in the conversation course in light of the conceptual consistency with Csikszentmihalyi's (1975/2000) flow theory and previous L2 flow research measuring the intensity of flow (e.g., Arai, 2025; Czimmermann & Piniel, 2016; Jia et al., 2024; Liu & Song, 2021; Payant & Zuniga, 2022). As for (3), this study aimed

to test the generalizability of Arai's study findings by examining the effects of learner-perceived, teacher-perceived, and objective C-S balances on the change of flow experience. Finally, to address (4), this study examined the effect of flow on the development of the students' EFL oral proficiency in the conversation course. While previous studies have suggested that L2 conversational tasks contributed to flow experience (e.g., Almukhailed & King, 2024; Aubrey, 2017a, 2017b; Piniel & Ritecz, 2022), little is known as to how the development of L2 oral proficiency can be explained by the longitudinal change flow experience.

The purpose of this study was to examine the relationship among flow antecedents (i.e., three types of C-S balance: learner-perceived, teacher-perceived, and objective ones), experience, and consequences (i.e., the development of L2 oral proficiency) by tracking Japanese EFL university students' flow experience in a semester-long conversation course. Research questions (RQs) were as follows:

**RQ1:** How does Japanese EFL university students' flow experience change through an English conversation course?

**RQ2:** Which type of C-S balance contributes to variability in their flow experience as flow antecedents: learner-perceived, teacher-perceived, and objective ones?

**RQ3:** To what extent is the change of their flow experience related to the development of their oral proficiency through the course as a flow consequence?

## RESEARCH METHODS

### Study Context and Participants

This study was conducted in an EFL conversation course at a private university in Japan where students aimed to develop conversation skills for general purposes (e.g., describing events). The course had five levels (Beginner, Basic, Intermediate, Upper Intermediate, and Advanced), corresponding respectively to the A1, A2, B1, B2, and C1/C2 levels of the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001). The participants were placed in one of the five course levels based on their scores on an in-house AI-based online speaking test named LANGX Speaking (for further details, see Saeki et al., 2024; Takatsu et al., 2026). Each course

session lasted 100 minutes and consisted of up to four students and one teacher who spoke English as a first language (L1) or was a proficient user. Most students attended the course twice a week for ten weeks in a semester (i.e., 20 sessions in total). In the course, students were supposed to achieve learning goals in each session, involving both linguistic (e.g., words and phrases) and functional targets (e.g., discussing a social issue), through form-focused activities and spontaneous interactive tasks, including role play. In every fifth session (i.e., fifth, tenth, fifteenth, and twentieth sessions), students were evaluated by their teachers while engaging in interactive tasks similar to those in the previous four sessions. This session was called an “assessment unit.”

The participants were 329 students (female: 186; male: 135; other gender: 1; no answer: 7). Their ages ranged from 18 to 32 ( $M = 18.7$ ;  $SD = 1.39$ ). All were first-year university students except for 12 participants (second year: 3; third year: 4; fourth year: 3; fifth year: 1; graduate student: 1). Except for several participants (Chinese: 6; Korean: 1), they spoke Japanese as their L1. They were from various faculties; most were non-English majors, while 19 participants were from English majors. Most faculties required students to take the course. Their course levels were Beginner ( $n = 8$ ), Basic ( $n = 138$ ), Intermediate ( $n = 101$ ), Upper Intermediate ( $n = 45$ ), and Advanced ( $n = 37$ ). Fifty-nine participants had experienced study abroad for more than a month before study participation.

## Materials

### *Measures of Flow Antecedents*

In light of the importance of C-S balance compared to other flow antecedents (e.g., immediate feedback, clear goals) in Csikszentmihalyi's (1975/2000, 1990) flow theory, the participants reported C-S balance in terms of one of the following five conditions: (1) “Session difficulty was much lower than ability,” (2) “Session difficulty was lower than ability,” (3) “Session difficulty and ability were in equilibrium,” (4) “Session difficulty was higher than ability,” and (5) “Session difficulty was much higher than ability.” This categorization was based on Barthelmäs and Keller (2021), who suggested that task challenge and person skills should be examined as a combined challenge-skills fit.

In order to empirically confirm Csikszentmihalyi's (1975) idea that it is the learner-perceived, not teacher-perceived or objective, C-S balance that affects flow experience, this study also included (a) teachers' perceptions of C-S balance, (b) LANGX Speaking pre-test scores, and (c) course levels. As a variable concerning teachers' perceptions of C-S balance, this study collected teacher judgments as to whether the participants were placed on appropriate course levels. Teacher judgments (see Arai, 2025, for more detailed information) were made in the fifth session based on the following three categories: (a) accurate placement (i.e., their students' proficiency matching the current course level); (b) false negative placement (i.e., their students' proficiency exceeding the current course level); and (c) false positive placement (i.e., the current course level exceeding their students' proficiency).

### *Measures of Flow Experience*

To measure flow that the participants experienced during a session, this study used an adapted version of the Flow Short Scale (Rheinberg & Engeser, 2018; Rheinberg et al., 2003), which consists of ten items on a six-point Likert scale (1 = “strongly disagree”; 6 = “strongly agree”). The aggregation of responses to each item represented the flow score (min: 10; max: 60), measuring flow as a unitary construct. The scale was developed to be “suitable for completed activities” (Rheinberg & Engeser, 2018, p. 606), allowing the present study to retrospectively measure flow experienced during the conversation course (see the Data Collection section), as Engeser and Rheinberg (2008) did in their study. Rossin et al. (2006) also suggested that people do not forget their flow experiences immediately (see also Bartholomeyczik et al., 2024). Previous research has measured flow reliably in a retrospective way (e.g., Jia et al., 2024). As for the validity of the scale, it has been validated in previous studies employing various research designs, including experimental, correlational, experience-sampling, and longitudinal design studies (e.g., Arai, 2025; Engeser & Rheinberg, 2008; Rheinberg et al., 2003; Schüler, 2007). Arai's study also provided partial evidence for the criterion-related validity of the scale in that the characteristics of flow experience measured by the scale were related to those elicited by a semi-structured interview.

The original scale was adapted in several ways. For example, the six-point Likert scale was used to avoid the

neutral option in the original seven-point Likert scale. Furthermore, the wordings of some items were modified. In addition, the tenth item in the original scale (“I am completely lost in thought;” Rheinberg & Engeser, 2018, p. 607) was replaced with a new item concerning intrinsic motivation, which is not tapped into in the original scale although this is an important flow characteristic (Csikszentmihalyi, 1975/2000). The adapted scale (available at <https://osf.io/n9g2f>) was translated into Japanese by the first author. Two Ph.D. students majoring in L2 acquisition reviewed the accuracy of the translation, the validity of the scale, and the wording. The Cronbach’s alpha results in this study suggested good reliability: .85, .90, .90, and .92, respectively, at each time of questionnaire administration.

### *Measures of Flow Consequences*

As a flow consequence, this study examined the development of the participants’ oral proficiency by adopting their scores in the LANGX Speaking test. More specifically, they took the test before the course for placement purposes and after the semester for achievement assessment purposes. The first score was used as the pre-test score, and the second as the post-test score in this study. The test had two sections: an interview and a conversation. In the interview section, the participants answered several questions from a conversational virtual agent. In the conversation section, they were asked to discuss a topic (e.g., social issues) with the virtual agent. The test took 20-30 minutes to complete, and the score was provided for test-takers immediately. The test scores were presented as the CEFR levels (for further details on the test, see Saeki et al., 2024; Takatsu et al., 2026).

### **Data Collection**

Students enrolled in the course were invited via email and were offered information about the research project and study procedure. They were also informed of the voluntary nature of study participation; that there was no relationship to course credits/grades; of their right to withdraw from the study at any time even after they had agreed to participate

in the study; and of the university’s research ethics committee’s approval of the research project. After that, those who were willing to participate completed a consent form and a background questionnaire.

The participants took the LANGX Speaking test twice: in the week before the first session (i.e., pre-test) and after the final session (i.e., post-test). During the course, they completed the flow questionnaire with the item concerning C-S balance four times in the assessment units (i.e., every fifth session). The questionnaire was administered immediately after each of the four sessions using Qualtrics. It took approximately one to two minutes to complete the questionnaire.

### **Analysis**

Tables 1 and 2 show the descriptive statistics on four occasions (i.e., four assessment units) for variables that either change over time (i.e., time-variant) or remain constant (i.e., time-invariant). To preliminarily validate the current data for analyzing changes in flow scores, a series of one-sample *t*-tests with Bonferroni correction was conducted to compare students’ flow scores against their mid-point (i.e., 35) at each time. The results demonstrated that flow scores at each time were significantly higher than the mid-point score with medium to large effect sizes (see Table 3; Plonsky & Oswald, 2014), indicating flow experience was unlikely to be completely absent. Because few students ( $n \leq 5$ ) chose categories 1 and 5 for the C-S balance item (i.e., session difficulty much lower/higher than ability), they were combined into categories 2 (i.e., session difficulty lower than ability; henceforth C<S) and 4 (i.e., session difficulty lower than ability; henceforth C>S), respectively. Table 4 presents the descriptive statistics on other time-variant variables across two occasions (i.e., pre- and post-test scores). As with C-S balance, the C2 level (pre-test:  $n = 1$ ; post-test:  $n = 5$ ) of the LANGX Speaking test was merged with C1 and was labeled C level due to the limited sample size.

**Table 1.** Descriptive Statistics of Time-Variant Flow Score and C-S Balance (N= 329)

Variables	Time1	Time2	Time3	Time4	Total
<b>Flow Score</b>					
M	43.980	43.177	43.722	44.387	43.799
SD	6.847	8.210	7.779	8.507	7.837
NA	30 (9.119%)	35 (10.638%)	48 (14.590%)	73 (22.188%)	186 (14.134%)
<b>C-S balance</b>					
Category 1	0 (0.000%)	4 (1.216%)	1 (0.304%)	2 (0.608%)	7 (0.532%)
Category 2	45 (13.678%)	29 (8.815%)	28 (8.511%)	26 (7.903%)	128 (9.726%)
Category 3	168 (51.064%)	173 (52.584%)	171 (51.976%)	147 (44.681%)	659 (50.076%)
Category 4	83 (25.228%)	83 (25.228%)	76 (23.100%)	77 (23.404%)	319 (24.240%)
Category 5	3 (0.912%)	5 (1.520%)	5 (1.520%)	4 (1.216%)	17 (1.292%)
NA	30 (9.119%)	35 (10.638%)	48 (14.590%)	73 (22.188%)	186 (14.134%)

Notes. M = mean; SD = standard deviation; NA = missing value

Other values indicate the frequency and percentage of the corresponding categories. Percentages were calculated based on the total number of categories for each occasion.

**Table 2.** Frequencies and Percentages of Time-Invariant Course Level and Accuracy

Variables	Category	Frequency	%
<b>Teacher Judgments of Placement Accuracy</b>	Accurate Placement	128	38.906
	False Positive	13	3.951
	False Negative	17	5.167
	NA	171	51.976
<b>Course Level</b>	Beginner	8	2.432
	Basic	138	41.945
	Intermediate	101	30.699
	Upper Intermediate	45	13.678
	Advanced	37	11.246

Notes. NA = missing value.

Percentages were calculated based on the total number of categories.

**Table 3.** Results of One-Sample T-Tests to Compare Flow Scores against Mid-Point Score

	t-statistics	df	p	Cohen's d
Time 1	22.677***	298	<.001	1.311
Time 2	17.077***	293	<.001	0.996
Time 3	18.796***	280	<.001	1.121
Time 4	17.655***	255	<.001	1.103

Notes. †p < .10; \*p < .05; \*\*p < .01; \*\*\*p < .001

p-values were adjusted using Bonferroni correction.

**Table 4.** *Frequencies and Percentages of Pre- and Post-LANGX Speaking Test Scores*

CEFR levels	Time1 (Pre-test)		Time2 (Post-test)		Total	
	Freq.	%	Freq.	%	Freq.	%
A1	14	4.255	18	5.471	32	4.863
A2	79	24.012	43	13.070	122	18.541
B1	54	16.413	90	27.356	144	21.884
B2	29	8.815	71	21.581	100	15.198
C1	25	7.599	52	15.805	77	11.702
C2	1	0.304	5	1.520	6	0.912
NA	127	38.602	50	15.198	177	26.900

*Notes.* NA = missing value.

Percentages were calculated based on the total number of CEFR levels for each time.

Since the current data contains numerous missing values, it is essential to conduct multiple imputation (Rubin, 1987) to prevent a loss of statistical power and biased analyses resulting from listwise deletion (Hox et al., 2018). To this end, multivariate imputation by chain equations (MICE; Buuren & Groothuis-Oudshoorn, 2011) was employed given its robustness to multilevel incomplete data (Hox et al., 2016). MICE iteratively predicts them based on observed values while adding random errors to consider the uncertainty of missingness. Since MICE assumes that the likelihood of missingness is independent of the variable itself but dependent on other variables (i.e., missing at random), this assumption was tested by visualizations and logistic regressions (see supplementary material: <https://osf.io/n9g2f>). The results showed that a missing pattern of post-test scores could be considered independent from any other variables (i.e., missing completely at random) in datasets for RQ1 and RQ2. Thus, listwise deletion was applied to handle incomplete post-test scores (Hox et al., 2016). Twenty imputed datasets were then generated for the subsequent analyses considering the uncertainty of missing value prediction. To compare the pooled results of twenty models fitted using imputed datasets, the mean values of the Akaike information criterion (AIC) and conditional  $R^2$  scores were calculated.

To address the RQs, a series of longitudinal analyses was performed using multilevel modeling (Hoffman, 2014). To explicitly model the nested structure of intra- and inter-individual variables, multilevel modeling was chosen over latent growth curve models (cf., Curran et al., 2010). First, to identify the baseline pattern of changes in flow experience over time (RQ1), four models were

compared: intercept, linear, quadratic, and cubic. Since there was no strong theoretical assumption of how flow experience can change longitudinally in the L2 context, linear and nonlinear models were explored given the possibility of the dynamically changing nature of flow experience (Ceja & Navarro, 2012). For nonlinear models, in order to systematically explore the complexity of flow experience trajectory (i.e., parabolic vs. more complex curve), polynomial models up to the cubic order were employed, given the availability of four observation occasions. In these models, the flow scores were used as the outcome variable, and the time (i.e., the occasion on which they were collected, from Time 1 to Time 4) was used as the predictor. Times 1, 2, 3, and 4 were coded as 0, 1, 2, and 3, respectively. In all models, participants were included as a random intercept. Random slopes of participants were also tested for each model, but singular fits were obtained, possibly due to an insufficient sample size, and thus they were excluded.

To identify factors affecting changes in flow experience (RQ2), multilevel models were expanded by incorporating both time-variant and time-invariant variables into the baseline model. As a first step, the time-variant variable (i.e., C-S balance on each occasion) was added to the model. To facilitate the interpretation of the results, the three C-S balance categories were transformed using simple contrast coding with a reference category of the C=S condition (i.e., session difficulty matching ability). Next, three time-invariant covariances were added to the model: (a) teacher judgments of placement accuracy; (b) course level; and (c) pre-test score. Teacher judgments were transformed using simple contrast coding with accurate placement as the reference category. Five course levels were handled using forward difference contrast

coding to account for their order. Pre-test scores were treated as a time-invariant covariance, as they were measured only once before the beginning of the course. Although pre-test scores were discrete ordinal variables, contrast coding for them was intentionally avoided to prevent an excessive number of parameters. To examine the contribution of pre- and post-test changes to flow scores, the final model included two more time-invariant variables: (a) post-test scores, which were processed similarly to pre-test scores but added in this step to account for the different timing of pre-test observations; and (b) test score changes, which were calculated based on the changes between the pre- and post-test scores, categorized as follows: pre>post (post-test score lower than pre-test), pre=post (post-test score same as pre-test), and pre<post (post-test score higher than pre-test). These variables were transformed into dummy variables using simple contrast coding, making it possible to compare the pre=post variable with the others. In the second set of analyses, this study did not include by-participant random slopes for the time-invariant variables due to potential convergence issues.

RQ3 addressed the contribution of flow to the development of oral proficiency. Therefore, new models were constructed to predict score change patterns. Given the ordinal scale of test scores (i.e.,  $A1 < A2 < B1 < B2 < C$ ), this study adopted multilevel cumulative link models. Since there were only two time points (i.e., pre- and post-test), to identify a best-fit model, this study compared intercept and linear models while incorporating linear time changes and time-invariant variables (i.e., unconditional and conditional models). Participants were included as a random intercept in all models. Meanwhile, the time-invariant variable of course level, which was determined by the pre-test score, was excluded from the models due to its potential circularity with pre-test scores. Considering differences in observation occasions, the flow scores were transformed into two kinds of time-invariant variables: (a) mean flow scores and (b) flow score changes at Time1–2, Time2–3, and Time3–4. Similarly, this study created two time-invariant variables for C-S balance by counting the number of categories (i.e.,  $N_{C>S}$ ,  $N_{C<S}$ ). Notably,  $N_{C=S}$  was

not included because it could be derived from the other two variables (i.e.,  $4 - N_{C>S} - N_{C<S}$ ), rendering it redundant information. R scripts for analyses and detailed results are available at <https://osf.io/n9g2f>.

## RESULTS

### Flow Experience Change through the Conversation Course

The results of the intercept, linear, quadratic, and cubic models are summarized in Table 5. First, the fits of the intercept and linear models were compared to examine whether the flow scores changed linearly over time. The differences in mean AIC ( $M_{AIC}$ ) and conditional mean  $R^2$  ( $M_{R^2}$ ) were minimal (only +1 and +.001, respectively). A likelihood ratio test indicated  $\chi^2(1) = 1.21, p = .27$ , suggesting a better fit of the intercept model. Next, nonlinear parabolic changes in flow scores were explored based on the quadratic model fit. Comparing the intercept and the quadratic model fits, there were slight differences in  $M_{AIC}$  and conditional  $M_{R^2}$  (a decrease of 13 and an increase of .004, respectively). A likelihood ratio test showed  $\chi^2(1) = 3.10, p = .045$ , indicating an improvement in model fit. The cubic model fit was also evaluated, but the cubic time effect was not statistically significant ( $p = .657$ ). Consequently, it was concluded that the changes in flow scores did not follow a complex cubic trajectory, and the quadratic change model was selected as the best fitting.

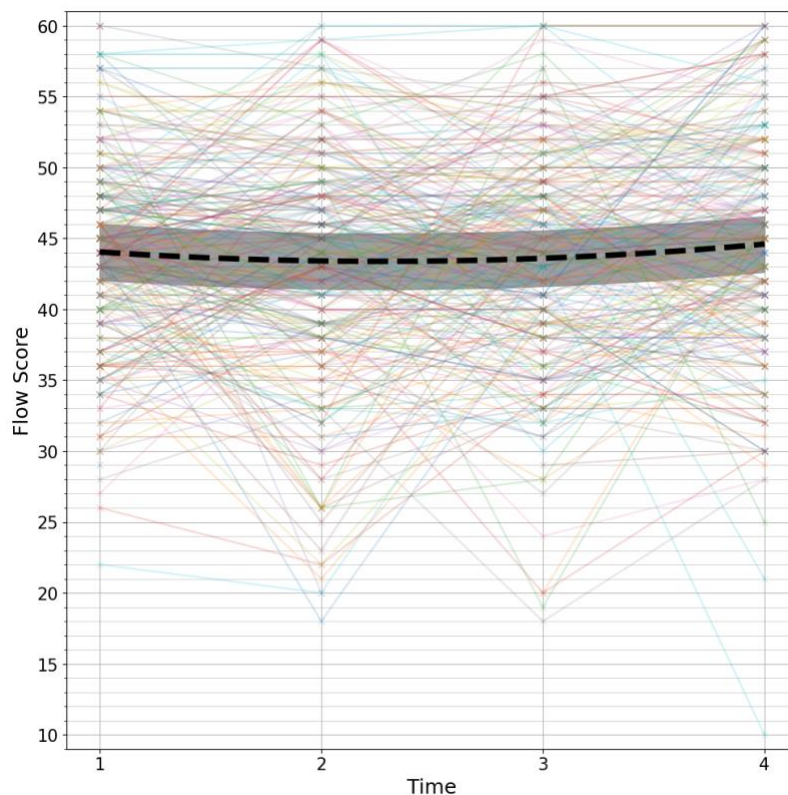
Figure 1 depicts the best-fit quadratic change model. The model demonstrated that the flow score change was significantly explained by the quadratic time effect ( $p = .025$ ) with an estimate of .388, suggesting the gradual acceleration of flow score change. In addition, the linear time effect marginally significantly contributed to the flow score change ( $p = .063$ ) with an estimate of -1.039. At Time 2 alone, coded as 1, the combination of linear and quadratic time effects fell below zero, exhibiting a gradual U-shaped pattern in flow scores over time, with a convex shape near Time 2.

**Table 5.** Summary of Intercept, Linear, Quadratic, and Cubic Models

Model	Estimates	CI	<i>p</i>	Mean AIC ± SD	Mean R <sup>2</sup> ± SD
<b>M1-1: Intercept</b>					
Intercept	43.903***	[43.196, 44.610]	<.001	7,502.354 ± 22.855	0.431 ± 0.011
<b>M1-2: Linear</b>					
Intercept	43.625***	[42.757, 44.493]	<.001	7,502.801 ± 22.880	0.432 ± 0.011
Time	0.185	[-0.145, 0.516]	0.271		
<b>M1-3: Quadratic</b>					
Intercept	44.033***	[43.112, 44.954]	<.001	7,499.266 ± 22.904	0.436 ± 0.011
Time	-1.039†	[-2.135, 0.056]	0.063		
Time <sup>2</sup>	0.408*	[0.056, 0.761]	0.023		
<b>M1-4: Cubic</b>					
Intercept	44.070***	[43.131, 45.010]	<.001	7,500.903 ± 22.813	0.436 ± 0.011
Time	-1.621	[-4.442, 1.200]	0.261		
Time <sup>2</sup>	0.965	[-1.526, 3.456]	0.448		
Time <sup>3</sup>	-0.124	[-0.670, 0.423]	0.657		

Notes. †*p* < .10; \**p* < .05; \*\**p* < .01; \*\*\**p* < .001

**Figure 1.** Model Estimated Quadratic Flow Score Change (Dashed Line) with 95% Confidence Interval and Individual Flow Score Change (Solid Line)



### Factors Affecting Flow Experience Change

Table 6 (see Appendix A) shows the results of the second set of multilevel models predicting flow score changes using both the time-invariant and time-variant predictor variables. Model 2-2 (baseline + C-S balance) yielded the metrics that were substantially different from the baseline ( $AIC = -111, R^2 = +.041$ ). A follow-up likelihood ratio test confirmed a significant improvement in model fit,  $\chi^2(1) = 50.08, p < .001$ . Meanwhile, Model 2-3, where time-invariant placement accuracy, course level, and pre-test score were added, did not outperform Model 2-2 ( $AIC = +2, R^2 = \pm 0, \chi^2(1) = .968, p = .453$ ).

### Flow Experience Change Impact on Pre- and Post-Test Scores

Table 7 (see Appendix B) presents the results of cumulative link models predicting pre- and post-test score changes from flow experience. The model fit improved when occasion and time-invariant variables were added to the intercept model. Specifically, the unconditional change model further reduced  $M_{AIC}$  (-140) and surpassed  $M_{R^2}$  (+.124) from the intercept model. Additionally, the conditional model including the time-invariance covariances achieved even lower  $M_{AIC}$  (-12) but equal  $M_{R^2}$  to the unconditional change model.

The conditional model was selected because its time-invariant variables had the lowest  $M_{AIC}$ . Three findings were worth noting. First, the model identified that linear time was a significant predictor ( $p < .001$ ), with an estimate of 2.358. This suggests that the post-test scores were approximately ten times more likely to be judged at the next level than pre-test scores. Second, the mean flow scores significantly explained test score changes ( $\beta = 0.179, p = 0.004$ ). Thus, students with overall flow scores one level higher are also 1.196 times more likely to have test scores one level higher. However, the changes in flow scores between adjacent time intervals across the four time points did not significantly predict the changes in proficiency scores. Finally, the other variables (i.e.,  $N_{C>S}, N_{C<S}$ , and placement accuracy) were not significant predictors.

### DISCUSSION

RQ1 addressed the change of flow experience over time. Descriptive statistics in Tables 1 and 3 showed relatively high mean flow scores throughout the conversation course. This is consistent with previous studies examining flow in EFL contexts (Czimmerman & Piniel, 2016; Dewaele & MacIntyre, 2024; Liu & Song, 2021; Payant & Zuniga, 2022), supporting the possibility that, on average, the conversation course encouraged the students to experience flow. It was also found that the flow experience trajectory through the course exhibited a U-shaped pattern. This finding is noteworthy in that cognitive psychology has paid attention to U-shaped learning behavior to understand learning and skill development (e.g., Strauss, 1982). The results also show that the participants improved their oral proficiency scores and that flow was a statistically significant predictor of that improvement. It is thus plausible that the flow experience trajectory observed in this study is aligned with the development of oral proficiency, supporting the idea that continuing to experience flow over time contributes to skill development in the long run. Given that the longitudinal change of flow experience in the language classroom has not frequently been explored (Jia et al., 2024), the findings cast new light on flow experience trajectories and their impact on language learning. However, it should be noted that there was the variability in the patterns of flow experience trajectories among individual learners, as shown in Figure 1 (see also Arai, 2025, who qualitatively examined potential reasons for the differences).

Although it is difficult to identify the reasons why the participants tended to report the lower intensity of flow around Time 2 and Time 3, a precursor qualitative study to the present study (Arai, 2025) found some classroom environmental factors affecting the low intensity of flow experience, including the change of class members and one-on-one sessions with teachers due to other students' absence. Given that flow is a product of the person-environment interdependency (Nakamura & Csikszentmihalyi, 2009), such environmental factors might have contributed to the change of flow.

RQ2 aimed to examine whether theoretically motivated flow antecedents affected variability in flow experience trajectories. More specifically, this study examined the roles of course levels, the LANGX Speaking test scores as a

proxy for oral proficiency, teacher judgments of placement accuracy, and C-S balance in predicting the flow experience trajectory through the conversation course. The results of the model comparison revealed that C-S balance was the only meaningful predictor of the change of flow experience over time. This concurs with Csikszentmihalyi's (1975/2000, 1990) flow theory in that C-S balance is an important flow antecedent (Nakamura & Csikszentmihalyi, 2009). In line with Csikszentmihalyi's flow theory, both the  $C < S$  and the  $C > S$  conditions were not related to flow experience compared to the  $C = S$  condition. However, the findings in this study were not consistent with some previous studies suggesting that the contribution of  $C < S$  or  $C > S$  to flow experience may depend on some learner-related factors (e.g., achievement motives; see Baumann et al., 2016; Bricteux et al., 2017; Engeser & Rheinberg, 2008). To understand this discrepancy, therefore, future L2 flow research can explore the relationship between such learner-related factors and flow experience in detail by conducting qualitative studies, which have been less frequent in the L2 flow research field than quantitative studies (Almukhail & King, 2024).

Furthermore, this study's findings quantitatively replicate Csikszentmihalyi's (1975) idea and Arai's (2025) qualitative findings that learner-perceived C-S balance is more closely related to the flow experience trajectory than objective (i.e., course level determined by the results of the placement test) and teacher-perceived C-S balance (i.e., teacher judgments of placement accuracy). The results also concur with Arai's suggestion that students' perspectives play an important role in understanding time-varying experiences of flow and skill development (note that nine students in Arai's study also participated in this study).

Finally, RQ3 addressed the extent to which the change of flow experience was related to the development of the participants' oral proficiency through the course. It was found that mean flow scores alone statistically predicted improvement in the oral proficiency scores. This provides positive evidence for the compatibility of Csikszentmihalyi's (1975/2000, 1990) flow theory with L2 learning contexts, suggesting that flow experience contributes to L2 development. The finding that C-S balance statistically failed to predict the test score improvement also confirms the roles of flow antecedents and consequences hypothesized in flow theory in that C-S balance is not a direct predictor of skill development but

indirectly contributes to skill development by enhancing flow experience. Furthermore, this study has confirmed that the test score improvement could be attributed to the U-shaped trajectory of flow experience. These findings are consistent with Engeser et al.'s (2005) finding that flow experience predicts language learning outcomes. While their study employed a cross-sectional research design, this study identified the change of flow experience over time and its relationship to the improvement in oral proficiency scores. Further longitudinal research should test the generalizability of the current findings in different learning contexts targeting various aspects of language proficiency.

## LIMITATIONS, IMPLICATIONS, AND CONCLUSION

There are some important limitations that should be addressed in future research. First, this study administered the flow questionnaire only in the assessment units of the target semester-long course. The flow scores may have been affected by the unique characteristics of assessment units (e.g., test-taking anxiety) rather than instructional activities, requiring cautious interpretations of the flow scores. Second, the characteristics of activities in the course (e.g., topics and difficulty) were not constant throughout the semester, which could have affected the change of flow experience as well as the comparability of flow experiences across each measurement point. Although tracking flow experiences in the existing EFL conversation course contributed to strengthening ecological validity, more attention should be paid in the future to the moment-by-moment change of flow experience during each task from a microgenetic perspective in an experimental context where task characteristics are strictly controlled. Third, the flow questionnaire was administered only four times during one semester. A longer time span for conducting the study with more observation time points within and between lessons would deepen the understanding of the dynamic change of flow experience. Fourth, a combination of different research tools (e.g., interviews) would have provided more information on the participants' flow experience. Fifth, the specific contextual nature of this study necessitates cautious generalization of the study's findings to other language learning contexts. For example, as discussed above, it is possible that task characteristics moderate the relationship between C-S balance and flow experience (e.g., Aubrey,

2017a, 2017b; Cho, 2018; Egbert, 2003). Furthermore, this study did not examine the possible roles of clear goals and immediate feedback despite these being regarded as flow antecedents in Csikszentmihalyi's (1975/2000, 1990) flow theory. Future research is encouraged to examine the interrelationship of different variables to flow experience in various study contexts. Finally, there were numerous missing values, which were found not to statistically affect the results of this study. It is challenging but important to understand why some participants failed to provide their responses at some time points.

This study has offered some important insights into future research and language education. Regarding implications for future research, more studies should employ a longitudinal research design (Jia et al., 2024) to theorize how flow experience contributes to language learning. Despite some longitudinal studies (Arai, 2025; Kirchoff, 2013), little is still known about how continuing to experience flow over time contributes to language learning. Future longitudinal research is encouraged to examine the relationship between flow experience and language learning outcomes. To this end, a longitudinal analysis using multilevel modeling has its strength in the understanding of both intra- and inter-individual changes in flow experience (Hiver & Al-Hoorie, 2020). Furthermore, the relationship between the change of flow experience and the development of other language proficiencies (e.g.,

reading) should be investigated in the future. As for pedagogical implications, this study found the effects of flow experience on oral proficiency development. The results in this study thus emphasized the importance of flow experience in the language learning classroom. To encourage flow, teachers need to understand individual learners, as the findings in this study suggested the importance of learner perspectives in flow experience. In particular, they should be aware of the importance of learner perceptions of C-S balance and provide due support and feedback to help learners experience flow. Measuring flow experience and its change over time may also deepen teachers' understanding of learners' subjective experiences as a conglomerate of cognition, motivation, and emotion (Dörnyei & Ushioda, 2011).

In conclusion, this study tracked EFL university students' flow experience in a conversation course over one academic semester. The findings supported flow theory in that (1) only learner-perceived C-S balance as a flow antecedent statistically predicted the U-shaped trajectory of flow experience as the best model; (2) flow experience predicted oral proficiency development as a flow consequence; and (3) flow antecedents (e.g., C-S balance) did not directly predict the flow consequence but indirectly through flow experience. This study encourages future longitudinal studies on flow experience trajectories and their effects on language learning.

### Acknowledgments

We would like to thank Professor Yasuyo Sawaki for her insightful comments on an earlier version of this paper. We are also grateful to Waseda University Academic Solutions Corporation for administrative support in conducting data collection. Our special thanks also go to the Tutorial English AI research team, including Fuma Kurata, Yoichi Matsuyama, Mao Saeki, Susumu Saito, Hiroaki Takatsu, Kotaro Takizawa, and Sadahiro Yoshikawa (in alphabetical order), for providing the study participants' LANGX Speaking test scores. This paper is based on results obtained from a project, JPNP20006 ("Online Language Learning AI Assistant that Grows with People"), subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

### Authors' Contributions

Yuya Arai conceptualized and designed the study, collected and interpreted the data, and wrote and revised the article. Ryuki Matsuura worked on data analysis and was involved in the writing of the article. Masaki Eguchi participated in data analysis and the revision of the manuscript. Shungo Suzuki participated in the design of the study and the revision of the manuscript. All authors read and approved the final manuscript.

### Ethics Approval & Consent to Participate

This study was approved by the Ethics Review Committee on Research with Human Subjects of Waseda University (approval no. 2024-199) on July 19, 2023. Respondents gave written informed consent for participation and publication and signature before starting study participation.

## Declaration of GenAI and AI-Assisted Technologies

As stated in the Method section, we used an AI-based speaking test called LANGX Speaking. However, we confirm that no part of the study conceptualization, original writing, analysis, or interpretation in this manuscript was generated using AI. Meanwhile, AI-assisted technologies were used to improve the clarity and flow of the manuscript's prose, to assist with language refinement, and for stylistic editing. We reviewed, verified, and take full responsibility for the content of the final manuscript.

## Funding

This paper is based on results obtained from project number JPNP20006 ("Online Language Learning AI Assistant that Grows with People") subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

## REFERENCES

- Abuhamdeh, S. (2020). Investigating the "flow" experience: Key conceptual and operational issues. *Frontiers in Psychology, 11*, 1–13. <https://doi.org/10.3389/fpsyg.2020.00158>
- Almukhailed, H., & King, J. (2024). Scaling the apex of engagement: Anxious language learners' positive flow experiences during speaking tasks. *System, 124*, Article 103373. <https://doi.org/10.1016/j.system.2024.103373>
- Amini, D., Ayari, S., & Amini, M. (2016). The effect of flow state on EFL learners' vocabulary learning. *International Journal of Multicultural and Multireligious Understanding, 3*(5), 9–18. <https://doi.org/10.18415/ijmmu.v3>
- Arai, Y. (2025). Exploring flow as the consequences of placement decisions in an EFL speaking course. *Language Testing in Asia, 15*, Article 13. <https://doi.org/10.1186/s40468-025-00347-y>
- Aubrey, S. (2017a). Inter-cultural contact and flow in a task-based Japanese EFL classroom. *Language Teaching Research, 21*(6), 717–734. <https://doi.org/10.1177/1362168816683563>
- Aubrey, S. (2017b). Measuring flow in the EFL classroom: Learners' perceptions of inter- and intra-cultural task-based interactions. *TESOL Quarterly, 51*(3), 661–692. <https://doi.org/10.1002/tesq.387>
- Barthelmäs, M., & Keller, J. (2021). Antecedents, boundary conditions and consequences of flow. In C. Peifer & S. Engeser (Eds.), *Advances in flow research* (2nd ed., pp. 71–107). Springer.
- Bartholomeyczik, K., Knierim, M. T., Weinhardt, C., Oettingen, G., & Ebner-Priemer, U. (2024). Capturing flow experiences in everyday life: A comparison of recall and momentary measurement. *Journal of Happiness Studies, 25*, Article 66. <https://doi.org/10.1007/s10902-024-00776-1>
- Baumann, N., Lürig, C., & Engeser, S. (2016). Flow and enjoyment beyond skill-demand balance: The role of game pacing curves and personality. *Motivation and Emotion, 40*, 507–519. <https://doi.org/10.1007/s11031-016-9549-7>
- Bricteux, C., Navarro, J., Ceja, L., & Fuerst, G. (2017). Interest as a moderator in the relationship between challenge/skills balance and flow at work: An analysis at within-individual level. *Journal of Happiness Studies, 18*(3), 861–880. <https://doi.org/10.1007/s10902-016-9755-8>
- Buuren, S. V., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- Ceja, L., & Navarro, J. (2012). 'Suddenly I get into the zone': Examining discontinuities and nonlinear changes in flow experiences at work. *Human Relations, 65*(9), 1101–1127. <https://doi.org/10.1177/0018726712447116>
- Cho, M. (2018). Task complexity and modality: Exploring learners' experience from the perspective of flow. *The Modern Language Journal, 102*(1), 162–180. <https://doi.org/10.1111/modl.12460>
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. <https://rm.coe.int/16802fc1bf>
- Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of Humanistic Psychology, 15*(3), 41–63. <https://doi.org/10.1177/002216787501500306>
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper and Row.

- Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety: Experiencing flow in work and play*. Jossey-Bass. (Original work published 1975).
- Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121–136. <https://doi.org/10.1080/15248371003699969>
- Czimmermann, É., & Piniel, K. (2016). Advanced language learners' experiences of flow in the Hungarian EFL classroom. In P. D. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive psychology in SLA* (pp. 193–214). Multilingual Matters.
- Dewaele, J.-M., & MacIntyre, P. D. (2024). “You can’t start a fire without a spark”. Enjoyment, anxiety, and the emergence of flow in foreign language classrooms. *Applied Linguistics Review, 15*(2), 403–426. <https://doi.org/10.1515/applirev-2021-0123>
- Dörnyei, Z., & Ushioda, E. (2011). *Teaching and researching motivation* (2nd ed.). Longman.
- Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal, 87*(4), 499–518. <https://doi.org/10.1111/1540-4781.00204>
- Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion, 32*, 158–172. <https://doi.org/10.1007/s11031-008-9102-4>
- Engeser, S., Rheinberg, F., Vollmeyer, R., & Bischoff, J. (2005). Motivation, Flow-Erleben und Lernleistung in universitären Lernsettings [Motivation, flow experience, and learning performance in university learning settings]. *Zeitschrift Fur Pädagogische Psychologie, 19*(3), 159–172. <https://doi.org/10.1024/1010-0652.19.3.159>
- Engeser, S., & Schiepe-Tiska, A. (2012). Historical lines and an overview of current research on flow. In S. Engeser (Ed.), *Advances in flow research* (pp. 1–22). Springer.
- Hiver, P., & Al-Hoorie, A. H. (2020). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.
- Hoffman, L. (2014). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.
- Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.
- Hox, J. J., van Buuren, S., & Jolani, S. (2016). Incomplete multilevel data: Problems and solutions. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 39–62). Information Age Publishing.
- 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 and Exercise Psychology, 24*(2), 133–150. <https://doi.org/10.1123/jsep.24.2.133>
- Jia, F., Meng, J., Ma, Y., & Mao, Y. (2024). Flow experience and self-efficacy in undergraduates' English learning: A weekly diary investigation with cross-lagged panel modeling approach. *System, 123*, Article 103312. <https://doi.org/10.1016/j.system.2024.103312>
- Kashimura, Y., & Nakamura, H. (2024). A scoping review of psychological flow in second language research. *HELES Journal, 23*, 20–35. [https://doi.org/10.24675/helesje.23.0\\_20](https://doi.org/10.24675/helesje.23.0_20)
- Kawabata, M., & Mallett, C. J. (2011). Flow experience in physical activity: Examination of the internal structure of flow from a process-related perspective. *Motivation and Emotion, 35*, 393–402. <https://doi.org/10.1007/s11031-011-9221-1>
- Kirchhoff, C. (2013). L2 extensive reading and flow: Clarifying the relationship. *Reading in a Foreign Language, 25*(2), 192–212. <https://nflrc.hawaii.edu/rfl/item/279>
- Lazarus, R. S. (2003). Does the positive psychology movement have legs? *Psychological Inquiry, 14*(2), 93–109. [https://doi.org/10.1207/S15327965PLI1402\\_02](https://doi.org/10.1207/S15327965PLI1402_02)
- Liu, H., & Song, X. (2021). Exploring “flow” in young Chinese EFL learners' online English learning activities. *System, 96*, Article 102425. <https://doi.org/10.1016/j.system.2020.102425>
- Nakamura, J., & Csikszentmihalyi, M. (2009). Flow theory and research. In S. J. Lopez & C. R. Snyder (Eds.), *The Oxford handbook of positive psychology* (2nd ed., pp. 195–206). Oxford University Press.

- Norworthy, C., Jackson, B., & Dimmock, J. A. (2021). Advancing our understanding of psychological flow: A scoping review of conceptualizations, measurements, and applications. *Psychological Bulletin*, 147(8), 806–827. <https://doi.org/10.1037/bul0000337>
- Payant, C., & Zuniga, M. (2022). Learners' flow experience during peer revision in a virtual writing course during the global pandemic. *System*, 105, Article 102715. <https://doi.org/10.1016/j.system.2021.102715>
- Peifer, C., & Engeser, S. (2021). Theoretical integration and future lines of flow research. In C. Peifer & S. Engeser (Eds.), *Advances in flow research* (2nd ed., pp. 417–439). Springer.
- Piniel, K., & Ritecz, F. (2022). Flow experiences during speaking tasks in the Hungarian English as a foreign language high school classroom. In G. Tanko & A. M. Wind (Eds.), *DEAL2022: Challenges and opportunities in contemporary English applied linguistics* (pp. 197–224). Eötvös Loránd University.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878–912. <https://doi.org/10.1111/lang.12079>
- Rheinberg, F., & Engeser, S. (2018). Intrinsic motivation and flow. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and action* (3rd ed.; pp. 579–622). Springer.
- Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). Die Erfassung des Flow-Erlebens [The assessment of flow experience]. In J. Stiensmeier-Pelster & F. Rheinberg (Eds.), *Diagnostik von Motivation und Selbstkonzept* (pp. 261–279). Hogrefe.
- Rossin, D., Ro, Y. K., Klein, B. D., & Guo, Y. M. (2008). The effects of flow on learning outcomes in an online information management course. *Journal of Information Systems Education*, 20(1), 87–98. <https://aisel.aisnet.org/jise/vol20/iss1/9>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley.
- Saeki, M., Takatsu, H., Kurata, F., Suzuki, S., Eguchi, M., Matsuura, R., Takizawa, K., Yoshikawa, S., & Matsuyama, Y. (2024). *InteLLA: Intelligent language learning assistant for assessing language proficiency through interviews and roleplays*. Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue. <https://aclanthology.org/2024.sigdial-1.34>
- Schüler, J. (2007). Arousal of flow experience in a learning setting and its effects on exam performance and affect. *Zeitschrift für Pädagogische Psychologie*, 21(3/4), 217–227. <https://doi.org/10.1024/1010-0652.21.3.217>
- Shin, N. (2006). Online learner's 'flow' experience: An empirical study. *British Journal of Educational Technology*, 37(5), 705–720. <https://doi.org/10.1111/j.1467-8535.2006.00641.x>
- Strauss, S. (1982). *U-shaped behavioral growth*. Academic Press.
- Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, 46(2), 195–207. <https://doi.org/10.1017/S0261444811000486>
- Takatsu, H., Suzuki, S., Eguchi, M., Matsuura, R., Saeki, M., & Matsuyama, Y. (2026). Gnowsisi: Multimodal multitask learning for oral proficiency assessments. *Computer Speech & Language*, 95, Article 101860. <https://doi.org/10.1016/j.csl.2025.101860>
- Waninge, F. (2015). Motivation, emotion and cognition: Attractor states in the classroom. In Z. Dörnyei, P. D. MacIntyre, & A. Henry (Eds.), *Motivational dynamics in language learning* (pp. 195–213). Multilingual Matters.

**APPENDIX A. Table 6.** *Summary of Regression Models for Time-Variant and Invariant Variables*

Model	M2-1: Baseline			M2-2: + C-S Balance			M2-3: + Course Level + Placement Accuracy + Pre-test Score			M2-4: + Post-test Score + Test Score Change		
	Estimates	95% CI	<i>p</i>	Estimates	95% CI	<i>p</i>	Estimates	95% CI	<i>p</i>	Estimates	95% CI	<i>p</i>
Intercept	44.033***	[43.112, 44.954]	<0.001	43.850***	[42.928, 44.772]	<0.001	41.923***	[38.510, 45.337]	<0.001	40.363***	[36.323, 44.402]	<0.001
Time	-1.039†	[-2.135, 0.056]	0.063	-0.946†	[-0.435, 0.897]	0.078	-0.950†	[-2.002, 0.103]	0.077	-0.956†	[-2.008, 0.097]	0.075
Time <sup>2</sup>	0.408*	[0.056, 0.761]	0.023	0.388*	[0.050, 0.727]	0.025	0.389*	[0.050, 0.728]	0.025	0.391*	[0.052, 0.729]	0.024
<b>C-S balance</b>												
C=S vs. C<S				0.587	[-0.874, 2.048]	0.432	0.513	[-0.966, 1.991]	0.497	0.410	[-1.070, 1.890]	0.588
C=S vs. C>S				-5.065***	[-6.052, -4.078]	<0.001	-5.092***	[-6.087, -4.097]	<0.001	-5.005***	[-6.001, -4.008]	<0.001
<b>Course Level</b>												
Beginner vs. Basic							1.476	[-2.532, 5.485]	0.471	2.118	[-1.951, 9.188]	0.308
Basic vs. Intermediate							-0.207	[-2.011, 1.596]	0.822	0.190	[-1.640, 2.019]	0.839
Intermediate vs. Upper Intermediate							-0.857	[-3.340, 1.625]	0.499	-0.558	[-3.087, 1.970]	0.654
Upper Intermediate vs. Advanced							1.221	[-1.487, 3.929]	0.377	1.189	[-1.531, 3.909]	0.392
<b>Placement Accuracy</b>												
Accurate Placement vs. False Positive							-1.731	[-5.371, 1.909]	0.356	-1.685	[-5.229, 1.859]	0.355
Accurate Placement vs. False Negative							0.563	[-1.998, 3.125]	0.668	-0.246	[-3.319, 2.826]	0.876
<b>LANGX Speaking Test Score</b>												
Pre-test Score							0.644	[-0.374, 1.661]	0.216	0.249	[-2.360, 2.857]	0.852
Post-test Score										0.848	[-1.900, 3.596]	0.546
Pre=Post vs. Pre>Post										1.373	[-2.739, 5.485]	0.516
Pre=Post vs. Pre<Post										0.617	[-2.716, 3.950]	0.717
Mean AIC ± SD	7,499.266 ± 22.904			7,388.734 ± 22.546			7,390.919 ± 21.863			7,391.553 ± 23.319		
Mean R <sup>2</sup> ± SD	0.436 ± 0.011			0.477 ± 0.011			0.477 ± 0.011			0.480 ± 0.011		

Notes. †*p* < .10, \**p* < .05, \*\**p* < .01, \*\*\**p* < .001

**APPENDIX B. Table 7. Summary of Multilevel Cumulative Models**

Model	M3-1: Intercept		M3-2: Unconditional (+ Time)		M3-3: Conditional (+ Placement Accuracy + Flow Score + C-S balance)					
	Estimates	<i>p</i>	Estimates	<i>p</i>	OR	95% CI for OR	Estimates	<i>p</i>	OR	95% CI for OR
Threshold										
A1-A2	-5.536		-6.535				1.434			
A2-B1	-2.068		-1.744				6.194			
B1-B2	0.749		2.223				10.168			
B2-C	3.201		5.718				13.676			
Time			2.360***	<.001	10.589	[5.043, 22.234]	2.358***	<.001	10.567	[5.117, 21.822]
Flow Score										
Mean Flow Score							0.179**	0.004	1.196	[1.064, 1.344]
Flow Score Change (1-2)							-0.059	0.171	0.942	[0.866, 1.026]
Flow Score Change (2-3)							-0.033	0.506	0.967	[0.877, 1.067]
Flow Score Change (3-4)							-0.052	0.269	0.949	[0.867, 1.040]
C-S balance										
Number of C<S							0.619†	0.066	1.857	[0.966, 3.570]
Number of C>S							0.346	0.178	1.414	[0.856, 2.335]
Placement Accuracy										
Accurate Placement vs. False Positive							2.099†	0.060	8.161	[0.936, 71.187]
Accurate Placement vs. False Negative							0.224	0.858	1.252	[0.108, 14.533]
Mean AIC ± SD	1,800.978 ± 21.591				1,661.474 ± 34.924				1,649.575 ± 36.609	
Mean R <sup>2</sup> ± SD	0.744 ± 0.026				0.868 ± 0.025				0.868 ± 0.025	

Notes. †*p*<.10, \* *p*<.05, \*\* *p*<.01, \*\*\* *p*<.001. OR indicates odds ratio.