Language Teaching (2026), 1–20
doi:10.1017/S0261444826101141

CAMBRIDGE UNIVERSITY PRESS

REPLICATION RESEARCH

Revisiting the intensity and perceived quality of L2 engagement scale in a different context: An approximate replication of Teravainen-Goff (2023)

Yohei Nakanishi[1] and Osamu Takeuchi[2]

[1]Graduate School of Foreign Language Education and Research, Kansai University, Osaka, Japan and [2]Faculty of Foreign Language Studies, Kansai University, Osaka, Japan
Corresponding author: Yohei Nakanishi; Email: yoheinakanishi23@gmail.com

(Received 28 February 2025; accepted 4 January 2026)

Abstract
This study conducted an approximate replication of Teravainen-Goff (2023) to validate the Intensity and Perceived Quality of Engagement Scale for university students in the Japanese EFL context. Teravainen-Goff (2023) developed this scale based on an action-oriented definition of engagement and proposed a novel approach to measuring engagement among secondary school language learners in the UK. The study identified an 18-item, five-factor structure from a pool of 36 items through exploratory factor analysis (EFA). In this replication, we examined the validity and reliability of Teravainen-Goff's scale in a different context, focusing on the replicability of the EFA results. We undertook this replication because engagement is context-dependent and EFA results can vary across samples. We compared the factorial structure with that of the initial study while modifying the target language and participant demographic. Results revealed a 22-item, six-factor structure with good fit. Although the same underlying factors emerged, several notable differences were observed. This approximate replication provided stronger evidence for the psychometric properties of the scale in a new context. Transparent documentation of modifications to the initial study and systematic comparison offered a promising approach to building robust evidence for engagement research and improving the rigour of questionnaire-based research overall.

Keywords: engagement; exploratory factor analysis; replication study

1. Introduction

The concept of engagement has garnered increasing scholarly attention in applied linguistics over the past few years. Recent publications (Hiver et al., 2021; Mercer & Dörnyei, 2020) and systematic reviews (Hiver et al., 2024) have highlighted this trend. Engagement has been regarded as 'the holy grail of learning' (Sinatra et al., 2015, p. 1). Research in educational psychology has demonstrated that higher levels of student engagement contribute to improved academic performance and student well-being (Reschly & Christenson, 2022). The field of second language acquisition (SLA) has also recognized the potential benefits of engagement in L2 reading comprehension (Khajavy, 2021) and L2 achievement (Eerdemutu et al., 2024). Furthermore, lower levels of L2 engagement have been found to negatively predict L2 writing procrastination (Zhou & Hiver, 2022). Despite its promising potential for holistically capturing the student learning process, the concept of engagement faces two key issues that must be addressed. The first concerns the overlapping nature of engagement (Hiver et al.,

2024). Engagement is multifaceted with behavioural, cognitive, and emotional engagement as its core components (Fredricks et al., 2004). Additionally, social engagement (Philp & Duchesne, 2016) and agentic engagement (Reeve, 2013) have been identified as important constructs. It is true that the multifaceted nature of engagement allows researchers to examine the learning process from a holistic perspective; however, it also implies conceptual overlap that ultimately leads to inconsistencies in the literature. Reschly and Christenson (2022) described this problem as the 'jingle-jangle issue' (p. 4), indicating that the same term may refer to a different construct of engagement. Conversely, different terms might describe the same construct. Hiver et al. (2024) also pointed out in their systematic review that fewer than 35% of the studies examined clearly defined engagement. The second issue relates to the blurred boundaries between engagement and associated concepts, particularly motivation (Teravainen-Goff, 2022; Vo, 2024). Given the growing body of research on the relationship between motivation and engagement (Khajavy, 2021; Noels et al., 2019; Oga-Baldwin et al., 2017), it is crucial to distinguish these two concepts clearly. According to Hiver et al. (2024), 'action is key in distinguishing engagement from motivation. Motivation represents initial intention and engagement is the subsequent action' (p. 23). In the field of educational psychology, according to Martin et al. (2017), 'motivation is defined as the inclination, energy, emotion, and drive relevant to learning, working effectively, and achieving; engagement is defined as the behaviors that reflect this inclination, energy, emotion, and drive' (p. 150). These notions highlight a fundamental difference between motivation and engagement.

SLA researchers have made significant efforts to address these two issues, particularly in developing valid and reliable measurement tools through questionnaires. For example, Oga-Baldwin and Nakata (2017) and Hiver et al. (2020) refined engagement scales adapted from general education, focusing equally on behavioural, emotional, and cognitive engagement. In contrast, the study conducted by Teravainen-Goff (2023), replicated in the current study, approached the aforementioned issue from a different perspective. Teravainen-Goff (2023) defined the concept of engagement as follows: 'engagement is an ultimately behavioral concept with underlying cognitive and affective facets' (p. 2). This operationalization aligns with the action-oriented definition of engagement as 'the amount (quantity) and type (quality) of learners' active participation and involvement in a language learning task or activity' (Hiver et al., 2024, p. 2). Teravainen-Goff's (2023) pioneering research addressed the issues surrounding the construct of engagement through a novel methodological approach based on a solid theoretical foundation, making this initial study worthy of replication for the further development of engagement research in SLA. Teravainen-Goff (2023) also emphasized the need for further investigation, stating that 'further research is therefore needed to test the Intensity and Perceived Quality of L2 Engagement Questionnaire in various language learning contexts' (p. 9). In response to this call, the present study aims to replicate Teravainen-Goff's (2023) study in the Japanese EFL context to contribute to a deeper understanding of engagement through meaningful comparisons with the initial study.

The validation of Teravainen-Goff's (2023) study through replication was also motivated by concerns raised by Sudina (2021) about questionnaire quality in applied linguistics. Among several recommendations for quality improvement, Sudina emphasized the need for transparency in questionnaire development, specifically stating that researchers should report 'whether scale items were adopted or adapted, and, in the case of the latter, what specific modifications (i.e. both the amount and type) were made' (p. 1183). Furthermore, Manapat et al. (2025) reported the replicability of exploratory factor analysis (EFA) in questionnaire development in the field of psychology, noting that 'although this is a relatively new area of methodological focus, there is empirical evidence in the factor analytic literature that can be considered issues of replication' (p. 1). Given these concerns about questionnaire quality (Sudina, 2021) and EFA replicability (Manapat et al., 2025), we determined that a replication framework best served our purpose, because replication research requires strict adherence to initial study protocols, yet it allows researchers to examine target variables systematically (Porte & McManus, 2019). We used the term 'initial' rather than 'original' to describe

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

the work by Teravainen-Goff (2023), which was based on the discussion by Marsden et al. (2018). Therefore, the present study leverages a replication framework to report comprehensive information (e.g. variable modification and psychometric properties) and revisit the validity of this newly developed L2 engagement scale (Teravainen-Goff, 2023) in the Japanese EFL context. We believe that this study contributes not only to deepening our understanding of engagement but also demonstrates the importance of questionnaire validation, particularly exploratory factor analysis (EFA), through a replication framework in applied linguistics (McManus, 2024).

2. Background

2.1. Literature review

Extensive research has been conducted on motivation (Boo et al., 2015); however, motivation alone cannot fully explain the complexities of the learning process (Teravainen-Goff, 2022). Therefore, a more comprehensive construct was required to capture the learning process. Scholarly attention consequently shifted to engagement, which offers a holistic perspective on the learning process (Hiver et al., 2024).

Engagement is a multifaceted construct encompassing behavioural, cognitive, and affective dimensions (Fredricks et al., 2004). Behavioural engagement refers to 'students' continuous performance in learning as determined by their expenditure of efforts on learning tasks, the quality of their participation, and their degree of active involvement in the learning process' (Sang & Hiver, 2021, p. 21). Cognitive engagement involves mental efforts such as connecting new concepts to existing knowledge (Hiver et al., 2021; Sang & Hiver, 2021). Affective engagement manifests through emotional experiences during the learning process (Sang & Hiver, 2021). Two additional dimensions have also been proposed: social engagement, which involves the collaborative aspects of learning with teachers and peers (Philp & Duchesne, 2016), and agentic engagement, which describes learners' active contributions to their learning environment (Reeve, 2013). The multifaceted nature of engagement provides a holistic view of the learning process; however, this multidimensionality necessitates a clearer operationalization of engagement within its dimensions, such as the behavioural and cognitive dimensions, and in relation to adjacent constructs, for example, motivation. A systematic review by Hiver et al. (2024) also revealed a notable gap in SLA research, reporting that fewer than 35% of the studies examined demonstrated a clear operationalization of engagement. This lack of conceptual clarity underscores the fundamental weakness in the current theoretical development of engagement.

The current lack of conceptual clarity could be a result of the relatively short history of engagement research in SLA. Nonetheless, SLA researchers have substantially tried to deepen our understanding of this construct. A crucial step in this line of research was the development of valid and reliable measurement tools to evaluate learner engagement based on clear operationalization. Among the various methods available to measure engagement, questionnaires are the most widely adopted approach (Fredricks, 2022). These questionnaires are generally categorized as either general engagement scales or domain-specific scales (Fredricks, 2022). Within the field of SLA, there is a pressing need for domain-specific engagement scales that address the unique aspects of language learning. Recent efforts to develop such scales have shown promising progress. For example, Eerdemutu et al. (2024) attempted to develop a valid, reliable, and domain-specific engagement scale based on Hiver et al.'s (2020) study, using rigorous analytical procedures for high school students in the Chinese EFL context and university students studying Japanese in China. Similarly, using a mixed-method approach, Guo et al. (2023) developed a foreign language classroom engagement scale in the Chinese EFL context. Teravainen-Goff's (2023) work replicated in this study aligns with this trend of developing domain-specific L2 engagement scales. Among the reviewed scales (Eerdemutu et al., 2024; Guo et al., 2023; Teravainen-Goff, 2023), we selected Teravainen-Goff's (2023) study for replication because of its innovative methodology, which is based on a strong theoretical foundation. A summary of the initial study and our rationale for this selection are presented in the following section.

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

2.2. Initial study: Teravainen-Goff (2023)

The initial study began by examining the literature on engagement, focusing on conceptual ambiguities in the previous definition. Building on a systematic review conducted by Hiver et al. (2024), this study conceptualized engagement as 'quantity and quality of learners' active participation' (Teravainen-Goff, 2023, p. 3), developing a new instrument, the 'Intensity and Perceived Quality of L2 Engagement' questionnaire. A total of 378 secondary school language learners (mean age = 13 years) from England participated in the study. Of these, 77.5% of students learned French, 46.3% learned German, and 18.3% learned Spanish as a foreign language in compulsory foreign language classes. Additionally, 40.1% of participants were learning a foreign language for a General Certificate of Secondary Education (GCSE) qualification and 1.9% for an A-level qualification.

The questionnaire development process followed a rigorous methodology. First, the researcher conducted a comprehensive review of engagement studies in both general education and applied linguistics. Second, the study incorporated the findings from a qualitative interview study (Teravainen-Goff, 2022), in which learners' and teachers' perceptions of factors influencing engagement in language learning had been explored. The researcher then carefully screened relevant items to avoid measuring pretense engagement or unintentionally tapping into the concept of motivation. Following these procedures, 60 items were compiled, with wording adapted from Hiver et al. (2020). The questionnaire comprised three constructs (intensity, perceived usefulness, and satisfaction of engagement) and four aspects (teachers, peers, activities, and teaching content). The three constructs included these four aspects, resulting in 12 distinct factors in the initial structure (e.g. intensity of engagement with peers, satisfaction of engagement with teachers, and usefulness of engagement with activities). Subsequently, the items underwent expert review by four researchers in this field, leading to a refined 36-item engagement questionnaire. Finally, data collected from the participants using this 36-item scale were analyzed. Exploratory factor analysis (EFA) with oblique rotation revealed six latent factors. The researcher proposed a five-factor model with 18 items through subsequent expert evaluation and reliability testing. Confirmatory factor analysis (CFA) demonstrated good fit indices; χ^2 (df = 125) 337.4, $p < .001$; comparative fit index (CFI) = .938; Tucker-Lewis index (TLI) = .916; root mean square error of approximation (RMSEA) (90%CI) = .067 (.059; .076).

2.3. Motivation for this replication study

This approximate replication study was driven by the increasing scholarly attention to L2 engagement and the need to address its conceptual challenges. Following Porte and McManus's (2019) framework, we classified this as an approximate replication study, which investigated 'the effects of two variables on outcomes' (p. 78).

The decision to replicate Teravainen-Goff's (2023) study is justified by its significant contribution to engagement research through a novel yet theoretically sound approach. This initial work (Teravainen-Goff, 2023) was recently published in *System* and demonstrates its value through its novel but theoretically robust approach. To our knowledge, this study was the first to reorganize related engagement literature, operationalize the concept of engagement with a focus on the behavioural construct (Hiver et al., 2024), and subsequently develop an L2 engagement scale. As discussed above, this newly developed scale was created based on a theory-supported operationalization that reorganized the constructs of engagement and clearly distinguished between engagement and motivation. Therefore, this novel but theoretically sound approach has the potential to address key concerns about conceptual ambiguities raised by several researchers (Martin et al., 2017; Nagy et al., 2022; Vo, 2024). Given that the initial study was conducted in a UK context with a sample of 378 secondary school students, this replication study offers additional insights into the concept of engagement and the validity and reliability of the newly developed scale in a different context. To facilitate meaningful comparisons, this approximate replication study modified the target language and participant demographics. This replication set out to confirm to what extent the initial study's findings hold in a new context

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

Table 1. Comparative summary of participants from the initial and present studies

	Initial study	Present study
Total number of participants	378	438
Context	UK	Japan
institution	Five state secondary schools	Six universities
Gender	Female: 75.1%, male: 17.5%, non-binary: 1.9%, not disclosed: 5.6%	Female: 55%, male: 43%, not disclosed: 2%
Age or grade	Mean, 13 years	First year: 314, second year: 111, third year: 6, fourth year: 6, others: 1
L1	Not mentioned	Japanese: 96%, others such as Korean: 4%
L2 (foreign language)	French, German, Spanish	English
Language test	40.1% studying for GCSE qualification	36% trying to take English tests

when these two key variables are changed. By doing so, this study enables us to examine whether the action-oriented engagement scale functions consistently across different learner populations and educational contexts. This, in turn, provides additional support for the scale's methodological foundation and contributes to advancing theoretical discussions on the multifaceted and context-dependent nature of engagement, a point also emphasized in the initial study by Teravainen-Goff (2023).

The first modification involved changing the target language variable from a UK context of learning various foreign languages to a Japanese context of learning English as a foreign language. The initial study was conducted in the UK, where the participants were learning foreign languages other than English such as French and German. In response to the following call in the initial study, 'it would be interesting to investigate whether the target language may have an impact on the findings, especially given that English as a Foreign Language (EFL) was not considered in this study' (Teravainen-Goff, 2023, p. 9), we conducted this replication study in Japan, where students learn English from primary school to university as a foreign language.

The second modification involved the demographic variables, particularly age. The initial study focused on UK secondary school students (mean age = 13 years), whereas our study targeted Japanese university students (typically ages 18–22; see details in Table 1). This age difference is considered a meaningful modification because previous research into motivation has shown that motivational factors and models vary across different age groups (Kormos & Csizer, 2008; Papi & Teimouri, 2012). Therefore, this replication study also modified the age of participants to examine whether the scale functions similarly across different developmental stages. This approach helps us gain a clearer understanding of the extent to which Teravainen-Goff's scale reflects age-specific engagement patterns or more universal features of engagement. This comparison also allows us to determine whether this scale is sensitive to developmental differences or whether it captures stable features of engagement that remain consistent across age groups. As regards other variables, we strictly adhered to the methodology of the initial study. Recognizing that scale translation could influence the findings (Dörnyei & Dewaele, 2023), we implemented a comprehensive translation process. The process involved working with a professional translation company to conduct back-translation and certify minimal semantic differences between the English and Japanese versions. Despite these rigorous measures, translation effects cannot be eliminated; thus, acknowledging this as a minor modification facilitates the careful interpretation of our findings.

3. The present study

This study was designed to address three out of the four aims of the initial study. Our modifications to the initial study aim are presented below in italics. We excluded the first aim of the initial study,

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

'provide a theoretical conceptualization of how the intensity and perceived quality of L2 engagement could be measured', because the present study was conducted based on Teravainen-Goff's (2023) theoretical conceptualization of engagement. Given the comparative nature of this study, we added aim (d): 'Investigate the similarities and differences in engagement between the initial and present studies'.

(a) Propose and test a structure for *the Japanese version* of the Intensity and Perceived Quality of L2 Engagement Questionnaire
(b) Examine the reliability and validity of the questionnaire in the *Japanese EFL learning context*
(c) Provide insight into the intensity and perceived quality of engagement for a sample of *EFL learners in universities* in *Japan*
(d) Investigate the similarities and differences in engagement between the initial and present studies

3.1. Method

All relevant and detailed information is shared on the OSF via this link (https://osf.io/an3ze/?view_only=064e30a1c4d4487dabde9496816f0d70), following our commitment to open science principles (Al-Hoorie et al., 2024; Liu, 2023; Liu et al., 2022) and facilitating future replication of this study. Following Soderberg's (2018) recommendation, we present the files below with descriptive names (e.g. file name: 'example') to facilitate easy access to relevant information.

3.2. Participants

In this study, we used convenience sampling to recruit participants, making every effort to include individuals from various academic backgrounds and different regions across Japan to enhance the generalizability of the findings. This study primarily used EFA and CFA; therefore, power analyses were not conducted to determine the ideal sample size. We endeavored to gather data from more than 300 participants, which was comparable to the initial study's sample size of 378. Given that learners' engagement can be influenced by teachers (Teravainen-Goff, 2023), we requested the participation of as many English teachers as possible to minimize this influence. Consequently, eight English teachers cooperated with this study.

Table 1 presents a comparative summary of the participants of this study and those of the initial study. A total of 438 participants were involved in this study, with two excluded for not having provided their consent. On the other hand, 378 participants joined the initial study. Participants for this study were recruited from six universities located in Eastern and Western Japan. The participants were from diverse academic fields, including foreign language, humanities, law, social welfare, education, psychology, health science, sociology, law and economics, textile science and technology, societal safety sciences, economics, engineering, international and English interdisciplinary studies, and international studies, whereas the initial study recruited participants from five secondary schools in the UK.

In terms of the gender distribution, the participants in this study showed a balanced distribution comprising 188 male and 240 female participants, with ten participants preferring not to disclose their gender. This contrasts with the initial study, which had mostly female participants (75.1%). The age groups also differed. In this study, 314 participants were in the first year, 111 were in the second year, six were in the third year, six were in the fourth year, and one classified as 'others'. Participants typically ranged from 18 to 22 years old within the Japanese education system. In contrast, the initial study involved participants with a mean age of 13 years. Regarding the target language, the participants in the initial study were learning foreign languages such as French, German, and Spanish, while participants in this study were learning English as a foreign language. All the participants in this study were enrolled in English classes at their respective universities. In terms of participants' first language,

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

96% of the participants in this study identified Japanese as their first language, which was not mentioned in the initial study. Additionally, 36% of participants in this study reported taking external English proficiency tests, such as the TOEIC, in addition to their regular English coursework. This is compared to 40.1% of UK students who are studying for GCSE qualifications.

3.3. Instruments

The questionnaire was carefully created through a rigorous procedure because the wording in a cross-linguistic context could influence the results (Dörnyei & Dewaele, 2023). Initially, the first author translated 36 items of the initial scale into Japanese, ensuring that the meaning of each word was maintained and that the wording retained the underlying action-focused meanings. Subsequently, two researchers in applied linguistics checked the wording for clarity and consistency. Based on their recommendations, the phrases of items were refined to enhance participants' comprehension of each item. To ensure cross-linguistic equivalence, Ulatus (https://www.ulatus.com/), an ISO-certified translation service provider (ISO17100), conducted back-translation. Minor refinements were made to improve clarity based on their suggestions. During the back-translation process, the company reported three minor discrepancies between the original English and the back-translated English versions, the details of which are shared below for transparency. The translation company certified that these variations maintained content equivalence. The certification documentation was uploaded to the OSF platform for transparency (see file name: 'certification').

(a) Original: 'I usually pay attention to what my teacher says'; back-translation: 'I usually listen carefully to my English teacher'
(b) Original: 'I usually ask my teacher questions when something is not clear'; back-translation: 'I usually ask the teacher questions during class if I do not understand something'
(c) Original: 'There is usually a good mix of different types of activities'; back-translation: 'There is a good balance of different types of activities incorporated in the regular classes'

Finally, the Japanese-translated version of this instrument was piloted with a sample of 37 university students. The analysis revealed high-reliability scores: McDonald's $\omega = 0.964$ [CI: 0.948, 0.978], Cronbach's $\alpha = 0.963$ [CI: 0.942, 0.978]. Based on these sound psychometric properties, we proceeded to use the Japanese version of the Intensity and Perceived Quality of L2 Engagement Questionnaire for the main study. The instrument used in this study is available on OSF (file name: 'instrument').

3.4. Data collection and analysis

Data were collected between June and July 2024. We obtained approval from our institutional ethics review board before data collection. We followed the data analysis procedures of the initial study as rigorously as possible. While the initial study used SPSS, we employed JASP (version 0.18), a free, open-source software package, for our statistical analyses, having confirmed beforehand that JASP could perform the same statistical analyses as SPSS. Prior to the primary analysis, we checked the assumptions for EFA. First, we assessed sample adequacy using the Kaiser-Meyer-Olkin (KMO) test. Second, we examined the correlation matrix for both multicollinearity (correlations > 0.9) and insufficient correlations (correlations < 0.3). Third, we determined the number of factors by examining the eigenvalues greater than one and inspecting the scree plot. Following the initial study, we retained items with factor loadings above 0.3. Finally, we conducted EFA using principal axis factor analysis with oblique rotation to examine the factor structure of the scale. Given that 'EFA is an inherently subjective process requiring a series of researcher judgments' (Plonsky & Gonulal, 2015, p. 19) and recognizing the critical need for methodological transparency, comprehensive documentation of every step of the EFA procedures is presented in a supplementary file available on OSF. Following EFA, we conducted CFA using maximum likelihood estimation. We assessed model fit using the same

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

criteria in the initial study: CFI > 0.9, TLI > 0.9, and RMSEA between 0.05 and 0.08. Within the CFA framework, we examined both the relationships between items and their intended factors, and the correlations among factors. We acknowledge that using the same dataset for both EFA and CFA is not ideal for inflated model fit values (Fokkema & Greiff, 2017). However, following van Prooijen and van der Kloot (2001) reasoning, we used the CFA fit indices as a reference: 'If a good fit is questionable when the factor structure is confirmatively tested on the same data, we cannot expect that a test of the factor structure in a confirmative follow-up study, that is, on different data, will lead to a good fit' (p. 790).

4. Results

Table 2 summarizes the results for each item. All items showed skewness and kurtosis values within the range ±2, indicating normal distribution (Hair et al., 2022). Before conducting EFA, we checked several assumptions. The KMO test and Bartlett's test of sphericity were conducted. The KMO test yielded an overall score of 0.939, with individual item scores ranging from 0.701 to 0.974, confirming adequate sampling (Hair et al., 2019). Bartlett's test of sphericity recorded $p < .001$, indicating that there were sufficient correlations (Hair et al., 2019). Subsequently, the correlation matrix was checked to minimize the multicollinearity issue (> .90) and eliminate the weak correlations with other items (< .30). Based on these criteria, Item 21 (without > 0.3 correlations with other items) was removed from the initial analysis. The complete correlation matrix is available on OSF (see file name: 'correlation matrix').

After checking the assumption, the main EFA was conducted. Initially, the number of factors to be retained was determined based on an eigenvalue greater than one (see Table 3) and a scree plot (see Figure 1). The results indicate that the seven-factor solution explained 59% of the variance. Table 3 lists the factor loadings for each item in descending order of value.

After verifying all assumptions for EFA, we conducted our analysis following the same procedure as Teravainen-Goff (2023). The cut-off value for retaining the item was set to be greater than 0.3. Consequently, four items (10, 11, 24, and 26) were removed. Next, Item 2 emerged as the only item loaded onto Factor 7 (see file name: 'EFA first stage'). In addition, the proportion of variance explained (PVE) of Factor 7 was 2.6%. Thus, Item 2 was removed in this stage. The following analysis indicated that Item 29 had cross-loading onto Factor 3 and Factor 4 (see file name 'EFA second stage'); therefore, this item was removed. The above-mentioned procedure was the pre-set protocol mentioned in the data analysis section. The aim of the initial study by Teravainen-Goff (2023) was to develop a questionnaire with reasonable numbers of items for each factor, with the author having mentioned that 'this was considered important to ensure the final questionnaire does not have too many items and thus will not take too long to complete in the classroom' (p. 6). Following the same procedure as the initial study, further examination was conducted to refine the questionnaire. The initial criteria for factor loading to retain the item were set to be > .30. This cut-off criterion was raised to > .40. Consequently, Item 12 and Item 28 were deleted (see file name 'EFA third stage'). At this stage, the number of items for Factor 1 to Factor 6 was 6, 6, 6, 4, 3, and 2, respectively (see file name; 'EFA fourth stage'). To develop a smaller set of items in each scale and with the initial study having a maximum of four items for each factor, Item 22 and Item 23, having the lowest loadings were removed. Further, Item 4 and Item 8, with the lowest loadings (see file name: 'EFA fifth stage'), were also removed. For transparency, both items had factor loadings of 0.580. These values were high; however, examining the items' contents revealed that they conveyed nearly identical meanings as other items within the same factor. Therefore, we concluded that these two items could be removed without losing the essential meaning of the intended factor. Lastly, Items 14 and 13 were considered (see file name: 'EFA sixth stage'). At this point, the target factor contained six items, including Items 14 and 13. Following our protocol of maintaining a maximum of four items per factor, we removed Item 14 because it had the lowest factor loading. Item 13, despite having the second-lowest loading, displayed a strong factor

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

Table 2. Descriptive statistics for all the items

Item	Mean	Std. Deviation	Variance	Skewness	Kurtosis
1	3.98	0.907	0.823	−0.754	0.243
2	3.03	1.191	1.420	0.025	−0.908
3	3.93	0.984	0.968	−0.835	0.252
4	3.93	0.938	0.880	−0.721	0.180
5	3.97	0.945	0.893	−0.878	0.643
6	4.02	0.950	0.903	−0.998	0.849
7	4.04	0.914	0.836	−0.792	0.216
8	4.04	0.892	0.796	−0.740	0.320
9	3.70	1.023	1.047	−0.478	−0.209
10	3.98	0.970	0.941	−0.913	0.541
11	4.09	0.894	0.800	−0.849	0.328
12	3.17	1.172	1.374	−0.151	−0.821
13	3.93	0.967	0.935	−0.745	0.126
14	4.22	0.883	0.780	−1.194	1.256
15	4.09	0.953	0.908	−0.915	0.306
16	3.90	1.032	1.064	−0.694	−0.131
17	3.82	1.101	1.212	−0.650	−0.395
18	4.00	0.942	0.887	−0.839	0.497
19	4.09	0.982	0.964	−0.897	0.040
20	3.97	0.908	0.825	−0.709	0.070
21	3.15	1.129	1.275	−0.200	−0.715
22	3.88	0.897	0.804	−0.657	0.345
23	3.72	0.980	0.961	−0.565	0.017
24	3.92	0.937	0.878	−0.696	0.147
25	3.48	1.187	1.410	−0.526	−0.603
26	3.88	0.872	0.760	−0.531	0.000
27	3.68	1.068	1.142	−0.596	−0.370
28	3.92	0.933	0.871	−0.615	−0.124
29	3.77	1.039	1.079	−0.578	−0.370
30	3.25	1.168	1.363	−0.121	−0.933
31	3.75	0.923	0.853	−0.536	0.149
32	3.91	0.917	0.841	−0.823	0.663
33	3.54	1.151	1.324	−0.586	−0.418
34	3.66	0.972	0.944	−0.413	−0.347
35	3.46	1.116	1.244	−0.514	−0.397
36	3.60	0.957	0.916	−0.429	−0.240

Note: $N = 427$. Eleven participants were excluded from the original sample of 438 due to potentially careless responding (i.e. selecting the same response option across all items).

loading (0.708), making its removal challenging to justify. Given that construct validity was prioritized over the number of items, we decided to retain Item 13. The final factor structure of the scale is presented in Table 4.

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

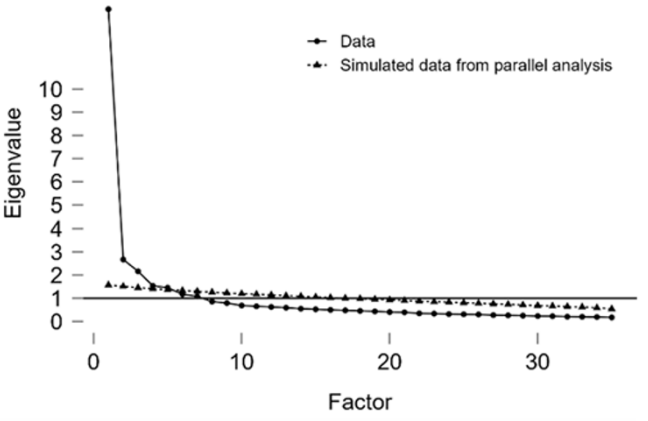

Figure 1. Scree plot.

The factors were named according to the initial study, except for Factor 6. Factor 1, containing items related to teacher support (e.g. 'My teacher usually helps me to succeed in my learning'), was named 'perceived quality of engagement with the teacher'. Factor 2, comprising items about peer relationships (e.g. 'I usually feel that I learn a lot from working with my classmates'), was called 'perceived quality of engagement with peers'. Factor 3, consisting of items addressing learning materials and topics (e.g. 'The topics we cover are useful for my learning'), was dubbed 'perceived usefulness of engagement with teaching content'. Factor 4, containing items related to student effort (e.g. 'I usually try my best when my teacher asks us to do something'), was identified as 'intensity of effort in learning'. Factor 5, comprising items related to student satisfaction with learning materials (e.g. 'I usually find the content boring'), was named 'perceived satisfaction of engagement with teaching content'. Factor 6, consisting of items addressing task difficulty (e.g. 'I usually feel the activities are at the right level of challenge for me'), was called 'perceived difficulty of teaching content and learning activities'.

Table 5 presents a comparative summary of the detailed factor structures of both the initial study and the present study. The initial study had five factors with 18 items, whereas the present study included six factors with 22 items. Both studies identified similar factors, such as perceived quality of engagement with teachers and peers, and the intensity of effort in learning. However, the present study had new factors, including perceived satisfaction of engagement with teaching content, perceived usefulness of teaching content, and perceived difficulty in teaching content and learning activities, while not including the intensity of social engagement and perceived quality of engagement with learning activities.

The factorial structure extracted through EFA proceeded to the CFA stage. Table 6 summarizes the model fit indices. The result shows that the chi-square value was significant ($p < .001$); however, this value was considered to have been influenced by the sample size (Hair et al., 2019). Therefore, we evaluated the model using other indices such as factor loadings and model fit indices. The CFI and TLI values were greater than 0.9, indicating a good fit. The RMSEA value also indicated a reasonable fit, and the SRMR value likewise demonstrated a good fit. The fit indices of the current study's model are nearly equal to those of the initial study's model described in Table 6.

The factorial structure detailed in Table 7 depicts all items loaded on their target factors with values between 0.611 and 0.95, supporting construct validity. Table 8 presents the details of factor covariance, demonstrating low to moderate relationships ranging from 0.159 to 0.699.

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

Table 3. Factor characteristics and loadings from the initial analysis

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Uniqueness
5	0.954							0.314
6	0.937							0.352
7	0.914							0.35
9	0.722							0.453
8	0.622							0.39
4	0.614							0.318
17		0.928						0.261
16		0.902						0.288
15		0.813						0.397
18		0.775						0.309
13		0.741						0.355
14		0.608						0.534
12		0.314					0.401	0.621
31			0.721					0.357
34			0.69					0.487
30			0.669					0.477
32			0.602					0.402
29			0.436	0.556				0.453
22			0.39					0.306
23			0.38					0.446
19				0.977				0.449
20				0.72				0.353
3				0.651				0.474
1				0.565				0.436
28				0.307				0.431

(*Continued*)

Table 3. (*Continued.*)

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Uniqueness
33					0.872			0.329
35					0.789			0.373
25					0.629			0.564
27						0.918		0.229
36						0.814		0.359
2							0.451	0.588
10								0.52
11								0.421
24								0.421
26								0.524
Eigenvalues	13.45	2.671	2.160	1.533	1.452	1.170	1.075	
PVE (%)	13.1	12.5	10.9	9.2	5.8	5.0	2.6	

Note: The applied rotation method is promax. PVE = proportion of variance explained.

Table 4. Final factor structure

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Uniqueness
6	0.905						0.229
5	0.895						0.249
7	0.794						0.325
9	0.586						0.466
17		0.919					0.380
16		0.895					0.306
18		0.786					0.297
15		0.703					0.387
13		0.678					0.479
31			0.883				0.259
34			0.711				0.479
30			0.584				0.586
32			0.545				0.441
19				0.867			0.416
3				0.669			0.457
20				0.595			0.377
1				0.526			0.472
33					0.873		0.308
35					0.8		0.359
25					0.607		0.586
36						0.878	0.221
27						0.822	0.287

Note: The applied rotation method is promax. Factors 1 and 2 were reordered to allow for easier comparison with the subsequent sections.

Table 5. Detailed comparison between the initial and present studies

	Initial study	Present study
Number of factors	5	6
Total number of items	18	22
Perceived quality of engagement with the teacher	4, 6, 7, 9	5, 6, 7, 9
Perceived quality of engagement with peers	13, 15, 16, 18	13, 15, 16, 17, 18
Intensity of effort in learning	3, 20, 28, 29	1, 3, 19, 20
Perceived quality of engagement with learning activities	22, 24, 25	None
Perceived satisfaction of engagement with teaching contents	None	25, 33, 35
Perceived usefulness of engagement with teaching contents	None	30, 31, 32, 34
Intensity of social engagement	2, 10, 12	None
Perceived difficulty in teaching content and learning activities	None	27, 36

Note: The numbers associated with the factors correspond to the items in the questionnaire.

4.1. Validity and reliability

Regarding construct validity, each factor in the model was well specified, as reflected in Table 7. For convergent validity, we examined the average variance extracted (AVE) values, using 0.5 as the

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

Table 6. Model fit indices

	$\chi 2$	df	p	CFI	TLI	RMSEA (90% CI)	SRMR
Present study	576.32	194	$< .001$	.924	.910	.068 (.062; .074)	.051
Initial study	337.4	125	$< .001$	.938	.916	.067 (.059; .076)	N/A

threshold. Factors 1 to 6 recorded values of 0.612, 0.661, 0.491, 0.521, 0.562, and 0.740, respectively. These results support convergent validity, although the AVE (0.491) for Factor 3 was marginally below the cut-off value. Regarding discriminant validity (see Table 9), we examined the heterotrait-monotrait (HTMT) values. All values were below 0.9 (Hair et al., 2019), supporting discriminant validity. As regards reliability, both McDonald's omega (ω) and Cronbach's alpha (α) were calculated. McDonald's omega is considered to provide a more robust estimate than Cronbach's alpha (Hayes & Coutts, 2020). Both coefficients (see Table 10) confirmed good reliability scores (coefficient $\omega = 0.943$; coefficient $\alpha = 0.909$). The initial study reported that each factor's coefficient alpha ranged from 0.7 to 0.9, which was nearly the same as the coefficient alpha in this study.

5. Discussion

This study endeavored to replicate Teravainen-Goff's (2023) study based on the aforementioned aims (a), (b), (c), and (d). Aims (a) and (b) are discussed separately, whereas aims (c) and (d) are discussed together to facilitate meaningful comparisons with the findings of the initial study. The first aim (a) was to propose and test the structural validity of the engagement scale. Building on the initial 36-item engagement scale and ensuring cultural equivalence through a rigorous translation process, we collected data from university students in the Japanese EFL context. As reported in the results section, EFA yielded a six-factor structure comprising 22 items. Subsequently, CFA verified the proposed structure, demonstrating a good model fit. The second aim (b) was to examine the validity and reliability of the scale. Multiple validity assessments were conducted to achieve this objective. Construct validity was evaluated through factor loadings and model fit indices. Convergent validity and discriminant validity were assessed using AVE and HTMT, respectively. For construct validity, all items demonstrated substantial factor loadings ranging from 0.61 to 0.95 on their intended factors, and model fit indices confirmed a well-defined factorial structure. Although Factor 3 fell slightly below the threshold value regarding convergent validity, the overall evidence supported convergent validity. Based on HTMT values, discriminant validity confirmed that all factors were distinctly separable.

The third and fourth aims, (c) and (d), were to gain insights from the engagement scale used in this study and compare the findings with those of the initial study. As Table 5 summarizes the key findings from both studies, the analyses revealed distinct structural differences: the initial study yielded a five-factor structure, whereas the present study identified a six-factor structure. Despite following identical procedures, the number of retained items differed between the studies. A detailed examination of the extracted factors revealed similarities and differences across studies. Regarding 'perceived quality of engagement with teachers', both scales demonstrated identical patterns with nearly identical items. This result suggests that teachers play a crucial role in promoting student engagement, regardless of the target language and participant demographics. Similarly, 'perceived quality of engagement with peers' demonstrated consistent patterns across studies. These findings highlight the significant roles of peers in student engagement. Furthermore, regarding 'intensity of effort in learning', similar patterns emerged across studies, although the specific items comprising these factors differed slightly. Despite these differences, all items retained the core semantic meaning of student effort.

However, notable differences between the two studies also emerged. The initial study identified 'perceived quality of engagement with learning activities' as a factor. In contrast, the present study did not extract the core meaning of learning activities; instead, the present study revealed

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

Table 7. Detailed factor structure

Factor	Item	Estimate	Std. error	z-value	p	95% CI lower	95 CI upper	Std. est. (all)
Factor 1	5	0.781	0.039	19.997	< .001	0.705	0.858	0.828
	7	0.72	0.039	18.6	< .001	0.644	0.795	0.788
	6	0.772	0.04	19.494	< .001	0.695	0.85	0.813
	9	0.724	0.045	15.965	< .001	0.635	0.813	0.708
Factor 2	17	0.95	0.043	21.851	< .001	0.865	1.035	0.864
	16	0.894	0.041	21.967	< .001	0.814	0.973	0.867
	18	0.762	0.039	19.765	< .001	0.687	0.838	0.81
	15	0.69	0.041	16.826	< .001	0.61	0.77	0.725
	13	0.737	0.041	18.07	< .001	0.657	0.817	0.764
Factor 3	31	0.762	0.039	19.564	< .001	0.685	0.838	0.826
	34	0.69	0.043	15.943	< .001	0.605	0.775	0.711
	32	0.695	0.04	17.34	< .001	0.616	0.773	0.759
	30	0.65	0.056	11.663	< .001	0.541	0.759	0.557
Factor 4	19	0.611	0.046	13.23	< .001	0.52	0.701	0.623
	3	0.687	0.045	15.306	< .001	0.599	0.775	0.699
	20	0.747	0.039	19.271	< .001	0.671	0.823	0.823
	1	0.677	0.04	16.878	< .001	0.598	0.755	0.747
Factor 5	33	0.901	0.054	16.792	< .001	0.796	1.007	0.784
	35	0.923	0.052	17.763	< .001	0.822	1.025	0.829
	25	0.753	0.057	13.157	< .001	0.64	0.865	0.635
Factor 6	27	0.917	0.049	18.566	< .001	0.82	1.013	0.859
	36	0.824	0.044	18.642	< .001	0.737	0.911	0.862

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

Table 8. Factor covariance

			Estimate	Std. Error	z-value	p	95% CI Lower	95% CI Upper
Factor 1	↔	Factor 2	0.49	0.043	11.44	< .001	0.406	0.574
Factor 1	↔	Factor 3	0.69	0.035	20.008	< .001	0.623	0.758
Factor 1	↔	Factor 4	0.646	0.037	17.36	< .001	0.573	0.719
Factor 1	↔	Factor 5	0.372	0.051	7.327	< .001	0.273	0.472
Factor 1	↔	Factor 6	0.47	0.046	10.256	< .001	0.38	0.559
Factor 2	↔	Factor 3	0.544	0.041	13.115	< .001	0.463	0.625
Factor 2	↔	Factor 4	0.55	0.041	13.364	< .001	0.47	0.631
Factor 2	↔	Factor 5	0.159	0.055	2.906	0.004	0.052	0.267
Factor 2	↔	Factor 6	0.372	0.048	7.687	< .001	0.277	0.467
Factor 3	↔	Factor 4	0.699	0.036	19.688	< .001	0.63	0.769
Factor 3	↔	Factor 5	0.292	0.055	5.341	< .001	0.185	0.399
Factor 3	↔	Factor 6	0.615	0.04	15.433	< .001	0.537	0.693
Factor 4	↔	Factor 5	0.288	0.055	5.252	< .001	0.181	0.395
Factor 4	↔	Factor 6	0.446	0.048	9.243	< .001	0.351	0.54
Factor 5	↔	Factor 6	0.206	0.056	3.672	< .001	0.096	0.316

Table 9. Heterotrait-monotrait values

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Factor 1	1	0.516	0.675	0.61	0.365	0.481
Factor 2		1	0.57	0.557	0.132	0.381
Factor 3			1	0.688	0.223	0.598
Factor 4				1	0.242	0.44
Factor 5					1	0.207
Factor 6						1

Table 10. Scale reliability

	Coefficient ω	Coefficient α
Factor 1	0.866	0.861
Factor 2	0.906	0.902
Factor 3	0.793	0.793
Factor 4	0.802	0.817
Factor 5	0.793	0.79
Factor 6	0.85	0.848
Total	0.943	0.909

'perceived satisfaction of engagement with teaching contents' as a factor, with only one item overlapping with the initial study's 'perceived quality of engagement with learning activities'. Additionally, the present study identified the 'perceived usefulness of engagement with teaching content' as a factor. These comparisons with the initial study suggest that university students in the Japanese EFL context tend to associate their engagement more with learning content rather than activities. This finding may also be related to 'the intensity of social engagement'. The initial study identified this factor with its core meaning of group work and class discussion. However, this factor was not revealed in the present study. This interesting difference could be explained by several factors, one of which is culturally bound learning styles. As Albertson (2020) noted, 'Japanese student perspectives toward class participation suggest they may incline toward silence due to aspects of the Japanese communication style' (p. 47). Furthermore, learning in Japanese contexts is frequently characterized by a teacher-centered approach (Albertson, 2020). These factors might influence the findings. In addition, the present study identified a unique factor that was not anticipated in the initial scale: 'perceived difficulty of teaching content and learning activities'. This factor, comprising two items with high factor loadings, suggests that university students in the Japanese EFL context tend to show high levels of engagement if the level of contents and activities match their level of English.

This discussion builds upon Teravainen-Goff's (2023) newly developed engagement scale, which focuses on the fundamental concept of engagement as 'action'. This primary focus on behavioural engagement allowed us to conduct our data interpretation and discussion with conceptual clarity without being complicated by other dimensions, such as emotional engagement and cognitive engagement, or related concepts, such as motivation (Martin et al., 2017; Nagy et al., 2022; Vo, 2024). Most importantly, this precise and clearer operationalization of engagement within the replication framework (Porte & McManus, 2019) provided a methodologically sound basis for conducting systematic comparisons across different contexts, thereby contributing to the applicability of this newly developed scale (Teravainen-Goff, 2023) through our findings.

https://doi.org/10.1017/S0261444826101141 Published online by Cambridge University Press

6. Conclusion

This study conducted an approximate replication of Teravainen-Goff (2023), having modified two variables: the target language and participant demographics. The present study identified a six-factor model with 22 items, compared to the initial study's five-factor structure with 18 items. Using the same engagement scale, we identified both common and unique factors across the studies. Both studies extracted three common factors: 'perceived quality of engagement with teachers,' 'perceived quality of engagement with peers', and 'intensity of effort in learning'. Different factors also emerged. The initial study identified 'perceived quality of engagement with learning activities' and 'intensity of social engagement'. In contrast, the present study revealed three different factors: 'perceived satisfaction with teaching content', 'perceived usefulness of teaching content', and 'perceived difficulty of teaching content and learning activities'. From these differences, it is inferred that participants from the initial study associated their engagement primarily with learning activities, whereas participants in the present study related their engagement to learning content.

In summary, this study provides interesting and meaningful findings that contribute to a better understanding of engagement from theoretical, methodological, and pedagogical perspectives. Theoretically, this study offers empirical evidence for further discussion to clarify the construct of engagement and its blurred boundaries with similar concepts such as motivation. Methodologically, we demonstrated the promising potential of scale development using a replication framework based on open science principles. The fundamental principle of identifying which elements to modify and which to maintain in replication studies enables the systematic accumulation of research findings across diverse contexts. Pedagogically, this study confirms the significant role of teachers and peers in learner engagement, suggesting that attention to these relationships is fundamental for promoting student engagement (Philp & Duchesne, 2016). In the Japanese EFL context with university participants, the learning content and its difficulty level play a crucial role in promoting engagement; therefore, teachers need to provide interesting and tailored (i.e. at an appropriate difficulty level) tasks for language practice.

Despite its valuable findings, this study has a few limitations. First, although we aimed to collect data from a diverse sample to ensure the generalizability of our findings, further research is needed with both a broader participant population and English teachers representing diverse educational backgrounds and instructional approaches. Second, as noted in the data analysis section, we used the same dataset for EFA and CFA, which may have led to inflated model fit indices. Future replication studies using different datasets for this questionnaire would provide more robust evidence for validity.

Acknowledgements. We would like to thank all the teachers and students who participated in this study. We are also grateful to the anonymous reviewers for their constructive comments, which helped improve the manuscript.

Funding statement. This work was supported by JST SPRING, Grant Number JPMJSP2150, granted to the first author of this article.

Competing interests. The authors declared that they have no conflicts of interest.

AI disclosure. We acknowledge the use of Claude 3.5 sonnet in the writing of this paper. The AI prompts used were 'Improve this English' or 'Check the grammatical errors.' The output from these prompts was used to write the following sections: introduction, literature review, data analysis, results, discussion, and conclusion. After using this AI tool, we reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Albertson, B. P. (2020). Promoting Japanese university students' participation in English classroom discussions: Towards a culturally-informed bottom-up approach. *Journal of Pan-Pacific Association of Applied Linguistics*, *24*(1), 45–66. https://doi.org/10.25256/paal.24.1.3

Al-Hoorie, A. H., Cinaglia, C., Hiver, P., Huensch, A., Isbell, D. R., Leung, C., & Sudina, E. (2024). Open science: Considerations and issues for TESOL research. *TESOL Quarterly*, *58*(1), 537–556. https://doi.org/10.1002/tesq.3304

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, *55*, 145–157. https://doi.org/10.1016/j.system.2015.10.006

Dörnyei, Z., & Dewaele, J.-M. (2023). *Questionnaires in second language research, construction, administration, and processing*. Routledge. https://doi.org/10.4324/9781003331926

Eerdemutu, L., Dewaele, J.-M., & Wang, J. (2024). Developing a short language classroom engagement scale (LCES) and linking it with needs satisfaction and achievement. *System*, *120*, 103189. https://doi.org/10.1016/j.system.2023.103189

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble. *European Journal of Psychological Assessment*, *33*(6), 399–402. https://doi.org/10.1027/1015-5759/a000460

Fredricks, J. A. (2022). The measurement of student engagement: Methodological advances and comparison of new self-report instruments. In A. L. Reschly, & S. L. Christenson (Eds.), *Handbook of research on student engagement* (2nd ed., pp. 597–616). Springer. https://doi.org/10.1007/978-3-031-07853-8_29

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Guo, Y., Xu, J., & Chen, C. (2023). Measurement of engagement in the foreign language classroom and its effect on language achievement: The case of Chinese college EFL students. *International Review of Applied Linguistics in Language Teaching*, *61*(3), 1225–1270. https://doi.org/10.1515/iral-2021-0118

Hair, J. F., Babin, B. J., Black, W. C., & Anderson, R. E. (2019). *Multivariate data analysis*. (8th ed.) Cengage.

Hair, J. F., Hair, J., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage.

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Hiver, P., Al-Hoorie, A. H., & Mercer, S. (Eds.). (2021). *Student engagement in the language classroom*. Multilingual Matters. https://doi.org/10.21832/9781788923613

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, *28*(1), 201–230. https://doi.org/10.1177/13621688211001289

Hiver, P., Zhou, S., Tahmouresi, S., Sang, Y., & Papi, M. (2020). Why stories matter: Exploring learner engagement and metacognition through narratives of the L2 learning experience. *System*, *91*, 102260. https://doi.org/10.1016/j.system.2020.102260

Khajavy, G. H. (2021). Modeling the relations between foreign language engagement, emotions, grit and reading achievement. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *Student engagement in the language classroom* (pp. 241–259). Multilingual Matters. https://doi.org/10.21832/9781788923613-016

Kormos, J., & Csizer, K. (2008). Age-related differences in the motivation of learning English as a foreign language: Attitudes, selves, and motivated learning behavior. *Language Learning*, *58*(2), 327–355. https://doi.org/10.1111/j.1467-9922.2008.00443.x

Liu, M. (2023). Whose open science are we talking about? From open science in psychology to open science in applied linguistics. *Language Teaching*, *56*(4), 443–450. https://doi.org/10.1017/S0261444823000307

Liu, M., Chong, S. W., Marsden, E., McManus, K., Morgan-Short, K., Al-Hoorie, A. H., Plonsky, L., Bolibaugh, C., Hiver, P., Winke, P., Huensch, A., & Hui, B. (2022). Open scholarship in applied linguistics: What, why, and how. *Language Teaching*, *56*(3), 432–437. https://doi.org/10.1017/s0261444822000349

Manapat, P. D., Anderson, S. F., & Edwards, M. C. (2025). Evaluating avoidable heterogeneity in exploratory factor analysis results. *Psychological Methods*, *30*(3), 660–677. https://doi.org/10.1037/met0000589

Marsden, E., Morgan-Short, K., Thompson, S., & Abugaber, D. (2018). Replication in second language research: Narrative and systematic reviews and recommendations for the field. *Language Learning*, *68*(2), 321–391. https://doi.org/10.1111/lang.12286

Martin, A. J., Ginns, P., & Papworth, B. (2017). Motivation and engagement: Same or different? Does it matter? *Learning and Individual Differences*, *55*, 150–162. https://doi.org/10.1016/j.lindif.2017.03.013

McManus, K. (2024). The future of replication in applied linguistics: Toward a standard for replication studies. *Annual Review of Applied Linguistics*, *44*, 133–149. https://doi.org/10.1017/S0267190524000011

Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press. https://doi.org/10.1017/9781009024563

Nagy, R. P., Martin, A. J., & Collie, R. J. (2022). Disentangling motivation and engagement: Exploring the role of effort in promoting greater conceptual and methodological clarity. *Frontiers in Psychology*, *13*, 1045717. https://doi.org/10.3389/fpsyg.2022.1045717

Noels, K. A., Lascano, D. I. V., & Saumure, K. (2019). The development of self-determination across the language course. *Studies in Second Language Acquisition*, *41*(4), 821–851. https://doi.org/10.1017/S0272263118000189

Oga-Baldwin, W. L. Q., & Nakata, Y. (2017). Engagement, gender, and motivation: A predictive model for Japanese young language learners. *System*, *65*, 151–163. https://doi.org/10.1016/j.system.2017.01.011

Oga-Baldwin, W. L. Q., Nakata, Y., Parker, P., & Ryan, R. M. (2017). Motivating young language learners: A longitudinal model of self-determined motivation in elementary school foreign language classes. *Contemporary Educational Psychology*, *49*, 140–150. https://doi.org/10.1016/j.cedpsych.2017.01.010

Papi, M., & Teimouri, Y. (2012). Dynamics of selves and motivation: A cross-sectional study in the EFL context of Iran. *International Journal of Applied Linguistics*, *22*(3), 287–309. https://doi.org/10.1111/j.1473-4192.2012.00312.x

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72. https://doi.org/10.1017/S0267190515000094

Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative L2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning*, *65*(S1), 9–36. https://doi.org/10.1111/lang.12111

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. *Journal of Educational Psychology*, *105*(3), 579–595. https://doi.org/10.1037/a0032690

Reschly, A. L., & Christenson, S. L. (Eds.). (2022). *Handbook of research on student engagement.* (2nd ed.) Springer.

Sang, Y., & Hiver, P. (2021). Engagement and companion constructs in language learning: Conceptualizing learners' involvement in the L2 classroom. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *Student engagement in the language classroom* (pp. 17–37). Multilingual Matters.

Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, *50*(1), 1–13. https://doi.org/10.1080/00461520.2014.1002924

Soderberg, C. K. (2018). Using OSF to share data: A step-by-step guide. *Advances in Methods and Practices in Psychological Science*, *1*(1), 115–120. https://doi.org/10.1177/2515245918757689

Sudina, E. (2021). Study and scale quality in second language survey research, 2009–2019: The case of anxiety and motivation. *Language Learning*, *71*(4), 1149–1193. https://doi.org/10.1111/lang.12468

Teravainen-Goff, A. (2022). Why motivated learners might not engage in language learning: An exploratory interview study of language learners and teachers. *Language Teaching Research*, 136216882211353. https://doi.org/10.1177/13621688221135399

Teravainen-Goff, A. (2023). Intensity and perceived quality of L2 engagement: Developing a questionnaire and exploring engagement of secondary school language learners in England. *System*, *112*, 102955. https://doi.org/10.1016/j.system.2022.102955

Van Prooijen, J.-W., & van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement*, *61*(5), 777–792. https://doi.org/10.1177/00131640121971518

Vo, H. (2024). Towards clarification of the second language learner engagement construct: Taking stock of its conceptual overlap and hierarchical structure. *Applied Linguistics*, *45*(6), 1030–1049. https://doi.org/10.1093/applin/amae012

Zhou, S. A., & Hiver, P. (2022). The effect of self-regulated writing strategies on students' L2 writing engagement and disengagement behaviors. *System*, *106*, 102768. https://doi.org/10.1016/j.system.2022.102768

Yohei Nakanishi is a Ph.D. student at the Graduate School of Foreign Language Education and Research, Kansai University. His current research interests include young language learners' motivation, foreign language enjoyment, and engagement. He has published articles in journals such as *Research Methods in Applied Linguistics*.

Osamu Takeuchi, Ph.D., is a professor in the Faculty of Foreign Language Studies and the Graduate School of Foreign Language Education and Research at Kansai University, Osaka, Japan. His current research interests include language learning strategies, self-regulation in L2 learning, L2 learning motivation, and the application of technology to language teaching. He has published articles in journals such as *Applied Linguistics, International Review of Applied Linguistics in Language Teaching (IRAL), Innovation in Language Learning and Teaching, RELC Journal, Research Methods in Applied Linguistics*, and *System*. Dr Takeuchi received the JACET Award for Outstanding Academic Achievement in 2004, the LET Award for Outstanding Academic Achievement in 2009, and the JLTA Award for Outstanding Academic Articles in 2024.

Cite this article: Nakanishi, Y., & Takeuchi, O. (2026). Revisiting the intensity and perceived quality of L2 engagement scale in a different context: An approximate replication of Teravainen-Goff (2023). *Language Teaching*, 1–20. https://doi.org/10.1017/S0261444826101141