



METHODS FORUM

Construction and validation of a questionnaire to study engagement in informal second language learning

Henriette L. Arndt 

Lund University, Lund, Sweden and University of Oxford, Oxford, UK
E-mail: henriette.arndt@humlab.lu.se

(Received 01 June 2022; Revised 01 November 2022; Accepted 27 December 2022)

Abstract

This article reports on the development and validation of the Informal Second Language Engagement questionnaire (ISLE) for capturing various aspects of learner engagement with informal second language practices. Whereas other questionnaires have primarily focused on learner behavior (the frequency, quantity, and diversity of informal activities in which learners engage), the ISLE additionally targets affective, cognitive, and linguistic aspects of engagement. Consequently, this questionnaire has the potential to help second language acquisition researchers better understand individual differences and variability in informal language engagement, how it changes over time, and the ways in which engagement affects language learning. The ISLE was developed and evaluated in three stages with a total of 382 German secondary school learners of English as a foreign language. The initial instrument development was informed by qualitative findings from focus group interviews. Subsequently, the questionnaire was piloted, and Exploratory Factor Analysis was applied to the data to uncover the underlying factor structures and reduce the number of items per construct. Finally, the revised instrument was evaluated further using Confirmatory Factor Analysis.

Introduction

Second Language Acquisition (SLA) researchers are becoming increasingly interested in studying learning outside of formal contexts (e.g., Benson & Reinders, 2011; Nunan & Richards, 2015; Sockett, 2014). Thanks to the internet, learners nowadays have more opportunities than ever to independently use and develop their second language (L2). In addition to materials developed to support independent study (e.g., podcast- or video-based lessons, grammar websites, and language learning apps), learners may also engage in more naturalistic, self-directed language use through activities not specifically designed for language learning. This includes, for example, online social

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

networking; reading news and entertainment websites; streaming music, online videos, TV series, or films; and playing video games. During the last decade, an increasing number of studies have been published that demonstrate that such informal second language practices (ISLPs) are indeed widespread among L2 learners (e.g., Lee, 2019; Sockett, 2014; Sundqvist & Sylvén, 2016). The majority of this research has focused on learners of English as a second language, which is reflected in the use of terms such as “online informal learning of English” (Sockett, 2014); “informal digital learning of English” (Lee, 2019; Lee & Dressman, 2017); and “extramural English” (Sundqvist, 2011). On the one hand, this mirrors a general predominance of studies on English language learning in the wider field of SLA. On the other hand, informal learning researchers have proposed that engagement with English in everyday life differs from engagement with any other L2 due to the strong influence of popular Anglo-American media and culture all over the world and the widespread use of English as a lingua franca, especially on the internet (Mair, 2020; Seidlhofer, 2011).

Many studies have found positive relationships between ISLPs and L2 proficiency, which is generally interpreted as evidence that informal activities help learners to improve their language skills (e.g., Kuppens, 2010; Lee & Dressman, 2017; Sundqvist, 2011; Verspoor et al., 2011). Similar associations have also been found between ISLPs and affective variables such as motivation, self-efficacy, confidence, and enjoyment (e.g., Cole, 2015; Lai et al., 2015; Lee, 2019; Lyrigkou, 2016). Nevertheless, considerable individual and group differences in proficiency and motivation associated with ISLPs have also been recorded (e.g., Cole, 2015; Sundqvist, 2011; Sundqvist & Sylvén, 2014), which seem to indicate that not all learners participate in and/or benefit from ISLPs in the same way.

This article presents a novel research instrument, the Informal Second Language Engagement questionnaire (ISLE), which was developed as part of a larger research project to investigate whether the aforementioned individual differences could, in some part, be explained by differences in the ways in which learners interact with informal L2 activities (Arndt, 2019). I begin by introducing the concept of student engagement, which was used in this study as a framework for investigating various aspects of learners’ interactions with ISLPs. Where prior research has primarily focused on learner behavior (the quantity and diversity of informal activities), the ISLE was designed to also capture affective, cognitive, and linguistic aspects of engagement in informal second language practices. Therefore, I suggest that the ISLE has the potential to help researchers better understand the previously observed individual differences and variability in ISLPs, how different aspects of informal L2 engagement change over time, and the ways in which they relate to language learning.

Student engagement

Engagement is a concept used in educational research to describe students’ levels of active participation, interest, and meaningful involvement in learning (Christenson et al., 2012). Engagement is always considered in a particular context, or with reference to a specific “object,” such as engagement in the community, at school, in a subject, or even a single learning activity (Hiver et al., 2021b; Skinner & Pitzer, 2012). That is, it describes the “observable and unobservable qualities of students’ interactions” with these objects (Wang et al., 2016, p. 17), or the ways in which learners act, think, and feel (Oga-Baldwin, 2019).

Seminal engagement studies explored the links between school dropout rates, students' sense of belonging, and participation in school-related activities (Finn, 1989). Since then, engagement has been the subject of extensive research, which has generally found it to be an important predictor of academic achievement (e.g., see contributions in the *Handbook of Research on Student Engagement*, Christenson et al., 2012, and review by Boekaerts, 2016). The vast majority of studies have focused on overall school engagement or engagement in particular subjects, whereas studies at the level of individual learning tasks, as well as those investigating engagement in learning outside of formal contexts, are less common (Eccles, 2016; Reschly & Christenson, 2012).

Although the term "engagement" is widely used and intuitively understood among educational practitioners, there is a lack of definitional clarity in the subject literature. This is due, in part, to the use of widely varying terminology, working definitions, and research tools (Reschly & Christenson, 2012), compounded by the diverse aims and contexts of individual studies. Nevertheless, there is general agreement that engagement is dynamic and malleable (i.e., it changes over time and can be affected through contextual changes, such as interventions) and that it comprises at least three different dimensions (or aspects) as identified by Fredericks, Blumenfeld and Paris (2004): behavioral, cognitive, and affective engagement. Suggestions for other dimensions include, for example, social, self-regulatory or agentic engagement (e.g., Boekaerts, 2016; Finn & Zimmer, 2012; Reeve & Tseng, 2011).

Behavioral engagement, broadly, refers to outwardly observable indicators of student involvement, such as time on task or the degree of persistence and active participation in learning (Finn & Zimmer, 2012). Cognitive engagement involves sustained attention, mental effort, and, on a task level, the application of strategies for self-regulation and learning (Appleton et al., 2006). Finally, affective engagement (also called emotional engagement) commonly refers to learners' affective reactions such as enthusiasm, interest, and enjoyment, as well as boredom, dislike, and frustration (Skinner et al., 2009). Other constructs such as perceived value, purposefulness, and autonomy have also been included in some definitions of affective engagement (Reinders and Nakamura, 2021), although others view them as precursors that affect learners' intentions and their decision to engage (Hiver et al., 2021a; Mercer, 2019).

Despite disagreements over the number of subconstructs and their definitions, there is a general consensus that the different aspects of engagement are interrelated and dynamically interact with each other across time (Reinders & Nakamura, 2021; Reschly & Christenson, 2012). For example, a learner might be more likely to participate actively (behavioral engagement) in a learning activity that they find interesting (affective engagement). Conversely, if a task requires high levels of sustained mental effort (cognitive engagement), this may negatively affect students' enjoyment of the activity (affective engagement) and they may not persist in it (behavioral engagement). The interconnected nature of the engagement dimensions means that they need to be studied together rather than in isolation (Janosz, 2012; Philp & Duchesne, 2016). It is because of this interconnectedness that several researchers have recently called for the application of the engagement framework to SLA research, to provide a more holistic, integrated way of studying behavioral, cognitive, and affective processes and their role in language acquisition (e.g., Hiver et al., 2021b; Oga-Baldwin, 2019; Mercer, 2019; Reinders & Nakamura, 2021). The following section contains a brief introduction to the prior literature on engagement in SLA.

Engagement and second language acquisition

During the last 20 years, SLA researchers have increasingly acknowledged engagement as a key variable in language learning (e.g., Dörnyei & Kormos, 2000; Hiver et al., 2021b; Mercer & Dörnyei, 2021; Philp & Duchesne, 2016). As Mercer (2019) notes, prior to this point SLA research had already investigated many aspects of engagement “under other guises” (p. 647), such as studies on effort (part of behavioral engagement), language learning strategies (cognitive engagement), and flow (which could be described as an overall heightened state of engagement). More recently, however, SLA researchers have begun to use the term “engagement” in a more deliberate and focused manner, building on the available body of work in educational science and extending it in domain-specific ways.

A number of theoretical introductions have been published that review foundational engagement research in education; discuss the applicability of definitions and findings in the context of SLA; and stress the value of the concept of engagement as a holistic framework that has the potential to unite diverse research strands in SLA to improve our understanding of the complex interactions between various affective, cognitive, behavioral, and social variables in the language learning process (e.g., Mercer, 2019; Oga-Baldwin, 2019; Reinders & Nakamura, 2021). An important issue in these discussions has also been to delineate the construct of engagement (what does and does not fall under the umbrella of this meta-construct) and consider how it connects to, and differs from, other common concepts in SLA, such as attention, self-regulation, anxiety, and collaboration (Philp and Duchesne, 2016) or mindsets, autonomy, and relatedness (Mercer, 2019). At the center of this discussion has been the issue of how engagement differs from, but may interact with, language learning motivation and related concepts such as attitudes and beliefs, investment, and interest (Mercer, 2019; Oga-Baldwin, 2019; Sang & Hiver, 2021). Consensus has emerged that motivation can be regarded as a precursor to engagement, the “force that energize[s] and direct[s] that behavior” that comprises engagement (Hiver et al., 2021b, p. 23).

This proposed relationship between motivation and engagement is at the center of the Contextual Model of Engagement (CME; Lam et al., 2012), which is widely used in educational research on student engagement. The model suggests that engagement, as a central mediator, can be used to understand the relationships between students’ past and future experiences and achievements, incorporating both personal characteristics (e.g., motivation, attitudes, and abilities) and features of the learning environment (e.g., resources and relationships). Oga-Baldwin (2019) suggested adapting the CME for research on formal language learning and teaching, to help understand and predict how students succeed in language learning in educational settings. Similarly, the model was used in this article to frame the role of engagement in informal second language learning. Figure 1 shows the proposed relationships between engagement in ISLPs and illustrative examples of a variety of potentially important environmental factors and learner characteristics. The larger study of which this article forms a part explored some of these relationships, namely between learners affective, behavioral, cognitive, and linguistic engagement in ISLPs; their prior and subsequent motivation and attitudes (language learning goals, attitudes toward informal and formal learning, and L2 selves); and general language proficiency (Arndt, 2019).

In addition to theoretical reviews, there have also been several empirical studies that have used interviews, observational methods, or self-report questionnaires adapted from educational studies to investigate the nature of engagement in SLA, its

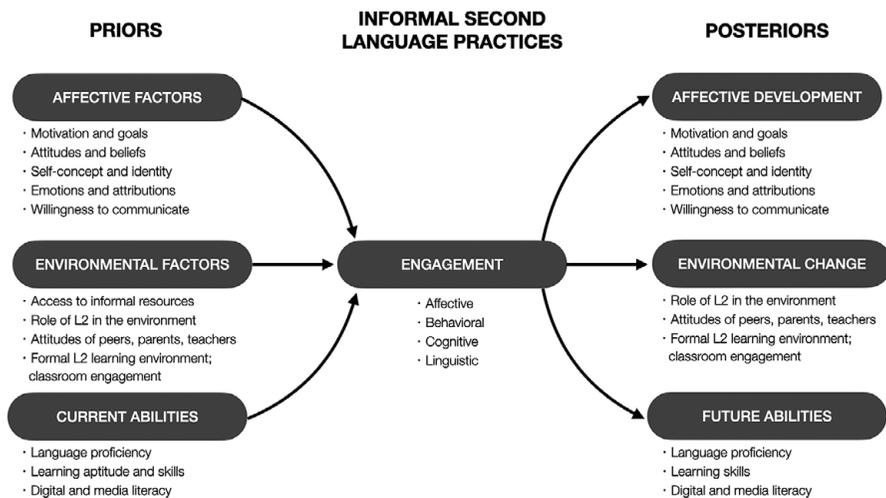


Figure 1. A CME in informal second language practices (adapted from Oga-Baldwin, 2019).

antecedents, and outcomes; how engagement changes over time; and strategies for increasing engagement among students (see recent reviews in Hiver et al., 2021a; Mercer, 2019; Zhou et al., 2021). Based on a systematic review of this literature, Hiver and colleagues (2021a) concluded that SLA researchers seem to find the multidimensional construct of engagement useful for framing how students think, act, and feel in formal instructional settings (language classrooms or virtual learning environments). By contrast, studies on engagement in informal settings are completely absent from the literature to date. In addition to this gap in language engagement research, Hiver et al. (2021a) also identified several areas for future development that hold relevance for the design of the research instrument introduced in this article: The authors point toward a need for more domain-specific measures of engagement in SLA, as well as “increasing the use of longitudinal and individual-based investigations uncovering the dynamic nature of engagement” (p. 25). In the same vein, the ISLE was developed as a tool to measure engagement in the specific context of informal second language learning and, when implemented as part of an Experience Sampling Method (ESM) approach (see the following text), enables researchers to gather rich, longitudinal data on learner engagement.

Extending engagement research into the realm of informal language learning requires careful consideration of this learning context because engagement is, by definition, context-dependent (cf. Christenson et al., 2012). That is, differences between informal and formal learning will determine “what counts as engagement” (Philp & Duchesne, 2016)—how it should be defined and measured—in each setting. Therefore, the next section discusses how engagement was defined in the current study for the context of informal second language learning.

A multidimensional framework of informal second language engagement

The current study used the multi-dimensional construct of engagement as a framework for studying the “qualities of students’ interactions” (Wang et al., 2016, p. 17) with

informal second language practices. For this purpose, the three core dimensions of engagement were defined as follows:

- (a) *Behavioral engagement*: the observable qualities of engagement in ISLPs, including duration, frequency, and the variety of informal activities
- (b) *Affective engagement*: learners' emotional states while engaging in ISLPs, for example, enjoyment and interest versus boredom and indifference
- (c) *Cognitive engagement*: the extent to which learners are concentrating on the activity at hand

The object of engagement in these definitions is the specific informal second language practice in which the learner participates. However, when thinking about engagement in the context of language learning (both formal and informal), we must also consider the extent to which an individual engages with the language. Because language is at the heart of most formal learning tasks (i.e., they are designed to support language learning, and many require focus on form), most task engagement in formal SLA settings is likely to entail some degree of engagement with language features. By contrast, research has observed that there is less focus on the language in informal L2 practices, in which most learners participate for entertainment, not to improve their language proficiency (e.g., Sockett, 2014; Sundqvist, 2011). Nevertheless, these learners might still be highly engaged in the task at hand—as noted by Svalberg (2009), individuals who are using the language “for purely communicative ends may be ... engaged but not *Engaged with language*” (p. 244; emphasis added). Given this distinction between engagement with language as the object, versus engagement in tasks with a communicative focus, the following fourth dimension was proposed as an additional variable to consider in the context of second language acquisition:

- (d) *Linguistic engagement*: the extent to which learners consciously focus on processing linguistic features and improving their language skills

Linguistic and cognitive engagement are closely related because they both describe the learner's mental state while engaging in a certain activity. Yet, whereas cognitive engagement, in the current study, describes the total amount of attention an individual brings to an informal activity (vs. other activities or stimuli), linguistic engagement refers to the part of this attention that is devoted to consciously noticing, decoding, and learning specific language features. The difference between the two concepts can be illustrated using the analogy of a mathematical vector, an arrow that represents both the strength and direction of a force. Cognitive engagement corresponds to the length of the vector, or the strength of participants' concentration on the activity at hand. Linguistic engagement describes the vector's direction, or where learners focus their conscious thought processes (i.e., focus on form vs. meaning). Reinders and Nakamura (2021) draw a similar distinction between the amount of attention learners bring to a task, versus their noticing particular features: “Words like ‘attention’ and ‘noticing’ are related in that they refer to the psychological processes by which cognitive resources are directed towards certain stimuli. Being engaged in a learning activity may lead to a greater likelihood of noticing something, but it is not a necessary outcome” (p. 137). Because learners who are not focused on the task at hand are unlikely to notice any particular language feature, cognitive engagement is likely a necessary precursor to linguistic engagement. Dörnyei and Kormos (2000) similarly argued that cognitive engagement is an initial condition for language processing to take place.

In proposing the addition of a linguistic engagement dimension, my aim is to increase the applicability of the engagement construct to the study of language learning in various contexts. The extent to which individuals pay conscious attention to linguistic forms during meaning-focused activities—and whether this is necessary or even beneficial to language learning—has been a topic of much debate, both in the field of informal second language learning (e.g., Cole, 2015; Hiver et al., 2021b; Sockett, 2014; Sundqvist, 2011) and in the broader SLA literature (e.g., Roehr-Brackin, 2018). Linguistic engagement thus creates a link to the vast bodies of work on *noticing, intentional and incidental language learning* (Hiver et al., 2021b), as well as research on *language learning strategies* (e.g., Oxford, 2016), that is, approaches that individuals employ to help them comprehend or learn new linguistic features. Although various taxonomies of strategies cover both mental processes (e.g., inferring the meaning of a new word from context or creating a mental image to aid memorization) as well as observable behaviors (e.g., looking up a word in the dictionary or taking notes) (see review in Grenfell & Macaro, 2007), it has been argued that the latter can be viewed as overt manifestations of underlying cognitive strategies (Macaro, 2006). According to the definitions employed in the current study, using strategies to try and learn new linguistic features is an expression of linguistic engagement.

Measuring informal second language engagement

Prior informal second language learning studies have largely employed retrospective questionnaires asking learners either to recall the informal activities in which they engaged during a specific time in the past, or to report their “typical” experience with ISLPs (e.g., Kuppens, 2010; Kusyk, 2017; Lee & Dressman, 2017). These methods imply that engagement in ISLPs is a relatively stable “trait variable” (Eccles, 2016; Oga-Baldwin, 2019). However, both in the broader field of SLA and in the study of engagement in particular, researchers are increasingly calling for more granularity to gain clearer understanding of language acquisition processes (Hiver et al., 2021b).

In addition, research participants may find it difficult to recall, summarize, and generalize their experiences in retrospective questionnaires, especially when their experiences vary considerably over time, and any small inaccuracies may quickly amplify when estimates are summed or averaged across many items targeting different informal activities (see discussion in Arndt, Granfeldt & Gullberg, 2021). What is more, it is well documented in the broader psychological literature that the accuracy of retrospective questionnaire responses can be affected by a range of cognitive biases, such as a general tendency to overestimate the duration and frequency of past events and to give more weight to more recent and more emotionally salient experiences (Shiffman et al., 2008; Wearden, 2008). To alleviate some of these concerns, several researchers have instead used daily surveys or diaries to collect data on ISLPs (e.g., Schwarz, 2020; Sundqvist, 2011). While better suited to capturing the dynamic nature of informal learning, they still require participants to recall and summarize their experiences (albeit across a shorter time span) and are therefore subject to similar retrospective survey biases.

In addition to concerns over measurement quality, prior research on informal second language learning has also been limited by its strong focus on frequency and quantity (behavioral engagement), while largely disregarding other aspects of engagement, such as learners’ thoughts and emotions. As noted by Briggs Baffoe-Djan and

Zhou (2020) in their discussion of approaches to measuring informal language contact during study abroad, focusing only on the net quantity of language use “is an unhelpful practice because it implies that all types of contact are equally beneficial (or not)” (p. 6).

The ISLE was therefore designed as an alternative instrument for the longitudinal study of engagement in informal second language practices that captures rich, situated data not just on the behavioral but also the affective, cognitive, and linguistic dimensions of engagement across time. The ISLE consists of an online survey that research participants complete on their mobile phones immediately after every ISLP in which they engage during the data-collection period. This measurement approach represents an “event-contingent” variant of an intensive repeated sampling design called the ESM (Hektner et al., 2011; also Ecological Momentary Assessment, Shiffman et al., 2008; or diary methods, Bolger et al., 2003). Psychologists originally developed the ESM to capture information about individuals’ daily behavior, mood, thoughts, feelings, and the physical and social context in which they occurred (Hektner et al., 2011). The method holds considerable promise for research in SLA because it offers a way to tap into various aspects of task engagement “in real time,” which can either be aggregated across time to provide a picture of learners’ typical experiences or analyzed longitudinally to investigate patterns and changes in engagement, both within and across individuals (Arndt, Granfeldt & Gullberg, 2022). There is also evidence to suggest that aggregated ESM data are more reliable and ecologically valid than retrospective survey data (cf. Bolger et al., 2003; Hektner et al., 2011), especially in regard to cognitive and emotional states, which are even more difficult to recall and generalize than behavior (Thigpen, 2019). I refer interested readers to Arndt, Granfeldt, and Gullberg (2021, 2022) for a more detailed review of the potential of the ESM for SLA research, including discussions of the strengths and limitations of this method compared to other survey approaches, and practical considerations in the implementation of the ESM.

The study

In this article, I report on the development and validation of the Informal Second Language Engagement questionnaire (ISLE), an instrument for gathering data on the behavioral, affective, cognitive, and linguistic dimensions of engagement in informal second language practices. This work was done as part of a larger research project that explored different aspects of engagement in informal second language learning among secondary school students in Germany and how they relate to changes in language proficiency, motivation, and attitudes toward language learning (Arndt, 2019).

As part of the evaluation process of the ISLE, the factor structure of the learner-internal (affective, cognitive, and linguistic) engagement scales was inspected using Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). Furthermore, as a means of assessing convergent validity, a Spearman rank-order correlation analysis was used to judge the extent to which the behavioral engagement data collected using the ISLE corresponded to that obtained through a one-off self-report questionnaire, such as those used in prior informal language learning research.

The analyses reported in this article were guided by the following research questions:

RQ1. What structural model best represents the learner-internal dimensions of informal second language engagement?

RQ2. How does behavioral engagement as captured by the ISLE compare to data obtained using a one-off, general self-report questionnaire?

Participants

Across the three stages of instrument development and validation, data were collected from a total of 382 students who were recruited through volunteer sampling from 21 tenth-grade classes at 18 schools in Germany. Both the students and their parents or guardians were provided with detailed information about the study and invited to contact the researcher with further questions before giving their consent for participation. Because subject curricula in Germany vary by state and school type, all participants were recruited from *Gymnasien*¹ in Lower Saxony to control for broad variations in formal language education. In addition, multilevel models were employed to account for teacher/class effects.

At the time of the study, the participating students were receiving 2¼ hours of English instruction per week (down from 3 hours in grades 5–9). In terms of language contact beyond the classroom, it is important to note that in Germany most popular media are dubbed (including films and TV programs, but also video games localised for the German market). That is to say, participating in informal English-language activities likely constitutes more of an active choice for learners in Germany than in countries where subtitled television programs constitute a major source of exposure to English (e.g., Sweden, the Netherlands, or Belgium). The students in this study all had access to internet-connected computers and/or mobile devices at home and, at 15 to 16 years old, had a high degree of autonomy in their recreational activities and their internet usage. Thus, they had plenty of opportunities to choose to engage in English-language leisure activities, particularly online.

In the first stage of this study, 47 pupils (two classes at one school) participated in semistructured focus group interviews which informed the initial ISLE questionnaire construction. Subsequently, the first questionnaire version was piloted with a group of 77 pupils (four classes at two schools). After some adjustments, data for further psychometric evaluation was collected from another sample of 258 pupils (15 classes at six schools). An overview of the demographic characteristics of the participants across the three stages is provided in Table 1.

Table 1. Demographic characteristics of participant samples across the three stages of data collection

	Stage 1: Focus group interviews informing instrument development	Stage 2: Instrument piloting	Stage 3: Instrument evaluation
Total number of participants	47	77	258
Girls	37 (79%)	55 (71%)	137 (53%)
Boys	10 (21%)	21 (27%)	119 (46%)
Gender nonbinary	—	1 (1%)	2 (1%)
Monolingually raised	41 (87%)	67 (87%)	231 (90%)
Multilingually raised	6 (12%)	10 (13%)	27 (11%)
Mean age (SD)	15.40 (.40)	15.38 (.52)	15.68 (.65)
Mean number of years learning English (SD)	7.64 (1.21)	7.70 (1.05)	7.89 (1.22)

¹*Gymnasien* are secondary schools comparable to grammar schools and sixth form colleges in the United Kingdom, whose primary purpose is to qualify students for entry into university or other forms of higher education (although 20–25% of final-year students report aiming for other paths postgraduation, e.g., entering employment or vocational training; Schneider & Franke, 2014). Nevertheless, given the increasingly important role that English plays in the German labor market and secondary education (Davydova, 2020), it seems likely that the majority of the students at these schools would consider English to be highly relevant for their personal future.

Altogether, 229 girls (60%) and 150 boys (39%) participated in the study, in addition to three students who self-identified as nonbinary. All learners were 15–16 years old ($M = 15 \text{ yrs } 8 \text{ mos}$, $SD = 8 \text{ mos}$), and most had grown up as monolingual speakers of German ($n = 339$; 89%). Of the pupils with multilingual backgrounds ($n = 43$; 11%), the vast majority grew up speaking German alongside Russian or Kurdish. One participant was excluded from the analysis in Stage 1 who was German-English bilingual-from-birth, as were two participants in Stage 3 who were nonnative speakers of German (one had recently come to Germany as a refugee, and one foreign exchange student).

Of the remaining 378 participants, the vast majority had learned English as a foreign language since the age of eight, or the third year of primary school, when the language is most commonly introduced in the state of Lower Saxony. Therefore, at the time of the study, they had been learning English for about eight years ($M = 7 \text{ yrs } 9 \text{ mos}$, $SD = 1 \text{ yr } 2 \text{ mos}$). The English teachers estimated that their pupils had, on average, B1-B2 proficiency according to the Common European Framework of Reference (CEFR; Council of Europe, 2001). This corresponds to what the curriculum for Gymnasien in Lower Saxony stipulates year-10 students should achieve (Böwing et al., 2006).

Procedure

Stage 1: Instrument development

It has been argued that using learner insights as the basis for item development increases the quality and suitability of research instruments (Dörnyei, 2010). Therefore, I invited learners from the target population to speak about their informal second language practices during exploratory focus group interviews. I expected that focus groups would elicit richer data than individual interviews because being part of a group (not feeling “put on the spot”) and listening to others’ opinions may help participants identify and reflect on their own experiences (Ho, 2006; Krueger & Casey, 2015).

Semistructured interviews were conducted with groups of four to six students. Each session lasted between 40 and 60 minutes. Wherever possible, the students were allowed to form their own interview groups, in the hope that being with their friends would make them feel more at ease and free to speak their mind. The interview language was German, and the questions (Arndt, 2023) were exploratory, to offer the researcher the chance to understand how and why these students engaged in ISLPs. Informal practices were defined as any activities involving L2 reading, writing, speaking, or listening in which the participants engaged in their free time, which were not explicitly designed for language learning (i.e., materials designed for independent study) and not assigned by others (e.g., teachers or parents). Although the learners were asked about activities that involved the use of any language other than their native language(s), almost all reported that they did not engage in activities in any L2 other than English (which thus became the focus of the remainder of study).

The interviews were audio-recorded and later transcribed by the researcher for *directed content analysis*, a process involving both deductive and inductive analysis, wherein a researcher “draw[s] on existing research and theory to inform the ... identification of emergent categories” (Renninger & Bachrach, 2015, p. 63). A primary focus of this analysis was to get an idea of the nature of these students’ behavioral, affective, and cognitive engagement in ISLPs. The findings were particularly valuable in showing the kind of language these students used to talk about their experiences, which subsequently helped to guide the selection and (re)phrasing (i.e., adaptation and

generation) of questionnaire items for the ISLE (see the following text). Furthermore, the qualitative findings inspired the proposal of linguistic engagement as a separate dimension from cognitive engagement because the vast majority of interviewed learners reported that they rarely or never paid attention to the language in informal L2 activities (low linguistic engagement), even though they often felt completely focused on or immersed in activities which they enjoyed (high cognitive and affective engagement).

Behavioral engagement was operationalized in the ISLE as the self-reported duration and type of activity in which the students had participated just before filling in the questionnaire, whereas cognitive, affective, and linguistic engagement were targeted using introspective self-report questions with semantic differential response scales (e.g., “All in all, this activity was [boring – entertaining]”). The scale labels were displayed in the online questionnaire above a 100-point slider scale, providing an easy-to-understand and intuitive way for participants to express the full range of their cognitive and emotional states (as opposed to being limited by a small number of scale steps in a Likert-style check-box design).

The 12 initial affective and cognitive questionnaire items were adapted from existing instruments for measuring engagement in educational studies (Appleton et al., 2006; Cox, 2014; Egbert, 2003; Flowerday & Schraw, 2003; Fredericks et al., 2005; Philp & Duchesne, 2016; Skinner et al., 2009; Storch, 2008; Svalberg, 2009). No one instrument could be adopted in full since only a few focused exclusively on engagement at the task level (as opposed to school- and subject-level engagement), and all contained items referring explicitly to features of formal learning environments (e.g., teachers, peers, learning tasks). Meanwhile, the six initial linguistic engagement items had to be newly formulated because linguistic engagement was a new dimension in the current study. The adaptation of existing items, and the formulation of new questions, were based on the working definitions of the different engagement dimensions used in this study, as well as insights from the focus group interviews, which gave an indication of how the learners conceptualized their thoughts and emotions in relation to ISLPs.

Two second language learning specialists were invited to examine the initial questionnaire. They indicated six items that they thought worked less well for operationalizing the constructs as defined in this study, leaving 12 items in the first ISLE draft (four each for affective, cognitive, and linguistic engagement). The questionnaire was constructed in English and translated to German before piloting. The translation was guided by the language that the focus group participants had used to speak about their experiences and improved through an iterative process (following best-practice guidelines by Dörnyei, 2010) involving back-translation and feedback from specialist and nonspecialist peer reviewers (i.e., researchers in SLA and in broader educational sciences).

Stage 2: Piloting

Pilot data were collected from 77 learners who, as part of the larger study, also completed a C-test and a questionnaire about their language learning motivation, attitudes toward informal and formal language learning, and general background information. After completing the language test and background questionnaire, the participants were instructed to save the hyperlink of the ISLE online survey to their mobile phones and to complete it after every time they used English during their free time over the next seven days.

The data collected in this way were analyzed in SPSS (version 25, IBM Corp, 2017) using Exploratory Factor Analysis (EFA) to uncover the underlying factor structures, that is, whether the items designed to measure different learner-internal aspects of engagement (affective, cognitive, and linguistic dimensions) loaded onto distinct latent constructs (e.g., Tavakol & Wetzel, 2020). EFA was also used to reduce the number of items by selecting those that most strongly represented each factor. The internal consistency of the updated scales was assessed using Cronbach's alpha. Behavioral engagement was not included in these analyses, or subsequent Confirmatory Factor Analysis (CFA), because it was operationalized through manifest, not latent, variables (i.e., the average time that participants reported spending on different types of ISLPs).

Instead, the behavioral engagement data were explored descriptively, for example to confirm that no common types of informal activities had been overlooked in the questionnaire construction. Furthermore, a closer look was taken at the timestamps that indicated when each survey response was submitted. These suggested that most participants were indeed recording informal L2 experiences at plausible times of the day (i.e., spread across nonschool hours). However, one pilot study participant submitted a single survey response at the end of each day, listing all of the ISLPs in which they had engaged that day and providing what were presumably aggregate estimates for the duration and affective, cognitive, and linguistic engagement. This was subsequently addressed through updating the participant instructions to stress that a new survey should be completed "**every time**" (emphasis in original) a student engaged with English outside of school. The analysis of the preliminary behavioral engagement data also brought to light the issue that missing data from cases in which a learner had not engaged in any informal L2 activities on a given day could not be distinguished from instances where they had forgotten, or purposefully neglected, to submit survey responses (for a more thorough discussion on the implication of missing data in this study, see Arndt & Rose, 2022). In the subsequent stage, the response option "*I didn't use English outside of school today*" was therefore added to the ISLE, so that participants who had not used English informally on a given day could report this.

Stage 3: Instrument evaluation

The revised version of the ISLE was used to gather data from another 254 students. As in the previous stage, these participants also completed a C-test and the questionnaire covering background information, language learning motivation, and attitudes toward formal and informal learning. As part of the background questionnaire, the participants were also asked to estimate their "typical" frequency of engagement in informal L2 activities (Appendix A), for comparison to the ISLE data.

The affective, cognitive, and linguistic engagement data collected during this stage were analyzed using CFA instead of EFA. EFA, as a "bottom-up" method, is used to explore the observed covariances in a dataset (e.g., Tavakol & Wetzel, 2020). Therefore, the extracted model can change drastically with small shifts in the data, so that EFA can only provide preliminary evidence of the factor structures underlying a set of measured indicators. By contrast, CFA is a "top-down" method, which assesses how consistent a set of data are with a specified factor structure (ibid.). In this study, CFA was therefore used to evaluate the extent to which the second, larger set of ISLE data reflected the hypothesized distinction between the latent variables affective, cognitive, and linguistic engagement. A two-factor model with a joint cognitive-linguistic engagement factor was also tested, specifically to evaluate the extent to which the proposed distinction between cognitive and linguistic engagement was supported in

the data. The CFA analyses were carried out using the *lavaan* package in R (version 0.6-3; Rosseel et al., 2018).

A range of fit indices were used to assess model fit, including the traditional chi-square test, which measures the difference between the predicted and observed item covariances observed. A nonsignificant chi-square statistic ($p > .05$) is desirable, indicating that the expected and observed covariances do not differ significantly. An advantage of the chi-square statistic is that it can be used to directly compare “nested” model solutions (cf. Blunch, 2013; Byrne, 2010). However, the test is extremely sensitive to sample size and almost always reaches significance in larger samples (Byrne, 2010). Therefore, the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) were also consulted, which compare the specified model to a saturated model (wherein all components are related to all other components, resulting in maximum fit) and an independence model (wherein all variables are unrelated, resulting in minimum fit). CFI and TLI values above .90 are generally considered “acceptable” and values above .95 “excellent” (Blunch, 2013; Byrne, 2010). Finally, the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Squared Residual (SRMR) were considered, which indicate how closely the specified model reproduces the covariances among the observed variables. Lower values denote smaller differences between the observed and estimated covariances (the residuals). Values below .10 are commonly cited as “acceptable” and values below .06 as “excellent” (Blunch, 2013; Byrne, 2010).

The factor loadings and residual variances derived from the CFA model were further used to calculate the composite reliability for the affective, cognitive, and linguistic engagement factors. This measure of internal consistency was chosen because the CFA showed that the assumption underlying traditional Cronbach’s alpha, that all items perfectly represent their latent factors (i.e., have factor loadings equal to one with no error covariances), did not hold true. Composite reliability takes into account the actual error variances, and the resulting coefficient can be interpreted similarly to Cronbach’s alpha, with scores closer to one indicating greater internal consistency (Hair et al., 2016).

Finally, Spearman rank order correlations were calculated to compare the behavioral engagement data obtained using the ISLE to the data from the background questionnaire about how frequently the same participants “typically” engaged in a variety of informal activities. Due to the novelty of the ISLE (and the ESM approach through which it was administered), this comparison was used to explore the extent to which the two data-collection methods provided a similar picture of the quantity of participants’ engagement in informal second language practices (i.e., examine their convergent validity).

Results

Affective, cognitive, and linguistic engagement

The analyses in this section focus on RQ1: What structural model best represents the learner-internal dimensions of informal second language engagement? The dimensionality of these engagement scales was inspected using EFA (Stage 2) and CFA (Stage 3). A particular focus during both analyses was whether the proposed theoretical distinction between cognitive and linguistic engagement was justified by the available data.

Exploratory Factor Analysis

The EFA was conducted using per-person, per-week average scores for each affective, cognitive, and linguistic engagement item (descriptive statistics in

Appendix B). Preliminary analyses showed that the majority of these scores were normally distributed, although two were significantly skewed ($|z_{\text{skewness}}| > 1.96$; cf. Kim, 2013). Thus, Principal Axis Factoring was chosen as an extraction method that is considered robust against nonnormality (Finney & DiStefano, 2008). Because it was expected that the various engagement factors would be highly intercorrelated, an oblique rotation (Direct Oblimin) was applied. The Kaiser–Meyer–Olkin measure was used to verify the sampling adequacy (overall $KMO = .80$ and all individual item $KMO > .60$) and suggested that, despite being comparatively small, the sample size ($n = 77$) was adequate for the planned EFA analysis (see also De Winter et al., 2009).

A preliminary examination revealed one cognitive engagement item (“It was [very easy – very difficult] to keep my mind on the activity”) that did not fit well within the overall item pool (few correlations > 0.3 with other items; lowest overall shared variance at 38%). After excluding this item from further analysis, three latent factors with eigenvalues above Kaiser’s criterion of 1 (Kaiser, 1970), were identified, which together explained 62.40% of the total variance in engagement scores. The scree plot supported the decision to extract these three factors. The factor loadings after rotation are listed in Table 2. Factor 1 explained 39.8% of the total variance, Factor 2 19.8%, and Factor 3 11.5%. The items loading most strongly onto each of these factors suggest that they represented cognitive, linguistic, and affective engagement respectively.

The EFA resulted in the deletion of three more items that significantly loaded ($\geq .30$) onto more than one factor (Items 9–11, Table 2). Of the remaining eight items, four clearly loaded onto cognitive engagement and two each onto affective and linguistic engagement. This imbalance resulted, in part, from the fact that two items intended to target these factors (Items 2–3, Table 2) loaded onto cognitive engagement instead.

Table 2. Exploratory factor analysis, factor loadings after rotation

Item	Factor 1 Cognitive Engmt.	Factor 2 Linguistic Engmt.	Factor 3 Affective Engmt.
1. All in all, I was [not at all – completely] focused on the content.	.84	-.16	.07
2. I was [not at all – completely] focused on trying to understand every single word.	.74	.28	-.18
3. The content <u>activity</u> was [not at all interesting – very interesting] to me.*	.73	-.03	.29
4. During this activity, I [did not concentrate at all – concentrated hard] on it.	.59	.20	.02
5. All in all, I thought [very little – a lot] about the language during this activity.	-.01	.89	.24
6. All in all, I was [not at all – completely] focused on the language.	.28	.55	-.11
7. All in all, this activity was [boring – entertaining].	-.18	.09	.86
8. I [really did not like – really liked] this activity.	.07	-.07	.84
9. I felt [not at all – completely] absorbed in this activity.	.32	.13	.49
10. During this activity, I felt [confident – overwhelmed]. (reverse-coded item)	.33	-.39	.46
11. This activity was [very – not at all] beneficial to improving my English. (reverse-coded item)	.30	.12	.42

*The wording of this item was modified after the pilot study.

While the three scales showed acceptable internal reliability² (Cronbach's $\alpha_{cog} = .85$; $\alpha_{ling} = .73$; and $\alpha_{aff} = .84$), it is generally recommended that multiitem scales should be comprised of a minimum of three items (e.g. Dörnyei, 2010). Therefore, in addition to rewording Item 3 to target affective engagement more clearly, a new item was also added that it was hoped would more directly probe linguistic engagement, or the extent to which participants focused on specific language features, versus the overall meaning, in their informal practices (“During this activity, I was primarily focused on [the general meaning – individual words/phrases]”). The measurement properties of the modified/new items were examined as part of the continued instrument evaluation using CFA in Stage 3. The final list of affective, cognitive, and linguistic engagement items is included in Appendix C, and the full version of the ISLE questionnaire (including behavioural engagement questions) can be downloaded from the IRIS database (Arndt, 2023).

Confirmatory Factor Analysis

To investigate whether the linguistic engagement dimension could justifiably be regarded as distinct from cognitive engagement, two models were compared using CFA (see Figure 2):

Model 1: A three-factor model with affective, cognitive, and linguistic engagement as distinct but correlated latent factors.

Model 2: A two-factor model with affective and cognitive-linguistic engagement as distinct but correlated latent factors.

To account for the hierarchical structure of the ISLE data, which consisted of a variable number of data points (total responses $n_{entries} = 1786$) nested within participants ($n_{students} = 258$), a Multi-level Confirmatory Factor Analyses (MCFA) was carried

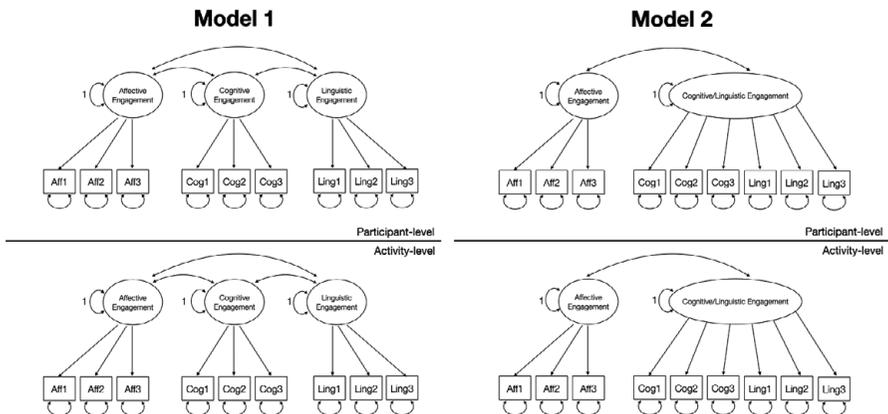


Figure 2. Hypothesized two- and three-factor correlated models of affective, cognitive and linguistic engagement.

²The appropriateness of Cronbach's alpha with two-item scales is disputed, as it “almost always underestimates the true reliability, sometimes rather substantially” (Eisinga et al., 2013, p. 641). However, this was not considered a problem in the current study because the statistic returned acceptable values for these scales.

out instead of using aggregated per-person engagement scores as in the preceding EFA. MCFA not only allows for the separate investigation of engagement patterns across two levels (activities and participants) but also the calculation of factor scores that accurately reflect each participants' general levels of engagement, which are less biased than simple item averages because they exclude activity-level variance.

Preliminary analyses (Appendix D) indicated that the engagement data were approximately normally distributed (absolute values³ of skewness and excess kurtosis <2), confirming that it was appropriate to use Maximum Likelihood estimation (Byrne, 2010). Two initial MCFAs were conducted to compare the fit of the two structural models shown in Figure 2. These models were distinct in the number of correlated, latent engagement factors (ovals) onto which the nine questionnaire items (rectangles) loaded. Each item was constrained to load only onto one factor (downward arrows from latent to observed variables) and the error terms (curved double-headed arrows next to each variable) were uncorrelated. The error variance of the latent variables was fixed to one to allow for model identification (Blunch, 2013; Byrne, 2010). The model structure was the same at both the activity and participant levels.

During the initial analysis, both models reached just below acceptable overall levels of fit (see Table 3), although Model 1 performed significantly better than Model 2 ($\Delta\chi^2[4] = 128.42, p < .001$). To identify the possible sources of misfit, the modification indices (MI) and item factor loadings were consulted. There is some disagreement about the extent to which it is appropriate to make changes to CFA models based on these indices: Some argue that researchers should return to EFA to find, and subsequently retest, a better fitting model (e.g., Browne, 2001) while others find it acceptable to exclude items identified as contributing to bad model fit, or to correlate some error variances within the same factor, but also warn that changes should be applied only where a theoretical explanation can be found (e.g., Blunch, 2013; Byrne, 2010; Schmitt, 2011). As limited time and resources in the current study precluded returning to the initial instrument design phase, the researcher decided to consult the MIs to identify sources of model misfit wherever necessary and to consider the suggested changes in light of the theoretical definitions of the latent constructs, the wording of the individual items, as well as the findings of the EFAs in Stage 2.

In both models, the third linguistic engagement item (Ling3: "During this activity, I focused on [the general meaning – individual words/phrases]") stood out due to its exceptionally low factor loadings at the activity level (.03 and .15) and negative loadings at the participant level. The MIs also indicated a high, unexpected cross-loading of this

Table 3. Goodness-of-fit statistics of the two- and three-factor MCFA models

Model	χ^2	<i>df</i>	<i>p</i>	CFI	TLI	RMSEA	SRMR _{a.l.} *	SRMR _{p.l.} *
<i>Model 1</i>	502.24	48	< .001	.85	.78	.09	.10	.25
<i>Model 1, excl. Ling3</i>	297.26	34	< .001	.92	.87	.08	.07	.19
<i>Model 1, Cog2 loading on linguistic engagement</i>	220.69	34	< .001	.95	.95	.05	.04	.06
<i>Model 2</i>	630.66	52	< .001	.80	.72	.10	.10	.25
<i>Model 2, excl. Ling3</i>	436.37	38	< .001	.87	.82	.10	.10	.20

*The subscripts *a.l.* and *p.l.* refer to the activity and participant levels, respectively.

³Absolute values were consulted instead of z-scores due to the very large number of data points (Kim, 2013).

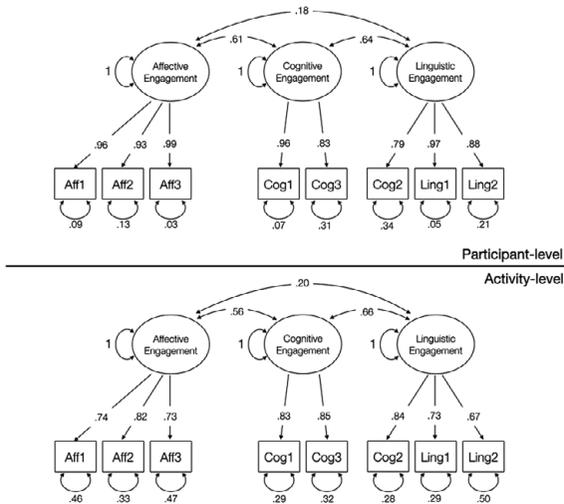


Figure 3. Final three-factor model of affective, cognitive, and linguistic engagement.

item onto affective engagement. Together, these findings suggest that this item was responsible for a large part of the model misfit and was not a good measure of linguistic engagement. A possible explanation for this may be that the item targeted only engagement with lexical features while ignoring conscious processing of, for example, grammar and phonology. After exclusion of this item, the three-factor model still achieved a significantly better fit than the two-factor model ($\Delta\chi^2[4] = 139.11, p < .001$). Together with the overall higher factor loadings in Model 1, this suggested a stronger structural validity of the three-factor solution.

While the model fit improved significantly after excluding Ling3 ($\Delta\chi^2[14] = 363, p < .001$), the *TLI* and *SRMR_{p.l.}* for Model 1 were still not quite within the range for good model fit. The MIs suggested that one cognitive engagement item (Cog2 “I was [not at all – completely] focused on trying to understand every single word”) exhibited a more substantive loading onto linguistic engagement. This was not surprising, given that this item had initially been designed to measure linguistic engagement (but had emerged as loading more strongly onto cognitive engagement in the EFA). Allowing this item to instead load onto linguistic engagement resulted in excellent model fit (see Table 3). Figure 3 shows the standardized factor loadings, error variances, and latent factor correlations of this final model.

Finally, the composite reliability coefficients for affective, cognitive, and linguistic engagement at both the activity and participant level (Table 4) were calculated using the standardized factor loadings and error variances mentioned in the preceding text. They indicated good internal consistency for the three scales (all > .80; Hair et al., 2016).

Table 4. Composite reliability of affective, cognitive, and linguistic engagement scales

	Affective engagement	Cognitive engagement	Linguistic engagement
Activity level	.81	.82	.82
Participant level	.97	.89	.92

Table 5. Spearman rank order correlations between self-reported frequency of engagement in different informal activities and frequencies derived from the ISLE data (mean number of entries per day)

Activity	r_s	p	n
Books	.22	< .001	254
Magazines	.09	.170	255
Music	.29	< .001	258
Films	.22	< .001	258
TV series	.42	< .001	257
Social media	.24	< .001	258
Video games	.48	< .001	254
Total	.46	< .001	258

Behavioral engagement

RQ2 was addressed using Spearman rank order correlations to assess the strength of association between the ISLE data (mean number of entries per day) and the participants' estimates of their frequency of engagement in ISLPs in the self-report questionnaire administered as part of the larger study (Table 5).

Overall, a moderate correlation was found between the frequency of engagement in ISLPs as reported through the two questionnaires ($r_s = .46, p < .001$). Furthermore, separate analyses by activity type revealed that the two measures corresponded more closely for watching TV series and playing video games ($r_s = .42$ and $.48$, respectively) than for reading books, watching films, listening to music, and social networking (r_s between $.22$ and $.29$). The correlations were statistically significant ($p < .001$) for all activities except reading magazines ($r_s = .09, p = .170$).

Discussion

Affective, cognitive, and linguistic engagement

The first part of this article focused on investigating the factor structure underlying the ISLE questionnaire items for affective, cognitive, and linguistic engagement (RQ1). A particular focus was on the proposed distinction between cognitive and linguistic engagement, as a new addition to the engagement construct, to increase its applicability in the SLA context. Both the EFA and MCFA results supported the conceptualization of these two dimensions as separate, but related variables. As hypothesized, a model with three factors (corresponding to affective, cognitive, and linguistic engagement) emerged as the best EFA solution. In the MCFA, too, the three-factor model fit the available data significantly better than the two-factor solution, in which cognitive and linguistic engagement were combined into a single factor.

Some items that strongly cross-loaded on both cognitive and linguistic engagement were consequently excluded from the analyses. Therefore, the final model (Figure 3) only included two cognitive engagement items. While it is essential that the ISLE contain only a small number of questions (to reduce burden and increase compliance when participants are required to fill in the same survey over and over again), latent variables should generally be represented by at least three items (Dörnyei, 2010; although it has also been shown that single- or dual-item measures can be reliable and valid if they are clear and unambiguous to the respondents; Ainley & Patrick, 2006). Thus, further questions targeting cognitive engagement should be trialed as part of future iterations of the ISLE.

The cognitive and linguistic engagement factors were found to correlate strongly in the MCFA, both at the activity and the participant level. However, similarly strong correlations were also found between cognitive and affective engagement, whereas linguistic and affective engagement only correlated weakly. The asymmetry of these correlations supports the conclusion that affective, cognitive, and linguistic engagement constitute three separate but highly interrelated dimensions of engagement, as hypothesized.

Indeed, the correlations between these factors provide some suggestions for how the different engagement dimensions may dynamically interact. The qualitative data from the focus group interviews in Stage 1 suggested that the amount of attention learners pay during ISLPs (cognitive engagement) is largely determined by how interesting and enjoyable they find these activities (affective engagement). A certain amount of attention, in turn, is required for noticing and actively decoding language features (linguistic engagement). These relationships seem to be reflected in the strong correlations between affective and cognitive engagement, and cognitive and linguistic engagement, respectively. Meanwhile, some (usually less proficient) learners reported in the interviews that, to reach a sufficient level of understanding, they had to focus closely on decoding certain language features—and these same learners also described deriving less enjoyment from engaging in informal activities than their peers. This implies a negative effect of linguistic engagement on affective engagement, which could explain the weaker correlation between these two factors in the MCFA.

Of course, relationships between the various engagement dimensions cannot be inferred from correlational data alone. Longitudinal studies are needed to gain a better understanding of how these factors interact and change (within the same informal activity, or over longer periods), the variables that influence these changes, and the ways in which they relate to informal second language learning outcomes.

Behavioral engagement

The second set of analyses focused on comparing the ISLE behavioral engagement data to the type of general self-report questionnaire commonly used in other studies of informal second language learning. Overall, the frequency of engagement in various ISLPs as indicated by the ISLE data correlated only moderately to the one-off questionnaire data, suggesting that both methods provided a slightly different picture of the participants' ISLPs. Although, without a reference criterion, we cannot know which instrument provided more accurate data, research has shown that general self-reports can be vulnerable to measurement errors from cognitive biases and recall failure (see reviews in Hektner et al., 2011; Shiffman et al., 2008). In turn, event-contingent methods, such as the ISLE, might be more affected by reactance (participants changing their behaviors as a result of taking part in the study, e.g., due to increased awareness) and habituation (decreased reflexivity and accuracy due to repeatedly completing the same survey) (Bolger et al., 2003).

Nevertheless, there is reason to believe that the ISLE might be well suited to capturing patterns of informal behaviors. For example, research on everyday internet use (comparing mobile tracking and survey data) found that individuals who went online frequently but briefly were less accurate in their retrospective survey reports (Araujo et al., 2017). In the current study the correlations between the ISLE and general questionnaire data were lower for browsing social media and listening to music (i.e., activities in which most learners engaged frequently but briefly) than for longer,

continuous practices such as watching TV or playing video games. Correlations were similarly lower for watching films and reading books, whereas reading magazines was the only activity for which the ISLE data did not significantly correlate with the self-report survey. All these were activities in which most of the learners in this study rarely or never engaged. As noted by Hektner and colleagues (2011), such infrequently occurring phenomena might be better captured by event-contingent methods (like the ISLE) because participants might deem them too insignificant to remember or report in general questionnaires. In light of these considerations, the findings of this study tentatively support the conclusion that the ISLE may better capture certain aspects of informal second language use than general self-report surveys, in addition to also providing insight into learners' concurrent cognitive and emotional states.

Conclusion

This article has reported on the development and validation of an instrument for measuring various dimensions of engagement in informal second language practices. In addition to the outwardly observable aspects of learners' interactions with informal L2 activities (behavioral engagement), the theoretical framework that formed the basis for this questionnaire specified three learner-internal aspects of engagement: affective, cognitive, and linguistic engagement. The results of the EFA and CFA suggested that these were indeed distinguishable, but closely correlated constructs. Linguistic engagement had been proposed in the current study as a new dimension of engagement, with the idea of being able to separate the overall strength of learners' task focus (cognitive engagement) from the direction of their attention (linguistic engagement, i.e., extent of focus on form). The results of the factor analyses tentatively supported the distinction between cognitive and linguistic engagement as separate dimensions, although there is a need to distinguish the two constructs more clearly in future iterations of the ISLE and/or other measures engagement in ISLPs and to investigate their role in language development. Future studies could also include other dimensions of engagement that have been suggested to be of importance in SLA but were not included in this questionnaire, particularly social engagement, which several researchers have argued to be a key dimension of engagement in (formal) language learning (Hiver et al., 2021a; Mercer, 2019).

In terms of behavioral engagement, this study also found that the data collected using the ISLE correlated moderately well to that gathered through a more general self-report questionnaire. This indicates that both instruments provided a somewhat different pictures of how frequently participants engaged with different types of ISLPs. Drawing on other research that has compared similar methodological approaches, it was suggested that the ISLE might more accurately capture engagement in very infrequently occurring activities (e.g., reading magazines), or in activities that are frequent but brief (e.g., browsing social media).

Of course, the limitations of the ISLE must also be acknowledged at this point. First and foremost, the instrument, by design, provides only a snapshot of learners' informal practices across the specific data-collection period—unlike most self-report questionnaires, which ask respondents for more generalized accounts of their “typical” or recent experiences. Previous research suggests that ISLPs can vary substantially from day to day as well as across longer periods (Kusyk, 2017; Schwarz, 2020), for example as personal interests and free time available for engaging in such activities change. Thus, researchers must carefully consider when to administer the ISLE. For example, in the

current study the data-collection period was scheduled during regular school term time, that is, not close to exam weeks in which students would likely have less time to participate in ISLPs. Data were collected across seven days to capture both weekday and weekend practices. ESM researchers tend to agree that a 7-day period is likely to yield a reasonably representative sample of typical daily activities without placing excessive burden on the research participants (Hektner et al., 2011).

Secondly, the instrument evaluation in the current study focused on the latent factor structure of the ISLE and comparing the behavioral components to a one-off questionnaire to explore convergent validity. More research is needed to fully examine the psychometric properties of the new instrument, including discriminant and criterion validity. Along similar lines, it must be stressed that the ISLE was developed and used to conduct research specifically with learners of L2 English in Germany. As mentioned previously, researchers have stressed that ISLPs in languages other than English are likely to look very different, due to the special social and cultural status of English around the world. In addition, the participants in this study were enrolled in a particular type of German secondary school (*Gymnasium*), in which most students aim for entry into higher education. This suggests that they might perhaps be more interested, and proficient, in English than other groups of secondary school students. Nevertheless, these specific sample characteristics do not inherently mean that the ISLE cannot be used to study the engagement patterns of other groups of learners. In fact, the tool might help researchers to better understand the ways in which L2 engagement could relate to the general cultural status of different target languages and the role they play in the learners' personal, academic, and social context. Of course, use of the ISLE (like any instrument) in contexts other than the one in which it was developed requires further validation through studies with different learner populations and target languages.

Finally, we must also consider participant compliance, that is, whether the students provided reports of all informal L2 activities in which they engaged (see discussion in Arndt & Rose, 2022). Compliance is affected by the amount of effort required to complete a questionnaire, which must be considered when choosing a research instrument. The ISLE requires substantial self-directed effort from participants, especially those who more frequently engage in informal activities. For this reason, the ISLE might be more appropriate for studying the practices of language learners who live in places where the target language is not commonly spoken and L2 contact is likely to occur less frequently, than for learners living in target-language communities (e.g., in study abroad). Similarly, if the aim is to get an overview of the types of informal activities that are popular among a particular learner population, or of the approximate amount of time that these learners typically come into contact with their L2 beyond the classroom, general self-report questionnaires may be sufficiently accurate. By contrast, the richer data that can be collected through the ISLE has the potential to help researchers better understand individual differences and variability in ISLL, how the various aspects of informal second language engagement interact and change over time, and their role in informal language learning.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/S0272263122000572>.

Data Availability Statement The final version of the Informal Second Language Engagement questionnaire and the focus group interview guide can be downloaded from the IRIS database: <https://www.iris-database.org/details/xFCkb-IxdcC>.

References

- Ainley, M., & Patrick, L. (2006). Measuring self-regulated learning processes through tracking patterns of student interaction with achievement activities. *Educational Psychology Review*, 18, 267–286. <https://doi.org/10.1007/s10648-006-9018-z>
- Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the student engagement instrument. *Journal of School Psychology*, 44, 427–445. <https://doi.org/10.1016/j.jsp.2006.04.002>
- Araujo, T., Wonneberger, A., Neijens, P., & de Vreese, C. (2017). How much time do you spend online? Understanding and improving the accuracy of self-reported measures of internet use. *Communication Methods and Measures*, 11, 173–190. <https://doi.org/10.1080/19312458.2017.1317337>
- Arndt, H.L. (2019). Informal second language learning: The role of engagement, proficiency, attitudes, and motivation [Doctoral thesis, University of Oxford, UK]. <https://ora.ox.ac.uk/objects/uuid:c579077d-61fd-4b94-bd57-de7063389122>
- Arndt, H.L., Granfeldt, J., & Gullberg, M. (2021). Reviewing the potential of the experience sampling method (ESM) for capturing second language exposure and use. *Second Language Research*, 39, 39–58. <https://doi.org/10.1177/02676583211020055>
- Arndt, H.L. & Rose, H. (2022). Capturing life as it is truly lived? Improving diary data in educational research. *International Journal of Research & Method in Education*. Advance online publication. <https://doi.org/10.1080/1743727X.2022.2094360>
- Arndt, H.L., Granfeldt, J. & Gullberg, M. (2022). The Lang-Track-App: Open-source tools for implementing the experience sampling method in second language acquisition research. *Language Learning*. Advance online publication. <https://doi.org/10.1111/lang.12555>
- Arndt, H.L. (2023). Informal second language engagement questionnaire and focus group interview guide [Collection of materials]. *IRIS Database*, University of York, UK. <https://doi.org/10.48316/xFCKb-IxdcC>
- Benson, P., & Reinders, H. (2011). *Beyond the language classroom*. Palgrave MacMillan. <https://doi.org/10.1057/9780230306790>
- Blunch, N. J. (2013). *Introduction to structural equation modeling using IBM SPSS Statistics and Amos* (2nd ed.). Sage Publications. <https://doi.org/10.4135/9781526402257>
- Boekaerts, M. (2016). Engagement as an inherent aspect of the learning process. *Learning and Instruction*, 43, 76–83. <https://doi.org/10.1016/j.learninstruc.2016.02.001>
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616. <https://doi.org/10.1146/annurev.psych.54.101601.145030>
- Böwing, C., Frome, W., Gerking, K., Haupt, D., & Schulte, W. (2006). *Kerncurriculum für das Gymnasium Schuljahrgänge 5-10*. [Core curriculum for Gymnasien grades 5–10]. <http://docplayer.org/10838846-Niedersaechsisches-kultusministerium-kerncurriculum-fuer-das-gymnasium-schuljahrgaenge-5-10-eng-lisch-niedersachsen.html>
- Briggs Baffoe-Djan, J., & Zhou, S. (2020). Close encounters of the third kind: Quantity, type and quality of language contact during study abroad. In M. Howard (Ed.), *Study abroad and the second language learner: Expectations, experiences and development* (pp. 69–89). Bloomsbury. <https://doi.org/10.5040/9781350104228.ch-004>
- Brown, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111–150. https://doi.org/10.1207/S15327906MBR3601_05
- Byrne, B. M. (2010). *Structural equation modeling with AMOS* (2nd ed.). Routledge.
- Christenson, S. L., Reschly, A., & Wylie, C. (2012). *Handbook of research on student engagement*. Springer. <https://doi.org/10.1007/978-1-4614-2018-7>
- Cole, J. (2015). *Foreign language learning in the age of the internet*. [Doctoral thesis]. University of Oxford, Oxford, UK.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. <https://rm.coe.int/1680459f97>
- Cox, C. B. (2014). *21st century skills and principles of flor in the foreign language classroom*. [Master's thesis]. Brigham Young University, Provo, Utah.
- Davydova, J. (2020). English in Germany: Evidence from domains of use and attitudes. *Russian Journal of Linguistics*, 24, 687–702. <https://doi.org/10.22363/2687-0088-2020-24-3-687-702>
- De Winter, J. C. F., Dodou, D. & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44, 147–181. <https://doi.org/10.1080/00273170902794206>

- Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing*. Routledge. <https://doi.org/10.4324/9780203864739>
- Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, 4, 275–300. <https://doi.org/10.1177/13621688000400305>
- Eccles, J. S. (2016). Engagement: Where to next? *Learning and Instruction*, 43, 71–75. <https://doi.org/10.1016/j.learninstruc.2016.02.003>
- Egbert, J. (2003). A study of flow theory in the foreign language classroom. *The Modern Language Journal*, 87, 499–518. <https://doi.org/10.1111/1540-4781.00204>
- Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58, 637–642. <https://doi.org/10.1007/s00038-012-0416-3>
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, 59, 117–142. <https://doi.org/10.3102/00346543059002117>
- Finn, J. D., & Zimmer, K.S. (2012). Student engagement: What is it? Why does it matter? In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 97–132). Springer. https://doi.org/10.1007/978-1-4614-2018-7_5
- Finney, S. J., & DiStefano, C. (2008). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. D. Mueller (Eds.), *Structural Equation Modeling: A second course* (pp. 269–314). Information Age Publishing.
- Flowerday, T., & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. *The Journal of Educational Research*, 96, 207–215. <https://doi.org/10.1080/00220670309598810>
- Fredericks, J. A., Blumenfeld, P. C., Friedel, J., & Paris, A. H. (2005). School engagement. In K.A. Moore & L. Lippman (Eds.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 305–321). Springer Science and Business Media. https://doi.org/10.1007/0-387-23823-9_19
- Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59–109. <https://doi.org/10.3102/00346543074001059>
- Grenfell, M., & Macaro, E. (2007). Claims and critiques. In A. D. Cohen & E. Macaro (Eds.), *Language learner strategies: Thirty years of research and practice* (pp. 9–28). Oxford University Press.
- Hair, J. F. J., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications.
- Hektner, J. M., Schmidt, J. A., & Csikszentimihályi, M. (2011). *Experience sampling method: Measuring the quality of everyday life*. Sage.
- Hiver, P., Al-Hoorie, A. H., Vitta, J. P. & Wu, J. (2021a). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 1–30. <https://doi.org/10.1177/13621688211001289>
- Hiver, P., Mercer, S., & Al-Hoorie, A. H. (2021b). Introduction. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *Student Engagement in the Language Classroom* (pp. 1–17). Multilingual Matters. <https://doi.org/10.21832/9781788923613>
- Ho, D. G. E. (2006). The focus group interview: Rising to the challenge of qualitative research methodology. *Australian Review of Applied Linguistics*, 29, 05.01–05.19. <https://doi.org/10.2104/ara10605>
- IBM Corp (2017). *IBM SPSS Statistics for Macintosh, Version 25.0* [computer software]. IBM Corp.
- Janosz, M. (2012). Part IV commentary: Outcomes of engagement and engagement as outcome: Some consensus, divergences, and unanswered questions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 695–703). Springer. https://doi.org/10.1007/978-1-4614-2018-7_33
- Kaiser, H.F. (1970). A second-generation little jiffy. *Psychometrika*, 30, 401–415. <https://doi.org/10.1007/BF02291817>
- Kim, H.-Y. (2013). Statistical notes for clinical researchers: Assessing normal distribution using skewness and kurtosis. *Restorative Dentistry & Endodontics*, 38, 52–54. <https://doi.org/10.5395/rde.2013.38.1.52>
- Krueger, R. A., & M. A. Casey (2015): *Focus groups: A practical guide for applied research* (5th ed.). Sage Publications.
- Kuppens, A. H. (2010). Individual foreign language acquisition from media exposure. *Learning, Media and Technology*, 35, 65–85. <https://doi.org/10.1080/17439880903561876>

- Kusyk, M. (2017). The development of complexity, accuracy, and fluency in L2 written production through informal participation in online activities. *CALICO Journal*, 34, 75–96. <https://doi.org/10.1558/cj.29513>
- Lai, C., Zhu, W., & Gong, G. (2015). Understanding the quality of out-of-class English learning. *TESOL Quarterly*, 49, 278–308. <https://doi.org/10.1002/tesq.171>
- Lam, S. F., Wong, B. P. H., Yang, H., & Liu, Y. (2012). Understanding student engagement with a contextual model. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 403–420). Springer. https://doi.org/10.1007/978-1-4614-2018-7_19
- Lee, J. S. (2019). Quantity and diversity of informal digital learning of English. *Language Learning & Technology*, 23, 114–126.
- Lee, J. S., & Dressman, M. (2017). When IDLE hands make an English workshop: Informal digital learning of English and language proficiency. *TESOL Quarterly*, 52, 435–445. <https://doi.org/10.1002/tesq.422>
- Lyrigkou, C. (2016). *Young learners' engagement with English in their free time through informal sources. How do frequency of exposure, effort and motivation affect oral production?* [Master's thesis]. University of Oxford, Oxford, UK.
- Macaro, E. (2006). Strategies for language learning and for language use: Revising the theoretical framework. *The Modern Language Journal*, 90, 320–337. <https://doi.org/10.1111/j.1540-4781.2006.00425.x>
- Mair, C. (2020). English in the German-speaking world: An inevitable presence. In R. Hickey (Ed.), *English in the German-speaking world* (pp. 13–30). Cambridge University Press. <https://doi.org/10.1017/9781108768924.002>
- Mercer, S. (2019). Language learner engagement: Setting the scene. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 643–660). Springer International Handbooks of Education. https://doi.org/10.1007/978-3-030-02899-2_40
- Mercer, S., & Dörnyei, Z. (2021). *Engaging language learners in contemporary classrooms*. Cambridge University Press. <https://doi.org/10.1017/97811009024563>
- Nunan, D., & Richards, J. C. (2015). *Language learning beyond the classroom*. Routledge. <https://doi.org/10.4324/9781315883472>
- Oga-Baldwin, W. L. Q. (2019). Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System*, 86, 1–10. <https://doi.org/10.1016/j.system.2019.102128>
- Oxford, R. (2016). *Teaching and researching language learning strategies*. Routledge. <https://doi.org/10.4324/9781315719146>
- 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>
- Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36, 257–267. <https://doi.org/10.1016/j.cedpsych.2011.05.002>
- Reinders, H., & Nakamura, S. (2021). Engagement. In T. Gregersen & S. Mercer (Eds.), *The Routledge handbook of the psychology of language learning and teaching* (pp. 137–148). Routledge. <https://doi.org/10.4324/9780429321498-14>
- Renninger, K. A., & Bachrach, J. E. (2015). Studying triggers for interest and engagement using observational methods. *Educational Psychologist*, 50, 58–69. <https://doi.org/10.1080/00461520.2014.999920>
- Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 3–20). Springer. https://doi.org/10.1007/978-1-4614-2018-7_1
- Roehr-Brackin, K. (2018). *Metalinguistic awareness and second language acquisition*. Routledge. <https://doi.org/10.4324/9781315661001>
- Rosseel, Y., Oberski, D., Byrnes, J., ... Jorgensen, R. D. (2018). *lavaan: Latent variable analysis* [computer software]. <https://cran.r-project.org/package=lavaan>
- 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. <https://doi.org/10.21832/9781788923613>
- Schneider, H., & B. Franke (2014). *Bildungsentscheidungen von Studienberechtigten* [Educational choices of persons with higher education entrance qualification]. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. https://www.dzhw.eu/pdf/pub_fh/fh-201406.pdf
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29, 304–321. <https://doi.org/10.1177/0734282911406653>

- Schwarz, M. (2020). *Beyond the walls: A mixed methods study of teenagers' extramural English practices and their vocabulary knowledge*. [Doctoral thesis]. University of Vienna, Vienna, Austria.
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press. <https://doi.org/10.1002/9781405198431.wbeal0243>
- Shiffman, S., Stone, A. A., & Hufford, M. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A motivational perspective on engagement and disaffection: Conceptualisation and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69, 493–525. <https://doi.org/10.1177/0013164408323233>
- Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 21–44). Springer. https://doi.org/10.1007/978-1-4614-2018-7_2
- Sockett, G. (2014). *The online informal learning of English*. Palgrave MacMillan. <https://doi.org/10.1057/9781137414885>
- Storch, N. (2008). Metatalk in pair work activity: Level of engagement and implications for language development. *Language Awareness*, 17, 95–114. <https://doi.org/10.1080/09658410802146644>
- Sundqvist, P. (2011). A possible path to progress: Out-of-school English language learners in Sweden. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 106–118). Palgrave MacMillan. https://doi.org/10.1057/9780230306790_9
- Sundqvist, P., & Sylvén, L. K. (2014). Language-related computer use: Focus on young L2 English learners in Sweden. *ReCall*, 26, 3–20. <https://doi.org/10.1017/S0958344013000232>
- Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning*. Palgrave MacMillan. <https://doi.org/10.1057/978-1-137-46048-6>
- Svalberg, A. M. (2009). Engagement with language: Interrogating a construct. *Language Awareness*, 18, 242–258. <https://doi.org/10.1080/09658410903197264>
- Tavakol, M., & Wetzell, A. (2020). Factor analysis: A means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 6, 245–247. <https://doi.org/10.5116/ijme.5f96.0f4a>
- Thigpen, C. (2019). Measurement validity of retrospective survey questions of bicycling use, attitude, and skill. *Transportation Research Part F*, 60, 453–461. <https://doi.org/10.1016/j.trf.2018.11.002>
- Verspoor, M. H., de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language development: Methods and techniques*. John Benjamins. <https://doi.org/10.1075/lllt.29>
- Wang, M.-T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Schall Lin, J. (2016). The math and science engagement scales: Development, validation, and psychometric properties. *Learning and Instruction*, 43, 16–26. <https://doi.org/10.1016/j.learninstruc.2016.01.008>
- Wearden, J. H. (2008). The perception of time: Basic research and some potential links to the study of language. *Language Learning*, 58, 149–171. <https://doi.org/10.1111/j.1467-9922.2008.00468.x>
- Zhou, S., Hiver, P., & Al-Hoorie, A.H. (2021). Measuring L2 engagement: A review of issues and applications. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *Student Engagement in the Language Classroom* (pp. 75–98). Multilingual Matters. <https://doi.org/10.21832/9781788923613>

Cite this article: Arndt, H. L. (2023). Construction and validation of a questionnaire to study engagement in informal second language learning. *Studies in Second Language Acquisition*, 45: 1456–1480. <https://doi.org/10.1017/S0272263122000572>