Learning and Individual Differences 80 (2020) 101850

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Supporting self-efficacy beliefs and interest as educational inputs and outcomes: Framing AI and Human partnered task experiences

Luke K. Fryer[a,*], Andrew Thompson[b], Kaori Nakao[c], Mark Howarth[b], Andrew Gallacher[b]

[a] The University of Hong Kong, Faculty of Education, CETL, Pokfulam Rd., CPD180, Hong Kong (SAR)
[b] Kyushu Sangyo University, Japan
[c] Seinan Gakuen University, Japan

ABSTRACT

Interest and self-efficacy are crucial to academic success. This study addresses two gaps in our understanding of their development and support during university courses: how prior self-efficacy and interest plays a role in, and how different classroom activities build toward the development of students' future interest and self-efficacy. In this study the interplay between ability-beliefs (self-efficacy/self-concept) and interest at three levels of specificity (Domain, Course and Task) were tested across a Japanese university language course ($n = 128$). Within this test, students' interest in two language practice tasks (i.e., Human and then Chatbot partners) were assessed and compared. Prior interest was a robust predictor of all future task/course interest. Only Human-Human task interest directly predicted future course self-efficacy, but was mediated by course interest for future domain interest. For future interest, Human practice partners are superior to AIs. Supporting prior domain and later course interest should be a focus for university educators.

1. Introduction

How do classroom experiences support students' interest and self-efficacy beliefs across a university course? Where do commonly employed teaching-learning activities such as those offered by educational technology fit into this support? Both questions are critical. The first, due to the central role of students' beliefs/motivations within academic success during higher education (Richardson, Abraham, & Bond, 2012). The second question, because of the omnipresent role of, and resulting complex decisions that need to be made regarding, educational technology uses in higher education (Hollands & Escueta, 2019).

Interest and self-efficacy are vital to enhancing learning experiences across formal education (for detailed overviews see Bandura, 1993; Renninger & Hidi, 2015). Separately and as a pair, these individual differences are critical drivers for the learning experience: Interest is the fundamental desire to reengage with a topic/object that drives learning onward (Renninger & Hidi, 2015). Self-efficacy is equally pivotal to learning, especially in the face of challenge and failure (Bandura, 1993). Understanding the links that build across learning experiences, the interest that students develop from task to task, across a course of study is an important and budding area of student learning research (Fryer, Ainley & Thompson, 2016; Dietrich, Viljaranta, Moeller, & Kracke, 2017; Fryer, Ainley, Thompson, Gibson, & Sherlock, 2017; Nuutila, Tuominen, Tapola, Vainikainen, & Niemivirta, 2018). Our understanding of the role of self-efficacy in this development is more nascent, but recent and longstanding evidence suggests its central function and potentially reciprocal relationship with interest (Fryer & Ainely, 2019; Bandura & Schunk, 1981; Hidi, Ainley, Berndorff, & Del Favero, 2006).

Modelling the practical connections between interest and self-efficacy opens doors to the second question: providing a frame for researching the popular push for increasing engagement through the many tools that educational technology offers (Oblinger, 2004). The use of technology (videos, games, polling, etc.) to "help students through" less interesting task materials used in a course is rarely questioned. However, it is well established that triggering interest (e.g., through technological tools) is in no way a guarantee that interest development will follow (Hidi, Renninger, & Krapp, 2004).

The current study was designed to build on research in the area of task interest experiences and extend our limited understanding of the critical connections between self-efficacy and interest. Concurrently, the current study sought to test how AI partnered learning, which is certain to grow in prominence in the years to come, compares to traditional learning experiences and how these partners fit into the broader individual differences for learning puzzle. To these ends, this study tested the longitudinal connections between self-efficacy, students' interest in two language learning tasks, the course of study and interest in the domain or self-efficacy for the course. This study employed an unobtrusive intervention and comparison in a university foreign language classroom: Human and AI (using a Chatbot; all

* Corresponding author.
E-mail address: lukefryer@yahoo.com (L.K. Fryer).

https://doi.org/10.1016/j.lindif.2020.101850
Received 5 June 2019; Received in revised form 15 February 2020; Accepted 27 February 2020

students engaging with both partner types) learning partners for conversation practice tasks. As part of the coordinated curriculum, these conversation practice tasks were already a regular part of students' weekly course. Task interest measurement after one Human-Human and the Human-Chatbot task was undertaken to compare the experiences and their longitudinal connections to key latent course outcomes (Domain interest and Course self-efficacy). Modelling was undertaken within a task-course-domain classroom model of interest relationships. This is a practical model for interest development across formal learning experiences. This model both builds and relies on the underlying reasons for potential development interest in an object or topic, as outlined in the Four-Phase model of interest development (Hidi & Renninger, 2006; Renninger & Hidi, 2011).

2. Background

2.1. Learning a new language within formal education

Foreign or second language learning is one of many sub-fields within education. Unlike other distinctive subfields (e.g., STEM, native language, music and physical education), however, foreign/second language learning found its way into its own almost completely separate pocket of individual differences research (e.g., motivation, beliefs and learning strategies; see Oga-Baldwin, Fryer, & Larson-Hall, 2019 and its accompanying Special Issue). This separation is due in small part to the implications arising from the complex socio-cultural trappings of language acquisition across a life span (Nikolov & Djigunović, 2006). The idiosyncrasies of a life-long learning journey, which is as much cultural as it is linguistic, are important for language acquisition across the life span. However, it is also critical to point out that for many individuals, a significant proportion (the entire proportion for many) of actual language acquisition takes place in highly structured, formal education settings. Applied linguistics researchers agree that theory needs to fit the language learning context (Boo, Dörnyei, & Ryan, 2015), suggesting that flexibility is necessary to meet the language learning needs of formal education.

To start with, in the specific context of formal education, *language acquisition* is more accurately framed as *language learning*. Within such a frame, it is a small step to suggesting that language learning/teaching can learn much from the individual differences research undertaken with other school subjects. For example, recent research (Oga-Baldwin & Fryer, 2020) has presented person-centered findings indicating that students' motivation to learn their native and a new language have considerable overlap.

Under this general premise, the programme of research the current study grows from has focused on two gaps in the language learning individual differences literature: interest and self-efficacy. Interest, because it is a unique developmental (i.e., integral part of the learning process) and potentially sustainable source of motivation that everyone (students, parents and educators) recognises as essential (Fryer, 2019). Self-efficacy, because of its robust relationship with achievement outcomes (e.g., Higher Education: Richardson et al., 2012; Schneider & Preckel, 2017). Self-efficacy is an essential source of motivation for students learning any subject, but particularly important for languages due to its role in overcoming challenge and failure (Bandura, 1993). Both sources of motivation draw on a rich and active body of theoretical and empirical research both within and outside formal education. This firm theoretical foundation puts both theories in a strong position to contribute to language learning and teaching as research in this area begins to grow.

2.2. Interest

Interest is widely recognised as crucial for both the initiation of, and persistence in, learning. Interest is perhaps best defined as a desire to reengage (Renninger & Hidi, 2015). Consistent with James' (1983/1892) original definition, this desire to reengage makes interest an essential and potentially sustainable source of motivation to engage and persist with the learning.

Through early education investigations of interest and its substantial role within reading research (e.g., Hidi, 2001), several decades of interest research slowly moved toward a broadly utilised model for interest development (Renninger & Hidi, 2015; Renninger, Hidi, & Krapp, 1992). Interest is acknowledged as being subject specific and at least two types are distinguished. The first is situational interest and is generally short-lived. It arises from an experience and ends with that experience. The second type is individual interest, which is the kind of interest developed over time. The Four-Phase model (Hidi & Renninger, 2006; Renninger & Hidi, 2011) is a refinement and extension of this dichotomy. This model describes how an individual's interest might develop across a series of re-engagements with a topic or object: 1) from triggered situational (e.g., a task or experience that elicits some emotion from the individual), 2) to maintained situational (e.g., an enjoyable series of tasks/experiences that continue to elicit some emotions such as fun, but also helping to build initial knowledge and value for the topic or object), 3) to emerging individual (e.g., learning experiences which continue to support value for and competence with the topic or object, but are increasingly self-directed) and finally 4) well-developed individual interest (e.g., strongly self-directed learning experiences with opportunities to build on existing knowledge of and value for the topic or object). According to this model, interest develops from short-term, chiefly affective experiences, which when repeated can deepen interest for and knowledge of the topic. This development can thereby yield a sustainable source of motivation, critical to language learning (Fryer, 2019).

2.2.1. Interest development within formal education

The Four-Phase model is a powerful source of hypotheses for the development of interest in a domain. The model describes the natural developmental process interest goes through as it builds and changes during reengagement with a domain. In highly structured environments like formal education, the four phases are more difficult to model. The learners' choices are often limited. Educators and curricula are the key determinants of how, when and how much an individual engages. In many countries, students are made to (re)engage with more or less the same subjects for at least 12 years (e.g., mathematics, native/foreign languages, science, history, etc.…), with the only meaningful curricular change often being the level of the content. This regimented approach to education, while critical to ensuring all students are exposed to what a nation deems vital to its citizens, can impinge on the interest driven reengagement the Four-Phase Model describes.

It is also a challenge to translate the four phases into concrete language that an educator can use to guide how they plan and instruct. In the context of higher education, where instructors often teach a single course, the Four-Phase model provides a model of the positive trajectory that students' interest might take. The Four-Phase Model however yields little feedback about specific tasks that contribute to students' interest in a specific lecture/tutorial, how that leads to interest in the course as a whole and the broader domain over semesters and years of study. These are questions that concern many educators.

To address this gap in our understanding of student interest within formal learning environments (with a specific focus on higher education) a practical model of interest development, built on the foundation of the Four-Phase model, was constructed and tested (Fryer, et al. 2016). The Task-Course-Domain model had a very specific and narrow aim: to test the longitudinal relationships between students' interest in specific tasks/activities, the specific course they were engaged in and the broader domain of study the course was nested in. For this model to be utilised meaningfully it was essential that students' task interest be assessed micro-analytically (i.e., measured immediately following specific learning experiences). The Task-Course-Domain model's initial hypotheses (Fryer et al., 2016) were that in the highly structured environment of university courses that interest development would flow from prior knowledge (actual and perceived) and interest (domain) through a myriad of courses task experiences (to greater and less degrees due to their nature and challenge) to interest in the course itself. It was further hypothesised that interest in the course would be a key predictor of future interest in the domain of study.

Longitudinal modelling within the Task-Course-Domain model of interest during language classes (Fryer et al., 2016) pointed to students' course interest as completely mediating the contribution of interest in tasks/activities for domain interest at the end of the course. These findings suggested that tasks/activities are critical (through course interest) for building domain interest. More recent evidence from research in Biochemistry, Mathematics and Organic Chemistry university courses indicates the very different role of different tasks, with some activities that aesthetically appear to be very engaging having little actual impact on students' future course interest (Fryer et al., 2018).

2.2.2. The role of ability-beliefs and tasks in interest development

Self-concept and self-efficacy are both important covariates of interest and need to be accounted for within any meaningful model of interest development. Academic self-concept is a retrospective belief about an individual's ability within a specific domain (Marsh, Martin, Yeung, & Craven, 2016). In contrast, self-efficacy is prospective, defined as an individual's belief in his/her ability to effect actions for the successful completion of a future task or goal (Bandura, 1997). More recently, Bandura (2012) has also noted that "Judgments of self-efficacy for pursuits like academic achievement, organizational productivity, entrepreneurship, and effecting social change encompass activities of broad scope, not just an isolated piece of work. Moreover, strength of self-efficacy is measured across a wide range of performances within an activity domain, not just performance on a specific item." (p.17).

Large-scale modelling in secondary school mathematics (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005) suggested a reciprocal, small relationship between self-concept and interest. Research in the specific context of language learning also suggested a consistent small role for self-concept predicting future interest (Fryer, 2015). Early correlational research (Bandura & Schunk, 1981) and more recent latent curve based examinations (Niemivirta & Tapola, 2007) position self-efficacy as an important correlate of an individual's interest in a domain. Research examining the relationship between self-efficacy and interest resulted in a hypothesis framing a reciprocal relationship between these variables, with interest as the chief driver of this cycle (Ainley, Buckley, & Chan, 2009; Hidi et al., 2006; Hidi, Berndorff, & Ainley, 2002). Longitudinal modelling of connections has highlighted the importance of self-efficacy for task interest generally (Fryer et al., 2016) and the triggering and maintaining of interest specifically (J. A. Chen et al., 2016). More recent tests employing reciprocal modelling have supported uni-directional (interest to self-efficacy; Grigg, Perera, McIlveen, & Svetleff, 2018) and reciprocal (Fryer, 2019) relations between interest and self-efficacy. Authors reported that after accounting for task-value, self-efficacy and prior achievement, that no statistically significant reciprocal relationships between interest and self-concept could be detected. While perhaps not a direct predictor of domain level interest after accounting for self-efficacy beliefs (Fryer, 2019), self-concept has presented an important moderating effect for self-efficacy's relationship with interest at the task level (Fryer et al., 2016).

The difficulty (e.g., in the recent case of reading specifically see Fulmer & Tulis, 2013; Fulmer, D'Mello, Strain, & Graesser, 2015; in the case of interest development more broadly see Hidi & Renninger, 2006; working within the Task-Course-Domain model of interest, Fryer et al., 2016) and nature (e.g., individual-cooperative, receptive-productive, achievement-related; Fryer et al., 2018; for a well-established review see Schiefele, 1991) of learning tasks can each play an important role in how or whether students develop interest for a topic or object. Students' strategies (e.g., Ainley et al., 2009) can also play an important role in the relationship between prior knowledge (both actual and perceived) and task interest. The complexity of this relationship between prior knowledge (actual and perceived) and interest is not always considered carefully when instructors seek new tasks that might enhance student interest in the course and domain of study. However, this relationship is critical as instructors reach for one of the growing number of educational technology supported tasks to encourage students' behavioural engagement (P.-S. D. Chen, Lambert, & Guidry, 2010; Oblinger, 2004). The technology instructors reach for might be something as simple as a series of animations to integrate into teaching materials (Parette, Hourcade, & Blum, 2011), or using classroom polling/quizzing (Sun, 2014), but it might be as involved as using chatbots as learning partners for students (Benotti, Martínez, & Schapachnik, 2014) or using Virtual Reality (Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014) to transport students into the world of the topic.

2.2.3. Applying the task-course-domain model to Chatbots in classrooms

Chatbots are precisely the kind of educational technology that many instructors reach for when seeking to increase students' behavioural engagement. Chatbots are quickly growing as a popular interface for the internet as a whole (Dale, 2016). In just a few years, they have gone from novelties, to online guides and merchants, to the growing focus of mobile phones and home assistants. For at least two decades researchers have noted their potential role as language partners, both supporting motivation for learning a new language (Fryer & Carpenter, 2006) and driving language interaction in general (Hill, Ford, & Farreras, 2015). While Chatbots' eventual place within foreign language learning seems obvious, their current potential in classrooms still needs to be clarified (Fryer, Coniam, Carpenter, & Lăpușneanu, 2020).

Applying the Task-Course-Domain model for interest in classrooms, Fryer et al. (2017) conducted a counter-balanced experiment to test whether there were differences in students' interest in human vs. Chatbot language practice partners (measured as relative task interest). The two primary findings were a novelty effect with Chatbot partners, but sustained interest in talking to a human partner. Furthermore, only task interest when practicing with human partners was a statistically significant predictor for students' future course interest. A follow-up study (Fryer, Nakao, & Thompson, 2019) with the same students (20 weeks later) indicated a statistically significant rebound (i.e., an increase after a decline) for students' task interest in the Chatbot partner condition. Regression analysis across the two semesters established that of the two partnered tasks, only task interest with the human partner measured in semester one, predicted future task interest for practicing with the chatbot partner in the subsequent semester. Coding of students' reported experiences with the two partners, indicated that students' perceptions of Chatbots as practice partners played a powerful role in task interest. Students who saw the Chatbot as a convenient tool for practice (e.g., I can use it anywhere.) were less interested in the task than students who saw the Chatbot as providing an opportunity to practice in a manner that a human partner was not willing to or cannot do: e.g., "I can learn new words when I practice with it (the Chatbot)"; "It (the Chatbot) does not get tired of talking to me."

An alternative (or additional) reason for the potential differences in interest inspired by AI and Human learning partners can be taken from evolutionary psychology (e.g., Greg, 2012). Cognitive load researchers working within this area suggested that psychological primary types of engagement (e.g., human to human communication) result in lower extraneous cognitive load relative to biologically secondary types of engagement (i.e., everything from writing to watching videos online) (Paas and Sweller, 2012). Lower extraneous cognitive load during the learning process can result in better learning outcomes (Sweller, 1994). In addition to affecting learning outcomes, cognitive load can also impinge on an individual's source of motivation to learn (e.g., interest and self-efficacy; Feldon et al., 2019).

These studies suggest that students perceive Chatbot practice partners as representing different kinds of learning experiences. Furthermore, findings suggest that interest in the Chatbot language partners might be enhanced by presenting Chatbots not as a replacement for human partners or even as a convenient tool, but instead as a powerful supplement to human-human practice. The Task/Course/Domain model was useful for examining how different kinds of tasks might affect students' course interest, which research indicates (Fryer et al., 2016) substantially contributes to domain interest.

The lines of research discussed have noted important differences in task interest under conditions of human and Chatbot conversation partners. Ability-beliefs have been highlighted as critical factors for both task (Ainley et al., 2009) and domain (Fryer & Ainley, 2019), longer-term (Renninger & Hidi, 2015) interest. Finally, research has demonstrated that the Task-Course-Domain model is an efficacious and practical means of testing the interplay between levels of interest and other important covariates in university courses (Fryer et al., 2016, 2017, 2018, 2019; Fryer & Bovee, 2020).

2.3. The current study

The current study extends previous research into the development of students' interest and course self-efficacy across a single university course (Fryer, 2015; Fryer et al., 2016; Fryer & Ainley, 2019). Furthermore, this research extends longitudinal tests of the relative efficacy of human-human vs. human-chatbot learning partners to predict longer-term (course and domain) interest (Fryer et al., 2017; Fryer et al., 2019).

The present study focuses on domain interest and course self-efficacy beliefs as two essential outcomes of the learning experience. Domain interest (closest to individual interest from the Four Phase Model) is the interest that students might take from course to course during higher education and is therefore an important course outcome. Course self-efficacy is the largest grain-size of academic self-efficacy that has meaning within education. Similar to domain interest, this is the confidence students might carry forward to future course experiences.

This study focuses on how students' prior interest (task, course, domain) and ability beliefs together contribute to these outcomes. To achieve these aims, latent modelling of longitudinal measures of these constructs was undertaken. At T1, domain self-concept and interest were modelled as predicting course self-efficacy (T1). All three constructs were modelled as predicting task interest in human (T2) and Chatbot (T3) conversation partners. In a fully-forward fashion, T1, T2 and T3 constructs were modelled together as predicting course interest (T4). Finally, T1-T4 were modelled as predicting the two outcomes (course self-efficacy and domain interest; T5). Modelling was undertaken separately for each of the two outcomes (domain interest and course self-efficacy beliefs) due to the limited sample size and preferred latent approach to analyses. Fig. 1 presents the overall design for the current study.

The current study addressed five research questions:

1) What was the longitudinal change in domain interest and course self-efficacy across the 12 weeks of the study (across a 15-week semester of the course)?

2) What if any were the differences in students' task interest in the two conversation partner conditions (Human and Chatbot at T2 and T3)?

3) Controlling for prior domain interest, what are the longitudinal relationships between prior course self-efficacy and domain self-concept with future task, course and domain interest?

4) Controlling for course self-efficacy and domain self-concept, what are the longitudinal connections between domain interest (T1), task interest (T2) and course interest (T4)?

5) Controlling for prior domain interest, domain self-concept and course self-efficacy, how do task and course interest predict future (A) course self-efficacy beliefs (B) domain interest?

3. Methods

3.1. Sample and context

The present study was undertaken with 128 first-year students (all participating students were either 18 or 19 years old) studying in a compulsory foreign language (English as a foreign language) course at a university in Western Japan. Students were studying in one of six courses (three teachers, two courses each). The three participating teachers each possessed a masters in applied linguistics and had between four and seven years experience teaching the course wherein the research was conducted. Class size ranged from 18 to 24. Students were of lower intermediate ability in the language studied (i.e., capable of basic daily communication in their foreign language). The courses were a part of a large ($n = 5000+$) coordinated program of foreign language learning wherein the tests (Stewart, Fryer, & Gibson, 2013; Stewart, Gibson, & Fryer, 2012), classroom learning materials (Fryer, Anderson, Stewart, Bovee & Gibson, 2010) and weekly e-learning assignments (Bovee & Fryer, 2011) were identical. The courses were all conducted during the same semester and each class lasted 90 min, once a week for 15 weeks. The sample for this study does not overlap with any previously published research or any other research currently under review.

The university where this research was undertaken does not have an institution-wide ethical review board. However, consistent with the institutional research ethics guidelines, prior to conducting the current study, the education Centre governing students' compulsory language studies reviewed this research's proposal. Based on the Centre's internal guidelines, and consistent with the 1964 Helsinki declaration and its later amendments for comparable ethical standards, permission for the study was granted and the study proceeded.

Before beginning with the study, informed consent was obtained from each participant. As part of the informed consent process students were informed of the overall research programme and its general aims. The instructors explained the study and the online consent form was prefaced by a detailed textual introduction to the research. In addition to outlining the project, this preface to the instruments stated that students' self-reports would remain anonymous and that their classroom teachers would not have access to any of the information given. All students who subsequently agreed to participate in the study were provided with the additional opportunity to opt out at each administration of the study's surveys.

3.2. Instrumentation

For the current study five scales were used to assess task interest, course interest, domain interest, course self-efficacy, domain self-concept (for English as a foreign language). Students' domain interest was assessed using four questions (e.g., I think English is always interesting; I know English arouses my curiosity). Students' course interest was measured by four items (e.g., I am fully focused on learning English in this course; This English course is interesting). Students' task interest was assessed by four items (e.g., This activity is personally meaningful; I enjoyed learning English in this activity). All three scales had demonstrated robust convergent/divergent validity (demonstrated through confirmatory factor analysis and intercorrelations) and reliability (Cronbach's Alpha >0.70; Devellis, 2012) in three previous task-centred (Fryer et al., 2016; Fryer et al., 2017; Fryer et al., 2019) and two previous domain-centred (Fryer, 2015; Fryer & Ainley, 2019) studies. Students' self-efficacy for the course was assessed by five items from the patterns of adaptive learning scales (Midgley et al., 2000) (e.g., I am certain I can master the skills taught in class this year; I am certain I can figure out how to do the most difficult class work). Students' self-concept for English as a foreign language was assessed by four questions (e.g., My English grade in high school was good.; I have a good memory for English) (from Ichihara & Arai, 2004). All scales were

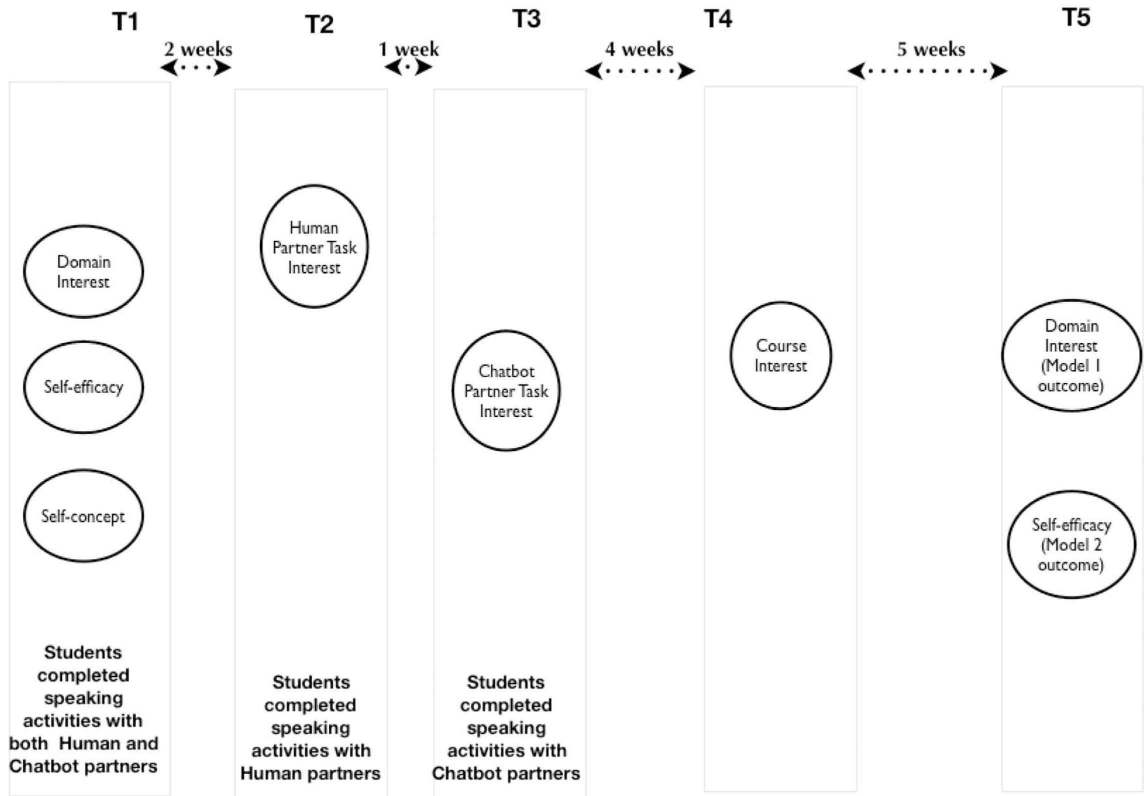

Fig. 1. Timing of questionnaire administration across 15-week semester.
Note: All participating students interacted with the Human partner and then Chatbot partner just prior to completing the task interest surveys at T2 and then T3.

used in a manner consistent with the past studies referenced here. No scale items were removed prior to the study or during the analyses phase of the study to improve fit or reliability. All surveys were presented in Japanese, the native language of all participating students.

All questions were assessed across a cumulative scale from one (not at all like me) to six (exactly like me). The questionnaire scales were completed during regular class time and took no >5 min to complete for each of the five data collections.

3.3. Procedures

As presented in Fig. 1, surveys were completed at five time points across one academic semester (15 classes). T1 surveys were conducted during the third class of the semester. The following surveys were completed across the 15-week term: Week 5, 6, 10 and 15, with the final surveys being completed during the final class of the semester. Surveys were completed on the tablets used for the Chatbot conversation tasks during regular class time.

The conversation tasks for which students' interest was tested, were developed directly from the class materials by the three instructors teaching the classes and consisted of a 15-min partially scripted conversation. Partially scripted refers to a scripted conversation to help get the pair (Human-Human and then Human-Chatbot) speaking, followed by suggestions for further conversation extensions. Conversation exercises were the same as those undertaken almost every week as a part of the overarching coordinated curriculum. Students were very familiar with the Human-Human condition (undertaken during the study with a random classmate), making the human-bot task condition the only actual intervention undertaken during this study. However, to ensure parity between the two tasks for which task interest was assessed, the two tasks were designed to be comparable in difficulty, procedure and length. Students' interest in the tasks was measured immediately after the tasks were completed (15 min to complete).

All students had an opportunity to pilot the Chatbot with a similar structured conversation at T1. This was done to ensure that the novelty due to the newness of the Chatbot partner did not result in a novelty effect, which would interfere with the study's results.

3.4. Analyses

All latent modelling was undertaken with M*plus* 7.0 (Muthén & Muthén, 2014). The missing data was low (< 4%), due to students missing class for one of the days of the study. Missing data was resolved through the Full Information Maximum Likelihood method (Enders, 2010).

Latent constructs' validity (CFA; confirmatory factor analysis), invariance and reliability (Raykov, 2009) were assessed. Structural model fit was assessed employing: Root Mean Square Error of Approximation (RMSEA), with values below 0.08 and 0.05 indicating acceptable and good fit respectively (Browne & Cudeck, 1992), Confirmatory Fit Index (CFI) with values above 0.90 and 0.95 indicating acceptable and good fit respectively (Marsh, Balla, & McDonald, 1988), and Standardized Root Mean Square Residual (SRMR) where values <0.08 are generally considered a good fit (Hu & Bentler, 1999).

For invariance testing the current study replicated Marsh, Nagengast, and Morin (2013). Invariance testing of domain interest and course self-efficacy (T1 and T5) therefore relied on CFI and RMSEA comparisons to assess the adequacy of the invariance between time points. Based on F. F. Chen (2007), the assumption of invariance is tenable if CFI does not change by >0.01 and the RMSEA increases no >0.015 for the invariant model.

The latent correlations of all variables modelled were calculated, followed by mean difference testing (*t*-test) (Research questions 1 and 2). Analyses concluded with the test of two structural equation models; the first model examined domain interest as an outcome and the second model examined self-efficacy beliefs for the course. Two model tests, with two different outcomes, rather than one with two outcomes were conducted due to the small sample size and complex fully-forward nature of the analyses (Research questions 3, 4 and 5a + b). For the current study the guidelines suggested by Keith (2015) were used for interpreting *β*s (i.e., standardized regression coefficients) for influences on learning: $\beta<0.05$ are interpreted as 'too small to be considered meaningful'; >0.05 are considered 'small but meaningful'; those >0.10 are considered 'moderate'; and > 0.25 are considered 'large'.

The latent analyses, which were at the center of this study, were conducted in time order, except at T1. At T1 the specificity of the constructs was deemed critical to understanding their interconnections. As a result, the two domain level constructs (domain self-concept and domain interest—both for learning English as a Foreign Language) were modelled as predictors of course self-efficacy all at T1. Prior to conducting regression modelling, a power analysis was conducted to assess whether the modest sample size was sufficient to confidently assert the results of the two models. Power analyses were conducted using G*power (Faul, Erdfelder, Lang, & Buchner, 2007), setting the a priori power minimum of 0.95 for each dependent variable (i.e., the standard setting).

4. Results

4.1. Descriptive and reliability results

Table 1 presents the latent correlations and descriptive statistics for all variables. The latent reliabilities (Raykov's Rho) for each survey utilised were considered to be acceptable (> 0.70; Devellis, 2012; Table 1).

4.2. Latent fit and preliminary difference testing

A CFA of all variables modelled together to assess convergent and divergent validity resulted in acceptable fit (see Table 2). Follow-up tests for invariance met the assessment guidelines set forth by Chen (2007) suggesting that the assumption of invariance was tenable; as a result, the scales (course self-efficacy and domain interest) can be assumed to be measuring the same construct at both measurement points.

Paired *t*-tests indicated significant increases ($p < .05$, Bonferroni corrected) for both domain interest ($t = 2.50$, $p < .05$, $d = 0.35$: T1 = 3.78, T5 = 4.11) and course self-efficacy ($t = 7.13$, $p < .05$, $d = 0.89$: T1 = 3.45, T5 = 4.31) (Research question 1). Paired *t*-test, however, indicated no significant difference in students' task interest between human and Chatbot partnered practice tasks ($t = 1.13$, $p = .78$, $d = 0.14$: Human = 4.31, Chatbot = 4.16) (Research question 2).

4.3. Longitudinal latent modelling findings

Given the high R^2 for the dependent variables, only a modest sample size was necessary to be confident of the models' outcomes. These results, along with the overall robust fit for both models, suggests that the sample size was sufficient to test the models presented. The longitudinal findings will be presented in three parts. First (research question 3), the relationships between prior ability-beliefs and task interest and course interest will be examined. Second (research question 4), the connections between task interest and course interest will be examined. Third (research questions 5a and 5b), domain interest as an outcome will be discussed followed by a separate analysis of course self-efficacy as an outcome. Both models were organised in a fully-forward fashion with prior variables predicting all later variables, with the exception of T1. At T1, domain self-concept and domain interest were modelled as predictors of course self-efficacy. These background connections will be presented briefly first. The differences between the two outcome models for all *β*s except for the T5 outcome (course self-efficacy and domain interest) are inconsequential. *β*s for T1-T2-T3-T4 connections will therefore be drawn from Model 2 exclusively.

Fig. 2 presents the model for the domain interest outcome. For both models, domain interest was a statistically significant ($p < .05$) predictor of course self-efficacy (T1; $\beta = 0.48$). No statistically significant relationship between domain self-concept and course self-efficacy was found.

After accounting for prior domain interest ($\beta = 0.45$), students' task interest for the Chatbot condition task interest was predicted by both course self-efficacy ($\beta = 0.60$) and domain self-concept ($\beta = -0.44$) with large *β*s. In contrast, after controlling for prior domain interest ($\beta = 0.45$) neither course self-efficacy nor domain self-concept presented a statistically significant β for task interest in the human condition. At T4, after accounting for prior domain interest and task interest for both conditions, neither course self-efficacy nor domain self-concept presented a statistically significant β for course interest (Research question 3).

After controlling for prior domain self-concept and course self-efficacy, domain interest predicted course interest ($\beta = 0.37$). Task interest for the human condition presented a statistically significant β (0.57), but task interest for the Chatbot condition did not (Research question 4).

The fully forward model (Fig. 2) presented a single statistically significant prediction (course interest T4; $\beta = 0.45$) for domain interest (T5) (Research question 5a). In contrast Fig. 3 resulted in two statistically significant *β*s from task interest (human condition, $\beta = 0.44$) and course interest ($\beta = 0.34$) to course self-efficacy(T5) (Research question 5b).

Table 1
Correlations, means, standard deviations and reliability.

	Domain Self-concept T1	Course Self-efficacy T1	Domain Interest T1	Human Task Interest T2	Chatbot Task Interest T3	Course Interest T4	Domain InterestT5	Course Self-efficacy T5
Domain Self-concept T1								
Course Self-efficacy T1	0.72							
Domain Interest T1	0.84	0.75						
Human Task Interest T2	0.59	0.50	0.62					
Chatbot Task Interest T3	0.37	0.63	0.53	0.66				
Course Interest T4	0.47	0.46	0.60	0.77	0.64			
Domain Interest T5	0.62	0.53	0.69	0.74	0.60	0.79		
Course Self-efficacy T5	0.53	0.51	0.56	0.71	0.50	0.69	0.89	
Raykov's RHO	0.85	0.87	0.93	0.90	0.94	0.92	0.95	0.91
Mean	3.04	3.45	3.78	4.31	4.16	4.33	4.11	4.31
Std Dev	1.00	0.85	1.16	0.80	0.95	0.79	0.93	1.05

Note: All *r*'s are significant ($p < .01$).

Table 2
Fit for all models.

	Chi-Square	RMSEA	CFI	SRMR
Configural Model	914(522)	0.079 (C.I. 90% 068–0.087)	0.91	0.056
Invariance Course Self-efficacy	935 (526)	0.081(C.I. 90% 069–0.089)	0.90	
Invariance Domain Interest	918,525)	0.079 (C.I. 90% 068–0.087)	0.91	
Course Self-efficacy Model	704(408)	0.078 (C.I. 90% 068–0.087)	0.91	0.060
Domain Interest Model	666(384)	0.078 (C.I. 90% 0.068–0.088)	0.92	0.052

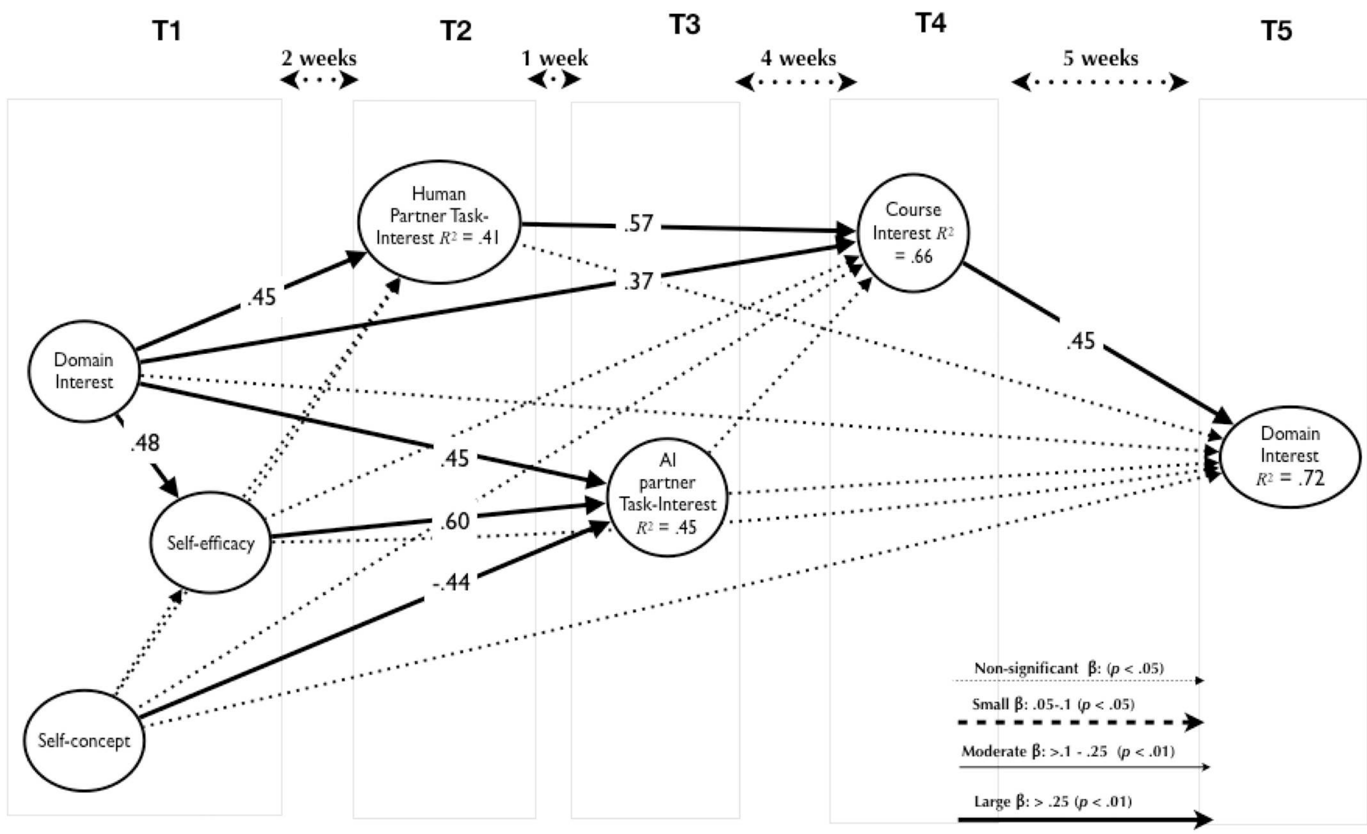

Fig. 2. Longitudinal prediction of Domain Interest.

5. Discussion

The current study set out to address five questions, examining the relationship between interest and ability-beliefs (domain self-concept and course self-efficacy); relationships between interest at task, course and domain levels of specificity; and finally, the relative predictive validity of these variables for course self-efficacy and domain interest as outcomes from a semester-long course. To this end, students' task interest for a human and Chatbot partners were assessed and modelled within a Task/Course/Domain model of interest. This modelling frame enabled an examination of the implications of AI vs. human learning partners for students' future motivation to learn.

Initial difference tests for the longitudinal variables (course self-efficacy and domain interest) indicated a small statistically significant increase for domain interest and large increase in course self-efficacy (Research question 1). No statistically significant difference for the two task interest conditions was observed (Research question 2).

Self-efficacy beliefs and self-concept presented statistically significant βs for task interest in the Chatbot, but not the human condition task (Research Question 3). The statistically significant relationship with task interest for the Chatbot, but not the human condition and their lack of (statistically) significant difference suggests that these tasks are not simply perceived by students as two similar means of conversation practice (despite reporting similar amounts of interest in them). The positive (self-efficacy) and negative (self-concept) βs for the task interest in the Chatbot condition was consistent with past predictive tests of these two ability beliefs for task interest (Fryer et al., 2016). This finding confirms that rearview beliefs of ones' abilities (self-concept) and thinking you can successfully do something in the future (self-efficacy) have very different implications for interest in classroom tasks.

Controlling for prior ability beliefs and domain interest (which predicted all T3 and T4 interest variables) only the human condition predicted (statistically significantly) course interest (Research question 4). This finding is consistent with a recent study (Fryer et al., 2017), which also observed that despite presenting similar levels of interest that human but not Chatbot partnered task interest predicted course interest. Students appear only to recognise the human-partnered learning as contributing to their course experience.

The semester-end domain interest outcome was only predicted (statistically significantly) by course interest at T4 (Research question 5a). The semester-end course self-efficacy presented statistically significant ($p < .05$) βs from task interest (human condition) and course interest (Research question 5b). These findings demonstrate that the development of students' interest in the domain flows from course activities, through course interest. Making course interest a critical nexus

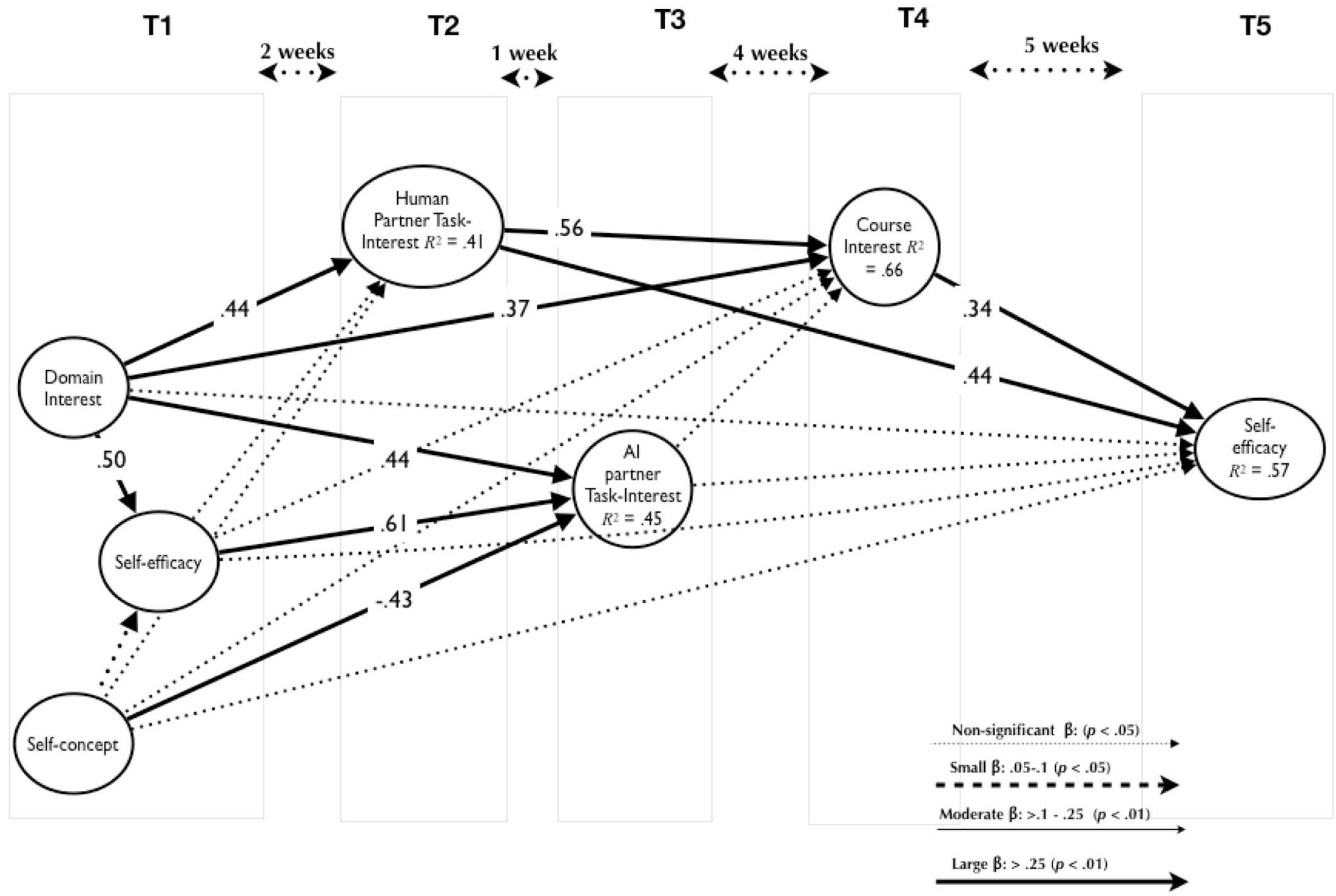

Fig. 3. Longitudinal prediction of Course Self-efficacy for the course of study.

for interest development across formal education.

5.1. Theoretical implications

The fact that ability-beliefs and interest development are strongly connected is well established empirically, and theoretical explanations are budding (Fryer et al., 2016; Bandura & Schunk, 1981; Grigg et al., 2018; Hidi et al., 2002; Hidi et al., 2006; Marsh et al., 2005). The specific contribution of self-concept and self-efficacy for future interest is less clear.

The current study's ability-belief related findings are consistent with past studies employing the same practical model of interest within formal education (task-course-domain; Fryer et al., 2016) and larger scale tests (Fryer & Ainley, 2019). As logic would suggest, rear (self-concept) and forward (self-efficacy) viewing ability beliefs predict interest at the domain and task level very differently. Self-efficacy for a course is consistently and strongly related to interest in the course's tasks. The role of self-concept is less clear.

Consistent with Fryer et al., 2016, modelling in the current study presented self-concept and self-efficacy as statistically significantly predicting future task interest, but only for the Chatbot condition task. Neither ability-belief presented a statistically significant β for the Human condition task. This difference was despite the fact that students' task interest in the two partners was not statistically significantly different. This suggests, first of all, that the nature and perhaps challenge (i.e., potentially due to its biologically secondary nature and resulting cognitive load: Geary, 2012; Paas & Sweller, 2012) of the task plays a pivotal role in how students' ability-beliefs interact with interest in the task.

As with research in Biochemistry, Mathematics and Organic Chemistry (Fryer et al., 2018) and language (Fryer et al., 2016) university courses, it is also reasonable to expect that how students perceive the task (in this case, interaction with the two different partners) and its difficulty each play important roles. The lack of prediction for the future human condition task could have been partially because the human partner conversation tasks were standard across regular class experiences. As a result of students' familiarity with the human-human task, ability-beliefs might have failed to play a role in students' task interest. For the same reason, this yields a hypothesis that human-partnered task was not seen as challenging enough to be impacted by students' self-efficacy (Bandura, 1997). The lack of significant regression from self-concept, in contrast, could be due to the psychological distance (from task experiences) and retrospective nature of self-concept (Marsh et al., 2016). The negative β for self-concept to the Chatbot-partnered task may have been due to students who felt particularly confident in their skills not seeing the Chatbot as a useful facsimile for human conversation and therefore not seeing it as an appropriate or interesting partner. The contrasting positive β for course self-efficacy might be explained by building on this supposition. Early research with Chatbot partners (Fryer & Carpenter, 2006) suggested that many students preferred Chatbots over human partners because Chatbots were perceived as less threatening: (i.e., students didn't mind making mistakes with Chatbots, as compared to a human partner). The present and past findings together generate a hypothesis indicating that perhaps students who have experienced less success with the target language in the past (i.e., low self-concept for the language), but remain positive about their prospective achievement in the course (i.e., high self-efficacy for the course), find the Chatbot more interesting.

The mediating role of course interest between task interest and students' domain interest, observed in Fryer et al., 2016, was replicated by the current study's modelling. Based on these findings, students' interest in their courses becomes a critical channel for interest development: i.e., course interest is the neck which students' experiences pass through to go on and impact their interest in the broader domain (domain interest) and thereby future courses and, eventually, potentially on to lifelong learning in that specific area.

Finally, results from the present study suggest much like Fryer et al., (2016) that self-efficacy beliefs for the course might in fact be more strongly tied to students' task interest than the broader level of course or domain interest. Self-efficacy beliefs have long been theoretically focused on task level experiences (Bandura, 1997), and it therefore makes sense that self-efficacy, even at the course level, would be strongly related to individual task experiences, rather than broader, aggregated sources of interest (i.e., domain).

5.2. Practical implications

When technological tools are employed to support classroom instruction, the aim is often at least partially to increase student engagement. While stimulating students' engagement during a specific class/lecture might be a proximal goal, it is reasonable to assume that most instructors hope for longer-term knock-on benefits from activities/tasks that stimulate students' interest and thereby engagement. The only means that teachers have for assessing the engagement stimulated is the increased activity students might physically exhibit. The current study emphasises previous findings (Fryer et al., 2017), suggesting that even when students' self-reported (task) interest is effectively equal for two tasks, it does not mean the tasks will have an equal longer-term impact on students' domain interest. Instructors need to take this point into consideration when they are trying to stimulate students' interest in course materials. Burgeoning evidence from past studies utilising the task-course-domain model (Fryer, et al, 2016; Fryer et al., 2017) and recent outcomes from an ongoing programme of research (Fryer et al., 2018) suggest that some tasks do not significantly contribute to students' course interest. It might be that students see additional tasks like conversations with Chatbots as a diversion rather than a “proper” part of the class. Fryer et al., 2018 observed that some group activities (e.g., building physical models) in a biochemistry lecture, which is usually lecture based, had a similar lack of contribution to students' interest in the class. This finding was set in sharp relief when compared with a well-structured 20-minute lecture, which *did* substantially predict students' interest in the same 90-minute class experience.

The current study confirms students' course interest as a crucial outcome of task experiences and mediator toward future domain interest. Students' course interest is related to self-efficacy beliefs across the course experience (but not directly) and consistently mediates task experiences impact on future domain interest (which is connected to lifelong learning; Hidi et al., 2004). These results suggest that instructors and their course can in fact stand between students leaving or persisting in the domain after the course is complete.

For language educators at universities the present study has two specific, practical recommendations. The first is that bringing educational technology into the classroom for a bit of novelty or to stimulate some alternative active learning might be appropriate, but if the aim is to build lasting interest, educators should consider reaching for authentic (biologically primary) learning experiences (e.g., face-to-face speaking activities). These experiences are more likely to be seen as being directly related to the purpose of the course, support interest in the course and thereby longer-term interest. In the context of language education specifically, a mixture of human and chatbot partners might be utilised, depending on the aims of the activity: a novel change of pace (chatbot partner) or one more step toward students building a lasting relationship with the language (human partner). Similar suggestions can be made for enhancing students' self-efficacy for language learning. The second point is that domain interest is critical and should be supported early and regularly. Orientation sessions that establish the value of foreign languages courses (especially if they are compulsory) should be a minimum and the best-case scenario would involve outreach to high school students, supporting their interest before they arrive at university.

6. Limitations and future directions

This study relied on self-reported experiences for the modelling presented. Self-report is still the most direct and efficient means of accessing individual differences like interest and self-efficacy. However, observed variables such as achievement, attendance and classroom observations are important covariates that should be included in future studies to extend and validate the research presented here.

Consistent with the current study, latent, longitudinal research designs are encouraged. For very small sample size research ($n = 50$–100), a Partial Least Squares SEM approach might support designs of this type.

Due to a survey implementation issue, we were unable to control for the sex of the respondents. Considerable evidence points to small, but consistent differences in individual differences like motivations and beliefs (Voyer & Voyer, 2014). We suggest future studies account for these differences when researching in this area.

7. Conclusions

Two conclusions can be confidently drawn from this study. First, the nature of the task, rather than the amount of interest it inspires, is key to its impact on students' broader interest experiences (i.e., in the course and the domain). This has potential implications for the higher education sector's push for active learning, often making observable student engagement a curricular aim (Barkley, 2010)–at least in part because students report enjoying it (Lumpkin, Achen, & Dodd, 2015). The nature, the challenge and even the cognitive load implications of tasks/activities students engage in can each play an important role in the development of students' interest. Some tasks despite engaging students can be completely disconnected from students' interest development. Second, helping students develop an interest in their courses is crucial. Course interest is a junction through which students' task interest flows onto their domain interest and course self-efficacy beliefs; both can support students' persistence in the course and the domain in the longer-term.

The current study lends one more voice, and additional direction, to a growing call (e.g., Harackiewicz & Priniski, 2017) to make ‘helping students develop the interest and beliefs about learning that they need to be persistent in a course and well beyond’ become a central focus for university courses.

Acknowledgement

We would like to acknowledge Mary Ainley for her careful review and feedback on an earlier draft of the manuscript. The present manuscript was completed during a Fellowship at Hughes Hall, Cambridge University by the first author. The fellowship was supported by the Doris Zimmerman endowment to The University of Hong Kong.

References

Ainley, M., Buckley, S., & Chan, J. (2009). Interest and efficacy beliefs in self-regulated learning. In S. K. M. Wosnitza, A. Efklides, & P. Nenniger (Eds.). *Contemporary motivation research: From global to local perspectives* (pp. 207–228). New York: Hogrefe.
Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117–148. https://doi.org/10.1207/s15326985ep2802_3.
Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: Freeman.
Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*, 9–44. https://doi.org/10.1177/0149206311410606.
Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology, 41*, 586–598. https://doi.org/10.1037/0022-3514.41.3.586.

Barkley, E. (2010). *Student engagement techniques: A handbook for college faculty.* San Francisco: Jossey-Bass.
Benotti, L., Martínez, M. C., & Schapachnik, F. (2014). Engaging high school students using chatbots. *Paper presented at the proceedings of the 2014 conference on innovation & technology in computer science education*.
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.
Bovee, H., & Fryer, L. K. (2011). オンライン*CALL*システムに関する開発経過報 告。九州産業大学 *COMMON*. Permanent Online Location:http://ci.nii.ac.jp/naid/40019173226.
Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*, 230–258. https://doi.org/10.1177/0049124192021002005.
Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*, 464–504. https://doi.org/10.1080/10705510701301834.
Chen, J. A., Tutwiler, M. S., Metcalf, S. J., Kamarainen, A., Grotzer, T., & Dede, C. (2016). A multi-user virtual environment to support students' self-efficacy and interest in science: A latent growth model analysis. *Learning and Instruction, 41*, 11–22. https://doi.org/10.1016/j.learninstruc.2015.09.007.
Chen, P.-S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners: The impact of Web-based learning technology on college student engagement. *Computers & Education, 54*, 1222–1232. https://doi.org/10.1016/j.compedu.2009.11.008.
Dale, R. (2016). The return of the chatbots. *Natural Language Engineering, 22*, 811–817. https://doi.org/10.1017/S1351324916000243.
Devellis, R. F. (2012). *Scale development: Theory and application* (3rd ed.). Thousand Oaks, CA: Sage.
Dietrich, J., Viljaranta, J., Moeller, J., & Kracke, B. (2017). Situational expectancies and task values: Associations with students' effort. *Learning and Instruction, 47*, 53–64. https://doi.org/10.1016/j.learninstruc.2016.10.009.
Enders, C. K. (2010). *Applied missing data analysis.* Guilford Press.
Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146.
Feldon, D. F., Callan, G., Juth, S., & Jeong, S. (2019). Cognitive load as motivational cost. *Educational Psychology Review, 31*, 1–19. https://doi.org/10.1007/s10648-019-09464-6.
Fryer, L. K., Anderson, C. J., Stewart, J., Bovee, H. N., & Gibson, A. (2010). *Coordinating a vocabulary curriculum: Exploration, pilot, trial and future directions. Paper presented at The Japan Association for Language Teaching national conference Nagoya.* Retrieved fromhttp://jalt-publications.org/proceedings/issues/2011-10_2010.1.
Fryer, L. K. (2015). Predicting self-concept, interest and achievement for first-year university students: The seeds of lifelong learning. *Learning and Individual Differences, 38*, 107–114. https://doi.org/10.1016/j.lindif.2015.01.007.
Fryer, L. K., Ainley, M., & Thompson, A. (2016). Modeling the links between students' interest in a domain, the tasks they experience and their interest in a course: Isn't i interest what university is all about? *Learning and Individual differences, 50*, 57–165. https://doi.org/10.1016/j.lindif.2016.08.011.
Fryer, L. K., Zeng, L., Wong, C. W., Ho, C. C., & Chiu, P. (2018). On-task student-to-teacher feedback: "Was that interesting?" &"Does it Matter" August, 2018 *A poster presented at EARLI SIG 4 Conference Topography of research on higher education: Promoting deep conversations,* 29–31. https://doi.org/10.13140/RG.2.2.29129.85604.
Fryer, L. K. (2019). Getting interested in learning a language at school: Developing a sustainable source of classroom engagement. *System, 86*, 102–120. https://doi.org/10.1016/j.system.2019.102120.
Fryer, L. K., & Ainley, M. (2019). Supporting interest in a study domain: A longitudinal test of the interplay between interest, utility-value, and competence beliefs. *Learning and Instruction, 60*, 252–262. https://doi.org/10.1016/j.learninstruc.2017.11.002.
Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior, 75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045.
Fryer, L. K., & Bovee, H. N (2020). Teaching for course interest. *Studies in Higher Education*. https://doi.org/10.1080/03075079.2020.17.
Fryer, L. K., & Carpenter, R. (2006). Bots as language learning tools. Language Learning & Technology, 10, 8–14. doi:http:// http://llt.msu.edu/vol10num3/emerging/default.html.
Fryer, L. K., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning and Technology, 24*(3).
Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior, 93*, 279–289. https://doi.org/10.1016/j.chb.2018.12.023.
Fulmer, S. M., D'Mello, S. K., Strain, A., & Graesser, A. C. (2015). Interest-based text preference moderates the effect of text difficulty on engagement and learning. *Contemporary Educational Psychology, 41*, 98–110. https://doi.org/10.1016/j.cedpsych.2014.12.005.
Fulmer, S. M., & Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction, 27*, 11–20. https://doi.org/10.1016/j.learninstruc.2013.02.001.
Geary, D. (2012). Evolutionary Educational Psychology. In K. Harris, S. Graham, & T. Urdan (Vol. Eds.), *APA educational psychology handbook. Vol. 1. APA educational psychology handbook* (pp. 597–621). Washington, D.C.: American Psychological Association.
Grigg, S., Perera, H. N., McIlveen, P., & Svetleff, Z. (2018). Relations among math self efficacy, interest, intentions, and achievement: A social cognitive perspective. *Contemporary Educational Psychology, 53*. https://doi.org/10.1016/j.cedpsych.2018.01.007.
Harackiewicz, J. M., & Priniski, S. J. (2017). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology, 69*. https://doi.org/10.1146/annurev-psych-122216-011725 annurev–psych–122216–011725–27.
Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review, 13*, 191–209. https://doi.org/10.1023/A:1016667621114.
Hidi, S., Ainley, M., Berndorff, D., & Del Favero, L. (2006). The role of interest and self-efficacy in science-related expository writing. In S. Hidi, & P. Boscolo (Eds.). *Writing and motivation* (pp. 203). UK: Emerald Group.
Hidi, S., Berndorff, D., & Ainley, M. (2002). Children's argument writing, interest and self-efficacy: An intervention study. *Learning and Instruction, 12*(4), 429–446. https://doi.org/10.1016/s0959-4752(01)00009-3.
Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111–127. https://doi.org/10.1207/s15326985ep4102_4.
Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, a motivational variable that combines affective and cognitive functioning. *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 89–115). .
Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior, 49*, 245–250.
Hollands, F., & Escueta, M. (2019). How research informs educational technology decision-making in higher education: The role of external research versus internal research. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-019-09678-z.
Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118.
Ichihara, M., & Arai, K. (2004). The development of academic perceived competence and intrinsic interest: A cross-sectional study in Grade 4 through 9 students. *Tsukuba Psychological Research, 27*, 43–50.
James, W. (1983/1892). *The principles of psychology.* Cambridge, MA: Harvard University Press.
Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modelling* (2nd ed.). New York: Routledge.
Lumpkin, A., Achen, R. M., & Dodd, R. K. (2015). Student perceptions of active learning. *College Student Journal, 49*, 121–133.
Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103*, 391.
Marsh, H. W., Martin, A. J., Yeung, A., & Craven, R. (2016). Competence self-perceptions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.). *Handbook of competence and motivation: Theory and application* (pp. 85–115). New York, NY, US: The Guilford Press.
Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Developmental Psychology, 49*, 1194–1218. https://doi.org/10.1037/a0026913.
Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x.
Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education, 70*, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033.
Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., & Urdan, T. (2000). Manual for the patterns of adaptive learning scales. *Ann Arbor, 1001* (48109-41259).
Muthén, L., & Muthén, B. (2014). *Mplus Version 7.2.* Los Angeles, CA: Mplus.
Niemivirta, M., & Tapola, A. (2007). Self-efficacy, interest, and task performance. *Zeitschrift Fur Padagogische Psychologie, 21*, 241–250. https://doi.org/10.1024/1010-0652.21.3.241.
Nikolov, M., & Djigunović, J. (2006). Recent research on age, second language acquisition, and early foreign language learning. *Annual Review of Applied Linguistics, 26*, 234–260. https://doi.org/10.1017/S0267190506000122.
Nuutila, K., Tuominen, H., Tapola, A., Vainikainen, M.-P., & Niemivirta, M. (2018). Consistency, longitudinal stability, and predictions of elementary school students' task interest, success expectancy, and performance in mathematics. *Learning and Instruction, 56*, 73–83. https://doi.org/10.1016/j.learninstruc.2018.04.003.
Oblinger, D. G. (2004). The next generation of educational engagement. *Journal of Interactive Media in Education, 1*. https://doi.org/10.5334/2004-8-oblinger.
Oga-Baldwin, W. L. Q., & Fryer, L. K. (2020). Profiles of language learning motivation: Are new and own languages different? *Learning and Individual Differences, 79*, 101852. https://doi.org/10.1016/j.lindif.2020.101852.
Oga-Baldwin, W. L. Q., Fryer, L. K., & Larson-Hall, J. (2019). The critical role of the individual in language education: New directions from the learning sciences. *System, 86*, 102118. https://doi.org/10.1016/j.system.2019.102118.
Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*, 27–45. https://doi.org/10.1007/s10648-011-9179-2.
Parette, H. P., Hourcade, J., & Blum, C. (2011). Using animation in microsoft powerpoint to enhance engagement and learning in young learners with developmental delay. *Teaching Exceptional Children, 43*, 58–67. https://doi.org/10.1177/004005991104300406.
Raykov, T. (2009). Evaluation of scale reliability for unidimensional measures using

latent variable modeling. *Measurement and Evaluation in Counseling and Development, 42*, 223–232. https://doi.org/10.1177/0748175609344096.
Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist, 46*, 168–184. https://doi.org/10.1080/00461520.2011.587723.
Renninger, K. A., & Hidi, S. (2015). *The power of interest for motivation and engagement.* New York: Routledge.
Renninger, K. A., Hidi, S., & Krapp, A. (1992). *The role of interest in learning and development.* New York: Psychology Press.
Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*, 353–387. https://doi.org/10.1037/a0026838.
Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist, 26*, 299–333. https://doi.org/10.1080/00461520.1991.9653136.
Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143*, 565–600. https://doi.org/10.1037/bul0000098.
Stewart, J., Fryer, L. K., & Gibson, A. (2013). Assessing the dimensionality of three hypothesized sub-skills of L2 vocabulary proficiency. *JACET JOURNAL, 51*, 51–71. Retrieved from http://ci.nii.ac.jp/naid/110009597363.
Stewart, J., Gibson, A., & Fryer, L. K. (2012). Examining the reliability of a TOEIC Bridge practice test under 1 and 3 parameter item response models. *Shiken research bulletin. 16*. Retrieved from teval.jalt.org/sites/teval.jalt.org/files/SRB-16-2-Full.pdf.
Sun, J. C.-Y. (2014). Influence of polling technologies on student engagement: An analysis of student motivation, academic performance, and brainwave data. *Computers & Education, 72*, 80–89. https://doi.org/10.1016/j.compedu.2013.10.010.
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312. https://doi.org/10.1016/0959-4752(94)90003-5.
Voyer, D., & Voyer, S. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*, 1174–1204. https://doi.org/10.1037/a0036620.