

# Language learning gains among users of English Liulishuo

Anthony Green, Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire

Barry O'Sullivan, British Council

## Abstract

This study investigated improvements in English language ability (as measured by the British Council Aptis test) among 746 users of the *English Liulishuo* app, the flagship mobile app produced by LAIX Inc. (NYSE:LAIX), taking courses at three levels over a period of approximately two months.

Users reported a very positive learning experience: almost 90% stated they had enjoyed studying with the app and that they planned to use *English Liulishuo* in the future. Over 80% believed that it had been more effective than other ways in which they had studied English.

Results showed 72% of users improved their test scores. Scores for those who only used *English Liulishuo* to study improved by an average of 11.45 points overall on the *Aptis* test with about 40% of users progressing at least to the next higher level of the Common European Framework of Reference for Languages (CEFR). Users made the greatest improvements in their Reading and Speaking skills.

Most users of the three course levels included in the study saw their English test scores improve. Those in the lowest-level course made the greatest improvements in their scores and those in the highest-level course made the smallest improvements.

## Introduction

This study seeks to establish the extent of proficiency gains in English following a period of study of approximately two months using the main course provided through the *English Liulishuo* mobile app. Launched in 2013, *English Liulishuo* integrates AI technologies, including automated speech recognition, with language learning pedagogy, gamified features and social elements. It is intended to provide a learning environment that covers a range of language skills and that motivates and engages users (see [www.liulishuo.com/en/liulishuo.html](http://www.liulishuo.com/en/liulishuo.html)).

The main course, offered at eight levels, is designed to improve users' English proficiency including their listening, speaking, reading and writing skills. There are two types of lesson: presentation lessons and support lessons. There are five types of presentation: Listening, Vocabulary, Dialogue, Reading and Letters & Numbers. These involve visual and auditory input and are designed to help users to learn new points of knowledge. Users can listen to and repeat sentences and receive instant feedback on their pronunciation. Interactive exercises test understanding of the input through multiple choice questions and fill-in the blank formats. Presentation lessons are followed by support lessons, which may focus on Listening, Grammar, Vocabulary, Speaking, Matching or Reading and Dictation. Different types of support lesson involve different types of exercise. These Support Lessons are presented in a gamified way and act as tests to reinforce learning of the key language points from the presentation lessons.

This study improves on similar research conducted for *Duolingo* (Vesselinov & Grego 2012) and *Babbel* (Vesselinov & Grego 2016) by taking account of other ways in which participants may have studied in addition to using the app during the research period; by relating score improvements to an internationally interpretable framework for language use: the Common European Framework of Reference for Languages or CEFR; and by using a test of functional language use (the British Council's *Aptis* test) rather than relying on a simple test of grammar and vocabulary.

## Methods

### Participants

A group of *English Liulishuo* main course users was recruited for the study from those signing up to use the app. A test administered at the beginning of the course places individual users into one of the eight levels, and provides content customized to their level. The participants were targeted at the first three levels of the *English Liulishuo* main course programme (Levels 1 to 3: see Table 1) as these are the most popular among users.

The proportion of male and female participants was broadly reflective of the overall population of users: 69.7% of users included in this study were female compared with 64.2% of all users. Demographic information on participants is given in the results section below.

*Table 1 Levels 1-3 of the English Liulishuo main course programme*

Level	Description
L1	No apparent knowledge of English, or limited to very basic vocabulary or set routines. Not able to communicate in English, even about time, numbers and simple objects. Limited or no reading or spelling ability.
L2	Oral skills limited to a few everyday expressions and basic phrases in English, including simple commands, time, numbers, and spelling. Can introduce him/herself and others and can answer questions about personal details such as family, nationality and job or school. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help. Knows some grammar, such as pronouns, subject-verb agreement and simple 'Wh' question formation. Can read and write simple sentences.
L3	Can converse about school or job, family, daily routine, weekly schedule, likes/ dislikes, seasons and weather, but not about life history or future plans. Can give simple instructions and descriptions. Sentences are short and limited to basic, general information about familiar or routine topics. Can read and follow basic instructions.

### The British Council *Aptis* test

All participants were registered to take one British Council *Aptis* test (see [www.britishcouncil.org/exams/aptis](http://www.britishcouncil.org/exams/aptis)) shortly before the beginning of their course and another

shortly after the end. The period between the two tests averaged 64 days, ranging between 51 and 87 days.

The key advantages of employing *Aptis* for this purpose were considered to be that:

- *Aptis* is not based on the content of *English Liulishuo* and therefore offers an independent indicator of English language proficiency.
- *Aptis* is professionally developed by a major international testing agency that publishes evidence for the qualities of the test (see [www.britishcouncil.org/exams/aptis](http://www.britishcouncil.org/exams/aptis)).
- *Aptis* measures across the traditional four skills or speaking, listening, reading and writing, allowing us to estimate the impact of studying with *English Liulishuo* on performance in all these areas.
- evidence is available to support the relationship between *Aptis* and a widely interpretable framework for language learning (the CEFR) (see O’Sullivan 2015).

*Aptis* includes five subtests, one covering knowledge of the systems of the language (*Grammar & Vocabulary*) and the other four focusing on the ability to use the language for communication (*Speaking, Writing, Reading and Listening*). The CEFR describes English language proficiency across seven broadly defined levels (pre-A1–C2). The test covers levels A1 to B2 of the CEFR in the tested content with additional reporting of a pre-A1 (or A0) level, and a C level (not distinguishing between CEFR C1 and C2) for the highest scoring test takers.

The Grammar & Vocabulary, Reading and Listening subtests employ selected-response formats such as multiple choice, gap fill and matching tasks. These subtests are scored automatically within the computer delivery system. Speaking and Writing subtests require test takers to provide samples of spoken and written performance and are scored by trained raters. The Speaking test is a semi-direct test in which test takers respond to pre-recorded prompts. The Writing test is designed to reflect online written communication.

*Aptis* test results for each subtest are reported on a numerical scale (0–50) and as a CEFR level. An overall score and CEFR level is also reported if all subtests are completed by the test taker.

## Questionnaires

Alongside the *Aptis* tests and data on the time users had spent using the app, students completed two questionnaires, one before beginning their course and a second after completing it. These provided data on factors that have been widely found to impact on language learning outcomes including users’ backgrounds, motivations for learning English and attitudes towards *English Liulishuo* as well as their language learning experiences prior to and during their *English Liulishuo* course.

### *The first (pre-course) questionnaire*

The questionnaire administered before users began studying with *English Liulishuo* focussed on demographic variables, prior learning experiences and language learning motivation.

#### **Demographic background and language learning experience (9 questions)**

Items in this section of the first (pre-course) questionnaire asked users about their gender, age, study/employment status and previous experience of learning English or other foreign languages. The latter included the following questions:

- What foreign languages (other than English) have you learned?
- As a child, did you live with a parent, guardian or other close relation who could speak English?
- At which stage(s) of your schooling school did you study English? (kindergarten/ primary school/ secondary school/ university/ extra-curricular language school)
- Did you study English as your major subject at university?
- Since leaving school/ university have you studied English? (if the answer is 'yes', After leaving, how long have you been studying English?)
- Have you stayed in any English-speaking countries for more than two weeks? (if the answer is 'yes', How long have you spent in English-speaking countries?).

### Language learning motivation (32 questions)

This section consisted of 32 statements about English language learning motivation designed by Dr Janina Iwaniec from the University of Bath (Shepherd and Ainsworth, 2017). This questionnaire was provided by the British Council and has been used in British Council *English Impact* studies (see <https://www.britishcouncil.org/exam/aptis/research/english-impact>). The section is based on research on language learning motivation in three key areas:

- the L2 Motivational Self System (Dörnyei and Ushioda, 2009) which consists of three constructs; ideal L2 self, ought-to L2 self and language learning experience
- international orientation (Yashima, 2009)
- self-concept (Bong and Skaalvik, 2003).

A brief explanation of each scale provided by the British Council is presented in Table 2 below. For the seventh of the eight scales (friends and family encouragement), the term 'friends and family' was substituted for 'parental' or 'parents' in the original scale and related items. This reflected the focus of the current study on adult learners.

Table 2. Motivation scales (adapted from Shepherd and Ainsworth 2017, pp.21-3)

Scale	Descriptive name	Brief description	Item	Item wording (English)
<b>Ideal L2 self (IDEAL)</b>	<i>Personal language goals</i>	<b>Ideal L2 self</b> is an image of oneself as a proficient speaker of a second language (Dörnyei's 2005). Though it relates to the future-self, <i>ideal L2 self</i> needs to be considered attainable to retain its motivational properties. English Impact employed the Iwaniec (2014) scale as it encompasses the four skills of language learning.	IDEAL1	I imagine myself speaking English fluently
			IDEAL2	I imagine myself comfortably reading in English on the internet
			IDEAL3	I imagine myself easily being able to follow what others say to me in English
			IDEAL4	I imagine myself writing emails in English with ease
<b>Ought-to L2 self (OUGHT)</b>	<i>Social expectations</i>	The <b>ought-to L2 self</b> is based on the external expectations placed on students and relates to the "attributes that one believes one ought to possess in order to avoid possible negative outcomes" (Dörnyei, 2005, pp. 105-106).	OUGHT1	I consider learning English important because the people I respect think that I should do it.
			OUGHT2	Studying English is important to me because other people will respect me more if I have knowledge of English.
			OUGHT3	Studying English is important to me

Scale	Descriptive name	Brief description	Item	Item wording (English)
				because an educated person is supposed to be able to speak English.
			OUGHT4	Learning English is necessary because people surrounding me expect me to do so.
<b>Language learning experience (EXPER)</b>	<i>Interest in learning English</i>	<b>Language learning experience</b> is concerned with the influence of the immediate environment on language learning (Dörnyei, 2005) and implies a strong focus on language learning attitudes.	EXPER1	Learning English is really great.
			EXPER2	I look forward to my English classes.
			EXPER3	I find learning English really interesting.
			EXPER4	I really enjoy learning English.
<b>Instrumentality (INSTR)</b>	<i>Future opportunities</i>	<b>Instrumentality</b> represents motivation stemming from the practical benefits of language (Gardner & Lambert, 1972). This scale measures the perceptions of usefulness of English on job markets and future prospects.	INSTR1	I need English for my future career.
			INSTR2	The things I want to do in the future require me to use English.
			INSTR3	I study English because it will facilitate my job hunt in the future.
			INSTR4	I study English as it will help me to earn good money.
<b>International orientation (INTL)</b>	<i>Global communication</i>	<b>International orientation</b> is a construct recently developed in response to the changing role of English. It denotes an "interest in foreign or international affairs... readiness to interact with intercultural partners" (Yashima, 2000, p. 57). The scale used, found in Iwaniec (2014), lends itself to adaptations that take into account the growth of online interaction rather than travelling abroad.	INTL1	Studying English will help me understand different people from other countries.
			INTL2	In the future, I would really like to communicate with people from other countries.
			INTL3	In the future, I would really like to communicate with people from other countries online.
			INTL4	If I could speak English well, I could get to know more people from other countries via the internet.
<b>English self-concept (SELF)</b>	<i>Self-confidence in English</i>	<b>Self-concept</b> is "a person's perception of himself" (Shavelson, Hubner, & Stanton, 1976) and this scale relates to self-evaluation in the students' ability to study English. The most common measurement of self-concept is Marsh's (Marsh, 1990) Academic Self-Description Questionnaire, adapted to language learning by Iwaniec (2014).	SELF1	I usually get good marks in English.
			SELF2	Compared to other students I'm good at English.
			SELF3	I have always done well in English.
			SELF4	Studying English comes easy to me.
<b>Friends/family encouragement (FANDF)</b>	<i>Family expectations</i>	Like the ought-to L2 self, <b>friends and family encouragement</b> focuses on external expectation. There is a potential for their motivation to be influenced by their parents, other family members or friends. Parents are considered to be one of the three groups of important others, together with teachers and peers (Burden & Williams, 1997).	FANDF1	My friends/family think I need to know English to be well-educated.
			FANDF2	My friends/family have stressed the importance English will have for me in the future.
			FANDF3	My friends/family feel that it is very important for me to learn English.
			FANDF4	My friends/family encourage me to practice my English as

Scale	Descriptive name	Brief description	Item	Item wording (English)
<b>Motivated learning behaviour (MOTIV)</b>	<i>Level of effort</i>	<b>Motivated learning behaviour</b> attempts to measure the behavioural component of motivation, i.e. the reported amount of effort student invests in English language learning.		much as possible.
			MOTIV1	I work hard at learning English.
			MOTIV2	I think I'm doing my best to learn English.
			MOTIV3	I put a lot of effort into learning English.
			MOTIV4	I spend lots of time studying English.

These statements were presented in a randomised order and users were asked to give a response to each on a six-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree).

#### *The second (post course) questionnaire*

On the second (post course) questionnaire, users were asked whether they felt that using *English Liulishuo* had improved their ability to use English, how they regarded the app, and whether they had combined use of *English Liulishuo* with any other form of English study.

#### **Improvements in English language skills (4 questions)**

The four questions in this section focussed on improvements in English language skills:

After studying with *English Liulishuo*, in the past two months,

- My ability to speak English has improved
- My ability to listen to English has improved
- My ability to write English has improved
- My ability to read English has improved

Again, users were asked to respond to each statement on a six-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree).

#### **Attitudes towards using *English Liulishuo* (5 questions)**

This section addressed the experience of using *English Liulishuo*. It included three selected response items with response options from 1 (strongly disagree) to 6 (strongly agree):

- I plan to continue using *English Liulishuo* in the future
- I enjoyed studying with *English Liulishuo*
- *English Liulishuo* was more effective than other ways I have studied English

The section also included two open-response items:

- What did you like about using *English Liulishuo*?
- What would you like to change about *English Liulishuo*?

#### **Other methods of studying English (14 questions)**

To establish the extent to which any study other than *English Liulishuo* had contributed to the score gains observed, users were asked to report details of other English language study they had undertaken over the period of the course.

This section posed the question:

- In the last 3 months, other than studying with Liulishuo did you use or study English in any other ways?

Where users indicated that they had studied using other methods, they were asked to estimate the number of hours per week spent on that activity. The items included the following:

- I attended an English class
- I studied using other apps
- I studied using textbooks or dictionaries
- I studied lists of English words
- I did grammar exercises
- I read books, magazines or webpages in English
- I listened to English (for example, songs/radio programmes/stories etc.)
- I watched English programmes
- I wrote in English
- I spoke in English
- I translated English sentences into Chinese
- I translated Chinese sentences into English
- I studied English in another way (please specify).

All questionnaire items were initially prepared in English then translated professionally into Chinese.

### Data set

Before proceeding with the analysis, 66 (8.07%) of the 818 users were excluded because of implausible results on one or both of their Aptis tests: results that appeared very unlikely to represent their true level of proficiency. User responses were excluded from the analysis if both of two conditions were satisfied:

- i) the user scored 0 on any component of Aptis on either occasion and
- ii) this score was discrepant with performance on the other modules of the test.

Discrepant results were those with unusually large score differences between the modules. Large score differences are much more likely to be caused factors such as the test taker giving up part way through a test or becoming distracted rather than by real differences in ability across test parts.

These discrepant scores were identified as outliers that exhibited a standardised residual greater than 2 when scores on that component were regressed on scores on the test as a whole, or when scores on the productive skills (Writing and Speaking) were regressed on scores on the other three components (Listening, Reading and Grammar & Vocabulary). It was assumed that a score of 0 under these circumstances was more likely to indicate a failure to attempt the subtest than a real lack of proficiency. A further six users were excluded because they were under the age of 16 and therefore below the target age range for *Aptis*. The remaining data for 746 participants were subsequently used for analysis.

## Results

### Questionnaire data

#### Demographics

Summaries of demographic data for the 746 *English Liulishuo* users included in the analysis are displayed in Tables 3 and 4 below. The majority of users were female (520: 69.7%) and most (407: 54.1%) were employees. Three-quarters of the users included in this study (565: 75.7%) were aged between 21 and 35. There were significant differences across levels by gender ( $\chi^2 = 8.61$ ,  $df = 2$ ,  $p = .014$ ), by age ( $\chi^2 = 28.43$ ,  $df = 14$ ,  $p = .012$ ) and by occupation ( $\chi^2 = 54.14$ ,  $df = 8$ ,  $p = .000$ ) with users at Level 3 more likely to be female, aged 25 or under and to be university students than those at Level 1. Over half of the participants in the freelance worker and school student categories used the Level 1 course.

Table 3 Participant age and gender by course level

Age	Gender	Level 1	Level 2	Level 3	Total
16 to 20	Male	9	9	12	30
	Female	8	17	20	45
21 to 25	Male	21	20	19	60
	Female	49	47	64	160
26 to 30	Male	27	17	17	61
	Female	42	40	49	131
31 to 35	Male	21	7	14	42
	Female	37	34	40	111
36 to 40	Male	9	8	2	19
	Female	14	14	13	41
over 40	Male	11	0	3	14
	Female	18	7	7	32
<b>Total</b>		266	220	260	746

Table 4 Participant occupation by course level

Occupation	Level 1	Level 2	Level 3	Total
School student	10	8	2	20
University student	24	50	76	150
Employee	158	114	134	406
Freelance worker	54	24	20	98
Other	20	24	28	72
<b>Total</b>	266	220	260	746

#### English language experience

Most users (691, 92.6%) reported that they had studied English at secondary school; 474 (63.5%) at university (but only 60, 8.0% were or had been English majors) and 310 (41.6%) had studied English at primary school. A small minority of 32 (4.3%) had studied English at kindergarten. 99 (13.3%) reported that they had taken English classes outside school and 107 (14.3%) reported that they had studied English since graduating. 55 (7.4%) had spent over two weeks in an English-speaking country. 99 (13.3%) reported that they had studied a foreign language other than English: Japanese, Korean, French, German and Russian being the most popular. Only 23 (3.1%) of users reported that during their childhood they had a parent or other close family member that could speak English.

#### English language learning motivation

Responses on the section that addressed the users' motivations for learning English are shown in Table 5 divided by course level.



Table 5 Participant English language learning motivation by course level (Maximum score for each scale: 24)

	Level 1 (N=266)		Level 2 (N=220)		Level 3 (N=260)	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
<i>Ideal L2 self</i>	16.32	5.796	16.38	5.832	16.48	5.261
<i>Ought-to L2 self</i>	15.50	4.086	16.05	3.400	16.52	3.461
<i>Language learning experience</i>	19.11	3.804	19.30	3.384	19.35	3.495
<i>Instrumentality</i>	17.13	4.251	17.59	4.326	17.59	3.906
<i>International orientation</i>	20.88	3.883	21.31	3.475	20.54	3.954
<i>English self-concept</i>	9.56	5.125	10.79	4.847	12.87	4.247
<i>Friends and family encouragement</i>	16.83	5.798	17.06	5.754	17.36	5.619
<i>Motivated learning behaviour</i>	14.02	5.390	13.71	5.028	14.22	4.605

When users across the three courses were all considered together as a single group, the highest mean scores were for *International orientation* (mean 20.89, standard deviation 3.800) and *Language learning experience* (mean 19.25, standard deviation 3.574) with relatively low scores for *English self-concept* (mean 11.08, standard deviation 4.949) and *Motivated learning behaviour* (mean 14.00, standard deviation 5.018). These results suggest that the users were generally motivated by their interest in learning English and in communicating with people from other cultures, but did not regard themselves as being particularly successful or dedicated language learners.

#### *The second (post-course) questionnaire*

On the second (post course) questionnaire, users were also asked to judge whether they felt that their abilities to speak, listen to, write and read English had improved as a result of using *English Liulishuo* (Tables 5 and 6 below show the figures divided by course level). Response options ranged from 1 (strongly disagree) to 6 (strongly agree). Across the three course levels a substantial majority of users felt that their abilities to speak and listen to English had improved as a result of using English Liulishuo: 481 (64.5%) agreed (i.e. responded 4, 5 or 6) that their speaking ability had improved and 545 (73.1%) that their listening ability had improved. A smaller majority of 404 (54.2%) agreed that their reading abilities had improved, but only a minority of 272 (36.5%) believed that their writing ability had improved as a result of using the app.

Table 6 Self-assessed improvement in English language skills (N=746): descriptive statistics (1, strongly disagree to 6, strongly agree)

	Level 1 (N=266)		Level 2 (N=220)		Level 3 (N=260)	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
<i>My ability to speak English has improved</i>	3.92	1.305	4.04	1.216	4.18	1.148
<i>My ability to listen to English has improved</i>	3.97	1.333	4.27	1.162	4.52	1.019
<i>My ability to write English has improved</i>	3.06	1.398	3.28	1.229	3.38	1.184
<i>My ability to read English has improved</i>	3.61	1.308	3.76	1.183	3.84	1.196

Table 7 Self-assessed improvement in English language skills by course level: Summary of responses

		Level 1	2	3	Total
My ability to speak English has improved	Agree	160	136	185	481
	Disagree	106	84	75	265
My ability to listen to English has improved	Agree	159	161	225	545
	Disagree	107	59	35	201
My ability to write English has improved	Agree	83	84	105	272
	Disagree	183	136	155	474
My ability to read English has improved	Agree	128	122	154	404
	Disagree	138	98	106	342

In the second (post-course) questionnaire, users reported on the experience of using *English Liulishuo*. Responses to the first three (selected response) items are set out in Table 7 below. These suggest very positive attitudes towards the course among these users with 657 (88.1%) agreeing (i.e. selecting response option 4, 5 or 6) that they planned to use *English Liulishuo* in the future, 645 (86.5%) that they had enjoyed studying with the app and 605 (81.1%) that it had been more effective than other ways in which they had studied English.

Table 8 Users' attitudes towards English Liulishuo (N=746)

		Level 1	2	3	Total	
<b>I plan to continue using English Liulishuo in the future</b>	<i>Strongly disagree</i>	1	0	0	1	
		2	9	5	20	
		3	24	17	68	
		4	52	45	164	
		5	70	51	186	
	<i>Strongly agree</i>	6	111	102	94	307
<b>I enjoyed studying with English Liulishuo</b>	<i>Strongly disagree</i>	1	3	0	2	5
		2	6	7	6	19
		3	34	19	24	77
		4	55	58	65	178
		5	61	52	87	200
	<i>Strongly agree</i>	6	107	84	76	267
<b>English Liulishuo was more effective than other ways I have studied English</b>	<i>Strongly disagree</i>	1	3	2	2	7
		2	13	8	10	31
		3	33	31	39	103
		4	65	47	74	186
		5	64	66	75	205
	<i>Strongly agree</i>	6	88	66	60	214

Overall, the users expressed mainly positive views of *English Liulishuo*. Positive features mentioned by users in response to the question, What did you like about using *English Liulishuo*? included the

convenience of learning with a mobile app, the attractive design, the daily reminders to study, the feeling of working collectively (and competitively) with classmates, AI features and the quality and variety of content.

Many of the comments on aspects that could be improved (What would you like to change about *English Liulishuo*?) focused on technical issues including poor internet speed and lost connections as well as difficulties associated with taking *Aptis*. A high proportion asked for greater support within the app including Chinese translations of texts, wordlists, transcripts for recordings and guidance on pronunciation and grammar features. There were some complaints about a lack of flexibility and frustration at not being able to progress to the next level.

In order to estimate the extent to which any study other than *English Liulishuo* might have contributed to the score gains observed, users were asked to report details of any other English language study they had undertaken over the period of the course. 186 users (24.9%) reported that they had studied English in other ways during the *English Liulishuo* course period. This was most popular for users on the Level 3 course (32.7% reporting that they had also studied in other ways) and least popular for those on the Level 1 course (18.0%). 24.1% of learners on the Level 2 course reported that they had studied English in other ways.

Among the alternatives, the most popular included using other apps (121, 65.1% of those studying English in other ways reported this: 58 at Level 3, 32 at Level 2 and 31 at Level 1). Studying lists of English words was also popular (104, 55.9%), while only 16 (8.6%) reported that they had studied by writing in English. 71 (38.2%) had taken an English class, with most (64.8% of those taking classes) attending the class for five hours or less each week. The five users who took English classes for ten or more hours each week were all university students. Translating English sentences into Chinese was the only activity that was more popular among users on the Level 1 course: 20 reported doing this compared with 12 at Level 2 and 14 at Level 3.

Table 9 Frequency of methods of studying English other than *English Liulishuo*

<i>Study method</i>	<i>N</i>	<i>Hrs/ week</i>	<i>Study method</i>	<i>N</i>	<i>Hrs/ week</i>	<i>Study method</i>	<i>N</i>	<i>Hrs/ week</i>
I attended an English class	71	5.02	I did grammar exercises	34	3.36	I wrote in English	16	2.38
I studied using other apps	121	3.91	I read books, magazines or webpages in English	82	3.15	I spoke in English	35	3.30
I studied using textbooks or dictionaries	68	3.41	I listened to English (for example, songs/radio programmes/stories etc.)	94	2.46	I translated English sentences into Chinese	46	1.99
I studied lists of English words	104	2.86	I watched English programmes	48	2.56	I translated Chinese sentences into English	25	3.83

Usage data revealed that participating users studied with *English Liulishuo* for an average of 53.86 days, ranging from 28 to 63 days. They studied for an average of 40.58 minutes each day: 36.88 hours over the duration of their *English Liulishuo* course. Those who studied English in other ways in

addition to using *English Liulishuo* reported spending on average over ten times as long on those other methods (626.3 hours) than on *English Liulishuo* (57.3 hours). However, it should be noted first that the data on use of other methods relies on users' own subjective reports of the time they spent on these (thus it is therefore not directly comparable to the more objective records we have of the time they spent logged on to the app) and second that some of time spent on the activities listed is likely to have overlapped. For example, users may have completed grammar exercises and translated sentences while using an app, or while studying textbooks as part of an English class.

## Score gains: descriptive statistics

### *Score gains and CEFR levels*

As any score differences between two test results might be attributable to factors such as measurement error, it is important to estimate the reliability of the test scores. Both *Aptis* tests proved to be highly reliable. The reliability coefficient (Cronbach's alpha) for the first *Aptis* test was .932 and for the second it was .923.

The average score gain on *Aptis* tests given before and after the course, was 12.58 points with a standard deviation of 25.469 (Table 10). Users who reported that they had also studied in other ways in addition to using *English Liulishuo* made greater improvements in their *Aptis* scores. We divided these users into two groups – those making extensive (over eight hours per week) use of other methods and those making more limited (8 hours or less) use of other methods. The 96 users who made the most use of other methods achieved the highest average improvements in *Aptis* scores: 17.07 points (standard deviation, 33.021) followed by the 90 who used other methods for 8 hours or less per week (13.94 points, standard deviation, 25.560). Among the three course levels, users placed into Level 1 and making extensive use of other methods improved their scores the most: an average of 24.19 points (standard deviation, 23.23 points). The 560 users who only studied with *English Liulishuo* improved by an average of 11.59 points (standard deviation, 23.884), although it should also be noted that those who used other methods alongside *English Liulishuo* studied for many more hours than those who used *English Liulishuo* alone. In the majority of cases, they studied for an additional 10 hours or more each week.

A more detailed analysis of the performance of the group that used *English Liulishuo* exclusively, investigating their rate of learning, is presented in the section headed *Rate of language learning* below.

Table 10 *Aptis* scores before and after using the *English Liulishuo* app

	Pre-course <i>Aptis</i> scores				Post-course <i>Aptis</i> scores			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
<i>Overall</i>	87.25	46.900	0.0	175.0	99.838	46.9538	0.0	187.0
<i>Listening</i>	22.34	10.614	0.0	48.0	23.788	9.5854	0.0	44.0
<i>Speaking</i>	19.29	14.148	0.0	45.0	22.685	14.3638	0.0	45.0
<i>Writing</i>	22.94	14.245	0.0	46.0	24.853	14.0118	0.0	48.0
<i>Reading</i>	22.68	13.163	0.0	50.0	28.512	14.8145	0.0	50.0
<i>Grammar &amp; vocab.</i>	25.48	10.704	0.0	48.0	26.300	10.9604	0.0	49.0

Paired sample *t*-tests (Table 11) indicated that score improvements were significant for total scores. Effect sizes (Cohen's *d*) for overall score differences were moderate: .495 for all users and .485 when

those using other methods were excluded. There were also significant differences between pre-course and post-course scores across all subtests ( $p < .001$ ) both for users who studied with *English Liulishuo* alone and for those also using other methods except on the *Aptis Grammar & Vocabulary* subtest. For this subtest, results were significant for the group of 746 as a whole ( $p = .001$ ), but not when those also studying by other methods were excluded ( $p = .055$ ). Effect sizes for the subtests were generally small. For all users the figures were .173 for Listening, .330 for Speaking, .200 for Writing and .119 for Grammar & Vocabulary. For those who only used *English Liulishuo*, they were .186 for Listening, .316 for Speaking, .168 for Writing. Moderate effects were observed for Reading: 564 for all users and .540 for those who only used *English Liulishuo*.

Table 11 t-tests for significance of Aptis score gains

	All English Liulishuo users (N=746)			Excluding users of other methods (N=560)		
	t	df	Sig.	t	df	Sig.
<b>Overall</b>	-13.494	745	.000	-11.441	558	.000
<b>Listening</b>	-4.728	745	.000	-4.343	558	.000
<b>Speaking</b>	-8.998	745	.000	-7.412	558	.000
<b>Writing</b>	-5.449	745	.000	-3.921	558	.000
<b>Reading</b>	-15.388	745	.000	-12.738	558	.000
<b>Gram. &amp; vocab.</b>	-3.243	745	.001	-1.921	558	.055

Table 12 below displays figures for users at each of the *English Liulishuo* course levels. This shows that users taking the Level 1 course made the greatest improvement in their *Aptis* scores and those taking the Level 3 course made the least. Among the five *Aptis* subtests, the largest average gains across all levels were observed for Reading (mean, 5.83; standard deviation, 10.345) and Speaking (mean, 3.39; 10.290) with only marginal average gains observed for Grammar & Vocabulary (mean, .82.; standard deviation 6.886).

Table 12 Score gain on Aptis by course level and subtest

	Level 1				Level 2				Level 3			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
<b>Overall</b>	15.02	22.320	-67	91	13.13	26.664	-70	106	9.63	27.201	-121	165
<b>Listening</b>	2.60	8.296	-30	30	2.13	7.930	-22	30	-0.31	8.527	-28	36
<b>Speaking</b>	3.33	9.332	-33	43	4.19	10.886	-33	38	2.78	10.694	-38	45
<b>Writing</b>	3.29	9.224	-22	40	1.15	10.788	-28	38	1.16	8.778	-30	40
<b>Reading</b>	5.80	9.086	-34	36	5.65	11.286	-28	32	6.00	10.752	-36	44
<b>Grammar &amp; vocabulary</b>	1.02	7.417	-29	30	0.10	6.662	-18	31	1.23	6.473	-23	26

537 users (72.0 %) achieved a higher score on the second *Aptis* test following the end of the course. Among the subtests, the highest proportion of users improving their scores was observed for the Reading subtest (539, 72.3%) and the lowest was for Listening (375, 50.3%). Across course levels, users made the highest average score gains in Reading (5.77 points) and the lowest in Grammar & Vocabulary (.80 points). 37.8% saw their test performance improve by at least one CEFR level. For 48.14%, outcomes remained at the same overall CEFR level while 14.1% saw a decline of one or more CEFR levels (Table 13).

Table 13 Number of learners placed at each CEFR level (based on *Aptis* test scores) before and after taking the English *Liulishuo* course.

Overall <i>Aptis</i> CEFR level (pre-course)	Overall <i>Aptis</i> CEFR level (post-course)											
	A0		A1		A2		B1		B2		C	
	N	Row %	N	Row %	N	Row %	N	Row %	N	Row %	N	Row %
A0			17	85.0%	3	15.0%						
A1	4	2.0%	102	51.0%	78	39.0%	15	7.5%	1	.5%		
A2	1	.6%	17	10.6%	64	40.0%	68	42.5%	10	6.3%		
B1			7	3.2%	30	13.7%	103	47.0%	78	35.6%	1	.5%
B2			1	.7%	2	1.4%	39	26.7%	91	62.3%	13	8.9%
C							3	42.9%	2	28.6%	2	28.6%

Figure 1 below shows the number of users classified at each CEFR level on the basis of their overall *Aptis* scores before and after their *English Liulishuo* courses.

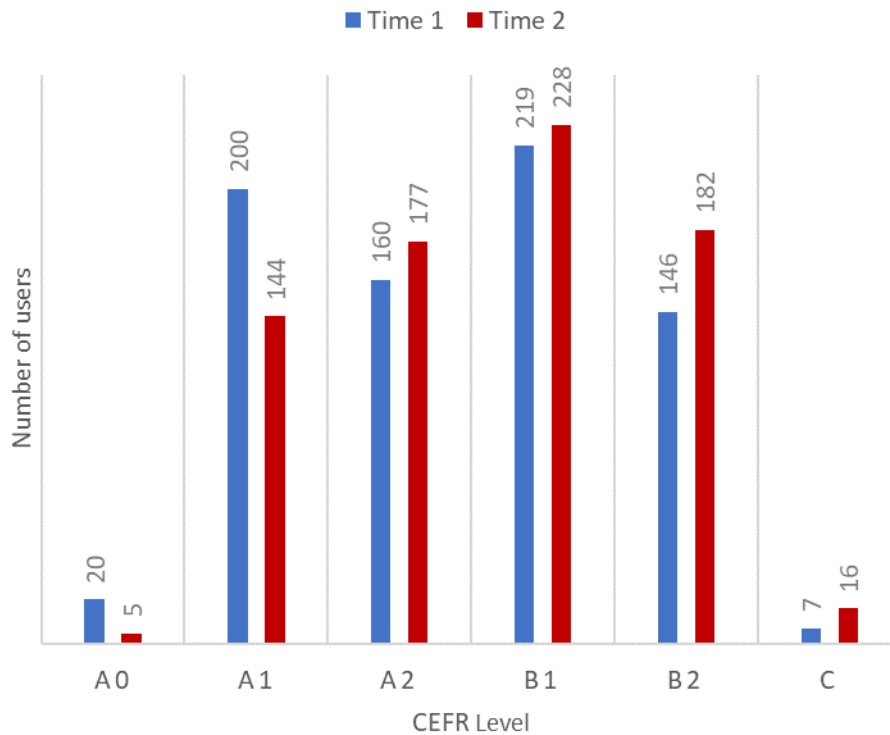


Figure 1. Number of learners at each CEFR level (based on Aptis scores) before (blue) and after (red) taking the English Liulishuo course (n=752)

### Rate of language learning

To take account of differences in the length of time that users had spent studying English, we calculated an index of English language learning rate. This involved dividing the observed gain in each user's overall *Aptis* score by the number of hours they had spent studying (the number of minutes spent logged on to the app added to any additional study they had reported). At this stage, one extreme outlier was identified and excluded from further analyses: this user scored 0 for Speaking before the course, but after studying for 28 days for an average of 28.6 minutes per day achieved 43 in Speaking following the course without making comparable gains in other skills. This left a data set of 559 users making average overall *Aptis* score gains of .0054 points per minute or 0.324 points per hour of English Liulishuo study (Figure 2). On the other hand, the group who studied English in other ways in addition to using *English Liulishuo* reported spending on average over ten times as long on those other methods (626.3 hours) than on *English Liulishuo* (57.3 hours) but only with similar scale of score gains (an average score gain of 15.6 points), leading to a learning rate of 0.03 points per hour of English study. Again, it should be noted that the data on use of other methods relies on users' own subjective reports of the time they spent on these and thus it is therefore not directly comparable to the more objective records we have of the time they spent logged on to the app. Nevertheless, the comparison of learning rates (0.324 versus 0.03) shows that learning with other methods does not appear to be as effective as using English Liulishuo alone.

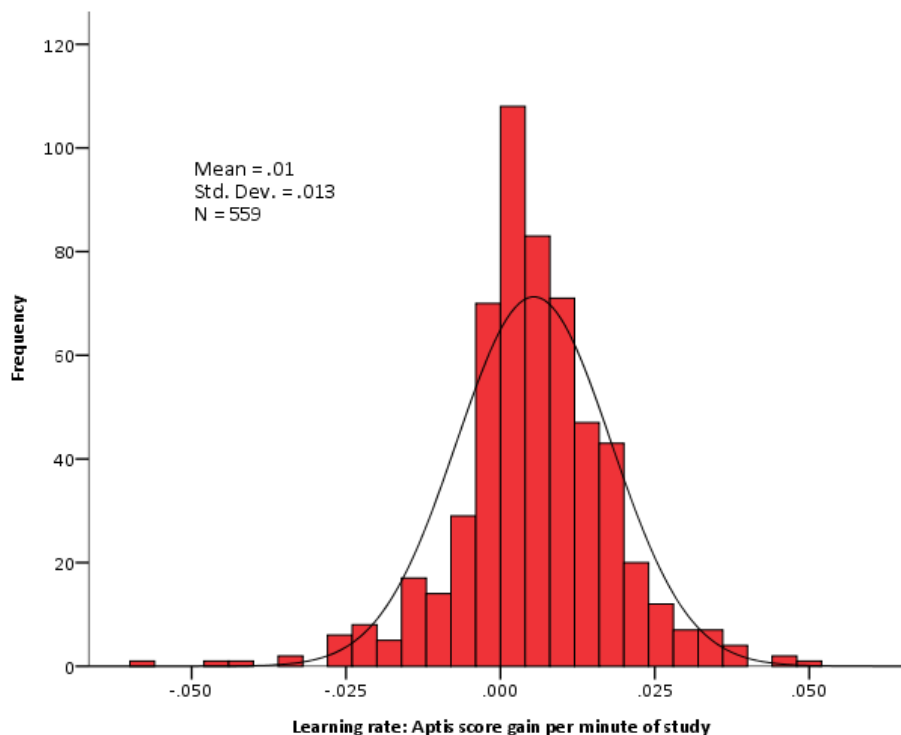


Figure 2 Distribution of learning rate scores (Aptis score gain per minute of study)

Addressing factors that have been found to impact on rates of learning, a one-way analysis of covariance was conducted to determine whether there were differences in learning rate between males and female users with age, *Aptis* score before the course, total time spent using the app and motivation scales as covariates.

Levene's test and normality checks were carried out and the assumptions met. There were significant effects for *Aptis* score before the course ( $F=24.152 (1,547)$ ,  $p=.000$ ,  $\eta^2=.042$ ) and for age ( $F=5.287 (1,547)$ ,  $p=.022$ ,  $\eta^2=.010$ ), but not for gender or for the length of time spent using the app. Older users tended to learn at a faster rate than their younger counterparts while those with higher scores before the course tended to improve at a slower rate. The non-significant result for length of time spent using *English Liulishuo* is encouraging as it suggests that, within the relatively short duration of the study, learners are able to sustain a consistent rate of learning and improvements in scores are unlikely to be attributable to factors other than language learning such as increased familiarity with *Aptis*.

Among the motivation scales, three produced significant results: *Language learning experience* ( $F=5.372 (1,547)$ ,  $p=.021$ ,  $\eta^2=.010$ ), *International orientation* ( $F= 4.658 (1,547)$ ,  $p= .031$ ,  $\eta^2=.008$ ) and *Motivated learning behaviour* ( $F= 7.038 (1,547)$ ,  $p= .008$ ,  $\eta^2=.013$ ). Higher scores on *Language learning experience* were associated with a faster learning rate, but higher scores on *International orientation* and *Motivated learning behaviour* were associated with slower learning rates. In short, those who generally enjoyed learning English seem to have benefitted more than those most interested in interacting with intercultural partners, or those who felt they were putting the most effort into their learning.

Finally, a series of ANOVA tests was conducted to explore the relationship between learning rate and users' judgements of their own improvement in language learning skills. These compared difference



scores on the four skills-based subtests among groups based on responses to the four questionnaire items, “After studying with *English Liulishuo*, in the past two months, my ability to speak/ listen to/ write/ read English has improved.” None of these tests yielded significant results, suggesting that users were not generally able to gauge the improvement they had made in using English.

## Conclusions

This study has demonstrated that a short period of study using *English Liulishuo* significantly improves users’ English language proficiency. Over 70% of the users improved their test scores. For over a third of users the improvement was sufficient to move them to a higher CEFR level on their second *Aptis* test. In common with most studies of language learning, we found the greatest score improvements for learners in the lowest level course, but users registered average gains across all three levels.

Improvements in proficiency as measured by the *Aptis* subtests were greatest for Reading and least for Grammar & Vocabulary. For users who only learned through *English Liulishuo*, score differences were significant for all four skills-based subtests, but not for Grammar & Vocabulary. The results suggest that a brief period of study with *English Liulishuo* is more effective at developing functional English language skills than at improving knowledge of grammar and vocabulary. Interestingly, the results contradicted users’ perceptions of their own improvement: they were more confident that their listening and speaking skills had improved rather than their reading.

It was not possible directly to compare the effects of studying with *English Liulishuo* with the effects of an equivalent period of study using alternative methods. However, based on experience and a review of the available literature, Benigno, deJong and van Moere (2017) have presented some general guidelines on expected rates of language learning (Figure 3), suggesting that it can take from 95 to 480 hours of active study for a beginner learner to reach A1 and up to 4,500 hours to reach C1. It can be seen from Figure 3 that the estimates are based on much longer periods than were involved in this study and that rates of learning are highly variable. Nonetheless, if the approximate rates of learning observed in this study could be sustained, users of *English Liulishuo* would, on average, progress through the CEFR levels at a relatively fast rate, even though English and Chinese are not closely related languages (a fact that makes learning English challenging for Chinese speakers).

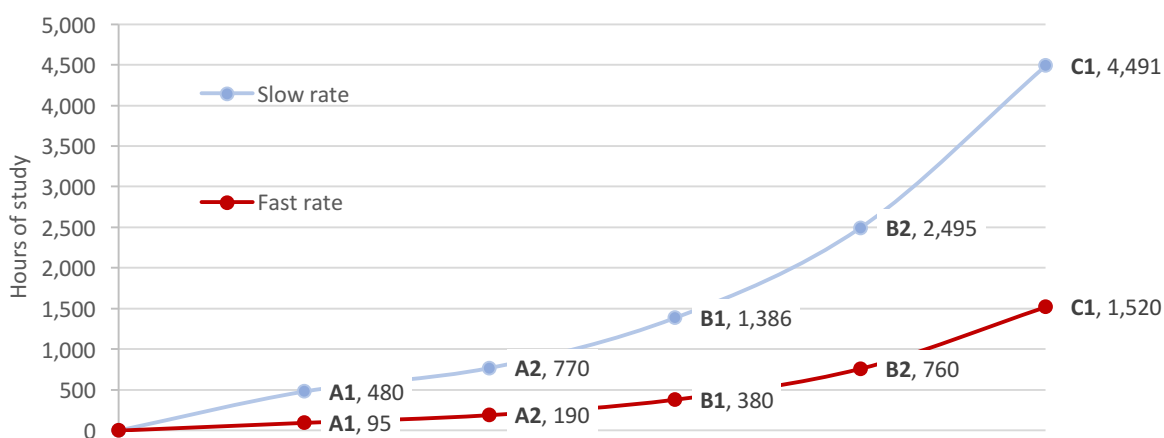


Figure 3 Pearson’s estimate of number of hours per increasing proficiency: from Benigno, deJong and van Moere (2017, p.8).

A quarter of the users made use of other methods of learning English during the study period. Although these users did make greater improvements in their *Aptis* scores than users who only studied with *English Liulishuo*, they appeared to achieve a much slower rate of learning than those using *Liulishuo* alone. This study has demonstrated that *Liulishuo English* is a popular and effective method for learning English that can bring relatively quick improvements.

## References

- Benigno, V., de Jong, J. and van Moere, A. (2017). *How long does it take to learn a language? Insights from research on language learning. Global Scale of English Research Series*. London: Pearson.
- Bong, M and Skaalvik, EM (2003) Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1–40.
- Burden, R. L., & Williams, M. (1997). *Psychology for language teachers: a social constructivist approach*. Cambridge: Cambridge University Press.
- Dörnyei, Z and Ushioda, E (eds) (2009). *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters.
- Iwaniec, J. (2014). Motivation of pupils from Southern Poland to learn English. DOI: 10.1016/j.system.2014.05.003
- O’Sullivan, B. (2015). *Linking the Aptis reporting scales to the CEFR*. Technical report. English Language Assessment Research Group, British Council.
- Shepherd, E. & Ainsworth, V. (2017). *English Impact Report – Madrid*. London: British Council.
- Yashima, T (2009) ‘International posture and the ideal L2 self in the Japanese EFL context’ in Z Dörnyei and E Ushioda (eds), *Motivation, language identity and the L2 self*. Bristol: Multilingual Matters, 215–228.
- Vesselinov, R., & Grego, J. (2012). Duolingo effectiveness study: Final Report. *City University of New York, USA*.
- Vesselinov, R., & Grego, J. (2016). The Babbel efficacy study: Final Report. *City University of New York, USA*.