Exploring the relationship between extramural English, self-efficacy, gender, and learning outcomes: A mixed-methods study in a Norwegian upper-secondary school☆

Alison Jones Rød [a], Raees Calafato [b,*,1]

[a] Department of Foreign Languages, University of Bergen, HF-bygget, Sydnesplassen 7, 5020 Bergen, Norway
[b] Department of Languages and Literature Studies, Faculty of Humanities, Sports, and Educational Science, University of South-Eastern Norway, Grønland 58, 3045 Drammen, Norway

A B S T R A C T

This article reports on a mixed-methods study regarding the extent to which the extramural English (EE), external attributions, self-efficacy (concerning EFL reading, speaking, writing, and listening skills), and gender of 42 students, learning English as a foreign language (EFL) in a Norwegian upper-secondary school, predicted their EFL learning outcomes. Data on participants' EE (receptive and productive), external attributions, and self-efficacy were collected through a questionnaire and language diaries, while their learning outcomes were measured through a language proficiency test, mock exam, in-depth project, and receptive and productive vocabulary tests. The data revealed several interesting findings, including participants' receptive EE statistically significantly and negatively predicting their productive vocabulary test scores, whereas their writing self-efficacy and attributions to specifically literature, TV, and film statistically significantly and positively predicted them. Moreover, neither receptive nor productive EE was found to mediate the relationship between self-efficacy, external attributions, and learning outcomes.

1. Introduction

Students in Norway need to acquire English language skills that will allow them to effectively interact with others nationally and globally in diverse contexts. The English curriculum for schools, published by the Norwegian Directorate for Education and Training (UDIR), states that English "is an important subject when it comes to cultural understanding, communication, all-round education, and identity development" and that it "shall help the pupils to develop an intercultural understanding of different ways of living, ways of thinking and communication patterns" (UDIR, 2022). English is also the only foreign language that all students are required to learn, starting in the second grade. Alongside their Scandinavian counterparts from Sweden and Denmark, students from Norway possess relatively higher levels of proficiency in English than those in most other European countries (EF, 2022), something which researchers (e.g., Bonnet, 2004; Busby, 2021; Olsson, 2012) have attributed to the time students in Scandinavia spend exposed to

English outside the classroom. At the same time, this exposure to English outside the classroom, referred to as 'extramural English' (EE) by Sundqvist (2009), has received comparatively little attention when compared "to the many thousands of published studies on classroom language learning" (Benson, 2011, p. 8). Indeed, even after a flurry of activity in the years following Benson's (2011) observation (e.g., Brevik, 2019; Peters, 2018; Sundqvist, 2019; Sundqvist & Wikström, 2015; Warnby, 2022), EE remains a relatively new field of research, having been mostly explored in connection with English vocabulary knowledge (e.g., Sundqvist & Wikström, 2015), motivation, and English reading proficiency (e.g., Brevik, 2016; Leona, et al., 2021), with few studies (e. g., Schurz et., 2022) venturing beyond these variables.

Given that a growing body of research recognizes an increase in EE amongst students, for instance, through social media use (Berns, 2007; Norwegian Media Authority, 2020), it becomes necessary to study how and, more importantly, what *types* of EE (e.g., gaming, watching TV and film, etc.) predict learning outcomes (e.g., writing, reading, vocabulary

---

knowledge, etc.) in learners, with these types preferably explored together in studies rather than in isolation. The value of such research lies in its ability to furnish us with data that would help us to better understand how learners are exposed to English in their lives outside the classroom and how this exposure affects their scholastic achievement, leading to the development of didactic approaches more in sync with EE. The resulting synergies between EE and classroom instruction thus achieved would contribute to making students' language experiences in formal learning contexts more meaningful (Henry et al., 2018). Of particular interest to OECD countries, including Norway, would be the development of EE-inspired didactic approaches that close the gender gap vis-à-vis scholastic achievement (for a meta-analysis, see Voyer & Voyer, 2014). Although lower levels of scholastic achievement amongst male students in OECD countries, when compared to female students, is not a recent phenomenon, it has become increasingly pronounced (Borgonovi et al., 2018). For example, PISA 2018, like its 2009 iteration, revealed that the gender achievement gap in reading in Norway was higher than the average gap across all OECD countries (see Borgonovi et al., 2018; Frønes et al., 2020). Data collected on students' EE and how it might predict their learning outcomes could thus provide us with insights to use in developing approaches aimed at improving outcomes for all students, especially males, who appear to be falling behind.

This article reports on a twenty-one-month mixed-methods study, conducted at an upper-secondary school in Norway, that examined the extent to which students' EE, gender, self-efficacy (related to their EFL reading, speaking, listening, and writing skills), and external attributions (to diverse activities, overall EE, and school experiences for contributing to their overall proficiency in English) predicted their learning outcomes, as measured via an English language proficiency test administered by the school, a mock exam, an in-depth project, and receptive and productive vocabulary tests. The study also investigated whether EE mediated the relationship between participants' self-efficacy, external attributions, gender, and learning outcomes. The decision to include self-efficacy and external attributions as variables was because of their purportedly positive links to EE and learning outcomes based on the social cognitive model of learning (Bandura, 1986; 1997; see also Henry, 2014), which this study used as its theoretic framework. The study's originality lies in its exploration of the relationship between different EE types, learning outcomes, as measured through multiple forms of formal assessment, and self-efficacy related to reading, speaking, listening, and writing, as well as gender, and external attributions, including the mediating effect that EE has on the relationship between learning outcomes and the other variables mentioned above. The study also placed a stronger emphasis on gathering quantitative data when compared to other EE studies conducted in Norway, where qualitative approaches have sometimes been given greater emphasis than quantitative ones (e.g., Brevik, 2016; 2019). This article also contains didactics-specific recommendations for educational institutions and teachers based on the findings presented herein and offers suggestions for EE research going forward.

2. Extramural English: contours and gender

In the past decade or so, researchers have proposed several terms and concepts to describe learning that occurs outside the conventional classroom, such as '*out-of-class', 'out-of-school', 'outside school', 'extra-curricular*', '*after school*', and '*extramural*', sometimes using these interchangeably and in the context of L2 English learning (e.g., Brevik & Holm, 2022; Lai et al., 2015; Peters, 2018; Sundqvist, 2009). The formality of learning has also been used by a few researchers to differentiate between in- and out-of-school learning, with formal learning typically associated with in-school learning and informal with out-of-school learning. Eshach, (2007), for example, presents a useful table outlining *formal, non-formal,* and *informal* learning when discussing school visits to a science museum and suggests that the degree of formality is not easily defined with respect to out-of-school settings.

Sandlin et al. (2010), in discussing Giroux's *public pedagogy*, observe that many public spaces, including museums, are effective learning spaces and they implore teachers to think further than the traditional classroom. Several other researchers have similarly encouraged thinking away from the conventional classroom and the 'in-and-out of school' dichotomy, introducing concepts such as *affinity spaces* (Gee, 2004) and *digital wilds* (Sauro & Zourou, 2019). They ask us to "look beyond contexts directly embedded within or linked to formal and highly familiar educational institutions and practices" (Sauro & Zourou, 2019, p. 1) and suggest that we should "begin thinking of space as a physical and virtual meld; begin dealing with spaces and groups as squishy and not well-bounded…" (Gee, 2017, p. 28).

To introduce some order and consistency within the field, Benson (2011) introduced the term '*language learning beyond the classroom*' to act as an umbrella term. Within this, he identified four major dimensions of language learning from the literature: location, formality, pedagogy, and locus of control, as well as two key analytical constructs – setting and mode of practice. According to Benson (2011, p. 9), the terms "'after-school', 'extracurricular' and 'extramural' usually refer to additional programs in school that are less formal than regular lessons and possibly organized by the students themselves". This differs slightly from Sundqvist's (2009, p. 25) definition of *extramural English*, which we use as our operational definition in this study, and which is outlined in her doctoral thesis as situations where "the learner comes in contact with or is involved in English outside the walls of the English classroom". Discussions surrounding the appropriateness and contours of the different terms mentioned above notwithstanding, there is a growing sense among researchers and the teaching community about the need to develop a more detailed understanding of learning beyond the conventional classroom, including what it involves and how it relates to dynamics within the classroom. Brevik and Holm (2022, p. 10) suggest that "drawing on students' language profiles allows integration of affinity spaces into L2 teaching in a way that connects with students' language learning outside school". In previous studies, Brevik (2016, 2019) identified three distinct profiles: *gamers, surfers,* and *social media users*. Students with a typical gamer profile have been the focus of several studies in Norway (see the references above), being first identified as 'outliers' because of their higher scores in L2 (English) reading proficiency in comparison to their mother tongue (Norwegian), subsequently attributed to extensive gaming.

Studies have also found a positive correlation between playing video games and motivation (e.g., Brevik, 2016; Brevik & Hellekjær, 2018), as well as vocabulary knowledge (primarily receptive vocabulary), although it should be noted that productive vocabulary tests have been used in some studies (e.g. Sundqvist, 2019; Sundqvist & Wikström, 2015). Aside from gaming, several researchers have argued that out-of-class extensive reading (e.g., Nation, 2015) and TV viewing (Kuppens, 2010; Webb, 2015) are effective means by which to acquire incidental vocabulary (see also Webb, 2020). Studies looking at the relationships between EE and vocabulary, as already mentioned, make up the bulk of research into EE and English language acquisition (e.g., Bollansée et al., 2020; Jensen, 2017; Peters, 2018; Sundqvist, 2009; Warnby, 2022). Indeed, the importance of vocabulary has long been recognized as a key component of L2 English acquisition (e.g., Coxhead, 2000; Nation, 2015) since a significant correlation has been found between general vocabulary knowledge and L2 English achievement, as well as between academic English lexis and academic achievement (Skjelde & Coxhead, 2020). However, as Peters (2018, p. 143) in his study of Flemish students' 'out-of-class exposure' to L2 English points out, little is known about "the relationship between different *types* of exposure and vocabulary knowledge". Puimège and Peters (2019, p. 3) also conclude that "research has not yet addressed the question of which words are more likely to be picked up from EE", and Schmitt (2019), amongst others, has called for more research looking at how EE can best facilitate vocabulary acquisition. Beyond vocabulary acquisition, however, it also becomes apparent that researchers should start examining

interactions between multiple EE types and learning outcomes more comprehensively (and other variables that have been found to predict learning outcomes) to better understand how learning takes place across contexts.

As for gender and EE, several studies have found that boys tend to play more video games and have heavier exposure to English media than girls (e.g., Brevik, 2016; Jensen, 2017; Olsson & Sylvén, 2015; Sundqvist & Wikström, 2015), even if girls outperform boys in scholastic achievement, especially concerning language learning (Voyer & Voyer, 2014). As already stated, few studies have explored the relationship between different types of EE and how these might relate to learning outcomes (beyond just vocabulary), including from a gender perspective, though the fact that boys tend to engage more heavily in gaming and with English media overall, yet still lag behind girls in language learning outcomes might mean that gaming and English media exposure (i.e., certain EE types) may not be the best predictors of learning outcomes. At the same time, other EE types, for example, reading for pleasure (e.g., literature), on which boys spend less time than girls outside school, could have stronger links with learning outcomes. Sundqvist (2009), in her study of the relationship between EE and oral proficiency and vocabulary size among 80 secondary school students in Sweden, found a weak, positive correlation between EE and vocabulary knowledge, as well as a moderate, positive correlation between oral proficiency and EE (albeit only for boys). She suggested that these correlations might have to do with the type of EE that boys typically engage in versus girls, which tend to be oriented more toward productive rather than receptive language skills. Similarly, in a more recent study, through differentiating between the scale of social interaction whilst gaming, Sundqvist (2019) claims that gamers (especially of massively multiplayer games) have a more advanced productive vocabulary than non-gamers, as shown in both vocabulary tests and essays. Interesting as this may be, more corroborative evidence is needed to determine whether productive language skills-oriented EE positively correlates with learning outcomes to a greater extent than receptive language skills-oriented EE does. As such, Sundqvist's findings appear to suggest that researchers can obtain deeper insights into the relationship between EE and learning outcomes by categorizing EE based on the type of language skills targeted, as well as a more nuanced categorization of the EE activity itself. The traditional classification of language skills into receptive and productive should be developed to reflect the evolving concept of *language learning beyond the classroom*, as discussed in this section.

2.1. Self-efficacy, attributions, and extramural English: a theoretical framework

Under a social cognitive model of learning, which emphasizes how people learn and interact with their environment (e.g., through observation and imitation), and how cognitive processes such as attention, perception, and memory shape their beliefs, motivation, and behavior (Bandura, 1986; 1997), self-efficacy and attributions are seen as playing important roles in individuals' regulation of their behavior in formal learning environments and outside the classroom (see Fig. 1) (Henry, 2014). Discussing the potential synergy between EE and classroom practices, Henry et al. (2018, p. 267) note that it can enable "students to experience flexibility, immediacy, and autonomy, and generates opportunities for them to speak and to act as themselves", lead to "positive emotion that, in turn, develops and sustains motivation to engage with learning tasks", and "increase experiences of self-efficacy… and self-authenticity." Based on the model, one can, likewise, hypothesize that students with high self-efficacy may be more likely to seek out and take advantage of opportunities to practice English outside the classroom, such as through gaming or social media, thereby further enhancing their language learning outcomes, including in formal contexts. Due to their self-efficacy, they might also initiate and persist in EE as they believe in their ability to succeed, creating, in this way, a reinforcing cycle of

higher self-efficacy, increased EE, and better performance.

Self-efficacy can be defined as beliefs that individuals harbor about their ability to perform specific tasks (Bandura, 1997). Concerning language learning, self-efficacy may be categorized according to language skill (Solheim, 2011; Sun & Wang, 2020), for instance, the extent to which students believe they can write an academic essay in English or read contemporary literary fiction. Research indicates a significant correlation between self-efficacy and learning outcomes (e.g., Chao et al., 2019; Graham, 2011; Mills et al., 2006), though few EE studies (we were only able to locate one where self-efficacy was an explicitly examined variable: Sundqvist, 2009; 2011) have included self-efficacy as a variable; and those that have, report inconsistent results for any correlations observed between the two. In the study by Sundqvist (2009, 2011), for instance, self-efficacy statistically significantly (medium strength) correlated with EE for boys but not girls, even if the type of self-efficacy explored in the study focused exclusively on learning English in general (i.e., the items were worded in broad terms, with no mention of language skills or tasks to accomplish). Had there been greater specificity regarding the participants' self-efficacy (e.g., what skill-specific tasks they felt they could accomplish with English), one might have obtained a different set of correlation results. Another predictor of learning outcomes, and one that is related to self-efficacy (Graham, 2011), is attributions. Attributions are individuals' beliefs regarding what causes outcomes (Weiner, 1979). With respect to language learning, attributions might take the form of a student crediting their native speaker friends or multiplayer games such as MMORPGs[2] for advancing their language skills in the target language. They can be internal, meaning that "the behavior is caused by a characteristic of the individual" (Brady & Woolfson, 2008, p. 529), for instance, learners attributing their learning outcomes to their intelligence, or external, meaning that behavior is based on external circumstances like the environment (e.g., EE). Attributions can also be stable or unstable over time, and controllable or uncontrollable (Schunk, 1994).

Attributions have been shown to correlate with motivation and outcomes in learners, though not always (Cochran et al., 2010; Nakamura, 2018), and their interactions with EE, as well as the effect of these interactions on learning outcomes, have been little studied, and then often with a focus on *teachers'* attributions regarding their students' EE practices (e.g., Schurz et., 2022; Schurz & Sundqvist, 2022; Sundqvist, 2019). Much like self-efficacy, obtaining a better understanding of students' attributions is useful because these can influence the extent to which they engage in certain activities when it comes to learning (Henry, 2014). For example, school students who attribute their success in learning English to playing video games are likely to put more effort into the activity because they feel they will get a strong return on their investment. In doing so, they would also learn English more rapidly, which could then lead to better performance in certain language assessments at school. Meanwhile, a student who feels that school does not contribute to their English ability may not participate actively in class and expend little energy on tasks because they do not view these as beneficial (and this might hurt their performance in assessments). Gender, too, might have a role to play. Henry (2014, p. 112) believes that "gender roles and gender stereotyping might be highly implicated in the formation of beliefs about learning English", suggesting that boys attribute their English proficiency to out-of-school learning to a greater degree than girls, which can have implications for self-directed learning, motivation, and, ultimately, achievement. These effects notwithstanding, attributions may not always accurately predict learning outcomes, including when EE is involved, though studies on the interactions between attributions, EE, and learning outcomes appear to be absent from the literature at present. As Henry (2014, p. 104) notes, "there may exist a dissonance between the actual effects on proficiency of out-of-school encounters with English, and what students attribute to

---

[2] massively multiplayer online role-playing games (MMORPGs)
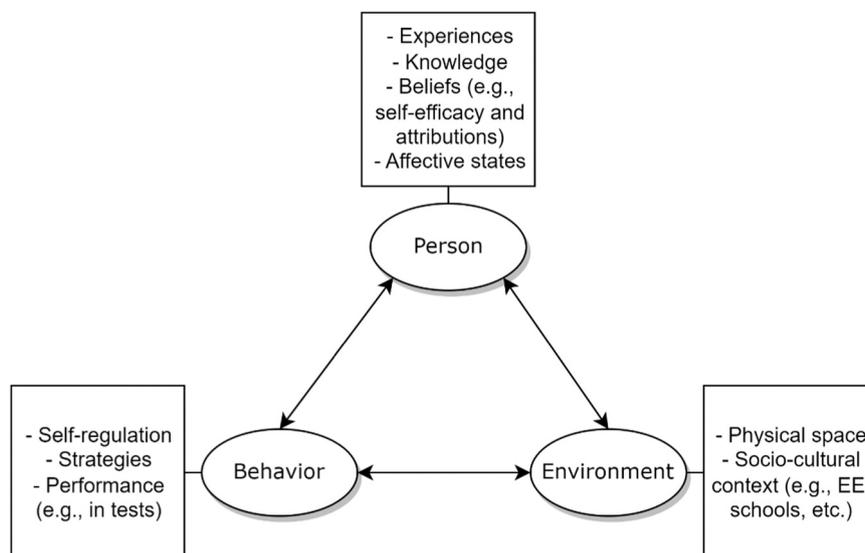
Fig. 1. The social cognitive model of learning.

such experiences", so that, "as with self-efficacy, research is needed on the nature and effects of the attributions students make in relation to the English language skills they develop".

2.2. Research questions

Given the limited number of studies on EE overall, especially those that have examined its relationship with self-efficacy, attributions, gender, and, more crucially, learning outcomes, this study investigated the following questions as part of its research focus:

1. Are there any gender differences in participants' EE, self-efficacy, external attributions, and learning outcomes?
2. How do participants' self-efficacy, EE, gender, and external attributions predict their learning outcomes?
3. Does participants' EE mediate the relationship between their self-efficacy, external attributions, gender, and learning outcomes?

3. Methods

3.1. Participants and data collection

Guidelines for this research project were provided by Norway's National Research Ethics Committees (NESH, 2021) and ethical approval was obtained from the Norwegian Data Protection Services (NSD). The participants, their guardians, and the school's principal (the study was conducted at one school) were informed of the study and presented with consent forms. It was made clear that participation was voluntary, and although guardian consent was not a requirement given the participants' age, they (i.e., the guardians) were nonetheless encouraged to co-sign the forms. Forty-two upper-secondary school students (19 males and 23 females), from two consecutive first-year groups taking the technical general studies course (TAF),[3] participated in the study. TAF is a hybrid course that contains elements of general and vocational studies. Students obtain a Specialized University College and Admissions Certificate and a Certificate of Completion of Apprenticeship within a four-year study course. TAF students follow the General Studies English Curriculum during their first year alongside General Studies students.

Proficiency test and summative assessment grades on completion are similar amongst the General Studies and TAF classes at the school, with grades roughly mirroring both the county and national averages for these subjects (UDIR, 2023).[4] The TAF course is therefore a suitable option for exploring the links between EE, self-efficacy, external attributions, and EFL learning outcomes since the findings would have relevance for all school learners, regardless of whether they chose academic or vocational English (due to the course's broad scope). Much like in Sundqvist's (2009) study, background information was collected to form an understanding of the participants 'cultural capital'; the study group was largely homogenous. Forty participants identified Norwegian as their first language (one also had Lao as a second language), whereas two indicated other first languages (Russian or Flemish). All participants reported using English during travel, which overwhelmingly involved visiting countries in Europe (10 participants had also been to Asia, mostly India and Thailand, or the MENA region, specifically, the United Arab Emirates and Tunisia). Interestingly, just one of the 42 participants had never traveled outside Norway. The research took place over 21 months and used an explanatory sequential approach that contained elements of convergent and case study design.

Data collection involved a mixed-methods approach, with more emphasis placed on quantitative data. Specifically, we used a questionnaire, as well as language diaries with a subset of participants, to explore their EE, self-efficacy, and external attributions, and the marks from an in-depth project, a mock exam, a language proficiency test, and receptive and productive vocabulary tests to assess their learning outcomes. The research timeline and the various instruments are detailed in Table 1. The questionnaire was designed, in part, based on previous studies (Olsson, 2012; Sylven, 2006; Sundqvist, 2009) together with feedback from a pilot study ($N = 14$ students). It contained 51 items, comprising multiple choice and open-ended questions, which elicited sociobiographical information, EE, self-efficacy regarding EFL reading, writing, listening, and speaking skills, and external attributions (to primary, lower-secondary, and upper-secondary school, as well as to textbooks, literature, TV and film, friends, strangers, YouTube, and EE overall for improvements in English proficiency). Reliability coefficients (McDonald's omega; ω) for EFL reading ($\omega = .81$), writing ($\omega = .90$), listening ($\omega = .83$), and speaking ($\omega = .82$) self-efficacy indicated acceptable internal consistency. Reliability was not calculated for

---

[3] TAF; also known as YSK. TAF is an acronym for *tekniske allmennfag*. YSK is an acronym for *Yrkes- og studiekompetanse* (vocational and general study competence).

[4] For example, the summative assessment grade for General Studies English in 2017–2018: National average 4.3; County average 4.3; School's average 4.4.

Table 1
Research timeline and instruments.

Participants	Group	N	Instrument	Timeline[c]	Content
1st year TAF group Research year 1	A	25	Proficiency test Questionnaire In-depth project Receptive + productive vocabulary tests Mock exam	M1 M7–9 M7–9 M7–9	Reading, writing, vocabulary Sociobiographical, EE, self-efficacy, external attributions, Speaking, writing, technical vocabulary - All language skills, vocabulary
Case-study	Subset Aa	7[a]	Language Diaries	M8	EE real-time frequency
1st year TAF group Research year 2	B	17	Proficiency test Questionnaire In-depth project Receptive + productive vocabulary tests Mock exam	M13 M19–21 M19–21 M19–21	Reading, writing, vocabulary Sociobiographical, EE, self-efficacy, external attributions, Speaking, writing, technical vocabulary - All language skills, vocabulary
Case-Study	Subset Bb	7[b]	Language Diaries	M20	EE real-time frequency

[a] While seven students kept language diaries, one diary had to be discarded.
[b] Seven students filled in language diaries. Two of these had to be discarded.
[c] In months (M1 = first month of the project)

attributions and EE because the items had unique meanings and did not measure the same underlying construct. For EE, when doing data analysis, we categorized the type of exposure based on whether it involved mostly receptive (12 items; e.g., watching TV and film, including news shows, listening to music, podcasts, reading digital or print comics, novels, and short stories) or productive language skills (eight items; e.g., Skype, blogging, tweeting, writing emails, being active on discussion forums and social media platforms such as Facebook and Snapchat) and used composite scores for these two categories when performing statistical tests. We kept gaming as a separate EE category due to data suggesting that it is primarily an activity in which male students engage (Brevik, 2016; Jensen, 2017). Example items from the questionnaire can be found in the Appendix. One of the authors was present when participants filled in the questionnaire to answer any queries.

Set A and Set B, which refer respectively to the two consecutive first-year TAF groups (see Table 1), were given the questionnaire in the spring semester, and all students completed it. Students in Sets A and B were also asked to write a language diary for a week. Fourteen students volunteered, but three of the diaries had to be discarded (the participants had not filled in enough days). All students from the same set started their diaries, which were intended to serve as a secondary source of data concerning participants' EE frequency and activity (i.e., they were asked to record hours spent on EE and give a detailed account of their activity in an ordinary week covering three school days, two workdays[5] and a weekend). However, due to the limited number of language diaries, the data obtained from them were not used during significance testing and are not reported in this article. As already mentioned, we gathered data on participants' learning outcomes via an English proficiency test[6] that all students at the school where the project was conducted must complete at the start of their first (upper-

secondary) year. The test, which is marked out of 150 points, gives a comprehensive overview of students' learning outcomes and is divided into three main sections: reading comprehension, writing, and vocabulary. Other sources of learning outcome data included the mock exam (part of the TAF course) and in-depth project (marks for both provided in percentages). The former was a good indicator of students' learning outcomes, including vocabulary knowledge, since it entailed writing a short text and a longer essay that revolve around sociocultural themes at the forefront of events in English-speaking countries (ca. 1200 words). The in-depth project involves students formulating a research question within their vocational field, conducting a literature search, and, if applicable, gathering data from fellow students, culminating in a written report and PowerPoint presentation. The 2000-word level receptive (Webb et al., 2017) and productive (Laufer & Nation, 1999) vocabulary tests were used to assess participants' receptive and productive vocabulary skills.

3.2. Data analysis

The data were analyzed in SPSS 28. Mediation analysis (see Fig. 2 for the mediation model) with multiple regression was performed using PROCESS (Hayes, 2017) to evaluate the effects of participants' gender (X_G in Fig. 1), self-efficacy (reading, writing, speaking, and listening), external attributions (covariates C_{SEB} and C_{ATTR} respectively), and EE (categorized as 'receptive', 'productive', and 'gaming' and represented by the mediator variable M_{EE} in Fig. 2) on their learning outcomes (i.e., the outcome variable: Y_{EA}), as measured by the school-administered proficiency test, mock exam, in-depth project, and vocabulary (receptive and productive) tests.

In performing the regression, we checked the data for multicollinearity by calculating the Variance Inflation Factor (VIF) per item, and autocorrelations via the Durbin-Watson statistic (d). A Mann-Whitney U test was done to check for statistically significant differences between participants' self-efficacy, EE, and external attributions based on gender, while a one-way ANOVA was performed to ascertain if there were gender differences concerning participants' learning outcomes, as measured by the proficiency test, mock exam, in-depth project, and vocabulary tests (Levene's test was conducted to determine if both samples had equal variance). An alpha level of .05 was used for all significance testing and the effect size is reported via Hedge's g due to the study's small sample size (partial eta-squared, η_p^2, is reported as an
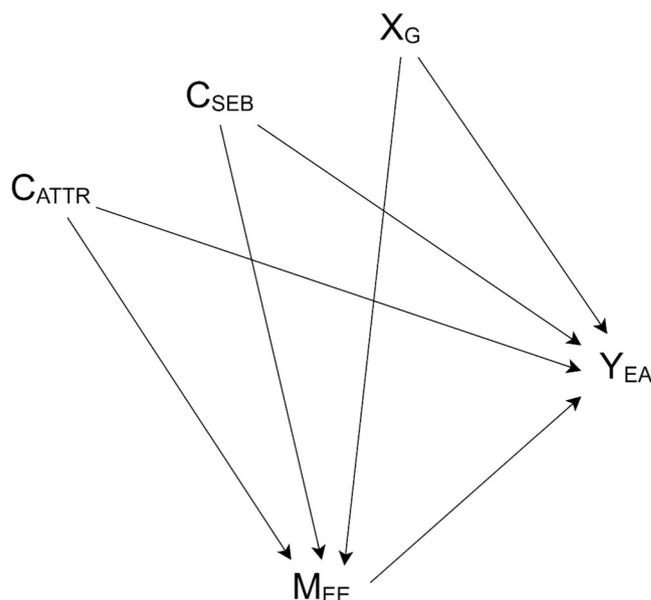
Fig. 2. Mediation model used in the study.

effect size for the regression). Observed power (*1-β*) is also provided for all results. When interpreting Hedge's *g*, .4 represents a weak effect size, .7 a medium effect size, and 1 or over a large effect size (Plonsky & Oswald, 2014). For η_p^2, .02 represents a small effect size, .13 a medium effect size, and .26 a large effect size (Miles & Shevlin, 2001).

4. Findings

Table 2 contains descriptive statistics for participants' productive (e. g., Skype), receptive (e.g., watching TV and film), and gaming EE, as well as their self-efficacy and external attributions. As can be seen from the table, participants' receptive EE was higher overall than their productive EE or gaming. As for gender, productive and receptive EE were similar for both male and female participants, whereas gaming was a more frequent activity for male participants. The data also indicated that

male participants had stronger self-efficacy for reading, listening, and speaking than female participants while the two group's writing self-efficacy was mostly analogous. Overall, participants were most confident in their listening skills (as evidenced by their self-efficacy), followed by reading, writing, and speaking (in that order). Gender differences can also be seen in participants' external attributions, with female participants crediting textbooks and literature for developing their English proficiency to a greater degree than male participants, who felt more strongly about YouTube helping them with their English proficiency than female participants. In general, participants felt that TV and film contributed to their English proficiency the most, whereas literature, encounters with strangers, and lower-secondary school had the least impact.

Mann-Whitney U test results indicated that gender differences were only statistically significant for gaming and attributions related to

Table 2
Participants' receptive, productive, and gaming EE based on gender.

		Gender	*n*	*M*	*SD*	*U*	*p*	*g*	*1-β*
EE[a]	Receptive (TV, films, music, podcasts, etc.)	Male	19	2.25	.46	232.00	.73	.06	.05
		Female	23	2.28	.53				
		Total	42	2.27	.50				
	Productive (emails, Skype, blogging, etc.)	Male	19	1.93	.44	205.50	.74	.11	.06
		Female	23	1.88	.46				
		Total	42	1.90	.45				
	Gaming	Male	19	1.97	1.17	132.00	.01	.70	.58
		Female	23	1.30	.75				
		Total	42	1.61	1.01				
Self-efficacy[b]	Reading	Male	19	4.37	.90	161.00	.10	.53	.37
		Female	23	3.91	.85				
		Total	42	4.12	.89				
	Listening	Male	19	5.00	.67	162.00	.13	.48	.31
		Female	23	4.57	1.04				
		Total	42	4.76	.91				
	Speaking	Male	19	3.63	1.01	168.50	.19	.49	.33
		Female	23	3.09	1.16				
		Total	42	3.33	1.12				
	Writing	Male	19	3.82	.62	200.50	.65	.20	.09
		Female	23	3.68	.75				
		Total	42	3.74	.69				
External attributions[c]	Textbooks	Male	19	3.16	.69	287.00	.06	.61	.47
		Female	23	3.65	.88				
		Total	42	3.43	.83				
	Literature (novels and short stories)	Male	19	2.42	1.26	264.00	.24	.30	.15
		Female	23	2.78	1.13				
		Total	42	2.62	1.19				
	TV and film	Male	19	4.00	.82	274.50	.11	.59	.44
		Female	23	4.39	.50				
		Total	42	4.21	.68				
	Friends	Male	19	3.42	1.22	229.00	.78	.14	.07
		Female	23	3.57	.90				
		Total	42	3.50	1.04				
	Encounters with strangers	Male	19	2.95	1.03	187.50	.42	.24	.11
		Female	23	2.70	1.02				
		Total	42	2.81	1.02				
	YouTube	Male	19	4.11	1.15	138.00	.03	.59	.44
		Female	23	3.48	.99				
		Total	42	3.76	1.10				
	Primary school	Male	19	3.05	1.08	243.50	.35	.29	.15
		Female	22	3.36	1.09				
		Total	41	3.22	1.08				
	Lower-secondary school	Male	19	3.00	1.05	200.00	.61	.10	.06
		Female	23	2.91	.67				
		Total	42	2.95	.85				
	Upper-secondary school	Male	19	3.74	.73	200.50	.62	.16	.08
		Female	23	3.61	.89				
		Total	42	3.67	.82				
	EE (Overall)	Male	19	3.11	.99	195.00	.53	.18	.09
		Female	23	2.96	.71				
		Total	42	3.02	.84				

Note. g = effect size; *1-β* = observed power
[a] 5-point scale (1 = never, 2 = up to an hour weekly, 3 = 1–5 h weekly, 4 = 5–10 h weekly, 5 = >10 h weekly)
[b] 6-point scale (1 = with great difficulty, 2 = with difficulty, 3 = with some difficulty, 4 = with some ease, 5 = with ease, 6 = with great ease)
[c] 5-point scale (1 = not relevant, 2 = almost nothing, 3 = somewhat, 4 = quite a lot, 5 = immensely)

YouTube (see Table 2), revealing that male participants engaged in gaming statistically significantly more frequently and believed statistically significantly more strongly that YouTube had contributed to the development of their English proficiency. The effect sizes were weak to medium in strength.

Table 3 provides descriptive statistics regarding the results of the various tests used to measure participants' learning outcomes. The data show a large gender difference for the mock exam and in-depth project (female participants performed better than their male counterparts), whereas results from the other tests suggest that both groups have similar learning outcomes. Moreover, participants scored higher on the receptive vocabulary test overall than on the productive vocabulary test.

One-way ANOVA test results revealed that gender differences were statistically significant for the in-depth project and mock exam but not for any of the other tests (see Table 3). Effect sizes were of medium strength. Levene's test indicated that there was homogeneity of variance for male and female participants in relation to their proficiency test ($p = .71$), mock exam ($p = .74$), in-depth project ($p = .28$), receptive vocabulary test ($p = .64$), and productive vocabulary test ($p = .15$).

Results from the regression, performed to ascertain the extent to which participants' gender, self-efficacy, external attributions (as independent variables), and EE (as a mediator variable) predicted their proficiency test, mock exam, in-depth project, and receptive and productive vocabulary test results (all outcome variables), showed that the model was statistically significant for the proficiency test [$R_N^2 = .67$, $F(22) = 2.60$, $p = .02$, $\eta_p^2 = .66$, $1\text{-}\beta = .84$, $d = 2.16$] and the receptive [$R_N^2 = .72$, $F(22) = 3.40$, $p < .001$, $\eta_p^2 = .74$, $1\text{-}\beta = .96$, $d = 2.20$] and productive vocabulary tests [$R_N^2 = .69$, $F(23) = 3.05$, $p = .01$, $\eta_p^2 = .70$, $1\text{-}\beta = .92$, $d = 2.39$], but not for the mock exam [$R_N^2 = .47$, $F(23) = 1.19$, $p = .34$, $\eta_p^2 = .47$, $1\text{-}\beta = .44$, $d = 1.76$] and in-depth project [$R_N^2 = .55$, $F(23) = 1.62$, $p = .14$, $\eta_p^2 = .57$, $1\text{-}\beta = .64$, $d = 1.68$]. Table 4 contains the parameter estimates for the regression model with the school-administered English proficiency test scores as the outcome variable. As can be seen from the table, participants' gender statistically significantly and positively predicted their proficiency test scores (i.e., being female correlated with higher scores), as did their writing and speaking self-efficacy.

Mediation analysis revealed a lack of statistically significant indirect effects from receptive EE [$\beta = -4.04$, $SE = 7.81$, 95%CI(-22.04, 8.00)], productive EE [$\beta = .61$, $SE = 4.86$, 95%CI(-8.83, 9.61)], or gaming [$\beta = -.20$, $SE = 4.61$, 95%CI(-9.05, 7.19)].

Table 5 contains the parameter estimates for the regression model with the receptive vocabulary test scores as the outcome variable. The data showed that reading self-efficacy and attributions to YouTube statistically significantly and positively predicted receptive vocabulary scores.

Here, too, neither receptive EE [$\beta = -.51$, $SE = .88$, 95%CI(-2.70, $.69$)], productive EE [$\beta = .08$, $SE = .68$, 95%CI($-.92$, 1.20)], nor gaming [$\beta = -.03$, $SE = .45$, 95%CI($-.88$, $.74$)] had any statistically significant indirect effects on vocabulary test scores.

Table 6 contains the parameter estimates for the regression model with the productive vocabulary test scores as the dependent variable.

As the table shows, participants' receptive EE statistically significantly and negatively predicted their productive vocabulary scores, whereas their writing self-efficacy and attributions to specifically literature, TV, and film statistically significantly and positively predicted them. Mediation analysis revealed an absence of statistically significant indirect effects from receptive EE [$\beta = -.99$, $SE = 1.81$, 95%CI(-4.97, 1.88)], productive EE [$\beta = -.09$, $SE = 1.08$, 95%CI(-2.00, 1.81)], or gaming [$\beta = -.08$, $SE = .91$, 95%CI(-1.49, 2.21)].

5. Discussion

This study explored whether there were any gender differences in participants' EE, self-efficacy, external attributions, and learning outcomes (as measured by multiple tests), the extent to which their gender, self-efficacy, EE, and external attributions predicted their learning outcomes, and whether EE mediated the relationship between their self-efficacy, external attributions, gender, and learning outcomes. Before embarking on a discussion of the findings, it is important to address the study's limitations. First, the study employed a small sample size, which affects the generalizability of the findings, though it is worth noting here that the study took place over 21 months and a relatively detailed EE profile of each participant was obtained. Second, our study relied primarily on quantitative data and analysis, unlike other EE studies from Norway (e.g., Brevik, 2016), which meant that we were not able to obtain more detailed insights into individual differences in participants' EE (e.g., whether they all watched TV and film in the same way, including their engagement with language form and function when doing so). Third, the study was conducted at one school in Norway, and it is possible, had other schools from other regions in the country been included, that the results would have been different. Finally, whilst this study chose to employ multiple data collection instruments, it recognizes that replicability of the research design may be difficult in some circumstances and require adjustments. The questionnaire has proved to be an effective means of collecting a broad spectrum of data, being consistently used in EE studies through the past three decades (e.g. Pickard, 1996; Sundqvist, 2009; Sundqvist, 2019) and its design is an important consideration. In contrast, language diaries may hold promise as a data collection instrument but can be problematic for participants to accurately and consistently maintain, especially younger learners. Ultimately, as previously mentioned, studies are needed that go beyond

Table 3
Performance on tests based on gender.

| | Gender | n | M | SD | ANOVA | | | | |
					F	df	p	g	1-β
School-administered proficiency test (out of 150)	Male	18	113.61	17.15	.93	1, 39	.34	.30	.15
	Female	23	118.61	15.87					
	Total	41	116.41	16.43					
Mock exam (out of 100)	Male	19	65.25	13.00	7.29	1, 40	.01	.84	.73
	Female	23	75.01	10.43					
	Total	42	70.60	12.52					
In-depth project (out of 100)	Male	19	67.98	11.15	7.92	1, 40	.008	.87	.76
	Female	23	78.30	12.35					
	Total	42	73.63	12.79					
Receptive vocabulary test (out of 30)	Male	19	28.45	2.06	.02	1, 39	.89	.05	.05
	Female	22	28.53	1.49					
	Total	41	28.49	1.76					
Productive vocabulary test (out of 18)	Male	19	11.34	3.86	.27	1, 40	.61	.16	.08
	Female	23	11.98	4.07					
	Total	42	11.69	3.94					

Note. g = effect size; 1-β = observed power

Table 4

Regression results for the relationship between gender, EE, self-efficacy, external attributions, and proficiency test scores.

		β	SE	t	p	95% CI LB	UB	η_p^2	1-β	Collinearity Tolerance	VIF
	Intercept	55.00	32.92	1.67	.11	-13.46	123.46	.11	.32		
	Gender	13.78	6.16	2.24	.04	.96	26.60	.19	.53	.45	2.22
EE	Receptive	-18.74	9.36	-2.00	.06	-38.21	.72	.16	.47	.19	5.17
	Productive	5.67	9.72	.58	.57	-14.55	25.90	.02	.08	.22	4.58
	Gaming	.81	2.94	.27	.79	-5.31	6.93	.00	.06	.48	2.07
Self-efficacy	Listening	-.07	3.73	-.02	.99	-7.83	7.69	.00	.05	.38	2.66
	Reading	-2.63	4.85	-.54	.59	-12.72	7.45	.02	.08	.23	4.38
	Speaking	6.29	3.07	2.05	.05	-.10	12.68	.16	.47	.35	2.83
	Writing	16.37	5.33	3.07	.01	5.29	27.45	.31	.82	.31	3.25
External attributions	Textbooks	-4.83	3.23	-1.49	.15	-11.55	1.89	.10	.29	.58	1.71
	Literature	3.71	3.09	1.20	.24	-2.72	10.14	.07	.21	.32	3.15
	TV and film	6.40	4.84	1.32	.20	-3.67	16.47	.07	.21	.54	1.87
	Friends	-5.75	3.09	-1.86	.08	-12.18	.68	.14	.42	.42	2.36
	Encounters with strangers	5.61	3.37	1.67	.11	-1.39	12.61	.12	.35	.42	2.40
	YouTube	1.39	2.97	.47	.65	-4.80	7.57	.01	.07	.40	2.48
	Primary school	-.89	2.59	-.34	.74	-6.28	4.50	.00	.06	.54	1.86
	Lower-secondary school	-1.15	4.92	-.23	.82	-11.37	9.08	.00	.05	.24	4.13
	Upper-secondary school	-2.52	4.19	-.60	.55	-11.23	6.20	.01	.08	.36	2.77
	EE (Overall)	-2.00	5.27	-.38	.71	-12.95	8.95	.01	.06	.22	4.56

Outcome Variable: Proficiency test

Note. SE = Standard Error; VIF = Variation Inflation Factor; CI = Confidence Interval; LB = Lower Bound; UB = Upper Bound; η_p^2 = effect size; 1-β = observed power

Table 5

Regression results for the relationship between gender, EE, self-efficacy, external attributions, and receptive vocabulary test scores.

		β	SE	t	p	95% CI LB	UB	η_p^2	1-β	Collinearity Tolerance	VIF
	Intercept	20.89	3.12	6.70	.00	14.41	27.38	.74	1.00		
	Gender	.87	.64	1.35	.19	-.47	2.21	.08	.24	.39	2.59
EE	Receptive	-1.74	.92	-1.90	.07	-3.65	.16	.15	.43	.19	5.29
	Productive	.49	.95	.51	.61	-1.48	2.45	.01	.08	.22	4.63
	Gaming	.19	.29	.66	.51	-.41	.79	.02	.09	.47	2.12
Self-efficacy	Listening	-.39	.38	-1.02	.32	-1.18	.40	.05	.16	.35	2.82
	Reading	1.89	.47	4.06	.00	.92	2.86	.42	.95	.23	4.29
	Speaking	-.24	.32	-.76	.45	-.91	.42	.03	.11	.32	3.09
	Writing	-.29	.50	-.58	.57	-1.33	.75	.00	.06	.34	2.94
External attributions	Textbooks	.14	.33	.43	.67	-.55	.83	.01	.08	.53	1.90
	Literature	.55	.30	1.82	.08	-.08	1.18	.13	.38	.31	3.19
	TV and film	.69	.43	1.62	.12	-.20	1.58	.16	.46	.47	2.11
	Friends	-.51	.29	-1.76	.09	-1.10	.09	.17	.49	.44	2.26
	Encounters with strangers	-.06	.33	-.18	.86	-.74	.62	.00	.05	.41	2.46
	YouTube	.63	.29	2.20	.04	.04	1.23	.17	.48	.40	2.50
	Primary school	-.13	.26	-.51	.62	-.66	.40	.01	.07	.56	1.79
	Lower-secondary school	-.01	.47	-.03	.98	-1.00	.97	.00	.05	.25	4.07
	Upper-secondary school	.20	.40	.50	.62	-.64	1.04	.02	.10	.36	2.75
	EE (Overall)	-.19	.50	-.38	.71	-1.23	.85	.02	.09	.23	4.34

Outcome Variable: Receptive vocabulary test

Note. SE = Standard Error; VIF = Variation Inflation Factor; CI = Confidence Interval; LB = Lower Bound; UB = Upper Bound; η_p^2 = effect size; 1-β = observed power

just looking at vocabulary and explore the relationship between different types of EE and other learning outcomes, which can similarly be indicative of language proficiency. These limitations notwithstanding, this study is one of the first to examine the extent to which EE, self-efficacy, external attributions, and gender are predictive of learning outcomes via multiple data collection instruments and an exploration of different EE types, EFL reading, writing, listening, and speaking self-efficacy, and external attributions for English proficiency to both EE and school. The following three subsections are organized according to the study's research questions.

5.1. Gender differences in participants' EE, self-efficacy, external attributions, and learning outcomes

The findings revealed that gender differences were present, yet statistically significant in only four instances: attributions regarding YouTube, gaming, and performance on the mock exam and in-depth project (the effect size was of medium strength). Specifically, male participants reported gaming to a greater extent and credited YouTube for their

English proficiency more strongly than female participants (see Table 2). Female participants, meanwhile, performed better on the mock exam and in-depth project. Regarding gaming, the findings support the evidence from other studies where boys have been found to play video games more regularly than girls (e.g., Brevik, 2016; Jensen, 2017). Overall, however, discounting gaming, there were no statistically significant differences between the two genders for their receptive *or* productive EE (the former type of EE being more frequent than the latter), meaning that both boys and girls at the upper-secondary level (at least at the school that participated in the project) had similar EE exposure. This contradicts the data from other studies in the Scandinavian context where boys have been reported to have heavier exposure to English media than girls on average (e.g., Brevik, 2016; Olsson, 2012). More interestingly, the various assessments implemented to gauge participants' learning outcomes revealed that male and female participants had similar levels of receptive and productive vocabulary knowledge (see Table 3). Overall, participants scored better on the receptive vocabulary test than on the productive vocabulary test, which is unsurprising given the general acceptance that language reception is easier

Table 6
Regression results for the relationship between gender, EE, self-efficacy, external attributions, and productive vocabulary test scores.

		β	SE	t	p	95% CI LB	UB	η_p^2	$1\text{-}\beta$	Collinearity Tolerance	VIF
	Intercept	-4.88	7.23	-.67	.51	-19.88	10.13	.18	.51		
	Gender	1.97	1.41	1.40	.18	-.95	4.90	.08	.24		
EE	Receptive	-4.59	2.14	-2.14	.04	-9.03	-.14	.18	.51	.44	2.28
	Productive	-.81	2.22	-.36	.72	-5.42	3.81	.01	.07	.19	5.21
	Gaming	-.33	.67	-.49	.63	-1.73	1.07	.01	.08	.22	4.62
Self-efficacy	Listening	1.20	.85	1.41	.17	-.57	2.97	.09	.27	.23	4.28
	Reading	-.23	1.08	-.21	.83	-2.48	2.02	.01	.07	.47	2.11
	Speaking	1.07	.70	1.52	.14	-.39	2.52	.10	.30	.37	2.71
	Writing	2.42	1.17	2.07	.05	.00	4.85	.21	.59	.35	2.85
External attributions	Textbooks	-1.38	.74	-1.87	.07	-2.92	.15	.13	.38	.33	3.03
	Literature	1.92	.71	2.73	.01	.46	3.39	.25	.69	.58	1.72
	TV and film	2.15	.99	2.17	.04	.10	4.20	.22	.62	.32	3.13
	Friends	-.39	.67	-.59	.56	-1.78	.99	.04	.15	.49	2.03
	Encounters with strangers	1.40	.76	1.83	.08	-.19	2.98	.16	.45	.45	2.21
	YouTube	.41	.67	.61	.55	-.98	1.80	.01	.07	.40	2.47
	Primary school	-.27	.59	-.45	.65	-1.50	.96	.01	.06	.40	2.49
	Lower-secondary school	-1.41	1.12	-1.26	.22	-3.72	.90	.05	.18	.54	1.86
	Upper-secondary school	.02	.94	.02	.98	-1.93	1.98	.00	.06	.25	4.06
	EE (Overall)	.09	1.18	.08	.94	-2.35	2.53	.00	.05	.37	2.72

Outcome Variable: Productive vocabulary test
Note. SE = Standard Error; VIF = Variation Inflation Factor; CI = Confidence Interval; LB = Lower Bound; UB = Upper Bound; η_p^2 = effect size; $1\text{-}\beta$ = observed power

than production. The participants also performed equally well in the school-administered proficiency test. The gender differences in the mock exam and in-depth project might, therefore, be due to the format of these assessments.

The in-depth project, for instance, involved research and organization skills while the mock exam required students to reflect on socio-cultural themes (see Section 3.1.). The female participants were likely more diligent and efficient during the research phase of the in-depth project and showed greater cultural awareness when tackling the mock exam. This argument finds partial support in the research literature, with studies indicating that females generally have higher levels of cultural empathy, self-control, and self-discipline (these last two would enhance performance in research-oriented tasks) than males (Cundiff & Komarraju, 2008; Solhaug & Osler, 2018), even if both genders are similar in their reflexivity (Thomson & Oppenheimer, 2016). Motivation could also have played a role: studies in Norway have shown wider gender gaps for classroom assessments versus national exams, which may relate to the stakes involved, whereby boys' motivation and effort increase as the stakes rise (as may have been the case with the proficiency test versus the in-depth project and mock exam) (Borgonovi et al., 2018). Meanwhile, statistically significant gender differences in external attributions were limited to YouTube (and participants attributed their English more to EE than to school, especially when considering lower-secondary; see Table 2), with male participants feeling statistically significantly more strongly that the platform contributed to their English proficiency. Jensen (2017), in her study of 107 Danish children, found that gaming correlated statistically significantly with vocabulary performance and that male participants engaged more frequently in gaming than female participants (much like in our study). She observed that male participants might "pay more attention to the language of the games they play and couple their gaming with walkthroughs of gameplay on YouTube as well as other clips, which are able to provide even more appropriate input" (p. 14), leading to improved learning outcomes.

5.2. Participants' self-efficacy, EE, gender, and external attributions as predictors of their learning outcomes

The findings revealed that gender and writing and speaking self-efficacy statistically significantly and positively predicted performance in the proficiency test, whereas reading self-efficacy and attributions to YouTube statistically significantly and positively predicted performance in the receptive vocabulary test. Only regarding the productive vocabulary test did EE (receptive) statistically significantly predict performance, and then negatively (and with a medium effect size). As with the proficiency test, writing self-efficacy again correlated statistically significantly and positively with productive vocabulary knowledge, as did attributions to literature, TV, and film. Taken together, the data show that self-efficacy, alongside external attributions, positively predicted performance in all three tests where the regression model was found to be statistically significant. In contrast, EE had no statistically significant relationship with performance, except for productive vocabulary. Gender's effects, too, were limited to the proficiency test, where differences were ultimately not statistically significant (see Table 3). As such, the most powerful predictors of performance in our study were self-efficacy and external attributions, with EE playing a less prominent role and gender being a statistically significant predictor of performance in only one instance. Self-efficacy has already been shown to correlate positively with language learning outcomes (Chao et al., 2019; Mills et al., 2006), with the social cognitive model of learning (Bandura, 1986; 1997) emphasizing that it is not enough for learners to use strategies in their learning or be exposed to a conducive learning environment; rather, they must also *believe* that they can accomplish the task at hand. Attributions, which, as already mentioned, are related to self-efficacy, function in much the same way (Graham, 2011), and so it was not unexpected to see both variables play a strong, predictive role when it came to participants' performance (even if teachers might not always be aware of these dynamics in students).

What was interesting, however, was that neither listening nor speaking self-efficacy was statistically significantly predictive of performance in any of the tests, perhaps reflecting the test formats, where listening and speaking were not given much weight. Another noteworthy finding was the statistically significant negative correlation found between receptive EE and productive vocabulary knowledge. Swain's (1985; 2001) research shows that students need to produce output alongside receiving comprehensible input to develop their communicative abilities in the target language. If we consider the fact that participants spent more time on receptive EE than productive EE and credited most of the improvement in their English proficiency to watching TV and film, then it comes as no surprise that their receptive EE negatively correlated with their productive vocabulary scores. Moreover, as Schmitt (2008) notes, "It is a commonsense notion that the more a learner engages with a new word, the more likely they are to learn it" (p. 338), which makes incidental exposure much less effective than intentional vocabulary learning where "explicit attention to learning the lexical items themselves" is incorporated into activities (p.

341). This is because, when incidental learning occurs, "learners who understand the overall message often do not pay attention to the precise meanings of individual words", and words that "are easily understood (guessed) from context may not generate enough engagement to be learned and remembered" (p. 341). As such, participants in this study were primarily honing their receptive skills in the language outside the school at the cost of less time allocated to actively and explicitly improving their productive vocabulary knowledge. The findings regarding the negative correlation between receptive EE and productive vocabulary also find partial support in the study by Bollansée et al. (2020), where the researchers found a weak, negative, statistically significant correlation between productive vocabulary knowledge and watching TV with L1 subtitles.

5.3. Participants' EE as mediating the relationship between their self-efficacy, external attributions, gender, and learning outcomes

The findings indicated, somewhat unexpectedly, that EE did not play a statistically significant mediating role for any of the measures of assessment used, suggesting that while participants' gender, self-efficacy, and external attributions were directly predictive of performance in some or all the assessments used to measure their learning outcomes, these variables (i.e., gender, self-efficacy, and external attributions) remained independent in this respect from any mediating influence exerted by EE. For instance, leaning on the social cognitive model of learning, we hypothesized that stronger self-efficacy and external attributions should positively correlate with EE (such beliefs would prompt students to practice their English outside the classroom more actively given their increased confidence in their abilities), which would then be reflected in an even stronger performance in the assessments (see Section 2.1.). However, higher self-efficacy regarding English reading, writing, speaking, and listening (and external attributions) did not lead to heightened EE (i.e., students who believed more strongly in their ability to accomplish tasks using their English language skills did not report more frequent EE), which, in turn, did not account for any heightened statistically significant indirect effect on their assessment scores. In other words, participants' exposure to English outside the classroom was perhaps not so much related to their beliefs about their English proficiency as it was to their desire to consume content that happened to be in English (and unavailable in Norwegian or the other languages they knew), that is, they had a pragmatic attitude towards the language during EE. This would corroborate the findings from the study by Brevik (2016), where she found that participants' EE was characterized by simply reading "the information they happened to come across in the language it appeared", so that, when gaming or engaging in other activities, their main focus was not on paying particular attention to the language itself. As Brevik (p. 53) notes, "Although these boys used English in their spare time by choice, they did not see how their English proficiency could be transferred to school activities unless specifically being presented with the idea."

6. Conclusion

This study explored the interactions between EE, self-efficacy, external attributions, gender, and learning outcomes among upper-secondary school students in Norway based on a social cognitive model of learning. From the study's findings, it becomes apparent that participants are frequently exposed to English outside the classroom yet might not be specifically attuned to noticing the language and using EE to enhance their language proficiency. Nevertheless, given their level of reported EE, it can become a significant source of active language learning if students' ability to notice language (and analyze it more critically) during EE is systematically developed. To support such development, teachers can implement activities in the classroom and as homework to actively monitor and help raise their students' EE-related language awareness while also encouraging them to be more attentive to

language outside the classroom. At the same time, before implementing activities, teachers should have a wide-ranging discussion (or a series of discussions) with their students regarding their EE to understand its contours, for example, whether their EE is mostly receptive or productive skills-oriented, as well as specific activities that students engage in. A fundamental part of this discussion should be whether the students consider this in-and-out-of-school synergy as an educational boon or an unwelcome impingement on their free time. If the students appear reluctant, then they need to be made aware of the benefits such an approach could have. In addition, a simple questionnaire can be used to collect this information. In terms of their EE contours, if teachers discover that students primarily engage in receptive EE, which negatively correlated with productive vocabulary knowledge in this study, they can prompt them to switch to more productive EE, with a stronger focus on the language itself (and not just content). Concerning gaming, for instance, teachers could encourage their students to record a segment of their play and then do a walkthrough of the segment where they must describe each element and action on the screen in English (they could either stream the walkthrough live or upload it to a site and present it before the class). These types of activities would prompt the students to focus more on not only *what* is being communicated but also on *how* it is being communicated in terms of the language (and how they communicate, in return).

In the classroom, teachers could show their students short clips from a film or TV show in English that they (i.e., the students) like, based on prior discussions with the students about their EE, and ask them to write subtitles in English for them (and then upload these to an online subtitle database). This activity, if done from time to time, would encourage students to more systematically and explicitly connect oral speech with text in English during EE and reflect more deeply on language form. Moreover, teachers do not even necessarily need to implement activities requiring active work with the language, at first; they could, as already mentioned, simply raise student's awareness of the benefits related to more actively noticing language during EE, for example, by discussing how spelling and pronunciation can be improved by paying attention to English subtitles when watching TV and noticing the lexical output, thereby rendering TV as an *edutainment* tool. The findings also underline the risks of assuming that students who report regularly being exposed to English outside the classroom are actually enhancing their language proficiency through such exposure. For language teachers, this means that they should be careful not to recommend that their students simply 'watch more TV shows in English' or 'listen to music' in the language without also emphasizing the need for students to reflect on (and pay attention to) the use of language in these instances (alongside content). Otherwise, students may not significantly benefit their language proficiency via EE and may incur negative effects in this respect. Furthermore, teachers should also raise students' awareness of the different types of vocabulary they are exposed to whilst involved in EE. Concerning future research directions, it is hoped that studies will start to approach EE through a more robust classification system based on the language skills targeted or other aspects of language learning (e.g., affective factors). This study applied rudimentary categorizations to EE that need to be developed into a more complex, nuanced system. Such a system will in turn lead to more sophisticated instruments for data collection. Future studies could also include an exploration of self-regulation effects and EE, specifically, if these two variables correlate (i.e., whether frequent EE accompanies greater self-regulation seeing as EE comprises generally planned activities where individuals have certain goals, even if these goals are not language-related).

Conflict of Interest

There are no conflicts of interest to declare.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.stueduc.2023.101302.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson, & H. Reinders (Eds.), *Beyond the Language Classroom* (pp. 7–16). Palgrave Macmillan.

Berns, M. (2007). The presence of English: Sociocultural, acquisitional, and media dimensions. In M. Berns, K. Bot, & U. Hasebrink (Eds.), *In the presence of English: Media and European youth* (pp. 1–14). Springer.

Bollansée, L., Puimège, E., & Peters, E. (2020). Watch out! Behind you is the enemy! An exploratory study into the relationship between extramural English and productive vocabulary knowledge. In V. Werner, & F. Tegge (Eds.), *Pop Culture in Language Education* (pp. 199–214). Routledge.

Bonnet, G. (2004). The assessment of pupils' skills in English in eight European countries. Ministère de l'Éducation Nationale.

Borgonovi, F., Ferrara, A., & S. Maghnouj, S. (2018). *The Gender Gap in Educational Outcomes in Norway*. OECD Publishing.

Brady, K., & Woolfson, L. (2008). What teacher factors influence their attributions for children's difficulties in learning? *British Journal of Educational Psychology, 78*(4), 527–544.

Brevik, L. M. (2016). The Gaming Outliers: Does out-of-school gaming improve boys' reading skills in English as a second language?. *Educational Technology and Polycontextual Bridging* (pp. 39–61). Brill,.

Brevik, L. M. (2019). Gamers, surfers, social media users: Unpacking the role of interest in English. *Journal of Computer Assisted Learning, 35*, 595–606.

Brevik, L. M., & Hellekjær, G. O. (2018). The Outliers: Upper secondary school students who read better in the L2 than in L1. *International Journal of Educational Research, 89*, 80–91.

Brevik, L. M., & Holm, T. (2022). Affinity and the classroom: Informal and formal L2 learning. *ELT Journal, ccac0*, 12.

Busby, N. L. (2021). Words from where? Predictors of L2 English vocabulary among Norwegian university students. *International Journal of Applied Linguistics, 172*(1), 58–84.

Chao, C. N., McInerney, D. M., & Bai, B. (2019). Self-efficacy and self-concept as predictors of language learning achievements in an Asian bilingual context. *The Asia-Pacific Education Researcher, 28*(2), 139–147.

Cochran, J. L., McCallum, R. S., & Bell, S. M. (2010). Three A's: How do attributions, attitudes, and aptitude contribute to foreign language learning? *Foreign Language Annals, 43*(4), 566–582.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238.

Cundiff, N. L., & Komarraju, M. (2008). Gender differences in ethnocultural empathy and attitudes toward men and women in authority. *Journal of Leadership & Organizational Studies, 15*(1), 5–15.

EF. (2022). EF English Proficiency Index. EF Education First.

Eshach, H. (2007). Bridging in-school and out-of-school learning: Formal, non-formal, and informal education. *Journal of Science Education and Technology, 16*(2), 171–188.

Frønes, T. S., Pettersen, A., Radišić, J., & Buchholtz, N. (2020). Equity, equality and diversity in the Nordic model of education: Contributions from large-scale studies. In T. S. Frønes, A. Pettersen, J. Radišić, & N. Buchholtz (Eds.), *Equity, equality and diversity in the Nordic model of education* (pp. 1–10). Springer.

Gee, J. P. (2004). *Situated learning and language: A critique of traditional schooling*. Routledge,.

Gee, J. P. (2017). Affinity spaces and 21st century learning. *Educational Technology, 57* (2), 27–31.

Graham, S. (2011). Self-efficacy and academic listening. *Journal of English for Academic Purposes, 10*(2), 113–117.

Hayes, A.F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford.

Henry, A. (2014). Swedish students' beliefs about learning English in and outside of school. In D. Lasagabaster, A. Doiz, & J. M. Sierra (Eds.), *Motivation and foreign language learning: From theory to practice* (pp. 93–116). John Benjamins.

Henry, A., Korp, H., Sundqvist, P., & Thorsen, C. (2018). Motivational strategies and the reframing of English: Activity design and challenges for teachers in contexts of extensive extramural encounters. *TESOL Quarterly, 52*(2), 247–273.

Jensen, S. H. (2017). Gaming as an English language learning resource among young children in Denmark. *Calico Journal, 34*(1), 1–19.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology, 35*(1), 65–85.

Lai, C., Zhu, W., & Gong, G. (2015). Understanding the quality of out-of-class English learning. *TESOL Quarterly, 49*(2), 278–308.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing, 16*(1), 33–51.

Leona, N. L., van Koert, M. J., van der Molen, M. W., Rispens, J. E., Tijms, J., & Snellings, P. (2021). Explaining individual differences in young English language learners' vocabulary knowledge: The role of Extramural English Exposure and motivation. *System, 96*, Article 102402.

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. Sage,.

Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals, 39*(2), 276–295.

Nakamura, S. (2018). How I see it: An exploratory study on attributions and emotions in L2 learning. *Studies in Second Language Learning and Teaching, 8*(3), 553–574.

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language, 27*(1), 136–145.

NESH. (2021). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora*. National Research Ethics Committee for the Social Sciences and Humanities.

Norwegian Media Authority. (2020). *Barn og Medier 2020: En kartlegging av 9–18-åringers digitale medievaner [Children and media 2020: A survey of 9–18 year olds' digital media habits]*. Norwegian Media Authority (Medietilsynet).

Olsson, E. (2012). *"Everything I read on the Internet is in English": On the impact of extramural English on Swedish 16-year-old pupils' writing proficiency*. Licentiate, University of Gothenburg, Department of Swedish.

Olsson, E., & Sylvén, L. (2015). Extramural English and academic vocabulary: A longitudinal study of CLIL and non-CLIL students in Sweden. *Journal of Applied Language Studies, 9*(2), 77–103.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *International Journal of Applied Linguistics, 169*(1), 142–168.

Pickard, N. (1996). Out-of-class language learning strategies. *ELT Journal, 50*(2), 150–159.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912.

Puimège, E., & Peters, E. (2019). Learners' English vocabulary knowledge prior to formal instruction: The role of learner-related and word-related variables. *Language Learning, 69*(4), 943–977.

Sandlin, J. A., Schultz, B. D., & Burdick, J. (2010). Understanding, mapping, and exploring the terrain of public pedagogy. In J. A. Sandlin, B. D. Schultz, & J. Burdick (Eds.), *Handbook of public pedagogy: Education and learning beyond schooling*. Routledge.

Sauro, S., & Zourou, K. (2019). What are the digital wilds? *Language Learning & Technology, 23*(1), 1–7.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching, 52*(2), 261–274.

Schunk, D. H. (1994). Self-regulation of self-efficacy and attributions in academic settings. In D. H. Schunk, & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational implications* (pp. 75–100). Erlbaum.

Schurz, A., & Sundqvist, P. (2022). Connecting extramural english with ELT: Teacher reports from Austria, Finland, France, and Sweden. *Applied Linguistics, 43*(5), 934–957.

Schurz, A., Coumel, M., & Hüttner, J. (2022). Accuracy and fluency teaching and the role of extramural English: A tale of three countries. *Languages, 7*(1), 35.

Skjelde, K. M., & Coxhead, A. (2020). Mind the gap: Academic vocabulary knowledge as a predictor of English grades. *Acta Didactica Norden, 14*(3), 1–20.

Solhaug, T., & Osler, A. (2018). Intercultural empathy among Norwegian students: An inclusive citizenship perspective. *International Journal of Inclusive Education, 22*(1), 89–110.

Solheim, O. J. (2011). The impact of reading self-efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology, 32*(1), 1–27.

Sun, T., & Wang, C. (2020). College students' writing self-efficacy and writing self-regulated learning strategies in learning English as a foreign language. *System, 90*, Article 102221.

Sundqvist, P. (2009). *Extramural English matters: Out-of-school English and its impact on Swedish ninth graders' oral proficiency and vocabulary*. Karlstad University Press,.

Sundqvist, P. (2011). A possible path to success out-of-school English language learners in Sweden. In P. Benson, & H. Reinders (Eds.), *Beyond the Language Classroom* (pp. 106–118). Palgrave Macmillan.

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology, 23*(1), 87–113.

Sundqvist, P., & Wikström, P. (2015). Out-of-school digital gameplay and in-school L2 English vocabulary outcomes. *System, 51*, 65–76.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition, 15*, 165–179.

Swain, M. (2001). Integrating language and content teaching through collaborative tasks. *Canadian Modern Language Review, 58*(1), 44–63.

Sylven, L. K. (2006). How is extramural exposure to English among Swedish school students used in the CLIL classroom? *Vienna English Working, Papers, 15*(3), 47–53.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making, 11*(1), 99–113.

UDIR. (2022). *Relevance and central values*. (ENG01–04). Norwegian Directorate for Education and Training. ⟨https://www.udir.no/lk20/eng01–04/om-faget/fagets-relevans-og-verdier?lang=eng⟩.

UDIR. (2023). *Karakterstatistikk for videregående skole*. Norwegian Directorate for Education and Training. ⟨https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/karakterer-vgs⟩.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174–1204.

Warnby, M. (2022). Receptive academic vocabulary knowledge and extramural English involvement–is there a correlation? *International Journal of Applied Linguistics, 173* (1), 120–152.

Webb, S. (2015). Extensive viewing: Language learning through watching television. In D. Nunan, & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 159–168). Routledge.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated vocabulary levels test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics, 168*(1), 33–69.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71*, 3–25.