The impact of Virtual Exchange on TPACK and foreign language acquisition: reviewing a large-scale implementation across 25 Virtual Exchanges

Bart Rienties, Tim Lewis, Robert O'Dowd, Irina Rets & Jekaterina Rogaten

Abstract

Several studies on Virtual Exchange (VE) have highlighted positive learning experiences, increases in technological and pedagogical skills (TPACK) and foreign language (FL) competence. However, most VE research to date use qualitative or descriptive case-studies of how VEs have been implemented, and what "might" have worked. In this large-scale quantitative two-study design, we explored how 622 pre-service teachers developed TPACK skills and (perceived) FL competence over time in 23 VEs across 34 institutions in 16 countries. In Study 1, we used a (quasi-) experimental design of 3 VEs in an experimental (n = 151) or control group (n = 77) to explore the impact on TPACK. In Study 2, we used a larger sample of 20 VEs and 394 participants to replicate and contrast the findings from Study 1 in a broader context. In contrast to our expectations, participants in the experimental condition did not have higher TPACK skills growth relative to the control condition in Study 1, which was further confirmed in Study 2. Nonetheless, in Study 2 pre-existing TPACK skills influenced the development of (perceived) FL competence over time, whereby those participants who further strengthened their TPACK skills during the VE were more likely to nurture FL competence. A major lesson from this large-scale implementation is that VEs do not generate TPACK skills and FL competence by osmosis. We encourage CALL researchers to carefully reflect on any positive or negative finding that something has "worked" when there is no comparison or control group included.

Keywords: foreign language competence; pre-service teachers; TPACK; virtual exchange

Introduction

In today's turbulent world with increased trade tensions, political instability, and the rise of populism, there is an increased call for people from different cultures to reconnect (Quinton, 2018). One of the most successful programmes to raise intercultural awareness and Foreign Language competence (FL) amongst students, by means of physical mobility, is the EU Erasmus programme (Cairns, 2017; Hernandez Nanclares, 2016). Since 1987 the Erasmus programme has brought together over 3.3 million students from 34 European

countries to study for 3-12 months in an Erasmus Exchange. A range of reasons have been suggested for students not going on Erasmus, namely financial, socio-economic, and time constraints (Cairns, 2017; O'Dowd, 2013; V3an Dorp, 2008), as well as disability (Seale, Georgeson, Mamas, & Swain, 2015), caring responsibilities, and increased requirements for students to sustain their study by working part/full-time (Darolia, 2014; Kudo, Volet, & Whitsed, 2017). One option that has been suggested to allow any student to experience some of the affordances of Erasmus exchange and to learn from different cultures is *Virtual Exchange* (VE).

Following O'Dowd and Lewis (2016) we define VE as a process of engaging students in online intercultural collaboration projects with partner classes within their programmes of study and under the guidance of teachers or trained facilitators. According to Baroni et al. (2019, pp. 8-9), VE "is based on student-centred, international, and collaborative approaches to learning where knowledge and understanding are constructed through interaction and negotiation with students from other cultures". While similar initiatives have been termed in different ways, such as telecollaboration (Dooly & Sadler, 2013; Luo & Yang, 2018), online cross-cultural exchanges (Lee, 2009; O'Dowd & Lewis, 2016), or Erasmus + Virtual Exchange, we opted to select the term VE given its currency among European policy makers.

Research has found that most VE participants have a positive learning experience (Antoniadou, 2011; Dooly & Sadler, 2013; Kohn & Hoffstaedter, 2017; O'Dowd & Lewis, 2016). Furthermore, VE has shown the potential to enhance not only the participants' FL competence (Tian & Wang, 2010) but also their technological pedagogical and content (TPACK) skills (Dooly & Sadler, 2013), intercultural competencies (O'Dowd & Lewis, 2016), and motivate them to build learning communities (Lee, 2009).

While research on VE has seen a steady growth in the past twenty years, a range of studies have identified several gaps in VE research (e.g., Baroni et al., 2019; Dooly, 2017).

One main gap concerns the scope of methods used to analyse VE data and their validity. Research on VE is mostly comprised of qualitative and/or descriptive case-studies of how VEs have been implemented, and what "might" have worked (Luo & Yang, 2018; O'Dowd & Lewis, 2016). Thus, as indicated by Baroni et al. (2019) there are neither evidence-based well designed studies nor large-scale studies that have tested whether VEs indeed increase FL competencies, and TPACK skills in particular, relative to a comparison or control group. This is in line with a wider critique on Computer Assisted Language Learning (CALL), whereby Golonka, Bowles, Frank, Richardson, and Freynik (2014, p. 71) reviewed 350 CALL studies and found that most of the reviewed studies were primarily descriptive case-studies, with relatively small sample sizes, and often had "poor choices of variables to be investigated … [and] lack of relevant data about participants".

Therefore, in this large-scale quantitative study we aim to explore how 622 students (i.e., pre-service teachers) developed TPACK skills and (perceived) FL competence over time in a large-scale implementation of 23 VEs across 34 institutions in 16 countries. In particular, we are interested to explore whether an increase in TPACK development due to participation in VEs influenced students' FL competence. Firstly, we will briefly review the literature on VE. Secondly, we will discuss how VEs might (not) increase TPACK skills of pre-service teachers, as well as FL competence. Thirdly, we will discuss a large (quasi-)experimental Study 1 of 228 participants who interacted in three VEs. Finally, we will discuss a replication Study 2 of 394 students amongst 20 VEs, and explore whether (or not) these interactions in those VEs led to increased TPACK skills and FL competence.

Virtual exchange, TPACK skills, and FLA

In many educational settings, students and pre-service teachers in particular develop their target FL and techno-pedagogical competencies through traditional classroom

instruction. Thus, many pre-service teachers who do not go on Erasmus do not have extensive direct contact with conversation partners abroad with whom they would have to communicate continuously in a foreign language (Dooly & Sadler, 2013; Luo & Yang, 2018; (O'Dowd & Lewis, 2016). CALL technologies make direct communication between FL students who share the same lingua franca more possible than ever before, while providing them with opportunities to reflect on the affordances of such technologies and their usefulness in teaching (Antoniadou, 2011; Kohn & Hoffstaedter, 2017; Luo & Yang, 2018).

Therefore, a growing number of scholars advocate VEs as an ideal environment for FL competence and the development of techno-pedagogical skills (TPACK) within formal institutional contexts (Baroni et al., 2019; Bueno-Alastuey, Villarreal, & García Esteban, 2018). Within such exchanges FL students and/or pre-service teachers from different geographical parts of the world can work together on authentic projects supported by teachers (Dooly & Sadler, 2013; O'Dowd & Lewis, 2016). For a recent elaborative review of 20 years of VE, we refer to Luo and Yang (2018) in this journal.

Despite the affordances that VE offers, it also poses a number of challenges on VE participants. For example, O'Dowd (2013) classified such challenges into four levels – individual, the classroom, the socio-institutional, and the interaction for failed or missed communication. On the socio-institutional level, O'Dowd (2013) specifically stressed that whether VEs are integrated into course syllabi, and whether VEs become a credit-carrying activity for students may influence how much value VE participants assign to it, and how much effort they put in the successful completion of the tasks. O'Dowd (2013) indicated that institutional recognition of VE and its assessment was the second-most cited challenge in VE. Notwithstanding the challenges that students and pre-service teachers in particular may encounter in VE, a considerable body of VE research has demonstrated the value of VE in

developing their participants' TPACK skills as well as enhancing their FL competence, which will be discussed in the next sections.

TPACK skills and Virtual Exchange

The importance of developing teachers' TPACK skills to meet the demands of the 21st century has been widely recognised (Kirschner, Wubbels, & Brekelmans, 2008; O'Dowd, 2013; Tondeur et al., 2012). It is evident that teachers' knowledge of subject matter does not automatically transfer to knowledge for integrating appropriate technologies (Niess, 2011; Rienties, Brouwer, & Lygo-Baker, 2013). Therefore, providing pre-service teachers with the necessary competencies to teach with technology has become one of the primary benchmarks for pre-service teaching programs (Kirschner et al., 2008; Seidel, Blomberg, & Renkl, 2013). The most widely used and validated framework (Graham, 2011; Rienties et al., 2013; Schmidt et al., 2009; Voogt, Fisser, Pareja Roblin, Tondeur, & van Braak, 2013) that identified the competencies (pre-service) teachers need to acquire in relation to the use of technology in teaching is the Technological Pedagogical Content Knowledge (TPACK) model developed by Mishra and Koehler (2006).

The TPACK model consists of three sources of knowledge, namely, pedagogical knowledge (PK), content knowledge (CK), and technological knowledge (TK). The knowledge that lies at the intersection of these three different competencies accounts for the understanding of effective technology integration into teaching. A review of 16 TPACK studies by Voogt et al. (2013) found that active involvement of pre-service teachers in technology-enhanced lessons, followed by modelling how to teach in a technology-rich environment, contributed to developing (self-reported) pre-service teachers' TPACK.

There are several studies that have looked into pre-service teachers' TPACK development as part of VE (e.g., Antoniadou, 2011; Bueno-Alastuey et al., 2018; Dooly &

Sadler, 2013; Hauck, Müller-Hartmann, Rienties, & Rogaten, 2020). Using the methods of thematic, conversational, and discourse analyses of student talk during the VE and student reflections, these studies have generally showed a positive contribution of VE to pre-service teachers' TPACK development. To exemplify, the seven pre-service teacher participants of a Spanish-American VE stated that VE helped them get acquainted with new technological tools, as well as helped them reflect on the affordances and constraints of these tools when designing a teaching unit as part of the VE (Antoniadou, 2011). At the same time, these seven participants also stated they did not necessarily develop a profound pedagogical understanding of the different technologies due to having insufficient prior experience of using them.

In a later study of Dooly and Sadler (2013) with an undisclosed number of participants of another Spanish-American VE, they stated that the opportunities of first-hand experiential use of technology provided them with enhanced self-efficacy in using technology in their own teaching. However, a rather different finding came from the study of Bueno-Alastuey et al. (2018). In their VE study between two Spanish universities they counted the number of episodes of different TPACK model domains that appeared in 55 participants' talk. Although, the high number of episodes found in the study indicated the potential that VE has in allowing pre-service teachers to focus on their TPACK, the study also showed that the participants predominantly focused on pedagogical content knowledge (PCK). Bueno-Alastuey et al. (2018) concluded that the pre-service teachers at the end of the VE still felt more knowledgeable in domains related to pedagogy, content or their intersection rather than technology.

In a recent mixed-method study on two VEs with 92 pre-service teachers Rets, Rienties, Lewis, 2020 Submitted), relative TPACK growth was calculated, and three clusters of VE participants with high, medium and low TPACK gain were identified. The thematic

analyses of the participants' answers at four different stages of the VE across the three clusters showed that cluster 1 with high TPACK gain were the least negative about the VE, and cluster 3 with low TPACK gain were the most negative. Moreover, what differentiated the participants in cluster 1 from the rest of the clusters was the frequent mentioning of the challenges or prior anxieties they were able to overcome through collaboration during the VE (Rets et al., 2020 Submitted).

TPACK, FL competence and VE

A second gap in the VE literature is whether it actually increases FL competence. For learners lacking opportunities for physical mobility, VE may be the only means of practising a foreign language (Darhower, 2008; Tian & Wang, 2010). VE is also intended to give pre-service teachers the skills needed to organise future exchanges for their students. FL competence therefore constitutes part of Pedagogic Knowledge (PK) that participants need to master, if they are to become successful organisers of VE. Finally, in Study 1 and Study 2 pre-service teachers studying foreign language education were the largest disciplinary group of participants. For them, FL competence was also part of Content Knowledge (CK). Accordingly, it made sense to gauge the extent to which participants felt they had made gains in FL competence through taking part in VE.

The concept of FL competence has a long and contested history (Bachman, 1990; Macintyre, Clement, Dornyel, & Noels, 1998l Young, 2008). For example, Chomsky (1965, p. 4) famously denied competence as the underlying knowledge of language possessed by a person, which allows that individual to understand and use a language, and drew "a fundamental distinction between *competence* (the speaker-hearer's knowledge of his language) and *performance* (the actual use of language in concrete situations)". Rebutting Chomskyan abstraction, Hymes (1972) viewed communicative competence as including

judgements about appropriateness and as involving features of actual, contextualised language use. Canale and Swain (1980, p. 28), further developed the concept of communicative competence as comprising "grammatical competence, sociolinguistic competence and strategic competence", as further expanded by Macintyre et al. (1998). Bachman (1990) proposed a theory of language competence which added a pragmatic dimension, in the shape of 'illocutionary competence', to the grammatical and sociolinguistic competencies identified by Canale and Swain (1980). Put simply, Bachman (1990) recognised that communication is a two-way process in which meaning is jointly constructed and where competence includes an awareness of the impact on one's utterances on an interlocutor.

Typically in many VEs, FL participants are predominantly rooted in a communicative paradigm (Dooly & Sadler, 2013; Kohn & Hofstaedter, 2017). Participants are asked to report on core linguistic competences such as grammar, pronunciation and vocabulary (Golonka et al., 2014; Kudo et al., 2017). In addition, their perceptions are sought of improvements in accuracy and fluency (Kudo et al., 2017; Luo & Yang, 2018; O'Dowd, 2013), the correlation between which has long been a staple of debate within the communicative approach (Brumfit, 1984). Therefore, Helm and Guth (2010) proposed a FL proficiency framework for the goals of VE in such domains as reading, listening, writing, spoken interaction and production, which align with the international standards that describe language ability such as CEFR (The Common European Framework of Reference for Languages). Indeed, several studies have explored the development of FL competences in VE (Darhower, 2008; Lee & Markey, 2014; Tian & Wang, 2010).

For example, Darhower (2008, p. 51) studied 9 FL Spanish and English learners engaged in a text chat exchange, using WebCT, and explored "learners' perceptions regarding linguistic affordances in their chat room discourse". While Darhower (2008)

viewed the affordances of text chat primarily as raising 'language awareness', rather than leading to second language development, he nonetheless reported a number of perceptions of second language development. Firstly, roughly half of his FL participants indicated that the chat experience helped them to be able to speak, write and read better. Second, most of participants believed that their ability to write longer sentences improved. Finally, Darhower (2008, p. 63) indicated that most participants also "felt more confident in communicating in their L2 and that their attitude to their L2 improved as a result of the chat experience".

In a study of a *Skype*-based eTandem exchange between 30 FL learners of Mandarin and English Tian and Wang (2010) adapted Darhower (2008) to identify the perceived linguistic benefits of participating in the exchange. Their two groups of learners reported rather different experiences, whereby FL learners of English identified fluency, pronunciation, intonation and spontaneous reflex replies as areas in which they felt they had made the highest gains. In contrast, FL learners of Mandarin identified listening skills, new words and expressions, and pronunciation and tones as their areas of highest gain. Neither group reported significant gains in grammatical competence. FL learners of English were much more positive than their Mandarin learning counterparts in reporting gains in confidence.

Lee and Markey (2014) built on these two studies in using an online questionnaire to collect 28 learners' views on their experience of a Spanish-English VE, using various Web 2.0 tools, including *Skype* and *Facebook*. Lee and Markey (2014, p. 294) conclude that "peer feedback as part of intercultural exchange helped learners increase lexical knowledge, prevent language fossilization and acquire native-sounding discourse". Similarly, Kohn and Hoffstaedter (2017) analysed non-native English and German lingua franca conversations in VEs between 18 learners in high schools in France, Germany, the Netherlands and Spain, and

found evidence that VE increased agency in connection with communicative participation as well as increased their sense of speaker satisfaction and speaker confidence.

Although each of these studies used different tools (and different modes), learners in all four studies reported a growth in confidence. In three of the four studies, learners perceived an increase in lexical, rather than grammatical competence. While the results reported in these publications broadly justify our focus on a particular set of competences, a serious concern is the relatively small sample size in all four studies.

Research questions

As highlighted in the literature review, most VE research on pre-service teachers' TPACK skills is still emergent, and primarily use descriptive or qualitative research methods with relatively small samples. In particular, to the best of our knowledge there is no study that has used an evidence-based approach, whereby a control or comparison group is included to compare whether (or not) VE increases TPACK skills. In order to explore whether VE can enhance TPACK skills, in this study we designed a two-study replication approach. In Study 1, we used a (quasi-)experimental design whereby in six classes in Hungary, Spain, and Sweden participants were assigned into either an experimental (i.e., a VE and continuation of normal classes) or control condition (i.e., no VE, continuation of normal classes) by their respective teacher. We expected that by allowing participants in the experimental condition from two distinct educational and cultural systems to work together in a VE for a period of 65+ days on three authentic tasks (see task design section) participants would enhance their TPACK skills relative to the control group. Therefore, our first research question is framed as follows:

RQ1 To what extent does a virtual exchange (VE) lead to an increase in TPACK skills of participants in the experimental condition relative to the control condition?

As indicated by several researchers (Baroni et al., 2019; Golonka et al., 2014; Torgerson & Torgerson, 2008), social science and language studies in particular often lack robust experimental and replication studies that test whether quantitative findings are replicated in wider, and perhaps different contexts and settings. Even if we find positive or negative effects in Study 1, it would be important to further validate these findings with a larger replication study to ensure sufficient generalisation of the findings. Therefore, in Study 2 we used a larger sample of 20 VEs to explore the impact of these VEs on TPACK. In particular, in line with recommendations by Bueno-Alastuey and Kleban (2016), we were keen to explore whether any of these TPACK increases were similar across the 20 VEs, or whether substantial differences were found.

RQ2 To what extent does VE lead to an increase in TPACK skills, and are these effects similar across 20 VEs?

Our third and final research question (RQ3) aimed to explore whether (or not) there was a positive effect of these VEs in terms of FL competence, and whether these were mediated by TPACK skills (as described above). Several VE studies (e.g., Antoniadou, 2011; Dooly & Sadler, 2013) indicated that participants were able to strengthen their TPACK skills due to participating in such VEs. Therefore, it would be reasonable to assume that TPACK skills and growth in these skills could influence the development of FL competence, as (self-) reported at the end of the VE.

RQ3 To what extent does VE lead to an increase in (perceived) FL competence, and how is this mediated by (growth in) TPACK skills.

Method

Setting and participants

This study took place in a EU-funded project called Evaluate (http://www.evaluateproject.eu/). In total 23 VEs were run, whereby 15 VEs ran in phase 1 in the winter semester 2017-18, while eight VEs took place in phase 2 in the spring semester 2018. In total, 34 institutions of initial teacher education from 16 countries were involved. Most institutions were from European countries, but the project also welcomed participation from teacher educators from Brazil, Canada, Macau, Israel, Turkey, and the United States (Baroni et al., 2019).

Altogether, 1,018 students were invited to take part (Baroni et al., 2019), whom were studying either undergraduate or Masters' degrees in initial teacher education, mostly in primary education, followed by secondary education and higher education. The largest group of participants (42%) were studying FL education. The average age of participants was 21.54 (SD = 3.31, range 16-44), and 64% of the participants were women. Classes were matched depending on different variables including subject being studied, language choice, their courses' start and end dates, and the themes their teachers wished to focus on. The majority of classes involved students who were studying English as a foreign language O'Dowd, 2017; O'Dowd & Lewis, 2016).

The treatment in Study 1 and Study 2 involved engaging these classes of initial teacher education in a period of intensive VE with partner classes in institutions in other countries based on specifically-designed tasks related to digital-pedagogical competences as well as intercultural competence (Baroni et al., 2019). The VE model was based on the

Progressive Exchange Model which has been widely used in VE research and practice to date (O'Dowd, 2013, 2017). As mentioned in O'Dowd (2017), the model involves "three interrelated tasks which move from information exchange to comparing and analysing cultural practices and finally to working on a collaborative product". Teacher trainers participating in both Studies were provided with three sets of tasks so they could choose from various tasks at each stage of the task sequence.

For a detailed discussion of the specific task designs and the innovative tools used for these VEs (e.g., Skype, Google Docs, WhatsApp), we refer to publications from the Evaluate project (Baroni et al., 2019). For example, Baroni et al. (2019) found that 64% of participants in this Evaluate project were positive about the VEs, and 75% of participants would recommend these VEs to other pre-service teachers. In this quantitative follow-up study, we specifically focussed on TPACK and (perceived) FL competence. Other studies have reported on intercultural skills development (Hauck et al., 2020) as well as more fine-grained qualitative analyses of specific VEs (Rets et al., 2020 Submitted).

Instruments

TPACK

TPACK was measured using an adjusted version of Schmidt et al. (2009). Given that participants had to fill in the TPACK twice as well as contributing to four diaries (Baroni et al., 2019; O'Dowd, 2017), the project team shortened the questionnaire to 17 items. The TPACK instrument consisted of one overall TPACK construct, and four main sub-constructs, whereby a Likert response scale of 1 (=totally disagree) to 5 (=totally agree) was used. For each construct, we illustrate the number of items per construct in brackets, the respective Cronbach alpha, and one exemplary item: (1) technology knowledge (TK: seven items, α_{pre} = .86, α_{post} = .87 "I can learn technology easily"); (2) technological pedagogical knowledge

(TPK: four items, α_{pre} = .73, α_{post} = .77, "I can choose technologies that enhance the teaching approaches for a lesson"); (3) technological pedagogical content knowledge (TPCK: five items, α_{pre} = .84, α_{post} = .85, "I can select technologies to use in my classroom that enhance what I teach, how I teach, and what students learn"); and (4) technological content knowledge (TCK: one item, "I can choose technologies that enhance the content for a lesson"). Pre- and post-test Cronbach alphas indicated reliable constructs, and subsequent explorative factor analyses indicated a reasonable fit in line with these four constructs.

Foreign language competence

(Perceived) FL competence was measured based on several decades of theoretical reflection on FL acquisition, development, and competence (e.g., Bachman, 1990; Canale & Swain, 1980; Darhower, 2008; Macintyre et al., 1998; Young, 2008), taking into consideration the specific context of VE (Baroni et al., 2019; O'Dowd, 2017). The overall question asked was: 'How (if at all) has your ability to use a foreign language developed in the course of the exchange?' Sub-questions were framed in plain language, to maximise understanding. A four-point Likert scale was used, to enable respondents to indicate whether each sub-competence had (1) improved much, (2) improved a little, (3) not improved at all, or (4) got worse. Seven questions covered language knowledge (grammatical accuracy, accuracy of pronunciation and range of vocabulary), skills (ability to understand, fluency in speaking), and pragmatic competence (ability to interact). A further question asked participants to indicate if they felt that their *confidence* in using FL had improved. For comparability with TPACK we reverse-coded the FL construct. Although this scale was a newly developed instrument, both screening of the items as well as explorative factor analysis indicated a reasonable fit. The Cronbach alpha indicated good to very good reliability (α = .892).

Data analysis

Study 1

In terms of RQ1, initially 228 participants were included in Study 1, whereby 151 participants were included in the experimental condition with English as a lingua franca (Spain-Sweden, Spain-Hungary, Spain-Germany). 77 participants were included in the control condition, who did not participate in the VE and continued like the experimental group to follow normal classes. Both the experimental and the control group followed classes in courses like *English language and didactica 2*, whereby both groups of participants were working both on FLA and TPACK concepts and learning materials in their respective course. 127 participants (84%) in the experimental condition also completed the post-test, while 63 participants (81%) in the control condition completed the post-test, which are considered good response rates. As participants were divided by their respective teachers in the respective condition, as a research team we had no control in terms of the randomization process of participants.

In terms of a potential imbalance in the two conditions, we checked whether participants had similar characteristics. Using ANOVA, no significant differences were found between the two conditions in terms of pre-TPACK scores, gender, specialization, or responding to pre- and post-tests. However, using ANOVA, participants in the control group were on average two years older than students in the experimental group ($F = 16.404$, $p < .01$), and using Chi-Square analysis relatively more participants were training for primary education than for secondary education ($\chi2 = 14.840$, $p < .01$). As the control group was not asked to complete the FL competence questionnaire, it was not possible to compare whether the experimental group perceived more gains in language skills over time relative to the control group. In follow-up analyses reported below, we used ANOVA analyses to compare

pre-, post-test, and average learning gain (i.e., post-test minus pre-test) of TPACK between the experimental and control group.

Study 2

In terms of RQ2, we selected all participants who completed the pre- and post-test of TPACK and who were not involved in the experimental-control design in Study 1. In total 790 participants were initially included in this dataset, whereby 394 (50%) participants completed the pre-test, and 442 (56%) completed the post test. For 394 participants we obtained both pre- and post-test results. In order to test whether we could aggregate participants in phase 1 who were not included in the experimental-control VE in Study 1 with participants from phase 2 of the data collection, using ANOVAs we compared the pre- and post-test TPACK and FL competence scores. Although in phase 2 the pre-test TK, TPK, TPCK and TPACK scores were significantly higher relative to phase 1, neither in the post-test TPACK and FL competence scores nor in the growth in TPACK scores did we find any significant differences, indicating that the two phases were comparable in TPACK development over time. Therefore, we were able to aggregate the 218 participants from phase 1 with 176 participants from phase 2, leading to 394 participants in total.

In terms of RQ3, we conducted correlation and subsequent regression analyses to determine which factors might influence FL competence. As previous studies (Erdogdu & Erdogdu, 2015; Harrison, 2012; Mittelmeier, Rienties, Rogaten, Gunter, & Raghuram, 2019) have found that demographics (e.g., age, gender) and educational programmes (e.g., specialisation, discipline) could influence students' behaviour and outcomes, we controlled for these factors in the regression models. Ethics approval for these two studies was provided by Human Research Ethics Community by the Open University, HREC-2655/Rogaten.

Results

TPACK skills

Study 1

In terms of RQ1, we first compared participants in experimental and control condition in three VEs with respect to their TPACK development over time in Study 1 using ANOVAs. As indicated by Figure 1, the control group (n=77) had a slightly higher score at the total TPACK pre-test, while at the post-test the experimental condition (n=127) had a slightly higher score. The average learning gain of TPACK for the experimental group was 0.29 (SD=0.53), while the average learning gain for the control condition was 0.18 (SD=0.51), although this effect was not statistically significant. In other words, over the (on average) 65 days of the VE, both the experimental and control group students increased their TPACK total scores as they were learning in their context, but there was a slightly higher effect for the experimental condition. However, in contrast to our expectations of RQ1 this increase in TPACK was not statistically significant, neither in terms of post-test TPACK ($F = 0.227$, $p > .05$) and learning gain of TPACK ($F = 1.890$, $p > .05$). Separate analyses across the three contexts indicated no differences between the experimental and control conditions (not illustrated).

➔ *Insert Figure 1 about here*

Subsequent analyses of the four TPACK sub-constructs did highlight significant differences between the treatment and control group in terms of TPK ($F = 4.884$, $p < .05$), with a small effect size. The other constructs were not significantly different from the control group. As visually illustrated in Figure 2, separate analysis per VE indicated that in two out of three VEs participants in the experimental condition acquired more TPACK skills,

although not at a statistically significant level. In one VE participants in the experimental condition had significantly lower pre-test TPACK scores ($F = 18.102$, $p < .01$), but there was no difference at the end of the VE relative to the control condition. In other words, in terms of TPACK there appeared to be no significant differences in development over time from a quantitative perspective when compared to the control group.

➔ *Insert Figure 2 about here*

One potential reason for the lack of superior learning gains of the experimental group relative to the control group was that in the control group participants also continued to learn via their normal face-to-face classes, and might (in)formally further developed their TPACK skills. Another potential reason could be related to educational contexts in which these three studies were conducted, as in all three VEs Spanish students were paired with non-native English speakers. Perhaps other VEs where participants had higher/lower TPACK skills and (perceived) FL competence the effects of VEs might become more pronounced, which we will explore next.

Study 2

In order to compare, validate, or perhaps contradict the initial findings from Study 1 across a wider set of VEs, in Study 2 we compared the TPACK developments over time amongst 394 students in 20 VEs. As indicated in Figure 3, at the pre-test 70% of participants were positive about their TPACK skills, taking a cut-off of 3.4. At the post-test 82%, of participants were positive about their TPACK skills. In terms of a paired-sample T-test, there was a significant increase in TPACK scores over time ($t = 11.844$, $p < .01$), with a medium effect size (Cohen $d = .59$). However, as highlighted in Figure 3, the control group from Study 1 used as a benchmark also made gains ($t = 2.785$, $p < .01$), with a medium effect size

(Cohen d = .62). In other words, we find some empirical support of RQ2 that VEs can help to increase TPACK skills, but this does not seem to lead to substantially larger gains over time relative to the control group.

➔ Insert Figure 3 about here

Afterwards, we compared pre- and post-test TPACK scores across the 20 VEs. As illustrated in Figure 4, there seemed to be substantial differences in both pre- and post-test TPACK scores across these VEs. Indeed, using ANOVAs we found significant differences in pre-test TPACK scores ($F = 2.508$, $p < .01$, $\eta^2 = .119$) and post-test TPACK scores ($F = 1.869$, $p < .05$, $\eta^2 = .091$), both with a medium effect size. However, in terms of learning gains of TPACK we did not find a significant difference between the 20 VEs ($F = 1.522$, $p > .05$, $\eta^2 = .076$), with the notable exception of TPCK ($F = 1.710$, $p < .05$, $\eta^2 = .084$). In other words, this seemed to suggest that for most participants there was some positive TPACK growth over time, irrespective of the specific VE they participated in. For a detailed breakdown of the pre-post TPACK scores per VE, we refer to Appendix Table 1.

➔ Insert Figure 4

While 19 VEs had a positive increase in TPACK, there were two VEs that had a negative TPACK trend (Macao – Portugal, Spain – Poland). Furthermore, as highlighted in Figure 4, there were several VEs (Germany – Poland, Spain – US) that had a significantly higher growth curve over time than the average TPACK line (dashed black line), while at the same time there were VEs that had a substantially lower trend line. Future qualitative research should unpack why these VEs might have been more or less successful in raising TPACK skills amongst participants.

TPACK and FL competence

In order to unpack RQ3, we first explored the overall FL competence construct and how participants reflected upon their (perceived) FL competence after completing their respective VE. Second, using correlation analyses we explored whether (pre & post-test) TPACK of participants was related to FLA. Thirdly, we conducted regression analyses to estimate how FLA was predicted by age, respective specialisation, and (pre & post-test) TPACK. Finally, we explored whether these found relations were comparable across all VEs, or whether substantial differences in TPACK and FL competence were found across the VEs.

First, of the 394 participants who completed the FL competence test at the end of the VE, 18.3% of participants indicated that their FL competence did not change while participating in the VE. The remainder of these participants indicated that their FL competence increased, from mostly a little to around 10% who indicated that it improved substantially. As illustrated in Table 1 subsequent analyses per FL competence category indicated that most improvement was found in terms of 'Confidence in using the foreign language', 'Ability to interact with foreign language speakers'. This goes in line with the evidence from the study of Kohn and Hoffstaedter (2017) who showed increased learner agency when communicating in English and contributed to emancipation of their non-native speaker identity.

➔ Insert Table 1 about here

As a second step, follow-up correlations in Table 2 indicated a positive correlation between pre-test TPACK scores and FL competence (rho = .126, $p < .05$), indicating that participants who were more familiar with technology before the VE had a higher level of self-reported FL competence at the end of the VE, although the size of the increase was small size. This is particularly in line with previous studies (Herodotou, Rienties, Boroowa,

Zdrahal, & Hlosta, 2019; Schmidt et al., 2009; Teo, 2010) who found that participants who have stronger technological skills might benefit more from participating in online learning activities, such as VE.

Subsequently, there was a stronger but small in size correlation between post-test TPACK scores and FL competence (rho = .229, $p < .01$), indicating that participants who at the end of the VE were positive about their TPACK skills were also more likely to report more FL competence. Similarly, in terms of growth in TPACK we again found a positive correlation with FL competence, indicating a small but positive learning effect of VE on FL competence. In other words, there seemed to be some initial support that VE can enhance TPACK skills of pre-services teachers, while at the same time also enhancing their FL skills. Nevertheless, the effects were relatively small and minor.

➔ Insert Table 2

Follow-up regression analyses of FL competence indicated a significant effect for age in Table 3, indicating that younger learners seemed to benefit more from VE in terms of their FL competence. Neither the specialisation that students were following (i.e., early child, primary, secondary education, other) nor whether they were studying FL education or another discipline significantly predicted FL competence outcomes. In Model 1, pre-test TPACK significantly predicted FL, indicating that participants with stronger TPACK skills benefited more from VE. When adding post-test TPACK in Model 2 participants post-TPACK skills significantly predicted FL competence. Finally, in Model 3 both pre-test TPACK and growth in TPACK over time significantly predicted FL competence. In other words, while pre-existing TPACK skills influenced the development of FL competence over time, those participants who further strengthened their TPACK skills during the VE were more likely to report gains in FL competence.

➔ Insert Table 3

Finally, as highlighted in Figure 5, there were relatively strong differences across the 20 VEs in terms of post-test TPACK and FL competence scores. While highlighted earlier the post-test TPACK scores were not significantly different across the VEs, we did find a significant difference in terms of FL competence development ($F = 3.442$, $p < .01$, $\eta^2 = .156$) across the 20 VEs using ANOVAs, with a medium effect size. In other words, while most participants increased their TPACK over time irrespective of the VE they took part in, FL competence development seemed to be significantly influenced by which VE was followed. For example, taking the three VEs with lowest reported increases in (perceived) FL competence, one VE had a relatively strong TPACK score but limited increase in FL competence, which was a result of a VE between US and Swedish participants using English language. Similarly, two respective VEs between Spanish and Dutch participants led to relatively low increase in English language skills.

Obviously there would be a ceiling effect in terms of language improvement for US participants, as well as for Dutch and Swedish participants who typically already have a strong mastery of English language, both in writing as well as speaking. In contrast, the three VEs with the highest (perceived) FL competence increase were between Turkish and Spanish participants, Polish and German participants, and between Spanish and Polish participants. In each of these three VEs, English was used as lingua franca, but one reasonable assumption might be that mastery of written and spoken English is typically lower than Dutch or Swedish students. Obviously, this does not necessarily mean that these US, Dutch and Sweden participants did not learn from the VEs, but in terms of self-reported FL competence improvements they were relatively less positive.

➔ Insert Figure 5 about here

Discussion

Widely accessible CALL technologies have allowed teacher educators to bring into the classroom opportunities such as Virtual Exchange (VE) for connecting communities of pre-service teachers from different language backgrounds to work on collaborative educational tasks while communicating in the target foreign language. An increasing number of institutions are currently exploring whether VEs can help to develop TPACK skills and FL competence. There is some emerging, but mostly descriptive and anecdotal evidence from mainly explorative and qualitative studies that VEs can actively increase pre-service teachers' TPACK skills (e.g., Antoniadou, 2011; Bueno-Alastuey et al., 2018; Dooly & Sadler, 2013; Hauck et al., 2020) and their FL skills (Darhower, 2008; Kohn & Hoffstaedter, 2017; Luo & Yang, 2018; Tian & Wang, 2010).

In line with calls by Golonka et al. (2014) and Baroni et al. (2019) to provide more robust empirical and evidence-based research on the impact of CALL on how pre-service teachers are actually learning in large scale settings, we used both a robust experimental design amongst 228 students, and a large-scale replication design amongst 394 pre-service teachers to explore whether VEs are actually leading to measurable positive increases in TPACK and (perceived) FL competence. To the best of our knowledge, this is the first large-scale study that has explored the impact of 23 VEs on TPACK and FL competence development.

In terms of RQ1, in contrast to our expectations in Study 1 we did not find that participants in the experimental condition had higher TPACK skills growth than those in the control condition. In both control and experimental conditions, participants became more confident in their TPACK skills over time, and indicated to be more aware of how they could

potentially use technology in their (future) classroom. At first, this might be surprising as one would assume that providing authentic training in a VE would help participants to become comfortable with using technology in a classroom setting. These findings confirm earlier results by Antoniadou (2011), who showed that while the pre-service teachers who participated in a VE identified this experience as positive, they did not develop a profound pedagogical understanding of the technologies they used during the VE due to their insufficient prior experience of working with those tools.

In order to test whether different contexts might lead to different results, in Study 2 we broadened the contexts substantially (e.g., combining more near-native English speakers) both in terms of the type of VE (e.g., beyond FL students) as well as reach (i.e., outside Europe) by extending the research to 20 additional VEs. In terms of RQ2, although most of the 394 participants indicated positive TPACK growth over time, their growth was not significantly larger than those in the control group.

There could be three reasons for this "null-finding". First of all, in terms of learning design (O'Dowd, 2017; Rienties, Lewis, McFarlane, Nguyen, & Toetenel, 2018) and quality assurance (O'Dowd & Lewis, 2016) of VE, there were several caveats in relation to how and when participants were contributing. For example, the duration and intensity of the VE was rather limited (6-8 weeks), whereby beyond the three tasks and some (diary) reflections the amount and intensity of actual interaction between the participants was relatively limited. Furthermore, in particular as the outcomes of the VEs were not formally assessed, participation was voluntary, and no formal ECTS credits were provided for successful completion, the motivation to participate might be mostly intrinsic.

As evidenced by both quantitative and qualitative data (e.g., Baroni et al., 2019; Hauck et al., 2020; Rets et al., 2020 Submitted), several participants really benefited from their active participation, but perhaps the lack of formal recognition could have limited some

participants to actively contribute as well. This goes in line with the findings of O'Dowd (2007, 2013) who showed that formal integration of VE into course syllabi and assessment of student participation and their completion of the tasks as part of the VE is one of the biggest challenges for the successful implementation of VE.

Second, the validity of using self-reporting scales is strongly dependent on the ability of participants to accurately reflect upon their own skills and knowledge. In line with Kruger and Dunning (1999) effect, a reduction or limited change in students' self-assessment of TPACK may actually indicate a greater awareness of the complexity of technology use in the classroom once students have actually experienced it rather than the initial, perhaps 'idealised', conception of technology and their ability to effectively work in such a context.

Thirdly, as students were randomised by teachers in the classroom in Study 1 into an experimental and control condition, and were aware of who were involved and who not, there were obvious issues with contamination, and Hawthorne's and Henry's effects (Torgerson & Torgerson, 2008). Both the control and the experimental group continued to learn in their FL and TPACK classes, so it would seem logical that both groups would continue to develop these skills. Furthermore, as all the three experimental studies included Spanish students working together with other European non-native English speakers, perhaps the composition of these VEs might have hampered potential learning gains in TPACK.

In terms of RQ 3, we found significant relations between (pre- and post-test) TPACK scores and (perceived) FL competence. Most participants indicated that their FL skills improved during the VE. Follow-up correlations and regression modelling indicated that pre-test TPACK significantly predicted FL competence, indicating that participants with stronger TPACK skills benefited more from VE. In our final model, we found that both pre-test TPACK and growth in TPACK over time significantly predicted reported gains in FL competence. In other words, while pre-existing TPACK skills influenced FL competence

development over time, those participants who further strengthened their TPACK skills during the VE were more likely to experience growth in FL competence.

Although the explained variance in the regression models was less than 10%, given the strong diversity in terms of students' backgrounds and the inevitable local differences in implementation of VEs, it is noteworthy that participants who developed stronger TPACK skills over time also became more confident in expressing themselves in a foreign language. Obviously, as we did not measure actual FL skills and ability (e.g., TOEFL), we have to be careful in stating that VEs can lead to increased FL skills, as these were just self-report measurements. As participants also composed a diverse range of written texts (e.g., discussion forums, diaries, individual reflections), it would be useful in future research to conduct a corpus analysis to explore whether (or not) their actual writing abilities improved over time in the VEs.

One potential reason why we found only small effects of TPACK on FL competence might be related to the relatively novel nature of VE. As the concept of VEs is relatively new (Baroni et al., 2019; O'Dowd, 2017) and the Evaluate project relied on enthusiastic (but perhaps occasionally inexperienced) VE teachers, perhaps it is not surprising that we found substantial differences within and across VEs. Future research should explore how teachers designed their respective VEs, and in particular why some VEs seemed to generate more opportunities to develop FL competence. Ideally, in future applications of VE, rather than assigning participants in all or nothing interventions, it would be useful to conduct more fine-grained designs where designers play with the duration (e.g., shorter vs. longer) and intensity of the treatment (e.g., more intensive, marked for credits activities vs. voluntary).

Limitations and ways forward

Although these two studies are (to the best of our knowledge) the firs large-scale evidence-based applications of VE, using robust (quasi)experimental and longitudinal designs, there are some limitations. One obvious limitation of our research is that we did not include more micro level fine-grained analyses of actual student behaviour (Rienties et al., 2018), and what students actually said and did in their VE, such as those done by Hauck et al. (2020) and Rets et al. (2020 Submitted). A second limitation limitation is that we solely based our analyses on self-report data, with its known self-selection, non-response, and Dunning Kruger biases. Given the context specifications, the instrument that was devised to gauge the reported development of (perceived) FL competence was a rudimentary one. FL competence was not the primary focus of the project team and space in the overall research methodology was extremely limited. We would welcome any future attempts to develop more sophisticated measures and hope that our preliminary efforts might somehow pave the way for these. Nonetheless, given the substantial sample size and robust evidence-based design the fact that the control group did equally well in terms of TPACK over time might indicate that VEs do not automatically lead to superior TPACK skills.

A major lesson from this large-scale implementation of VE is that, even with the best intentions from teachers and students, learning in online settings do not generate TPACK skills and FL competence by osmosis. In particular, as many studies in CALL do not use robust evidence-based approaches, we caution readers to conclude that something has worked when students report a positive or negative "effect" at the end of a CALL "intervention". Without appropriate pre- and post-testing and an inclusion of a control or comparison group, it might be difficult to calculate the value-added of such intervention. Obviously beyond the numerical data, we encourage researchers to critically analyse more fine-grained behavioural, cognitive and qualitative data of VE implementations. Hopefully more robust evidence can

be gathered from VEs of technology skills and FL competence development using mixed method approaches.

Acknowledgement

First of all, we would like to acknowledge the excellent contributions and feedback from Reviewer 1-2, which substantially improved the quality of this manuscript. Second, we are extremely grateful for all teachers and students who volunteered to participate in the 23 virtual exchanges. Third, we are very grateful for all the members of the Evaluate team who have contributed tirelessly on these exchanges. Evaluating and Upscaling Telecollaborative Teacher Education (EVALUATE) (582934-EPP-1-2016-2-ES-EPPKA3-PI-POLICY) is funded by Erasmus+ Key Action 3 (EACEA No 34/2015): European policy experimentations in the fields of Education, Training, and Youth led by high-level public authorities. The views reflected in this presentation are the authors' alone and the commission cannot be held responsible for any use which may be made of the information contained therein.

References

Antoniadou, V. (2011). Using Activity Theory to understand the contradictions in an online transatlantic collaboration between student-teachers of English as a Foreign Language. *ReCALL, 23*(3), 233-251. doi:10.1017/S0958344011000164

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford university press.

Baroni, A., Dooly, M., Garcia, P. G., Guth, S., Hauck, M., Helm, F., … Rogaten, J. (2019). *Evaluating the impact of virtual exchange on initial teacher education: A European policy experiment*. Leon, Spain: Research-publishing.net.

Brumfit, C. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy* (Vol. 129). Cambridge: Cambridge University Press Cambridge.

Bueno-Alastuey, M., & Kleban, M. (2016). Matching linguistic and pedagogical objectives in a telecollaboration project: A case study. *Computer Assisted Language Learning, 29*(1), 148-166. doi:10.1080/09588221.2014.904360

Bueno-Alastuey, M., Villarreal, I., & Garcia Esteban, S. (2018). Can telecollaboration contribute to the TPACK development of pre-service teachers? *Technology, Pedagogy and Education, 27*(3), 367-380. doi:10.1080/1475939X.2018.1471000

Cairns, D. (2017). The Erasmus undergraduate exchange programme: A highly qualified success story? *Children's Geographies, 15*(6), 728-740. doi:10.1080/14733285.2017.1328485

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47. doi:10.1093/applin/1.1.1

Chomsky, N. (1965). *Aspects of the theory of syntax* (Vol. 11). Cambridge, MA: MIT press.

Darhower, M. (2008). The role of linguistic affordances in telecollaborative chat. *CALICO Journal, 26*(1), 48-69.

Darolia, R. (2014). Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review, 38*(Supplement C), 38-50. doi:10.1016/j.econedurev.2013.10.004

Dooly, M. (2017). Telecollaboration. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 169–183). Hoboken, NJ: Wiley Blackwell.

Dooly, M., & Sadler, R. (2013). Filling in the gaps: Linking theory and practice through telecollaboration in teacher education. *ReCALL, 25*(1), 4-29. doi:10.1017/S0958344012000237

Erdogdu, F., & Erdogdu, E. (2015). The impact of access to ICT, student background and school/home environment on academic success of students in Turkey: An international comparative analysis. *Computers & Education, 82,* 26-49. doi:10.1016/j.compedu.2014.10.023

European Commission. (2015). Erasmus: Facts, figures and trends. *The European Union support for student and staff exchanges and university cooperation in 2013-14* (pp. 1-40). Luxembourg.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70-105. doi:10.1080/09588221.2012.700315

Graham, C. R. (2011). Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education, 57*(3), 1953-1960. doi:10.1016/j.compedu.2011.04.010

Harrison, N. (2012). Investigating the impact of personality and early life experiences on intercultural interaction in internationalised universities. *International Journal of Intercultural Relations, 36*(2), 224-237. doi:10.1016/j.ijintrel.2011.03.007

Hauck, M., Muller-Hartmann, A., Rienties, B., & Rogaten, J. (2020). Approaches to researching digital-pedagogical competence development in VE-based Teacher Education. *Journal of Virtual Education*.

Helm, F., & Guth, S. (2010). The multifarious goals of telecollaboration 2.0: Theoretical and practical implications. *Telecollaboration, 2,* 69-106.

Hernandez Nanclares, N. (2016). The transition processes of Erasmus students: Motivation, social networks and academic performrance. In D. Jindal-Snape & B. Rienties (Eds.), *Multi-dimensional transitions of international students to higher education* (pp. 240-256). London: Routledge.

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., & Hlosta, M. (2019). A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. *Educational Technology Research and Development, 67*(5), 1273-1306. doi:10.1007/s11423-019-09685-0

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269-293). Harmondsworth: Penguin.

Kirschner, P., Wubbels, T., & Brekelmans, M. (2008). Benchmarks for teacher education programs in the pedagogical use of ICT. In J. Voogt & G. Knezek (Eds.) *International handbook of information technology in primary and secondary education* (pp. 435-447). Boston, MA: Springer US.

Kohn, K., & Hoffstaedter, P. (2017). Learner agency and non-native speaker identity in pedagogical lingua franca conversations: Insights from intercultural telecollaboration in foreign language education. *Computer Assisted Language Learning, 30*(5), 351-367. doi:10.1080/09588221.2017.1304966

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134. doi:10.1037//0022-3514.77.6.1121

Kudo, K., Volet, S., & Whitsed, C. (2017). Intercultural relationship development at university: A systematic literature review from an ecological and person-in-context

perspective. *Educational Research Review, 20,* 99-116. doi:10.1016/j.edurev.2017.01.001

Lee, L. (2009). Promoting intercultural exchanges with blogs and podcasting: A study of Spanish-American telecollaboration. *Computer Assisted Language Learning, 22*(5), 425-443. doi:10.1080/09588220903345184

Lee, L., & Markey, A. (2014). A study of learners' perceptions of online intercultural exchange through Web 2.0 technologies. *ReCALL, 26*(3), 281-297. doi:10.1017/S0958344014000111

Luo, H., & Yang, C. (2018). Twenty years of telecollaborative practice: Implications for teaching Chinese as a foreign language. *Computer Assisted Language Learning, 31*(5-6), 546-571. doi:10.1080/09588221.2017.1420083

Macintryre, P. D., Clement, R., Dornyel, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal, 82*(4), 545-562. doi:10.1111/j.1540-4781.1998.tb05543.x

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017-1054. doi:10.1111/j.1467-9620.2006.00684.x

Mittelmeier, J., Rienties, B., Rogaten, J., Gunter, A., & Raghuram, P. (2019). Internationalisation at a distance and at home: Academic and social adjustment in a South African distance learning context. *International Journal of Intercultural Relations, 72,* 1-12. doi:10.1016/j.ijintrel.2019.06.001

Niess, M. L. (2011). Investigating TPACK: Knowledge growth in teaching with technology. *Journal of Educational Computing Research, 44*(3), 299-317. doi:10.2190/EC.44.3.c

O'Dowd, R. (2007). Evaluating the outcomes of online intercultural exchange. *ELT Journal, 61*(2), 144-152. doi:10.1093/elt/ccm007

O'Dowd, R. (2013). Telecollaborative networks in university higher education: Overcoming barriers to integration. *The Internet and Higher Education, 18,* 47-53. doi:10.1016/j.iheduc.2013.02.001

O'Dowd, R. (2017). Exploring the impact of telecollaboration in initial teacher education: The EVALUATE project. *The Eurocall Review, 25*(2), 38. doi:10.4995/eurocall.2017.7636

O'Dowd, R., & Lewis, T. (2016). *Online intercultural exchange: Policy, pedagogy, practice.* New York: Routledge.

Quinton, W. J. (2018). Unwelcome on campus? Predictors of prejudice against international students. *Journal of Diversity in Higher Education, 12*(2), 156-169. doi:10.1037/dhe0000091

Rets, I., Rienties, B., & Lewis, T. (Submitted). Transforming pre-service teacher education through Virtual Exchange: A mixed-method analysis of perceived TPACK development.

Rienties, B., Brouwer, N., & Lygo-Baker, S. (2013). The effects of online professional development on higher education teachers' beliefs and intentions towards learning facilitation and technology. *Teaching and Teacher Education, 29,* 122-131. doi:10.1016/j.tate.2012.09.002

Rienties, B., Lewis, T., McFarlane, R., Nguyen, Q., & Toetenel, L. (2018). Analytics in online and offline language learning environments: The role of learning design to understand student online engagement. *Computer Assisted Language Learning, 31*(3), 273-293. doi:10.1080/09588221.2017.1401548

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological Pedagogical Content Knowledge (TPACK): The development and validation of an assessment instrument for preservice teachers. *Journal of Research on Technology in Education, 42*(2), 123-149. doi:10.1080/15391523.2009.10782544

Seale, J., Georgeson, J., Mamas, C., & Swain, J. (2015). Not the right kind of 'digital capital'? An examination of the complex relationship between disabled students, their technologies and higher education institutions. *Computers & Education, 82,* 118-128. doi:10.1016/j.compedu.2014.11.007

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education, 34,* 56-65. doi:10.1016/j.tate.2013.03.004

Teo, T. (2010). A path analysis of pre-service teachers' attitudes to computer use: Applying and extending the technology acceptance model in an educational context. *Interactive Learning Environments, 18*(1), 65-79. doi:10.1080/10494820802231327

Tian, J., & Wang, Y. (2010). Taking language learning outside the classroom: Learners' perspectives of eTandem learning via Skype. *Innovation in Language Learning and Teaching, 4*(3), 181-197. doi:10.1080/17501229.2010.513443

Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education, 59*(1), 134-144. doi:10.1016/j.compedu.2011.10.009

Torgerson, D. J., & Torgerson, C. (2008). *Designing randomised trials in health, education and the social sciences: An introduction*. London, UK: Palgrave Macmillan.

Van Dorp, C. A. (2008). A premier European platform for clearing e-internships. *British Journal of Educational Technology, 39*(1), 175-179. doi:10.1111/j.1467-8535.2007.00731.x

Voogt, J., Fisser, P., Pareja Roblin, N., Tondeur, J., & van Braak, J. (2013). Technological pedagogical content knowledge – a review of the literature. *Journal of Computer Assisted Learning, 29*(2), 109-121. doi:10.1111/j.1365-2729.2012.00487.x

Young, R. F. (2008). *Language and interaction: An advanced resource book.* London: Routledge.

Table 1 Foreign Language Acquisition at the end of the virtual exchange

	Much improved	Improved a little	No improvement	Has got worse
Ability to understand	16.8	43.1	23.1	0.3
Fluency in speaking	14.7	33.2	51.8	0.3
Grammatical accuracy	10.7	44.9	43.9	0.5
Accuracy of pronunciation	10.7	27.9	60.4	1.0
Range of vocabulary	14.0	54.1	31.7	0.3
Confidence in using the foreign language	24.4	37.3	37.8	0.5
Ability to interact with foreign language speakers	24.6	43.1	32.2	0.0

n = 394

Table 2 Correlation matrix of TPACK and Foreign Language Acquisition

	M	SD	1	2	3
1. Pre-Test TPACK	3.62	0.52			
2. Post-Test TPACK	3.92	0.51	.510**		
3. Δ TPACK	0.30	0.51	-.505**	.485**	
4. FLA	2.75	0.57	.126*	.229**	.102*

n = 394, * p < .05, ** p < .01

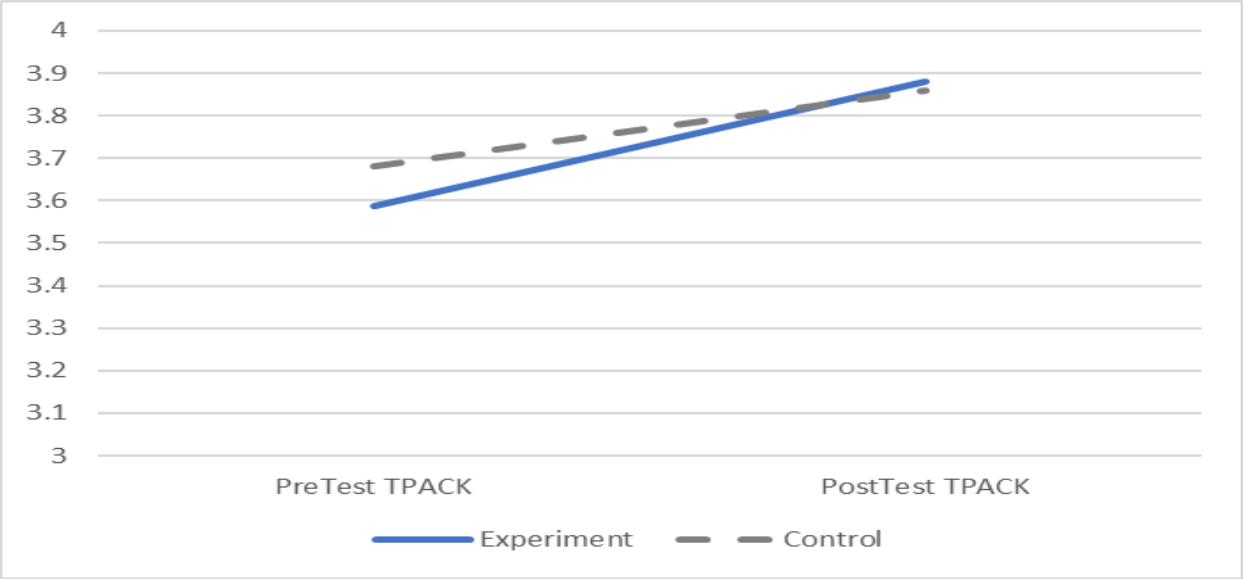
Table 3 Regression analysis of Foreign Language Acquisition

	Model 1	Model 2	Model 3
(Constant)			
Age	-.162**	-.153**	-.153**
Gender	.017	.017	.017
Early child specialisation	-.005	-.028	-.028
Secondary education specialisation	-.079	-.115	-.115
Other specialisation	.100	.057	.057
Studying Foreign Language	.096	.086	.086
Pre-Test TPACK	.106*	.006	.212**
Post-Test TPACK		.203**	
Δ TPACK			.202**
Rsq adj.	6.70%	9.40%	9.40%

a Dependent Variable: FLA, n = 385, primary education as benchmark

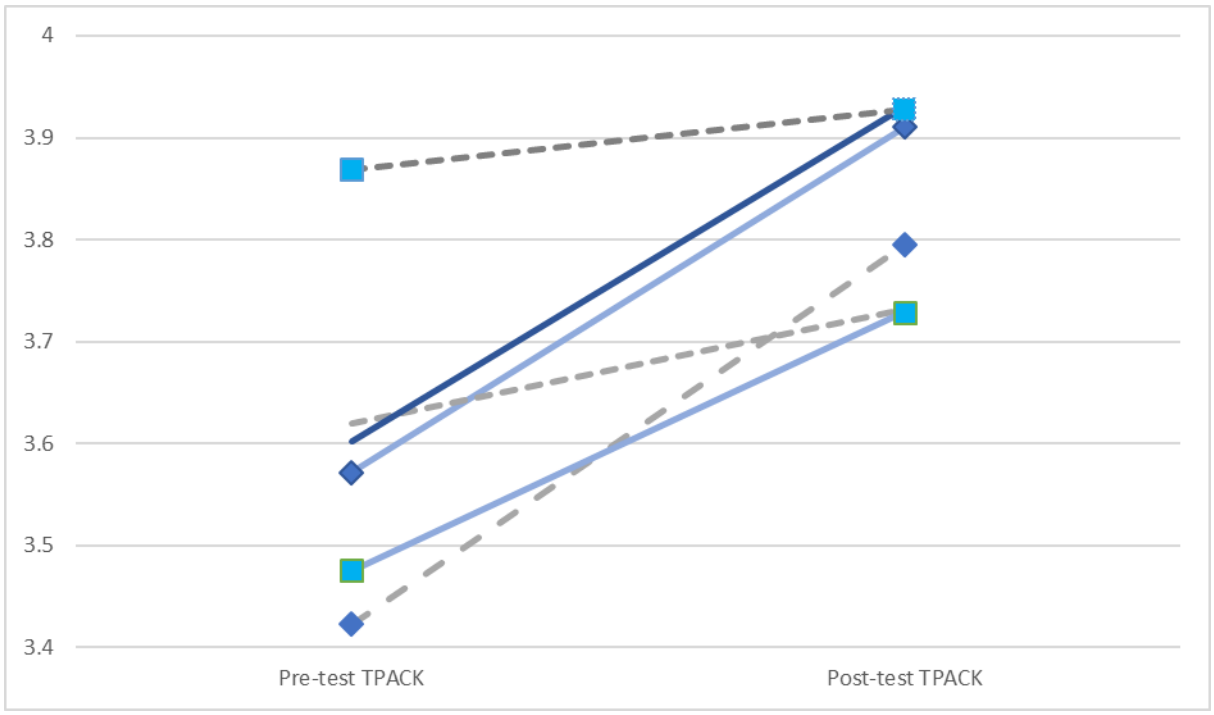
* p < .05, ** p < .01

Figure 1 Pre- and post-test TPACK comparison between experimental and control group



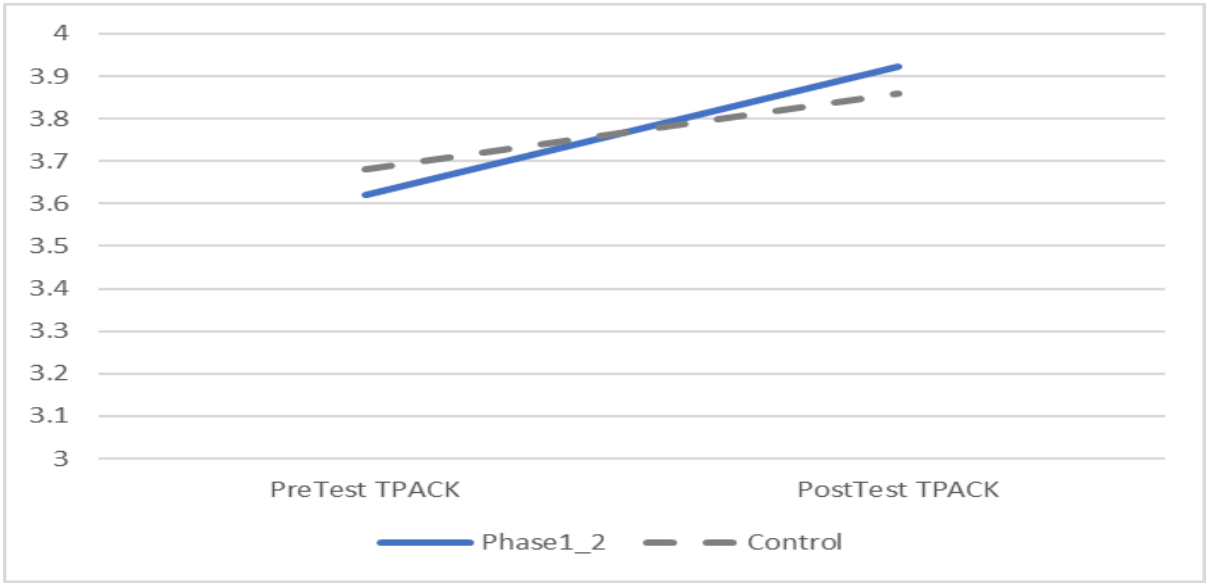
n exp = 122, n con = 63, 3 virtual exchanges

Figure 2 Pre- and post-test TPACK comparison between experimental and control group (per Virtual Exchange)



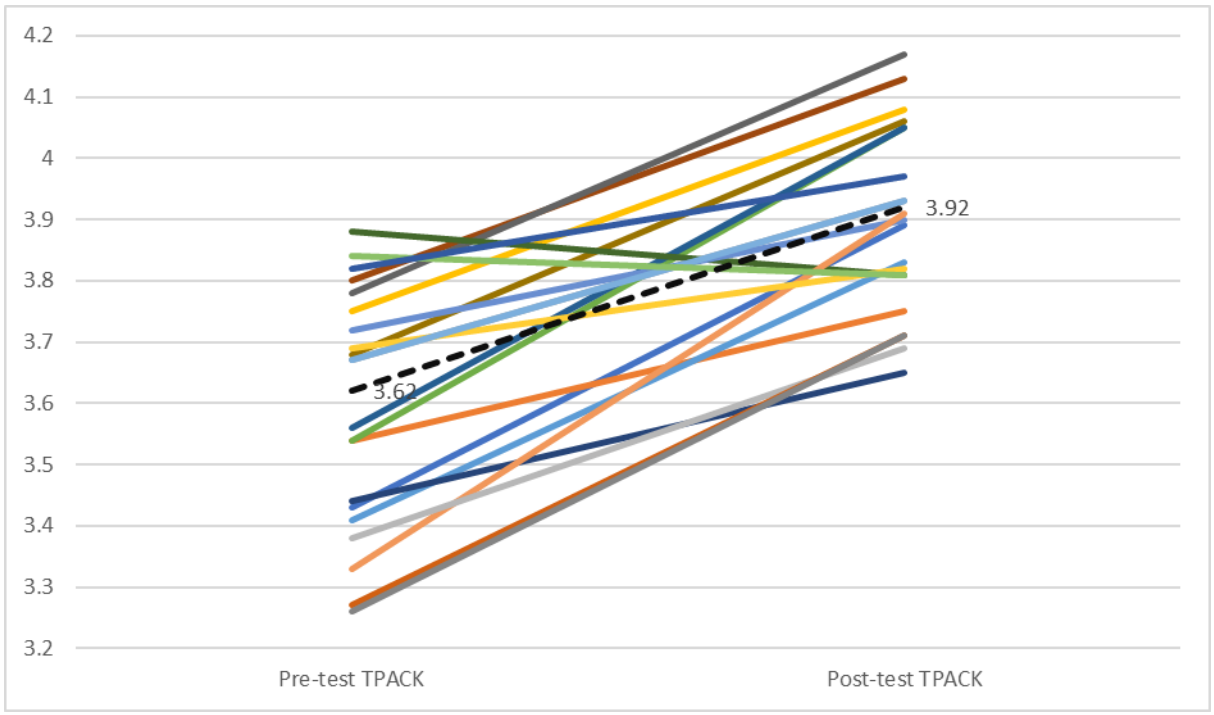
Three virtual exchanges (i.e., diamond, square, no label) indicated in blue, relative to three control conditions in dashed grey (i.e., diamond, square, no label)

Figure 3 TPACK gains over time during Phase 1_2 (control group as comparison)



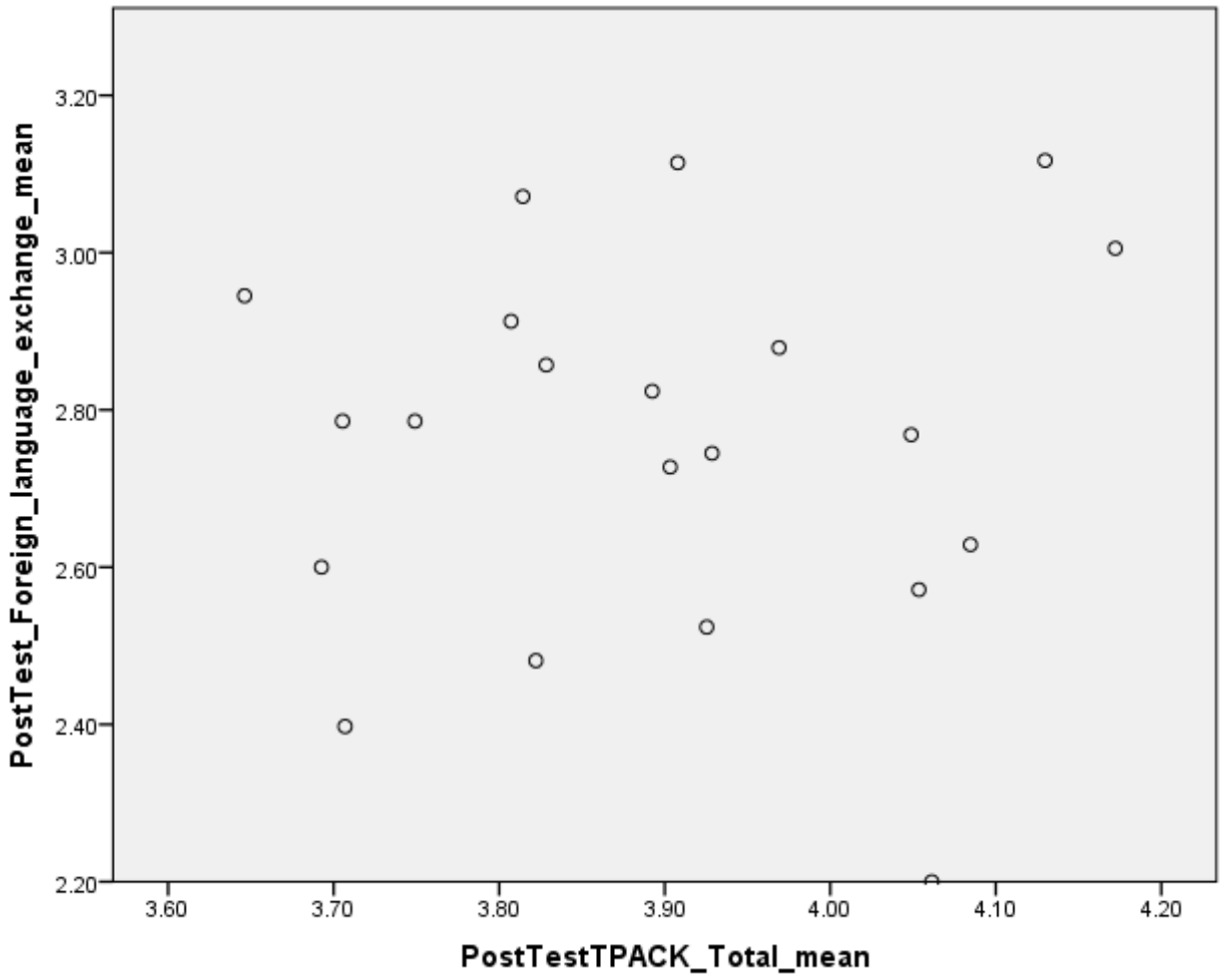
n phase 1_2 = 394, 21 exchanges; n con = 63, 3 exchanges

Figure 4 Pre- and Post-test TPACK skills of 21 virtual exchanges



n phase 1_2 = 394, 21 exchanges

Figure 5 Post TPACK and Foreign Language Acquisition across the 21 virtual exchanges



Appendix Table 1 Pre-post test results of Study 2 per exchange

Table A1. Pre-post test results of Study 2 per exchange.

VE	n	Pre-Test TPACK	SD	Post-Test TPACK	SD	Gain	T-test	VE country	Language used in VE
1	15	3.33	0.63	3.91	0.41	0.57	3.687**	POL-GER	EN
2	15	3.38	0.47	3.69	0.48	0.31	2.878*	POR-BRA	PT
3	19	3.69	0.43	3.82	0.46	0.13	1.507	SPA-NED	EN
4	28	3.67	0.40	3.93	0.35	0.26	3.567**	SPA-POL	EN
5	24	3.84	0.01	3.81	0.17	−0.02	−0.172	SPA-POL	EN
6	13	3.82	0.56	3.97	0.63	0.15	0.971	ISR-SPA	EN
7	23	3.27	0.54	3.71	0.34	0.43	3.662**	NED-SPA	EN
8	4	3.26	0.38	3.71	0.16	0.45	2.843	POL-SPA	EN
9	30	3.43	0.47	3.89	0.47	0.46	5.186**	GER-ISR	EN
10	8	3.54	0.27	3.75	0.46	0.21	1.259	CAN-ISR	EN
11	9	3.67	0.45	3.93	0.52	0.26	2.531*	GER-POR	EN
12	40	3.75	0.61	4.08	0.57	0.33	4.451**	SWE-ISR	EN
13	12	3.41	0.47	3.83	0.55	0.42	2.497*	POR-BRA	PT
14	23	3.44	0.50	3.65	0.55	0.21	2.252*	SPA-SWE	EN
15	13	3.80	0.36	4.13	0.55	0.33	2.334*	SPA-TUR	EN
16	40	3.78	0.65	4.17	0.63	0.39	4.073**	ISR-MEX	EN
17	20	3.68	0.47	4.06	0.55	0.38	2.741*	SWE-USA	EN
18	11	3.56	0.47	4.05	0.52	0.49	4.064**	USA-SWE	EN
19	15	3.88	0.38	3.81	0.60	−0.07	−0.482	POR-MAC	EN
20	53	3.72	0.40	3.90	0.39	0.18	2.733**	SPA-POL	EN

Paired T-test per VE, $^{*}p < .05$, $^{**}p < .01$