

UC Merced

UC Merced Previously Published Works

Title

Changes in interlanguage complexity during study abroad: A meta-analysis

Permalink

<https://escholarship.org/uc/item/8107f6xr>

Author

Xu, Yiran

Publication Date

2019-02-01

DOI

10.1016/j.system.2018.11.008

Peer reviewed

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

Changes in Interlanguage Complexity During Study Abroad: A Meta-Analysis

Yiran Xu¹

Georgetown University

Abstract

Despite the growing interest in interlanguage complexity development in study abroad (SA) research, no clear-cut conclusions can be made as to whether and to what extent learners' interlanguage complexity increases following a sojourn abroad. The current study meta-analyzed the overall effects of study abroad on measured oral and written complexity, as well as the moderator effects (i.e., learner demographics, SA contextual features, and outcome measures) on the variability of interlanguage complexity effect sizes (Cohen's *d*). A comprehensive search was conducted to obtain studies that have quantitatively documented lexical and syntactic complexity changes during SA through a pre-and-post SA design. A total of 30 independent samples from 28 primary studies involving 602 participants were retrieved and coded for gains and for moderator variables. Results show an overall small effect of study abroad on language complexity development ($d = 0.37$). In addition, moderator analyses suggest that larger effects are associated with (a) learners at an intermediate proficiency level, (b) learners enrolled in a language study program while SA, (c) programs that implemented a language pledge, or (d) programs with Mandarin Chinese as the target language. More fine-grained and systematic reporting practices are proposed for future research.

Keywords: interlanguage complexity, meta-analysis, study abroad, moderators

¹ I would like to thank Dr. Lourdes Ortega and Dr. Luke Plonsky for their continuous support and guidance on the project. I would also specifically thank Dr. Carmen Pérez-Vidal and her research team for contributing their longitudinal data to this project. The author email address: yx110@georgetown.edu.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

1. Introduction

The prevailing impression among language learners and study abroad (SA) practitioners is that study abroad provides learners a springboard which allows individuals to pick up a target language rapidly and effortlessly. It is often assumed that the immersion of the target language in a natural environment, combined with instruction in a classroom setting, can significantly expedite the learning process and lead to considerable proficiency changes. These beliefs are not without empirical support, as several studies have shown, learners make measurable linguistic gains while abroad (e.g., Collentine & Freed, 2004; Freed, 1995; Segalowitz & Freed, 2004; Grey, Cox, Serafini, & Sanz, 2015). More recent research also shows SA learners improved greatly in terms of neurocognitive behavior after studying abroad (Faretta-Stutenberg & Morgan-Short; 2018).

Nevertheless, previous SA literature has reported conflicting results with regard to the different dimensions of language development (i.e., complexity, accuracy, and fluency). In general, learners were seen to make measurable gains in oral fluency to varying degrees (e.g., Freed, 1995; Juan-Garau & Pérez-Vidal, 2007; Segalowitz & Freed 2004; Tullock & Ortega, 2017), but their development in complexity yields inconsistent results (e.g., Llanes & Muñoz, 2013; Llanes & Serrano, 2014). The ‘complexity’ in the present study refers to linguistic complexity, defined as the structural complexity of the underlying interlanguage system developed (Skehan, 2003), which is operationalized as a wide variety of both basic and sophisticated structures and words available to learners in oral and written production (Wolfe-Quintero, Inagaki, & Kim, 1998). Despite the growing interest in language development in the SA literature, no cumulative evidence has been gathered with regard to the trajectory of complexity changes among SA learners. Therefore, whether learners are able to produce

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

complex expressions, in writing or in speaking, after a sojourn, and how language complexity is moderated by the learners' demographic, SA contextual, and outcome measure variables remains mostly unclear.

This meta-analysis investigates the overall effects of study abroad on interlanguage complexity changes and the variations that result from a number of moderators, such as age, L2 proficiency, and length of stay. It then examines the reporting practices in the SA domain. The study aims to contribute to the collective efforts made in previous SA syntheses and meta-analyses (DeKeyser, 2014; Kinginger, 2009; Llanes, 2011; Pérez-Vidal, 2017; Sanz, 2014; Tullock & Ortega, 2017; Xiao, 2015) and to establish a common ground for the understanding of language learning in SA contexts.

2. Literature Review

In this section I first provide a brief outline of the research designs typically found in SA research. This is followed by a review of the potential factors (moderators) that have been found in previous literature to influence or predict the development of linguistic complexity during SA.

2.1 Experimental Design

A typical quantitative study that investigates language development in the SA domain commonly adopts either a within-group design, a between-group design, or a combination of the two. A within-group study compares the same learners' language performance prior to and immediately after the SA experience (e.g., Llanes, Tragent, & Serrano, 2015; Pérez-Vidal & Barquin, 2014). By contrast, a study adopting a between-group design usually compares learners' post-sojourn language performance in a SA context and in an At Home (AH) context

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

(e.g., Mora & Valls-Ferrer, 2012; Serrano, Llanes, & Tragant, 2016). There are also a small number of studies that have examined the retention of linguistic gains after learners returned to their home countries through a delayed posttest (e.g., Huensch & Tracy-Ventura, 2017; Pérez-Vidal & Barquin, 2014). Despite the different experimental designs adopted in previous research, the within-group comparison design is more reliable in terms of eliminating possible confounding factors related to individual differences (Sanz, 2014) and, for this reason, only studies with within-groups (pre-post) designs are included to ensure sufficient comparability across studies.

With regard to the experimental protocols, the majority of studies administered the same task across testing periods (i.e., pretest, posttest, and/or delayed posttest). A typical writing task involves learners producing a comparative or argumentative essay with prompts, and the same task was usually administered consistently across testing periods. Learners were not timed in the writing activity, but they were restricted from external assistance such as dictionaries and textbooks. Regarding the oral tasks, picture-elicited narrative tasks and guided interviews were two commonly adopted elicitation tasks. Data elicited from the written or oral tasks were subsequently coded and analyzed in terms of syntactic complexity, lexical complexity, or both.

2.2 Moderators of Complexity Development in SA Context

Despite a similar design adopted by SA researchers, studies vary on a number of dimensions which may contribute to different pace in complexity growth. This study investigates three categories of moderating variables: (a) learner demographics (i.e., age and L2 proficiency), (b) contextual features (i.e., length of stay, types of coursework, and language pledge), and (c) outcome measures (i.e., modality and operationalization of complexity).

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

2.2.1 Learner Demographics

Empirical evidence collected from non-SA settings has shown that in a natural environment adults and older children enjoy an initial advantage over younger children in language learning due to a higher level of cognitive maturity, while younger children tend to perform better in the long run (e.g., Krashen, Long, & Scarcella, 1979; Long, 1990). Thus, one empirical question of interest to SA researchers and practitioners is: To what extent do adults differ from young learners in terms of linguistic gains during study abroad where natural learning is afforded. While a growing body of research has been conducted with adult learners, few studies are dedicated to young learners enrolled in an SA program under the age of 18 and even fewer have compared the linguistic development of children to that of adult learners in SA settings (e.g., Serrano, Llanes, & Tragant, 2016). In addition, Llanes (2011) has noted that other contextual factors are dependent on and intertwined with participants' age, such as living arrangements, the amount of instruction, and the degree of language contact while abroad. For example, children who study abroad may be exposed to higher-level input, in terms of both quantity and quality, because children are usually arranged to stay with host families. A comprehensive and systematic analysis of any outcome differences between SA adults and children will expand our understanding of age as a potential moderator for language development in the SA context.

Another demographic characteristic that might moderate the benefits of SA is the proficiency of learners' target language prior to study abroad. DeKeyser (2007) has noted that SA learners need to be equipped with sufficient declarative knowledge before studying abroad in order to take full advantage of their language experiences. Therefore, advanced learners are more likely to maximize interactions with native speakers and develop deeper connections in the host

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

culture compared to less-proficient learners, whose utterances are constrained by their limited proficiency (e.g., Segalowitz & Freed, 2004; DeKeyser, 2014; Dewey, Bown, Baker, Martinsen, Gold, & Eggett, 2014). However, this does not mean that advanced learners would necessarily make more linguistic progress. In fact, empirical evidence has shown that learners with lower-initial proficiency often made greater progress compared to advanced learners after studying abroad, because whereas the less proficient learners enjoy rapid growth thanks to the catch-up effect, the more advanced learners may find it difficult to improve substantially due to the diminishing returns (e.g., Brecht, Davidson, & Ginsberg, 1990; Freed, 1995; Llanes & Muñoz, 2009). The synthetic nature of the current study allows for a comprehensive comparison of complexity gains after SA across learners of different initial L2 proficiencies, thus providing more insights about the role of initial L2 proficiency in interlanguage complexity development.

2.2.2 SA Contextual Features

The length of stay while abroad has been investigated extensively in SA literature. The fundamental question is how long is considered long enough to observe a relatively significant change in language proficiency. Most studies found longer stays to be associated with greater language gains (e.g., Llanes & Muñoz, 2009; Sasaki, 2009), yet some studies show otherwise (Lara, Mora, & Pérez-Vidal, 2015; Llanes & Serrano, 2011). This pattern of conflicting findings may be partially accounted for by the fact that the length of stay abroad may be confounded by other program features, such as the intensity of instruction, the amount of language contact, and curriculum structure. Nevertheless, whether a more extended stay affords more learning opportunities and linguistic gains remains inconclusive.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

Another potential moderating factor that has been largely overlooked in the existing literature is the type of coursework available to learners during study abroad. Learners studying abroad either take exclusively language classes (e.g., Xu, 2016; Shi, 2016) or develop their content knowledge through an L2 (e.g., Mora & Valls-Ferrer, 2012; Zalbidea & Martin, 2016). In this sense, these two groups of learners should be distinguished since their SA motivations, purposes, and foci of classroom instruction are substantially different.

Finally, a language pledge that has been implemented by some SA programs may affect the language gains from SA. Upon signing a language pledge, learners promise to use only the target language during their entire stay while abroad. Therefore, a language pledge is often considered to create a fuller immersion environment for the learners, thus affording more language interactions. Several studies have explored the gains in SA programs that feature a pledge, and the majority of them have reported improvements in different aspects of language development to varying degrees (e.g., Du, 2013; Grey et al., 2015; García-Amaya, 2012). The present study takes advantage of the meta-analytical method and directly compare gains reported by programs with and without a language pledge.

2.2.3. Outcome Features

Several studies have suggested that oral proficiency is more sensitive to the SA context than writing proficiency (e.g., Freed 1995; Freed, Segalowitz, & Dewey, 2004). This is supported by some empirical evidence that SA learners have demonstrated measurable improvements in oral fluency (e.g., Freed, 1995; Mora & Valls-Ferrer, 2012; Segalowitz & Freed, 2004), while they tend to take more time and effort to develop writing proficiency, especially in terms of writing syntactic complexity. Ortega's (2003) synthetic study suggested

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

that a minimum of one year of college-level instruction was needed for substantial improvements in the syntactic complexity of L2 writing to be observed. One primary goal of the current study is to compare the complexity development between the two modes based on aggregated lexical and syntactic complexity scores.

Furthermore, scholars of complexity development in non-SA contexts have noted that findings with regard to complexity development are often inconsistent, partly because the definition of complexity as a descriptor of L2 performance is ambiguous and thus operationalized differently across studies (Housen & Simoens, 2016; Ortega, 2003; Skehan, 2003; Wolfe-Quintero et al. 1998). For example, in their review of L2 studies using complexity measurement, Bulté and Housen (2012) found that 40 different measures were employed in 40 studies. Therefore, the current study incorporates the organic approach to syntactic complexity proposed by Norris and Ortega (2009) that combine the redundant measures and categorize different measures into four dimensions of interlanguage complexity such as length, the amount of subordination, and the amount of coordination.

2.3 Previous Meta-Analyses of the SA Domain

So far, two meta-analyses have undertaken a synthetic approach to SA gains. Yang (2016) compared SA learners' posttest performance between available SA groups and AH groups in 11 studies. She found that SA learners overall outperformed AH learners ($d = 0.75$). She examined the length of stay as a moderator and showed that shorter stays overall yielded better language performance. However, the results should be interpreted with caution, since the effect sizes were calculated based on posttest scores exclusively. In addition, some studies in the sample reported enormous effect sizes: $d = 5.45$ (Dewey, 2008) and $d = 7.80$ (O'Brien,

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

Segalowitz, Freed, & Collentine, 2007), which may have skewed the results. The other meta-analysis was conducted by Tullock and Ortega (2017) and it focused exclusively on oral fluency changes. Employing a different approach from Yang, they looked into oral fluency development within the SA contexts and how the effect was moderated by construct measurements. Overall, they found students became substantially more fluent after SA experiences, with effect sizes of speech rate ranging mostly from $d = 0.5$ to $d = 1.2$. Looking closely at different measures of fluency, their results showed that character- and syllable-based units of analysis were more reliable than word-based units for speech rate, if all other variables were held constant. In gauging complexity, too, the choice of construct measurement should not be taken for granted.

3. Research Questions

The present study addresses the following three questions:

1. To what extent does learners' interlanguage complexity increase during study abroad, based on the aggregated lexical and syntactic complexity scores extracted across individual studies?
2. To what extent are the effects of study abroad on target language complexity moderated by learner demographics, SA contextual features, and outcome measures?
3. What are the reporting practices of SA research on complexity development, and how can they be improved in future research?

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

4. Methods

4.1 Study Identification and Retrieval

The literature search, which ended in July 2017, followed the searching practices recommended by Cooper (2016) as well as the approaches in several previous meta-analyses of L2 research (e.g., Norris & Ortega, 2000; Plonsky, 2011). Unpublished studies from Ph.D. dissertations and conference presentations were considered in the search process in order to minimize publication bias (Borenstein, Hedges, Higgins, & Rothstein, 2009). Seven searching channels were utilized in order to obtain a comprehensive and exhaustive collection of studies on learners' language complexity changes during SA: (a) domain-specific databases—Linguistics and Language Behavior Abstracts (LLBA), the Educational Information Resource Center (ERIC), PsycINFO, ProQuest; (b) non-domain-specific engines—Google search engine and Google Scholar; (c) online searches through SA-specific journals (e.g., *Frontier*) and 14 applied linguistics journals; (d) manual searches in two conference programs and proceedings published before 2017 (e.g., the Conference of the American Association for Applied Linguistics; the Second Language search Forum); (e) manual searches on the websites of prominent scholars in this domain; (f) backward searches in the reference lists of included studies; and (g) personal contacts with active scholars.

In order to ensure search reliability, the following set of keywords was applied consistently across databases: [study abroad OR sojourn abroad OR overseas study] AND [Second language OR foreign language OR Oral OR writ*] AND [Complexity]. Minor adjustments were made to accommodate different features built in each database.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

4.2 Inclusion and Exclusion Criteria

Studies had to meet the following screening criteria in order to be included in the analyses:

1. Learners had to study abroad to fulfill degree requirements or to participate in exchange programs for educational purposes (Kinginger, 2009).
2. The sample had to be original and unique. Studies using duplicate samples or a subsample of a larger sample were eliminated (e.g., Barquin, 2012; Pérez-Vidal, 2015).
3. Studies had to report basic descriptive statistics (i.e., mean and standard deviation) that allow for the calculation of Cohen's *d* effect sizes. Missing data were obtained through direct contact with the authors.
4. Studies with non-retrievable data after contacting the author were excluded.
5. The study had to adopt a pre-post design for within-group comparison. Cross-sectional studies that compared an SA group with an AH group but that did not include a pre-post contrast were eliminated (e.g., Serrano, Tragant, & Llanes, 2014).
6. Language complexity had to be treated as a dependent variable that describes learners' interlanguage. Studies that only investigated other aspects of language development, such as fluency and accuracy, were excluded (e.g., Collentine, 2004).
7. The study had to utilize analytic measures to gauge complexity development. Studies that used holistic measures, such as ratings obtained from the Oral Proficiency Interview (e.g., Di Silvio, Donovan, & Malone, 2014; Sasaki, 2009) or holistic ratings elicited from human raters (Hardison, 2014) were not considered.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

8. For studies that investigated writing complexity, the major tasks administered had to be essay-writing related tasks. Studies using other tasks, such as a grammatical judgment task, were excluded (e.g., Isabelli, 2001).
9. The study had to be reported in English.

4.3 Coding Procedures

The coding scheme (shown in Appendix A of the Supplementary Data File online) was developed through an iterative process of repeated revisions. All the studies were coded following the optimal meta-analytic practices (Lipsey & Wilson, 2001). Three types of descriptors were coded as possible moderators: learner demographics, SA contextual features, and outcome measures. In addition, combining the recommendation of APA guideline and the study by Plonsky (2013), this study also examined the quality of reporting practices in SA research. Specifically, 17 study features were coded as ‘reported’ or ‘not reported’: sample size, home institution, SA destination, L1 background, age, L2 proficiency, standard of L2 proficiency, target language, length of stay, accommodation, types of coursework, use of language pledge, language contact, report of mean and standard deviation, effect size of test scores, and operationalization of outcome constructs. The frequency of reporting practices for each feature was counted and evaluated for quality analysis. To ensure the coding reliability, all studies were first coded twice by the same rater. Then, a second rater coded 20% of the sample with an inter-coder agreement of 96.7%.

To investigate the moderating effect of complexity measure on outcome, different operationalizations of complexity construct were categorized into seven sub-dimensions based

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

on Read's (2000) analysis on lexical complexity and Norris and Ortega's (2009)

multidimensional analysis on syntactic complexity (see Table 1).

Table 1. Lexical and syntactic complexity construct operationalization

Construct Dimensions	Operationalization
Lexical Complexity	
<i>Lexical density</i>	Academic word density Lexical word density Functional word density Content word density
<i>Lexical diversity</i>	Guiraud's index Type and token ratio
<i>Lexical sophistication</i>	Greco-Latin index Percentage of words in Academic Word List Lexical sophistication
Syntactic Complexity	
<i>Length</i>	No. of words per clause/AS-unit/T-unit
<i>Amount of subordination</i>	No. of clauses per sentence/T-unit /AS-unit No. of dependent clause per T-unit/AS-unit No. of subordinate clauses per clause No. of T-units per sentence
<i>Amount of coordination</i>	Coordination index
<i>Total frequency of certain forms considered to be sophisticated</i>	No. of verb phrases per clause Coordinate phrases per clause/T-unit The Ratio of complex T-unit

4.4 Analysis

Following previous practices and suggestions in conducting meta-analyses (Norris & Ortega, 2000; 2006; Oswald & Plonsky, 2010; Plonsky & Oswald, 2014), Cohen's *d* was used to calculate the effect size in representing gains in complexity. Specifically, the overall effect sizes were calculated based on the pretest (prior to SA) and the posttest (immediately after SA) scores of aggregated oral and written complexity. Delayed posttest scores were not considered because only one study in the sample investigated learning retention (Pérez-Vidal & Barquin, 2014). A

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

random-effects model was chosen for aggregating study effects because the sample was highly heterogeneous in terms of the study features. In such cases, a fixed-effects model may lead to a Type I error and falsely narrow the confidence intervals (Borenstein, Hedges, Higgins, & Rothstein, 2010). All the effect sizes (d) were calculated based on the descriptive data (i.e., mean, standard deviation) and weighted by sample size using the Effect Size Determination Program (Wilson, 2001). Furthermore, one effect size was generated for each independent sample for the main analyses. As a result, 30 effect sizes were produced from a total number of 30 independent samples (from 28 studies) and 602 SA learners.

Then, a series of the moderator analyses were performed in order to understand how the observed changes may be potentially moderated by internal and external factors. Three categories of moderators were coded in the current study: (a) participant demographics; (b) contextual features and (c) outcome measures. The analyses of moderator variables were consistent with the procedure in calculating the overall effect size, except that the sample was further divided into different comparison subgroups based on the features of the moderators.

5. Results

5.1 Characteristics of Study Sample

A total of 30 independent samples were retrieved from 28 primary studies (see the references in Appendix C of the Supplementary Data File online), including 19 journal articles, one book chapter, three unpublished doctoral dissertations, four conference presentations, and one unpublished manuscript. These samples are consisted of a total of 602 SA learners. The sample sizes ranged from 4 to 101, with an average of 31 ($SD = 21.7$), and a median size of 23

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

participants per study. The studies were all published between 2003 and 2017, with 21 (70%) of them published after 2013.

In terms of the demographic features, nearly half of the studies investigated adult Catalan and Spanish bilingual learners (43.3%, $k = 13$), and the majority of the studies examined English as a target language (76.7%, $k = 23$). Four studies investigated Chinese as a target language (Kim et. al, 2015; Shi, 2016; Wright & Zhang, 2014; Xu, 2016), two Spanish (Jensen & Howard, 2014; Zalbidea & Martin, 2016), and one French (Godfrey, Treacy, & Tarone, 2014). The L2 proficiency was highly variable regarding both reported level and evaluation criteria. Three studies examined learners of intermediate level, seven studies advanced learners, and the remaining a mixed level of proficiency among learners. Most studies referred to the International English Language Testing System (IELTS) and the Common European Framework of Reference for Languages (CEFR), and only a small number of studies referred to other standardized tests such as American Council on the Teaching of Foreign Languages (ACTFL) and Test of English as a Foreign Language (TOEFL). Five studies reported L2 proficiency using local measures such as placement or diagnostic tests, or self-reported data. Therefore, while many studies were consistent in terms of age, L1 background, and target language, other studies featured distinct characteristics, which may contribute to variations in effect size (d).

Similar patterns were observed with the coded contextual features. The majority of the participants were adult learners enrolled in short-term study abroad programs that lasted from 5 weeks to 4 months ($k = 24$, 80%), and most study abroad programs took place in English-speaking countries (76.7%, $k = 23$). Three programs implemented a language pledge that encouraged or enforced learners to use the target language exclusively throughout the duration of

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

the study abroad (Zalbidea & Martin, 2016; Shi, 2016; Xu, 2016). It also shows that SA learners took both language courses and content courses delivered in the target language while abroad.

Finally, outcome measures also varied to different degrees. For example, in terms of the writing assessment tasks, most programs adopted narrative or argumentative essay tasks. In contrast, tasks used in oral assessment show a lower degree of consistency: nine studies used picture-elicitation tasks, and the other tasks included a role play, a semi-guided interview, and an oral diary. In addition, the construct of complexity was operationalized differently across studies. Some measures, such as Guiraud's index (GUI) and clause per T-Unit (CL/TU), tended to dominate among certain teams of researchers.

5.2 The Overall Effects of SA on Complexity Development

Research Question 1 (RQ1) addressed the main effect of study abroad on interlanguage complexity development. Table 2 shows the sample size (N), unweighted and weighted effect sizes (d , d_w), variance, and a 95% confidence interval for each study sample. The overall weighted effect size is $d = 0.37$, which means that the effects of study abroad on language complexity is overall positive yet small based on Plonsky and Oswald's (2014) SLA domain-specific benchmarks. As shown in the table, the effect sizes of the majority of studies clustered around the overall weighted effect size ($d = 0.37$), except for four studies (i.e., Kim et al., 2014; Wright & Zhang, 2014; Shi, 2016; Xu, 2016), which featured computed effect sizes that are higher than 1.5. Interestingly, all the four studies investigated Chinese as a target language. Despite the fact that these effect sizes might be considered as statistical outliers, they were retained for the remaining analysis in order to preserve the integrity of the sample. However, an

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

additional moderator analysis on target language was created to account for the outliers in subsequent analyses.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

Table 2. Aggregated Effect Sizes

Authors (year)	<i>N</i>	<i>(d)</i>	<i>(d_w)</i>	<i>Variance</i>	95% Confidence Interval	
					Lower	Upper
Lara (2014) sample 2	14	-0.19	-0.18	0.06	-0.72	0.34
Lara (2014) sample 1	33	-0.14	-0.14	0.03	-0.49	0.20
Zaytseva (2016)	30	-0.04	-0.04	0.03	-0.40	0.32
Robson (2015)	23	0.00	0.00	0.04	-0.41	0.41
Mora & Valls-Ferrer (2012)	30	0.02	0.02	0.03	-0.34	0.38
Knoch, Rouhshad, Oon & Storch (2015)	31	0.03	0.03	0.03	-0.32	0.38
Longcope (2003)	7	0.08	0.07	0.11	-0.67	0.82
Serrano, Llanes & Tragant (2016)	16	0.19	0.18	0.06	-0.31	0.68
Knoch, Rouhshad & Storch (2014)	101	0.13	0.13	0.00	-0.07	0.33
Godfrey, Treacy & Tarone (2014)	4	0.15	0.13	0.12	-0.66	0.95
Llanes & Muñoz (2013)	46	0.23	0.23	0.02	-0.07	0.52
Llanes, Tragant & Serrano (2012)	24	0.23	0.22	0.04	-0.17	0.64
Storch (2009)	25	0.26	0.25	0.04	-0.14	0.66
Llanes, Tragant & Serrano (2015)	64	0.28	0.28	0.02	0.03	0.53
Serrano, Llanes & Tragant (2011)	25	0.29	0.28	0.04	-0.11	0.69
Romanova (2016)	21	0.32	0.31	0.05	-0.12	0.76
Llanes & Serrano (2014) sample 2	24	0.37	0.36	0.04	-0.05	0.78
Deng et al. (2010)	31	0.38	0.37	0.03	0.02	0.74
Leonard & Shea (2017)	39	0.40	0.39	0.03	0.07	0.72
Pérez-Vidal & Barquin (2014)	73	0.43	0.43	0.01	0.19	0.67
Llanes & Serrano (2014) sample 1	50	0.43	0.42	0.02	0.14	0.72
Zalbidea & Martin (2016)	18	0.56	0.54	0.06	0.06	1.05
Karlin (2016)	38	0.59	0.58	0.03	0.24	0.93
Jensen & Howard (2014)	18	0.61	0.58	0.06	0.10	1.11
Serrano, Tragant & Llanes (2012)	14	0.64	0.60	0.08	0.05	1.21
Storch & Tapper (2009)	69	0.72	0.71	0.02	0.45	0.98
Xu (2016)	21	1.59	1.53	0.10	0.93	2.23
Wright & Zhang (2014)	10	1.67	1.53	0.20	0.67	2.63
Kim et al. (2015)	22	1.75	1.69	0.11	1.07	2.41
Shi (2016)	12	1.84	1.71	0.19	0.88	0.77
Overall	602	0.38	0.37	0.00	0.24	0.49

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

5.3 Moderator Analyses

Research Question (RQ2) addressed the effect of the moderators. Table 3 summarizes the demographic and contextual features. Output for outcome measures is presented in Figures 1 through 3.

5.3.1 Learner Demographics

One goal of the current analysis was to examine whether ages and L2 proficiency levels are correlated to interlanguage complexity changes. Result shows that there was no substantial difference in complexity changes between adults and young sojourners under the age of 18: the effects were small for both groups. Although the conclusion can only be drawn tentatively due to the unbalanced sample sizes, it appears that age has no substantial impact learners' complexity development in SA context. As for the role of L2 proficiency, results show a clear advantage for intermediate-level learners ($d = 0.64$, $k = 3$) over advanced learners ($d = 0.17$, $k = 7$). However, since the sample sizes were unbalanced and the variance in the mean effect size of the intermediate level ($\sigma^2 = 0.17$) was considerably larger than that of the advanced level group ($\sigma^2 = 0.01$), the conclusion should be interpreted with caution. A closer examination of study features was carried out to check whether other contextual features were correlated. The results show that all the learners of the advanced level proficiency took content courses, whereas most intermediate level learners were enrolled in language courses. In addition, two studies investigating intermediate level learners studying Mandarin reported effect sizes greater than 1.5, which may partially account for the better performance in the intermediate level subgroup.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

5.3.2 SA Contextual Features

Turning to the effects of the contextual features, results show similar effects for short-term SA programs ($d = 0.38$, $k = 24$) and long-term SA programs ($d = 0.32$, $k = 6$), which indicate that studying abroad for an additional semester may not yield a substantial gain on linguistic complexity. Learners who only enrolled in language courses while abroad (mostly of intermediate-level proficiency) made substantially more progress ($d = 0.7$, $k = 10$) compared to those who enrolled exclusively in content courses (mostly of advanced-level proficiency) ($d = 0.24$, $k = 10$). Only three of the studies in the sample investigated a program that required learners to sign a pledge, and the weighted effect size was large ($d = 1.27$, $k = 3$). In contrast, the rest of programs that did not implement a language pledge only revealed small gains in language complexity ($d = 0.3$, $k = 27$). Taking a closer examination of the individual studies that implemented a language pledge, two of the three were Mandarin programs (Shi, 2016; Xu, 2016), and the other one was Spanish (Zalbidea & Martin, 2016). A subsequent analysis was conducted on target language as a potential moderator, which clearly showed that learners who participated in the Mandarin SA programs made the largest gains ($d = 1.61$), followed by moderate gains by learners in Spanish programs ($d = 0.56$), and learners in English and French programs only made modest gains ($d = 0.26$; $d = 0.15$).

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

Table 3. The effectiveness of SA on L2 complexity by demographics and contextual moderators

Group	Subgroup	Value	<i>d</i>	<i>k</i>	<i>N</i>	Variance	95% Confidence interval	
							Lower	Upper
Participant characteristics	Age	Minor	0.3	4	154	0.007	0.14	0.46
		Adults	0.39	26	779	0.006	0.24	0.54
	L2 proficiency	Intermediate	0.64	3	73	0.167	-0.159	1.44
		Advanced	0.17	7	216	0.014	-0.069	0.4
		Mixed	0.42	15	474	0.009	0.238	0.6
SA contextual features	Length of stay	0-16 weeks	0.38	24	817	0.004	0.24	0.49
		17-32 weeks	0.32	6	116	0.03	0.003	0.63
	Coursework type	Language course	0.7	10	252	0.03	0.36	1.05
		Content course	0.24	10	377	0.007	0.18	0.5
		Mixed	0.23	10	304	0.014	0.005	0.46
	Language pledge	No	0.3	27	881	0.003	0.19	0.41
		Yes	1.27	3	52	0.19	0.41	2.12
	Target Language	Mandarin	1.61	4	65	0.034	1.25	1.97
		English	0.26	23	846	0.002	0.168	0.36
		Spanish	0.56	1	18			
French		0.15	1	4				

5.3.3 Outcome Measures

Since the mean effect size of the overall language complexity was calculated based on the combined oral and written modalities, the current moderator analysis parses out the effects of each modality and construct. Figure 1 shows the means and 95% confidence intervals of this

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

analysis. As can be seen, a relatively small effect was found for both written ($d = 0.31, k = 14$) and oral ($d = 0.41, k = 20$) complexity development, although the latter shows a slight advantage. As the general assessing construct, learners made slightly larger gains in lexical complexity ($d = 0.29, k = 18$) than syntactic complexity ($d = 0.2, k = 17$). However, both effect sizes were small.

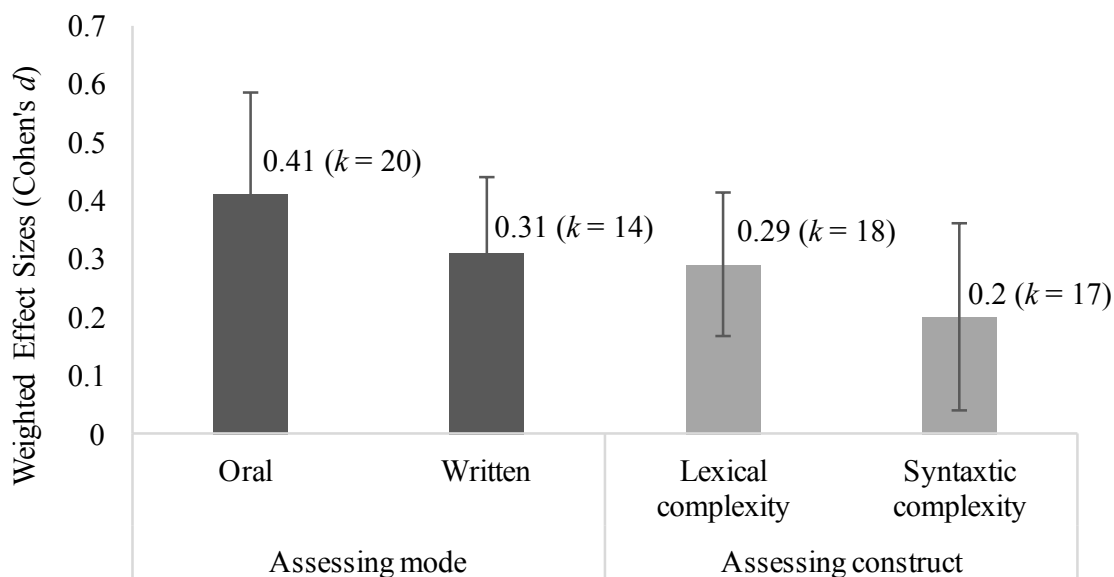


Figure 1. Means (d) and 96% confidence intervals of assessment mode and construct

Figures 2 and 3 further show a detailed analysis of the operationalization of lexical and syntactic complexity in both the oral and the written modalities. Overall, learners demonstrate positive growth in lexical density, diversity, and sophistication in the written mode over time. In contrast, their levels of lexical density and sophistication are smaller in the speaking mode, although the degrees vary by measures. However, lexical diversity shows notable changes in both written and oral modes. In general, SA learners demonstrated modest progress in syntactic complexity comparing to lexical complexity. The overall length of production unit, the amount of subordination, and the frequency of sophisticated forms of SA learners did not seem to improve remarkably.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

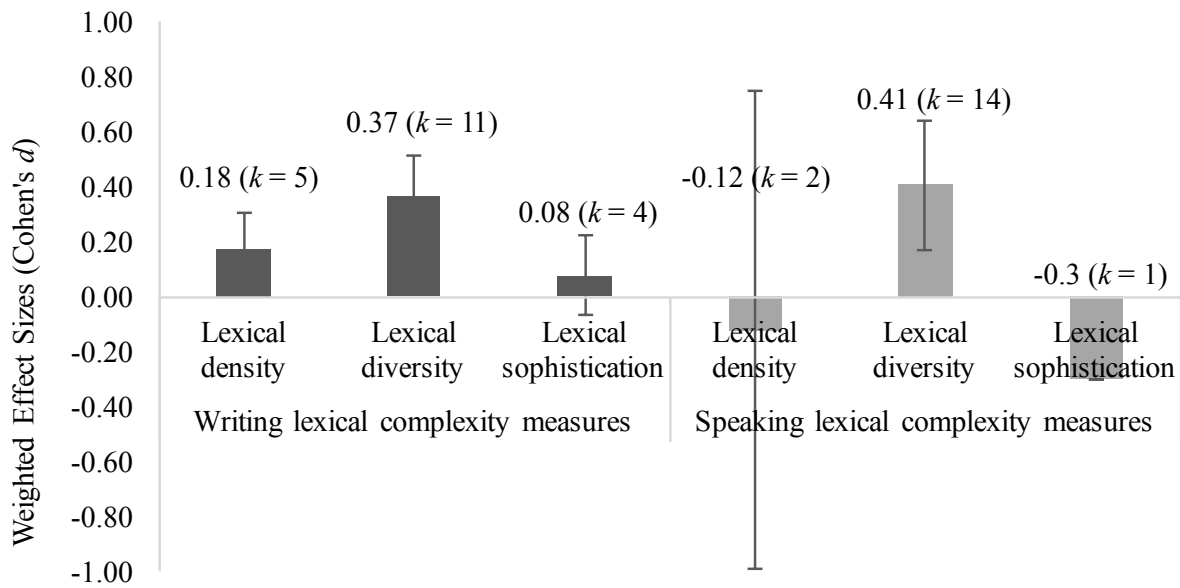


Figure 2. Means (*d*) and 95% confidence intervals for development of lexical complexity in writing and speaking

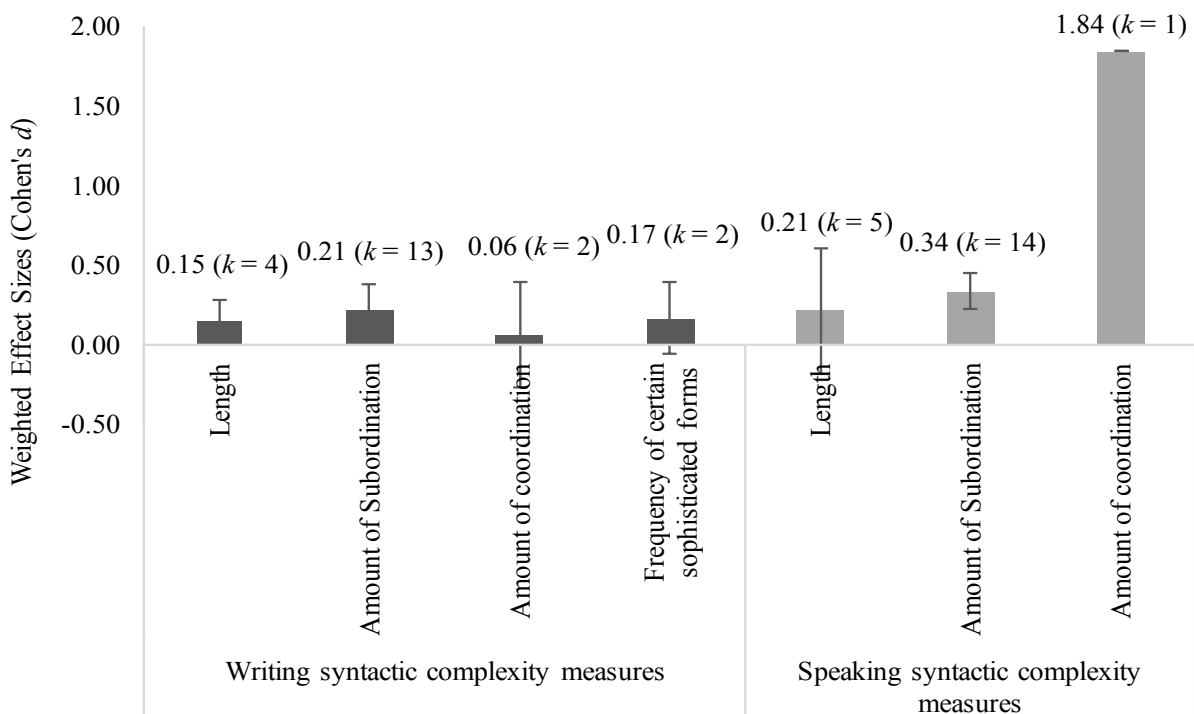


Figure 3. Means (*d*) and 95% confidence intervals of writing and speaking syntactic complexity

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

5.4 Assessment of Transparency

Research Question (RQ3) involved an assessment of the quality of studies and reporting practices in the SA domain represented by 30 samples. Study quality is defined in the field of meta-analysis by Plonsky (2013) in terms of adherence to standards of appropriate contextualization, rigorousness of methodology, and transparency and completeness of reporting. The results show that the majority of features, such as sample size, learner age, and home country, were consistently reported in the present sample. Some other features, however, were inadequately reported. Figure 4 illustrates the reporting practices in the current sample in terms of the participant demographics, contextual features, and outcome measures. First, it can be seen that the majority of the studies reported L1 background ($k = 28$), L2 proficiency ($k = 22$) and living arrangements ($k = 28$). However, it is somewhat perplexing that two studies did not report L1 backgrounds and living arrangements, and eight studies did not mention L2 proficiency. Then, in 11 studies there was no specific mention of the language proficiency criteria, making it difficult to interpret results or to compare learner language profiles across studies. In addition, nearly half of the studies failed to report the intensity of language instruction, which influence learners' the actual amount of language exposure on a daily or weekly basis. Similarly, more than half of the samples ($k = 19$) neglected to report the learners' language contact while abroad, defined as hours of exposure to language-related activities that learners receive outside of classroom settings. In addition, more than half of the sample ($k = 16$) failed to report previous SA experience. Finally, only eight studies reported effect sizes, and only three out of them reported effect sizes of results that yielded a statistical significance. The majority of the reports ($k = 20$) did not provide an effect size.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

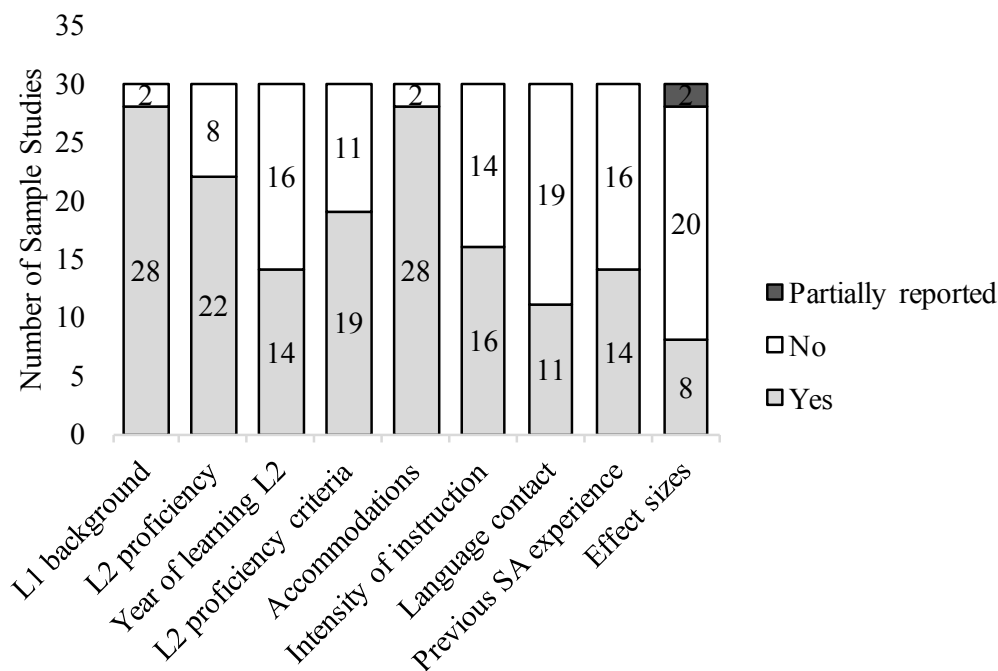


Figure 4. Reporting Practices in SA Research on Language Complexity

6. Discussion

This meta-analysis investigated the overall effects of study abroad on interlanguage complexity changes and sought to identify the variables that may moderate those effects. It also aimed to evaluate the reporting practices of the sample so as to propose a reporting guideline to promote transparency and consistency in this field.

Results show that students who studied abroad were able to produce more complex utterances, although such improvement remained quite modest ($d = 0.37$). This result, in general, concurs with the results reported in previous SA literature that learners did not show substantial improvements in complexity (DeKeyser, 2014; Kinginger, 2009; Llanes, 2011; Pérez-Vidal, 2017; Sanz, 2014; Xiao, 2015). This can be attributed to a number of factors. First, the overall length of learners' study abroad under investigation was rather short, ranging from five weeks to four months, which may have been too brief to observe substantial language complexity

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

development. This is especially true in the case of writing syntactic complexity, which usually requires at least 12 months of language instructions for learners to demonstrate noticeable improvements (Ortega, 2003). In addition, only one-third of the programs were language programs with explicit language learning goals, while the rest of the programs focused on content learning mediated through the target language. In such cases, the classroom learning primarily focused on the content, rather than the language per se. As a result, they might receive little feedback on their language usage. Furthermore, a trade-off among complexity, accuracy, and fluency during language development should also be considered. Skehan (1998) noted that meaning is often reflected in fluency whereas form is reflected in language accuracy and complexity. Accordingly, he proposes that a primary focus on fluency may compromise the development in complexity and accuracy due to the limited processing capacity of learners. Following his reasoning, the cumulative evidence provided in Tullock and Ortega (2017) on learners' moderate to large oral fluency development in SA settings may partially explain the modest gains in language complexity. This is further supported by a number of individual studies which have shown that learners demonstrate measurable improvements in oral fluency during study abroad but not necessarily in language complexity (Freed, 1995; Freed et al. 2004; Segalowitz & Freed 2004; Juan-Garau & Pérez-Vidal 2007; Mora & Valls-Ferrer, 2012; Valls-Ferrer, 2011; Llanes & Muñoz, 2009). However, the modest gains in linguistic complexity should not be interpreted as evidence against study abroad activities. To be sure, the results of the current study do indicate that learners make certain improvement after study abroad. Nevertheless, the results would perhaps caution learners not to overestimate the benefits of studying abroad.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

In addition, the moderator results show that L2 proficiency level, the use of a language pledge, type of coursework, and target language studied tend to moderate the observed effects, while other features exert negligible, if any, influence on learner's linguistic complexity development. In particular, several findings are worthy of further discussion. First, there is a tendency that intermediate-level learners benefit more than advanced-level learners in complexity gains, which concurs with the results of a number of previous studies (e.g., Brecht & Robinson, 1995; Freed, 1995; Llanes & Muñoz, 2009). It is possible that advanced learners have reached a point of diminishing returns and therefore make less progress during a short stay (DeKeyser, 2014). By comparison, intermediate-level learners, who are equipped with basic language skills prior to their departure, may be able to take full advantage of their stay and absorb as much linguistic knowledge as they could from their immediate exposures.

Second, the results suggest that studying abroad for a few weeks to one or two semesters does not make a remarkable difference in linguistic complexity, which is in agreement with Ortega's (2003) synthesis that a minimum of one year of college-level instruction is needed for substantial changes in the syntactic complexity of L2 writing. Therefore, the development of language complexity abroad should not be understood in a linear way.

Furthermore, learners in the three studies that implemented a language pledge showed a clear advantage ($d = 1.27$) over those in studies that did not ($d = 0.3$). However, given that only three studies have reported using a language pledge in the program, the results may not be generalizable. This moderator, however, is closely associated with specific language programs (i.e., Mandarin and Spanish) in which the learners, at least within this limited sample, significantly outperform learners in SA programs of other languages. A closer examination of these studies on Chinese/Spanish language complexity development reveals that most of them

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

featured high intensity of instruction, rich language activities, and a high degree of involvement of the language instructors. For example, Kim et al. (2014) investigated a Chinese SA program with an average reported language instruction of 20 hours per week, with a focus on learning vocabulary, grammar, and culture from language textbooks. Zalbidea and Martin (2016) examined a 5-week intensive Spanish program that required at least 28 to 32 hours of exposure to the target language every week, which included three content courses, in addition to fieldwork and conversation exchanges. Compared to the rest of the studies that have reported average hours of instruction between 8 to 15 hours per week (e.g., Serrano, Llanes, & Tragant, 2016; Pérez-Vidal & Barquin, 2014; Deng et al., 2010), these programs are significantly more intensive and involve extra curriculum activities, research activities, and out-of-class interactions. While such intensive SA programs help learners improve language skills efficiently, it is important to acknowledge that such intensity may not suit all learners. After all, studying abroad is not only about learning the target language; participants also learn to negotiate their identity and to discover and appreciate the target cultures in their encounters.

It is important to acknowledge that these features are likely to interact with each other. For example, as noted above, most advanced-level learners were enrolled in content courses and did not sign a language pledge, whereas intermediate-level learners were mostly enrolled in language courses in programs that required a pledge. Nevertheless, the analyses provided some rudimentary evidence of potential interaction among the moderating variables. Further research is needed in order to tease out the covariance in the correlated variables that contribute to the observed effects.

Finally, the reporting practices in this area of SA research require standardization, since a number of variables were inappropriately reported or loosely defined. However, what constitutes

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

high quality of reporting remains to be firmly established. For example, although L2 proficiency was largely reported across studies, the evaluation standards were seldom mentioned. Moreover, effect sizes and standard deviations were rarely reported in association with the means reported. The lack of consistency and transparency in reporting practices would hinder both practitioners and researchers from contextualizing and interpreting the findings, comparing and contrasting the results across studies, and replicating or meta-analyzing the existing research. The current study, therefore, calls for better reporting practices in study abroad research and proposes a concrete guideline for future studies (see Appendix B of the Supplementary Data File online). Additionally, researchers should also refer to the basic reporting practice in the APA guidelines as well as those proposed by Plonsky (2013, 2014) and Norris, Plonsky, Ross, and Schoonen (2015) to achieve maximum transparency.

Some limitations need to be acknowledged. First, at the current early stage in this research domain, the present findings are constrained by a limited sample size and inconsistent reporting practices. As a result, some potential moderators could not be examined as thoroughly as one may wish, such as the intensity of instruction, language contact, and the types of tasks. In addition, considering some learners were more balanced bilinguals (e.g., Spanish/Catalan), while some were emerging bilinguals who just started to learn a second language abroad, it would be worth investigating their L1 backgrounds and language repertoire in relation to interlanguage complexity development, had the language background was reported more consistently with further detail in the original studies. Another limitation concerns the completeness of sampling. Results in Jensen and Howard (2014) were only partially included in this analysis because the descriptive data for two measures (clauses per AS unit & words per clause) could not be retrieved. One study (Tavakoli, 2018) was excluded from the analysis because it was published

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

after the research period. Furthermore, the interactions between study features should be further addressed to better disentangle the contribution of each moderator. Finally, as Tullock and Ortega (2017) have argued, most study abroad research has been framed in monolingual terms. As a result, it still remains to be addressed how language complexity development is influenced by the multilingual contexts for SA or the multilingual demographics of the participants.

7. Conclusion

This study was designed to explore the overall effects of study abroad on learners' language complexity changes as well as to identify some of the variables that may potentially moderate such effects. Thirty samples in 28 studies published between 2003 and 2017 were reviewed and meta-analyzed. Overall, learners' interlanguage complexity was seen to increase after short-term study abroad (i.e., for periods that ranged from five weeks to eight months), although the improvement was modest. Larger complexity gains over time tended to be seen among learners who: (a) were identified to be at the intermediate level of proficiency, (b) were enrolled in a language course during SA, (c) followed a language pledge, or (d) studied Mandarin Chinese. It was found that some aspects of complexity tend to develop better than others after SA, such as lexical diversity. Therefore, fine-grained investigation of the various constructs and dimensions associated with linguistic complexity should be prioritized in future research. This meta-analysis also identified several areas in reporting practices that need improvements and proposed a checklist that researchers in this domain may find helpful.

Several promising lines of research have emerged from the current review. An important lesson is that future studies should be more cautious in assessing and reporting L2 proficiency by investigating the potential interactions between L2 proficiency and contextual features. It would

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

also be worthwhile to explore advanced learners' complexity development from a systemic functional perspective and take into account the influence of genre and task (Byrnes, 2014; Ryshina-Pankova, 2015). In addition, the potential benefit of a language pledge should be further investigated in order to unravel the extent to which it facilitates language learning in an SA context (García-Amaya, 2012). Finally, given that the current study showed a significant advantage in all the four Mandarin programs included the 30-study sample, future studies should explore whether this pattern is related to the nature of target language or to a shared instructional approach followed by Mandarin SA programs.

Finally, the study sheds light on some practice issues. For example, learners should evaluate program features and personal goals when studying abroad, and that program directors should assess the pros and cons brought about by the implementation of certain features such as the language pledge within a program. However, the decision with regard to whether or not to study abroad, or to build a new feature in the program, should not be solely driven by learning outcomes. Some invisible aspects should also be valued, such as personal experiences and the acquisition of global perspectives. Nevertheless, the focus of the present meta-analysis is outcomes with regard to linguistic complexity, with the hope that this synthetic effort affords some insights that can benefit both learners and practitioners engaged in language learning during studying abroad.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

References

Barquin, E. L. (2012). *Writing development in a study abroad context* (Doctoral dissertation).

Retrieved from <https://www.tdx.cat/bitstream/handle/10803/96196/teb.pdf?sequence=1>.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). Publication bias. *Introduction to meta-analysis*, 277-292.

Brecht, R. D., & Robinson, J. L. (1995). On the value of formal instruction in study abroad. *Second Language Acquisition in a Study Abroad Context*, 318-334.

Brecht, R. D., Davidson, D., & Ginsberg, R. (1990). The empirical study of proficiency gain in study abroad environments of American students of Russian. In *American contributions to the VII International Congress of MAPRIAL* (pp. 123-152). Washington, DC: American Council of Teachers of Russian.

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26, 42-65.

Bulté, B., & Housen, A. (2012). *Defining and operationalising L2 complexity* (pp. 21-46). Amsterdam/Philadelphia, PA: John Benjamins.

Byrnes, H. (2014). Theorizing language development at the intersection of ‘task’ and L2 writing: Reconsidering complexity. In H. Byrnes, & R.M. Manchón (Eds.), *Task-based language learning—Insights from and for L2 writing* (pp. 79–106). Amsterdam: John Benjamins.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26(2), 227-248.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, 26(2), 153-171.
- Cooper, H. (2016). *Research synthesis and meta-analysis: A step-by-step approach*. Sage Publications.
- DeKeyser, R. M. (2014). *Research on language development during study abroad* (pp. 313-326). Amsterdam: John Benjamins.
- DeKeyser, R. M. (2007). Study abroad as foreign language practice. *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*, 208-226.
- Deng, X., Cheng, L. K., Varaprasad, C., & Leng, L. M. (2010). Academic writing development of ESL/EFL graduate students in NUS. *Reflections on English Language Teaching*, 9(2), 119-138.
- Dewey, D. P., Bown, J., Baker, W., Martinsen, R. A., Gold, C., & Eggett, D. (2014). Language use in six study abroad programs: An exploratory analysis of possible predictors. *Language Learning*, 64(1), 36-71.
- Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 15, 127-148.
- Di Silvio, F., Donovan, A., & Malone, M. E. (2014). The effect of study abroad homestay placements: Participant perspectives and oral proficiency gains. *Foreign Language Annals*, 47(1), 168-188.
- Du, H. (2013). The development of Chinese fluency during study abroad in China. *The Modern Language Journal*, 97(1), 131-143.
- Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics*, 30(4), 474-509.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System, 80*, 199-211.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59-84.
- Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research, 34*(1), 67-101.
- Freed, B., Dewey, D., Segalowitz, N., & Halter, R. (2004). The Language Contact Profile. *Studies in Second Language Acquisition, 26*(2), 349-356.
- Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition, 26*(2), 275-301.
- Freed, B. F. (1995). What makes us think that students who study abroad become fluent. *Second Language Acquisition in a Study Abroad Context, 9*, 123-148.
- Garcia-Amaya, L. J. (2012). *Second language fluency and cognition: The study of Spanish second language development in an overseas immersion program and an at-home foreign language classroom* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (3509912).
- Godfrey, L., Treacy, C., & Tarone, E. (2014). Change in French second language writing in study abroad and domestic contexts. *Foreign Language Annals, 47*(1), 48-65.
- Grey, S., Cox, J. G., Serafini, E. J., & Sanz, C. (2015). The role of individual differences in the study abroad context: Cognitive capacity and language development during Short-Term intensive language exposure. *The Modern Language Journal, 99*(1), 137-157.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Hardison, D. M. (2014). Changes in second-language learners' oral skills and socio-affective profiles following study abroad: A mixed-methods approach. *Canadian Modern Language Review*, 70(4), 415-444.
- Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, 38(2), 163-175.
- Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.
- Huensch, A., & Tracy-Ventura, N. (2017). Understanding second language fluency behavior: The effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Applied Psycholinguistics*, 38(4), 755-785.
- Isabelli, C. A. (2001). *The impact of a study-abroad experience on the acquisition of L2 Spanish syntax: The null subject parameter* (Published doctoral dissertation). Retrieved from: <http://hdl.handle.net/2142/86111>.
- Jensen, J., & Howard, M. (2014). The effects of time in the development of complexity and accuracy during study abroad: A study of French and Chinese learners of English. *EUROSLA Yearbook*, 14(1), 31-64.
- Juan-Garau, M. J., & Pérez-Vidal, C. P. (2007). The effect of context and contact on oral performance in students who go on a stay abroad. *Vigo International Journal of Applied Linguistics*, (4), 117-134.
- Karlin, O. (2015). *Assessing English Environment Personality and its role in oral proficiency* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (3745833).

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Kim, J., Dewey, D. P., Baker-Smemoe, W., Ring, S., Westover, A., & Eggett, D. L. (2015). L2 development during study abroad in China. *System*, 55, 123-133.
- Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. Springer.
- Knoch, U., Rouhshad, A., & Storch, N. (2014). Does the writing of undergraduate ESL students develop after one year of study in an English-medium university? *Assessing Writing*, 21, 1-17.
- Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, 28, 39-52.
- Krashen, S. D., Long, M. A., & Scarcella, R. C. (1979). Age, rate and eventual attainment in second language acquisition. *TESOL Quarterly*, 573-582.
- Lara, R., Mora, J. C., & Pérez-Vidal, C. (2015). How long is long enough? L2 English development through study abroad programmes varying in duration. *Innovation in Language Learning and Teaching*, 9(1), 46-57.
- Leonard, K. R., & Shea, C. E. (2017). L2 Speaking Development During Study Abroad: Fluency, Accuracy, Complexity, and Underlying Cognitive Factors. *The Modern Language Journal*, 101(1), 179-193.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.
- Llanes, À., Tragant, E., & Serrano, R. (2015). Examining the role of learning context and individual differences in gains in L2 writing performance: The case of teenagers on an intensive study-abroad program. *The Language Learning Journal*, 1-18.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Llanes, À., & Serrano, R. (2014). The effectiveness of classroom instruction 'at home' versus study abroad for learners of English as a foreign language attending primary school, secondary school and university. *The Language Learning Journal*, 1-13.
- Llanes, À., & Muñoz, C. (2013). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning*, 63(1), 63-90.
- Llanes, À., Tragant, E., & Serrano, R. (2012). The role of individual differences in a study abroad experience: The case of Erasmus students. *International Journal of Multilingualism*, 9(3), 318-342.
- Llanes, À., & Serrano, R. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española de Lingüística Aplicada*, 24, 95-110.
- Llanes, À. (2011). The many faces of study abroad: An update on the research on L2 gains emerged during a study abroad experience. *International Journal of Multilingualism*, 8(3), 189-215.
- Llanes, A., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37(3), 353-365.
- Longcope, P. D. (2003). *What is the impact of study abroad on L2 learning? A descriptive study of contexts, conditions, and outcomes* (Doctor dissertation). Retrieved from ProQuest Dissertations & Theses Global. (305308689).
- Long, M. H. (1990). Maturation constraints on language development. *Studies in Second Language Acquisition*, 12(3), 251-285.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36-62.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, 46(4), 610-641.
- Norris, J. M., Plonsky, L., Ross, S. J., & Schoonen, R. (2015). Guidelines for reporting quantitative methods and results in primary research. *Language Learning*, 65, 470-476.
- Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578.
- Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching*. John Benjamins Publishing.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29(4), 557-581.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
- Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, 30, 85-110.
- Pérez-Vidal, C. (2017). Study abroad and ISLA. *The Routledge handbook of instructed second language acquisition*, 338-360.
- Pérez-Vidal, C. (2015). Practice makes best: Contrasting learning contexts, comparing learner progress. *International Journal of Multilingualism*, 12(4), 453-470.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Pérez-Vidal, C., & Barquin, E. (2014). Comparing progress in academic writing after formal instruction and study abroad. *Language Acquisition in Study Abroad and Formal Instruction Contexts*, 217-234.
- Plonsky, L. (2014). Study quality in quantitative L2 research (1990-2010): A methodological synthesis and call for reform. *Modern Language Journal*, 98, 450-470.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912.
- Plonsky, L. (2013). Study quality in SLA. *Studies in Second Language Acquisition*, 35(4), 655-687.
- Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, 61(4), 993-1038.
- Robson, G. G. (2015). The Relationship between WTC and Oral Proficiency Measurements in the Study Abroad Context. *International Education Studies*, 8(12), 56.
- Romanova, N. (2015, March). *Second Language Writing Development in a Study-Abroad Context*. Paper presentation at the American Association of Applied Linguistics, Toronto, Canada.
- Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing*, 29, 51-63.
- Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcomes: The SALA project. *Second Language Acquisition in Study Abroad and Formal Instruction Contexts*. Amsterdam/Philadelphia: John Benjamins.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Sasaki, M. (2009). Changes in English as a foreign language students' writing over 3.5 years: A sociocognitive account. *Writing in Foreign Language Contexts: Learning, Teaching, and Research*, 49-76.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173-199.
- Serrano, R., Llanes, À., & Tragant, E. (2016). Examining L2 development in two short-term intensive programs for teenagers: Study abroad vs. "at home." *System*, 57, 43-54.
- Serrano, R., Tragant, E., & Llanes, À. (2014). Summer English courses abroad versus 'at home'. *Elt Journal*, 68(4), 397-409.
- Serrano, R., Tragant, E., & Llanes, À. (2012). A longitudinal analysis of the effects of one year abroad. *Canadian Modern Language Review*, 68(2), 138-163.
- Serrano, R., Llanes, À., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System*, 39(2), 133-143.
- Shi, Z., & Ning, J. (2016, September). *A longitudinal study on second language development in study abroad context*. Paper presentation at the 35th Second Language Research Forum, the Teachers College of Columbia University. NY.
- Skehan, P. (2003). Task-based instruction. *Language Teaching*, 36(1), 1-14.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Storch, N. (2009). The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*, 18(2), 103-118.

- Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.
- Storch, N., & Tapper, J. (2009). The impact of an EAP course on postgraduate writing. *Journal of English for Academic Purposes*, 8(3), 207-223.
- Tavakoli, P. (2018). L2 development in an intensive Study Abroad EAP context. *System*, 72, 62-74.
- Tulloch, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, 71, 7-21.
- Wilson, D. (2001). Effect Size Determination Program. University of Maryland, College Park.
- Wright, C., & Zhang, C. (2014). *Examining the effects of study abroad on Mandarin Chinese language development among UK university learners*. Newcastle and Northumbria Working Papers in Linguistics, 20, 67-83.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. University of Hawaii Press.
- Xiao, F. (2015). Adult second language learners' pragmatic development in the study-abroad context: A review. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 25, 132-149.
- Xu, Y. (2016, September). *Changes in complexity, accuracy, and fluency of L2 Chinese writing during study abroad*. Paper presentation at the 35th Second Language Research Forum, the Teachers College of Columbia University. NY.
- Yang, J. (2016). The effectiveness of study-abroad on second language learning: A meta-analysis. *Canadian Modern Language Review*, 72(1), 66-94.
- Zaytseva, V. (2016). *Vocabulary acquisition in study abroad and formal instruction: an investigation on oral and written lexical development* (Unpublished doctoral dissertation). Universitat Pompeu Fabra, Barcelona, Spain.

Xu, Y. (2019). Changes in interlanguage complexity during study abroad: A meta-analysis. *System*, 80, 199-211.

Zalbidea, J., & Martin., A. (2016, September). *Towards a more multidimensional analysis of accuracy and complexity development during study abroad*. The 35th Second Language Research Forum, the Teachers College of Columbia University. NY.