

DEVELOPMENT OF COMPLEXITY, ACCURACY, AND FLUENCY IN HIGH SCHOOL STUDENTS' WRITTEN FOREIGN LANGUAGE PRODUCTION

Bouchaib Benzehaf

Chouaib Doukkali University, El Jadida, Morocco

bbenz841@gmail.com

Abstract: The present study aims to longitudinally depict the dynamic and interactive development of Complexity, Accuracy, and Fluency (CAF) in multilingual learners' L2 and L3 writing. The data sources include free writing tasks written in L2 French and L3 English by 45 high school participants over a period of four semesters. CAF dimensions are measured using a variation of Hunt's T-units (1964). Analysis of the quantitative data obtained suggests that CAF measures develop differently for learners' L2 French and L3 English. They increase more persistently in L3 English, and they display the characteristics of a dynamic, non-linear system characterized by ups and downs particularly in L2 French. In light of the results, we suggest more and denser longitudinal data to explore the nature of interactions between these dimensions in foreign language development, particularly at the individual level.

Keywords: CAF, proficiency, interaction.

INTRODUCTION

In the quest for insights into language development, researchers have suggested different tools to measure learners' language development. At first, they borrowed length-based measures from the field of first language (L1) acquisition, the most common ones being the mean length of particular structures (Norris & Ortega, 2009) which have been widely adopted in the second and third language acquisition research enterprise. But these measures

proved to be fraught with problems. For instance, beginner learners rely much on rote-learned formulaic sequences to complement their nascent grammar (Myles, 2012), and, therefore, perceived longer production of such structures which gives false impressions of increased proficiency. To solve the problem, Larsen-Freeman (1978) proposed an Index of Development which was further operationalized as measures of Complexity, Accuracy, and Fluency (CAF). CAF measures were meant to indicate the level of a learner's proficiency but this index, in turn, is not without problems as proficiency is hard to pin down to a definition.

Although researchers do not agree on definitions of proficiency in a language, it can generally be claimed that it refers to a person's ability to use the language in an appropriate way in different contexts either in writing or in speaking. Writing and speaking are two modes that can represent a person's proficiency level. Thus, studies targeting language development should rely on "concrete realizations"; that is, what learners can do in their language productions (Buisse & De Clercq, 2014). To meet this end, CAF measures have been introduced as qualitative dimensions that capture the development of language (Housen, Kuiken & Vedder 2012a). The present study focuses on CAF dimensions in the written mode of foreign language production, namely in L2 French and L3 English in high school.

The study is motivated by a noticeable scarcity of research comparing L2 and L3. A review of the literature shows that research into language acquisition has almost exclusively been concentrating on L1 and L2 development. Research into L3 development is still a "very young" field and little has been done to observe L3 development (Jessner, 2008). There is also a scarcity of studies which holistically take into account learners' L2 and L3 development (see Kobayashi & Rinnert, 2013). In light of the qualitative differences between second language acquisition and third language acquisition, the present study is intended to inquire into learners' L2 and L3 writings simultaneously via analysis of CAF dimensions.

The purpose of the study is twofold: first, to examine the nature of development of CAF dimensions in written foreign language production in high school students in English and French, and secondly, to explore the process of interaction between the three dimensions.

LITERATURE REVIEW

The debate on the nature of the interactions in the CAF triad carries on into empirical research. For instance, VanPatten (1990) investigated learners' capacity to pay attention to both form and content simultaneously, and indicated that comprehension levels went down when learners had to pay attention to both form and content, and that this was even more problematic in the framework of second language learning. Based on these findings, Skehan and Foster argued that complexity and accuracy compete for attention and that the learner is incapable of attending to more than one area of language, particularly if the task is cognitively difficult and demanding. Thus, concurrent attention to different areas of L2 is considered difficult.

Verspoor, Lowie, and van Dijk (2008) conducted a longitudinal study (over a period of 3 years) observing the academic writing of an advanced learner of English. The researchers reported that the sentence length measure and the type token ratio did not develop concurrently and that there was a competitive relationship between them, pointing to an absence of the ability to allocate attentional resources equally on the part of the language learner. The study also showed that the learner's language development was characterised by much variability and non-linearity and thus a dynamic nature.

Adopting a case study approach, Ferrari (2012) longitudinally observed one participant's language development. In line with the previous study, Ferrari reported traces of trade-off effects between complexity and accuracy at least in a certain time period. Another study which also lent support to the trade-off hypothesis was conducted by Myles (2012). This study reported interactions not only

among the CAF dimensions but also between the triad and the learners' communicative adequacy.

Robinson (1995), Robinson (2007) and Gilabert (2007) compared cognitively simple and complex interactive performances. Using simple here-and-now tasks and difficult there-and-then tasks, these studies looked into the effects of increased task difficulty on L2 task performance. The results of these different studies indicated that the difficult task did promote accuracy and complexity at a significant level, thereby confirming the cognition hypothesis.

Spoelman and Verspoor (2010) also investigated the nature of interaction between accuracy rates and complexity measures in a Dutch student learning Finnish for a lengthy period of 3 years. The researchers observed that accuracy rates went up and down in early stages but settled down as the system relaxed. They also noted that interaction between accuracy and complexity was not stable and that it changed over time, suggesting a dynamic system that neither supports the trade-off hypothesis nor the cognition hypothesis. Another study which also disconfirmed both hypotheses was done by Gunnarson (2012). She found neither competition between complexity and accuracy nor any significant interactions between syntactic complexity and fluency.

Vyatkina (2012) examined the longitudinal and cross-sectional development of lexicogrammatical complexity in learners' written production at college level. The findings confirmed that length-based complexity measures correlated well with proficiency levels. Vyatkina reported a rising trend in the development of lexico-grammatical complexity measures. However, significant variability at the individual level was also reported, with each participant's developmental pattern being highly dynamic and idiosyncratic.

Similarly, Polat and Kim (2013) looked into the dynamics of complexity and accuracy in L2 development of a Turkish immigrant in the USA. They conducted a longitudinal observation of the development of CAF constructs in a naturalistic context, not in a classroom. The findings showed that while their participant's

syntactic complexity and lexical diversity developed well, accuracy did not. The participant's interlanguage was, thus, found to be highly variable.

Using a case study approach, Rosmawati (2013) investigated the nature of interactions between complexity and accuracy in L2 writing. The study targeted an advanced female L2 learner's academic writing during her postgraduate study. The results suggested that complexity and accuracy measures showed the characteristics of a dynamic system. Also, their development was highly variable and non-linear although a moderate negative association was observed between complexity and accuracy which did not reach a statistically significant level. It was concluded that the developmental patterns of complexity and accuracy are highly dynamic and idiosyncratic.

Yang and Sun (2015) investigated the development of fluency, accuracy and complexity from the perspective of the dynamic systems theory in 5 learners over a period of one academic year. The study was centered on the development of CAF constructs across L1 Chinese, L2 English and L3 French writing. Results showed that the developmental patterns of CAF in multilingual learners' writing did not follow one clear trajectory path as they were non-linear, recurrent and quite chaotic particularly at the individual level. However, CAF constructs were also integratively and interactively correlated with each other in the participants' writing over time.

The divergent results obtained in different studies indicate the multidimensional facets of L2 development. This situation underscores also the fact that CAF constructs are not straightforward but highly dynamic and complex constructs. Norris and Ortega (2009) indicated that CAF is a dynamic and interrelated set of constantly changing subsystems, and that only longitudinal observations can capture the nature of the CAF development and interactions. Hence, the present study attempted to longitudinally observe students' foreign language development over a period of 4 semesters.

THEORETICAL FRAMEWORK

CAF Measures

It is widely believed that L2 proficiency constructs are multi-componential in nature, and that the notions of complexity, accuracy and fluency can satisfactorily capture their principal dimensions (e.g. Skehan 1998; Ellis 2003). Though they do not constitute a theory in themselves, complexity, accuracy and fluency (henceforth CAF) have figured as major research variables in research into acquisition of second and third language. They have figured as dimensions for describing oral and written performance and for measuring progress in language learning. As such, they have succeeded in passing as a conceptual framework within which language development can be benchmarked.

CAF have been suggested as dimensions that describe language performance. They are usually employed to determine variation among individual students. Researchers agree on the validity and usefulness of these constructs, but they do not agree as to their operationalization. According to researchers, the best measures we can adopt to investigate, distinguish between individual students, and track language development are those that adequately represent their underlying constructs and also allow for different levels to clearly come into view. The literature shows that fluency and accuracy were constructs utilized to investigate the development of L2 proficiency in classroom contexts in the 1980's.

Brumfit (1984) distinguished fluency-based activities from accuracy-based activities stating that the former increase spontaneous oral L2 production and the latter focus on form. Fluency may also be defined as "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen 2005, p. 139). In other words, it is the ability to process language with native-like speed. Accuracy refers to the degree of conformity to certain norms. More specifically, it means use of grammatically correct linguistic forms, or the ability to produce error-free speech. In the 1990's, Skehan added

the dimension of complexity which he incorporated in a CAF-based L2 model.

Complexity retains multiple meanings. One such operationalization of complexity refers to use of more elaborate and varied language (Ellis, 2003) while another one refers to the increase over time of structural complexity (use of complex grammatical structures) (Spada & Tomita 2008, p. 229). Bergman and Abrahamsson (2004, p. 611) proposed a three-level scale to describe the syntactic structures in L2. At the beginner level, sentence structures are characterized by simplicity and only basic linking elements (such as and, but, then) are present. At the intermediate level, complexity begins to grow with variation in the use of linking elements and the appearance of dependent clauses and non-finite clauses in the learners' writing. Complexity further increases at the advanced level as language production becomes rich in different sentence structures which consist of multiple dependent and non-finite clauses.

The Trade-off Hypothesis Vs the Cognition Hypothesis

Researchers have also studied the interaction among CAF constructs. Considering the issue of interdependency between CAF measures, Skehan came up with his Trade-off Hypothesis (also known as the Limited Attentional Capacity model) which states that the dimensions are interdependent such that increased performance in one area may occur at the expense of performance in the other areas. In other words, working memory, which is responsible for attention allocation, is under pressure when it is faced with multiple stimuli. Therefore, and due to limited attentional capacity (Skehan, 1996, 2009; Skehan & Foster, 2001), attending to one particular area may take attention away from the other two areas.

Skehan and Foster argue that as L2 learners focus on the communicative goal, prioritizing meaning over form (VanPatten 1990), the attention that is left for form is distributed between complexity, accuracy, and fluency. Particularly cognitively complex

tasks put L2 learners under attentional pressure most obviously between linguistic complexity and accuracy (Skehan 1996, 2009; Skehan & Foster 2001).

In contrast to Skehan's Trade-off Hypothesis, Robinson (2001, 2005) proposes the Cognition Hypothesis stating that not every complex task necessarily causes trade-off effects. The fundamental pedagogic claim of the Cognition Hypothesis is that the more cognitively and functionally demanding the task is, the more encouraged the learner is to produce more complex and more accurate language production. Such a claim is underpinned by the idea that L2 learners can rely on multiple pools of attention because different processes may draw on various attentional pools. Thus, concurrent attention to different areas of L2 is considered not only possible, but also natural.

METHOD

Research Design

The study is a quantitative investigation, based on a longitudinal observation of a 45 participants' written production over two academic years. It examines the development of the constructs of complexity, fluency and accuracy. The data are collected and coded using a quantitative approach and submitted to statistical analyses to answer the research questions.

Participants and Setting

The participants are 45 high school students tracked over two years, first and second year in high school. They are 25 girls and 20 boys, and they are all students in 6 November high school situated in Ouled Frej in El Jadida. Their age range is between 16 and 18. They studied in their first year and passed to second year which they also completed successfully. Some of these students were introduced to English in their last year of primary education but with no more than two hours a week mostly dedicated to oral communication. In high school, all the students started studying English with three hours a

week. By contrast, they have completed seven years of French education, with an average of 5 hours a day. Hence, their French is supposed to be stronger than their English.

Data

The data were collected twice a year, at the end of each semester (2014-2015 and 2015-2016). The rationale for choosing to collect the data at the end of every semester was underpinned by the assumption that students needed at least one semester to be able to produce a writing task in English as they only started studying it in first year of high school. Therefore, the corpora consisted of 4 different pieces in French and in English and the approach was a time-series one which allowed for benchmarking the development of complexity, fluency and accuracy. The topics across L2 and L3 writing were the same. The topics were (a film that everyone should see, where and how you spent your latest holidays, how you spend time, a book that everyone should read). Albeit seemingly different, the topics unanimously fall under the umbrella of the genre of personal narrative essays. The rationale behind such kind of uniformity in genre is to make the comparative inquiry of the longitudinal written data of distinct topics feasible.

Sampling and Coding

CAF indices have figured in much research as important criteria to assess learners' written and oral productions. Thus, the data were coded for complexity, fluency and accuracy constructs. Given that the participants range from beginner learners to pre-intermediate, and given that language learners learn to use cognitively demanding material rather late in their learning process, the coding was simplified. Thus, for complexity which can be broken down into length, amount of embedding, and frequency of certain sophisticated structures (e.g. non-finite clauses), we considered only the quantitative aspect of the definition, namely, the length of the T-unit excluding the qualitative aspects (amount of embedding, and

frequency of certain sophisticated structures). For fluency, we counted the number of T-units written by the participants. And for accuracy, we calculated error free T-units per total number of T-units ratio. The data were coded as follows:

Table 1: CAF coding and measurement

Complexity	Mean length of T-units
Fluency	Total number of T-units
Accuracy	Error free T-units per total number of T-units ratio

The choice of the T-unit (defined as the minimal terminable unit consisting of one main clause and any subordinate clauses and non-clausal units or sentence fragments attached to it) as a unit of measurement of learner language is empirically motivated. It is easily computable, and hence allows for high inter-rater reliability. It also does not pose punctuation problems as sentence boundaries are important. Lastly, it best captures linguistic maturity by charting obvious increases in length and complexity.

Inter-coder reliability

The participants' written texts were submitted to two coders, the author as coder 1, and a French teacher with 6 years of teaching experience as coder 2 who was given coding information prior to doing the coding. I, the author coded the English texts and the French teacher coded the French texts. However, initially, we each coded 10 same French texts to check for inter-rater reliability which reached 0.92. Then, we discussed discrepancies, and attained 100% agreement. We finally plotted the quantitative data in Microsoft Excel charts and transformed them into line graphs to allow for visualizing the complex and dynamic development of CAF in the participants' L2 and L3 writing.

RESULTS

The development of complexity, fluency and accuracy measures in the *participants'* writing in the observation period showed a great deal of variability. The data collected were analysed and the results are presented below.

Development of CAF constructs in foreign language production

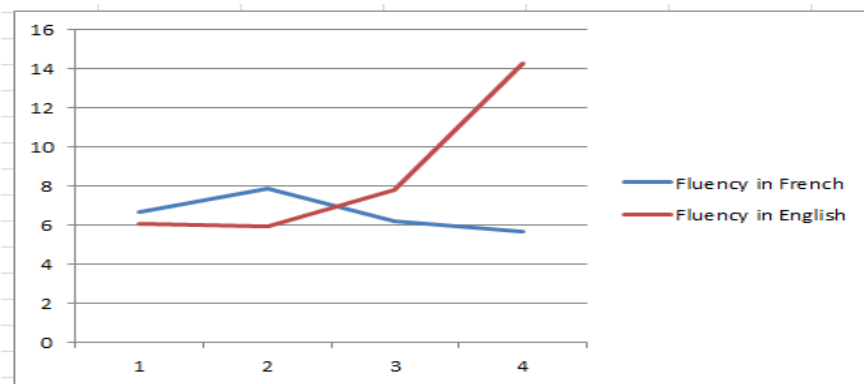


Figure 1: Group averages in fluency over 4 semesters

Figure 1 indicates that group averages of fluency in French and English writing underwent entirely different developmental paths. Over the 4-semester period, French fluency first increased sharply (from 6,3 T-units in semester 1 to 7,9 in semester 2), and then decreased substantially (from 7,9 in semester 2 to 6,1 in semester 3) to below 6 in semester 4. In contrast, the level of fluency in English started below that in French (6 T-units) and remained almost stable in semester 2 (5,9 T-units), but then, it increased sharply in semester 3 (an average of 7,8 T-units) and continued to grow more sharply in semester 4 reaching an average of 14,2 T-units.

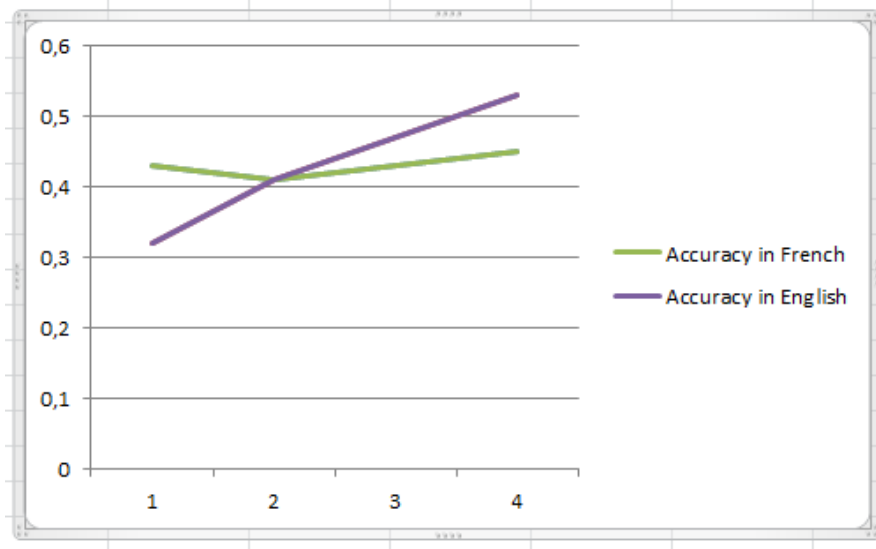


Figure 2: Group averages in accuracy over 4 semesters

Regarding development of accuracy in written foreign language production, the trajectory is slightly different from that in fluency. As figure 2 above shows, accuracy as represented by error-free T-units to total number of T-units ratio was 0,44 in French in semester 1; then, it decreased slightly to 0,42 in semester 2 and increased again to reach 0,44 in semester 3 and finally 0,46 in semester 4. By contrast, in English the trend was different. Accuracy in English was below that in French in semester 1 (0,32) and then it increased to 0,42 in semester 2. It continued to rise in semester 3, reaching 0,46, and again in semester 4 scoring 0,53.

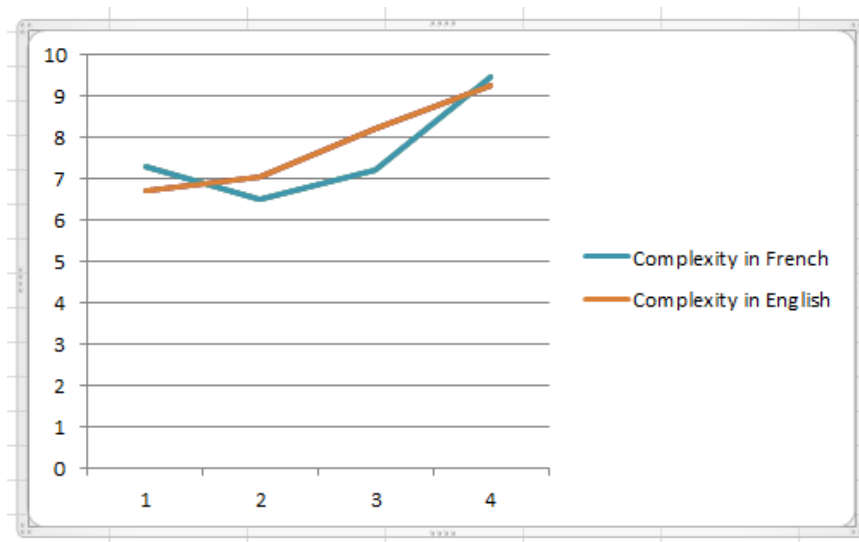


Figure 3: Group averages in complexity over 4 semesters

The figure above benchmarks the development of complexity in written foreign language production as measured by mean length of T-units. It is evident that complexity levels in French started higher than complexity levels in English. In semester 1, it was 7,3 but it went down in semester 2 scoring 6,5. In semester 3, it started rising once again to reach 7,2 and finally 9, 4 in semester 4. In English, the trajectory was slightly different. The mean length of T-units was 6,7 in semester 1 and it rose to 7 in semester 2. It continued to rise scoring 8,2 in semester 3 and 9,2 in semester 4.

Interaction of CAF constructs in foreign language production

To observe the three constructs and examine how they interact with each other across each language, we had to normalize the performance measures by recalculating the data to values from 0-1 so as to guarantee the comparability across the different constructs and represent all of them together within a single graph. Thus, we adjusted the values measured on different scales to a notionally common scale putting everything on a 0-100% scale by dividing each measure by the maximum value of that measure. Thus, we obtained the following new values in English written production:

Table 2: CAF values in English written production

Semester	Fluency in English	Accuracy in English	Complexity in English
1	0,42	0,60	0,72
2	0,41	0,79	0,76
3	0,54	0,86	0,89
4	1	1	1

Similarly, we obtained the following new values in French written production:

Table 3: CAF values in French written production

Semester	Fluency in French	Accuracy in French	Complexity in French
1	0,79	0,95	0,77
2	1	0,91	0,69
3	0,77	0,95	0,76
4	0,73	1	1

These new values enabled us to represent CAF constructs in one single graph for English as follows:

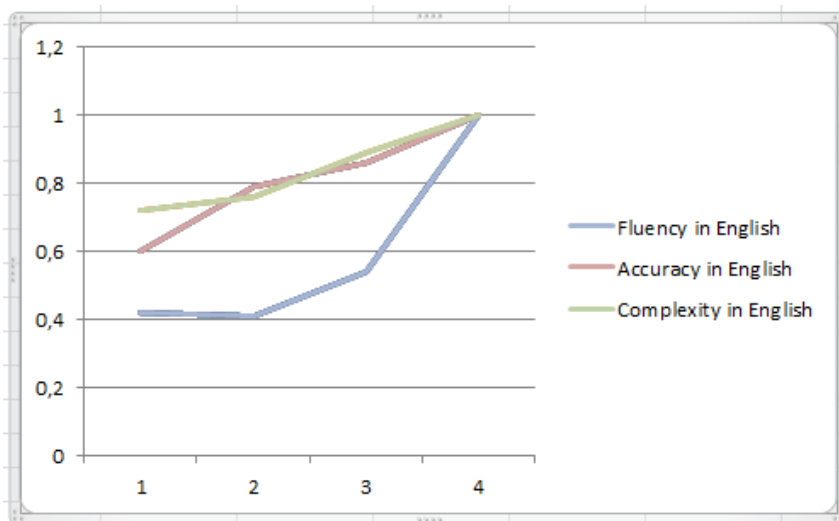


Figure 4: CAF in English written production

The plotted raw data points show that the three lines are moving in the same direction, i.e. they develop concurrently although not to the same extent.

We also represented CAF for French written production in one single graph as follows:

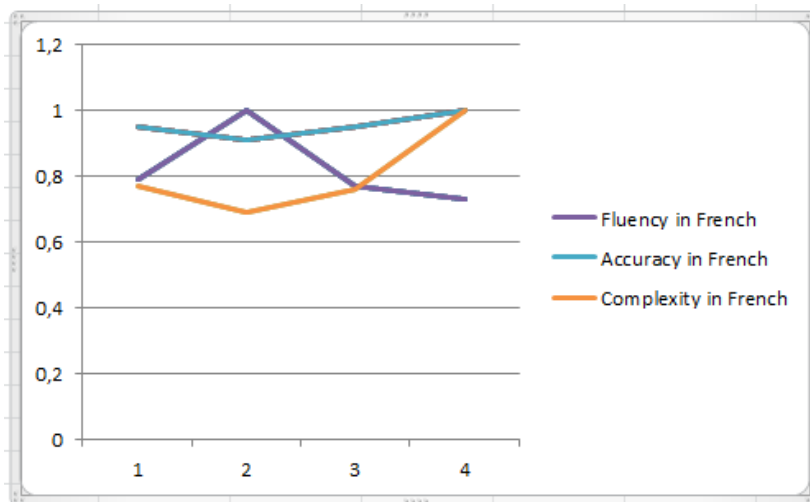


Figure 5: CAF in French written production

The plotted raw data points show that complexity and fluency are moving in opposite directions. As fluency increases, complexity decreases. Accuracy develops in the same direction as complexity but in opposite direction with fluency.

In addition, a correlation analysis was performed the result of which supported the existence of positive association in English between all constructs. Between fluency and accuracy, the correlation was statistically significant ($r = 0,826$ $p > .05$). Between fluency and complexity, it was even more significant ($r = 0,919$ $p > .05$), and also between accuracy and complexity with a significant value ($0,936$ $p > .05$).

In French, negative association was noted between fluency and accuracy. The correlation analysis supported such observation significantly ($-0,876$ $p > .05$). Negative association was also observed between fluency and complexity, and it was supported by the

correlation analysis, though not to a very statistically significant level ($r = -.698$, $p > .05$). However, the correlation was positive between accuracy and complexity at a statistically significant level ($0,962$ $p > .05$).

DISCUSSION

The results obtained from data analysis indicate that at the group level, the participants failed to show stable patterns in their L2 French writing development. They demonstrated neither general linear downward trends nor smooth upward trajectories development in terms of CAF analyzed. In reality, CAF in group learners' L2 French writing all developed in non-linear and dynamic fashions, with ups and downs from time to time. Further, the constructs measured suggested a supportive relationship between accuracy and complexity, thereby lending support to Robinson's Cognition hypothesis (1995) which states that the learner is encouraged to produce more complex and more accurate language production, particularly if the task is cognitively demanding. Fluency, however, appeared to move in opposite direction of accuracy and complexity, suggesting a complex interaction between the three constructs. This finding is in conflict with that obtained in Yang and Sun's study (2015) which suggested that the three constructs were integratively and interactively correlated with each other in their participants' writing over time. The present study showed correlation only between accuracy and complexity in French L2 writing.

Accuracy and complexity grew side by side to reach their peak in semester four, and the correlation was positive at a statistically significant level. However, rather a relationship of competitiveness appeared between accuracy and complexity on one part and fluency on the other part. While fluency goes up, accuracy and complexity go down and vice versa indicating that the participants could not attend to the three constructs concurrently. This finding is also consistent with the finding obtained in Verspoor et al.'s study (2008). These researchers have also reported that the measures do not develop

concurrently and that there is a competitive relationship between them. Verspoor concluded that the learner cannot allocate attentional resources equally.

In contrast with French, growth was more salient and persistent in L3 English in all three constructs marking an absence of competitiveness. All three constructs persistently increased over time, particularly fluency which reached its peak in semester four. Thus, CAF constructs were integratively and interactively correlated with each other in the participants' L3 writing over time, much in the same way that Yang and Sun (2015) reported about their participant's writing over time. This growth is also consistent with Jessner's model of multilingual development (Jessner, 2008) according to which multilingual learners' L3 undergoes constant increase. Jessner's model also accounts for the backsliding of proficiency in L2 French particularly in terms of fluency which was characterized by a steep decrease starting from semester two. According to Jessner (2008), the persistent growth of L3 occurs in sharp contrast to the decline of L2, resulting in a gradual attrition or loss of L2.

The participant multilingual learners in the present study are just taking up the study of English and instructional and learning contexts are expected to vary resulting in such discrepancy in terms of proficiency levels in the two languages. It is suggested in this context that English is taught in a more active and efficient way than French, though such a suggestion needs to be research based. Previous studies conducted in Morocco also showed that English is increasing at the expense of French. As early as 1991, Sadiqi reported an increase in the number of university graduates in English attributing it to the general policy adopted by both decision makers and educationalists in Morocco. Not to forget that French is also the language of the ex-coloniser for Moroccans, and thus it is regarded as a symbol of colonialism. By contrast, and according to Zouhir (2013), English is the only foreign language with no colonial overtones for Moroccans. English is also associated with opportunities in Moroccans' thinking and it is the language that allows them to go global. Hence, they have

positive attitudes to English and are motivated to learn it more than French. These facts are likely responsible for such apparent backsliding of L2 French and salient progress in L3 English over time.

Interestingly, the findings obtained in this study also suggest that factors exist which override Lenneberg's critical period hypothesis (CPH) (1967). This hypothesis posits that language acquisition is successful only if it occurs before cerebral lateralization is complete, thereby linking language acquisition with maturational constraints. In spite of the fact that the participant learners of L3 English in this study are beyond the critical period, they could display signs of effective learning of English. These learners, therefore, teach us that the statement that "the older one becomes, the more difficult acquisition is" is not that correct. This is in line with some studies conducted over the latest decades. For instance, Birdsong (2014) concludes that age of onset of learning additional languages and ultimate attainment levels are not straightforward. He also cites Singleton (2005) who explored the literature related to the CPH and concluded that "the CPH cannot plausibly be regarded as a scientific hypothesis" (Singleton, 2005, p. 280, quoted on p. 44). In another recent study which failed to confirm the CPH, Fei and Li-qin (2016) analyzed the effect of CPH on English teaching in China and determined that the influence of the CPH on second language acquisition and foreign language learning is still unclear.

CONCLUSION AND RECOMMENDATIONS

The present study set out to test the nature of development and interaction of CAF constructs in high school multilingual learners' L2 French and L3 English in Morocco. The study followed a longitudinal observation design over a period of four semesters. Detailed analysis of the quantitative data showed that the developmental patterns of CAF in multilingual learners' L2 French and L3 English writing did not follow the same trajectory. In French, the general trend was downward but in English it was upward with an absence of clear consistent linearity in either language. Particularly

in French, the development of CAF constructs was characterized by recurrent ups and downs, and by complex interactions. In English, the development trajectory was persistently upward but not at the same rate all through the observation period. The progress was sometimes fast and sometimes slow. Supportive relations between some measurements and competitive relations between other measures were evidenced in students' writing over time. Also, at different times, certain indices developed faster and more remarkable than others.

An important implication that can be drawn from this study is that multilingual development is indeed a dynamic and complicated process, which may provide us with insight into multilingual development. Besides, it was evident that CAF dimensions have the potential to provide a conceptual framework capable of capturing the dynamics of multilingual learners' language development.

However, the results are yielded from mean analysis of group learners, thereby sketching the dynamics of multilingual development from a collective perspective which disguises individual variations. Given that there are abundant individual differences in language acquisition, case study research is required which places particular stress on individual developmental aspects. Expanding the measures to include other aspects of each construct is likely to further uncover the active dynamism underpinning the behaviour of the constructs. Further, adding qualitative analysis to the quantitative findings will enrich the discussion regarding the development of foreign language production.

Lastly, another important area worthy of investigation is motivation, attitudes and instructional environment. Since the adolescent participants of this study showed that they can still learn additional languages successfully beyond the critical period (their L3 English was developing quite well in terms of the three CAF constructs), factors responsible for this success are worthy of attention and research.

REFERENCES

- Bergman, P. & Abrahamsson, T. (2004). Bedömning av språkfärdigheten hos andraspråkselever. I: Hyltenstam, K.& Lindberg (red.), *Svenska som andraspråk- i forskning, undervisning och samhälle*. Lund: Studentlitteratur. S. 597-626.
- Birdsong, D. (2014). The critical period hypothesis for second language acquisition: Tailoring the coat of many colors. In M. Pawlak & L. Aronin (eds.), *Essential Topics in Applied Linguistics and Multilingualism, Second Language Learning and Teaching* (pp. 43-50). Switzerland: Springer International Publishing.
- Brumfit, C.J. (1984). *Communicative methodology in language teaching: The roles of fluency and accuracy*. Cambridge: Cambridge University Press.
- Buyse, M. & De Clercq, B. (2014). Matching the data: Developing a multilingual corpus of language development. *JADT: 12es Journées internationales d'Analyse statistique des Données Textuelles*, 729-738.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.
- Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford: Oxford University Press.
- Fei, D. & Li-qin, Z. (2016). An analysis of critical period hypothesis in English teaching. *Sino-US English Teaching*, 13(2), 116-122.
- Ferrari, S. (2012). A longitudinal study of complexity, accuracy, and fluency variation in second language development. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 277-297). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Benzehaf, Development of Complexity, Accuracy, and Fluency in High School Students' Written Foreign Language Production

- Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics*, 45, 215-240.
- Gunnarson, C. (2012). The development of complexity, accuracy and fluency in the written production of L2 French. In A. Housen, F. Kuiken & I. Vedder (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Housen A., F. Kuiken & I. Vedder. (2012a). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken and I. Vedder (eds) *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA* (pp. 1-20). Amsterdam: Benjamins, 1-20.
- Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732-39.
- Jessner, U. (2008). A DST model of multilingualism and the role of metalinguistic awareness. *The Modern Language Journal*, 92, 270-283.
- Kobayashi, H., & Rinnert, C. (2013). L1/L2/L3 writing development: Longitudinal case study of a Japanese multicompetent writer. *Journal of Second Language Writing*, 22, 4-33.
- Larsen-freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12, 439-448.
- Lenneberg, E. (1967). *The biological foundations of language*. New York: John Wiley & Sons.
- Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 171-197).

Amsterdam/Philadelphia: John Benjamins Publishing Company.

- Myles, F. (2012). Complexity, accuracy, and fluency: The role played by formulaic sequences in early interlanguage development. In A. Housen, F. Kuiken & I. Vedder (eds.) *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 71-94). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Norris, J. M. & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30, 555- 578.
- Polat, B., & Kim, Y. (2013). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics*, 1-25. doi: 10.1093/applin/amt013.
- Robinson, P. (1995). Attention, memory and the 'noticing' hypothesis. *Language Learning*, 45, 283-331.
- Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (ed.) *Cognition and second language instruction* (287-318). Cambridge: Cambridge University Press.
- Robinson, P. (2003). Attention and memory during SLA. In C. M. Doughty & M. H. Long (eds.) *The handbook of second language acquisition* (631-78). Oxford: Blackwell.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, 43(1), 1-32
- Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching*, 45(3), 193-213.

Benzehaf, Development of Complexity, Accuracy, and Fluency in High School Students' Written Foreign Language Production

- Rosmawati, R. (2013). A case study in complexity and accuracy development in ESL academic writing: A dynamic perspective. The Asian Conference on Education. Osaka, Japan.
- Sadiqi, F. (1991). The spread of English in Morocco. *International Journal of the Sociology of Language*, 87, 99-114.
- Skehan, P. (1996). Second language acquisition research and task-based instruction". In J. Willis and D. Willis (Eds.), *Challenge and change in language teaching* (pp. 17-30). Oxford: Heinemann.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2009). Modeling second language development: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30(4), 510-532.
- Skehan, P. and P. Foster. (2001). Cognition and tasks. In P. Robinson (ed.) *Cognition and Second Language Instruction* (183-205). Cambridge: Cambridge University Press.
- Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, 49(1), 93-120.
- Skehan, P., & Foster, P. (2007). Complexity, accuracy, fluency and lexis in task-based performances: A meta-analysis of the Ealing research. In S. V. Daele, A. Housen, F. Kuiken, M. Pierrard & I. Vedder (Eds.), *Complexity, accuracy and fluency in second language use, learning and teaching* (pp. 207-226). Brussels: Contactforum.
- Spada, N. & Tomita, Y. (2008). The complexities of selecting complex (and simple) forms in instructed SLA research. In A. Housen and F. Kuiken (Eds.), *Proceedings of the complexity, accuracy and fluency (CAF) conference* (pp. 227-254) Belgium: University of Brussels.

- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(1), 1-22.
- VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, 12, 287-301.
- Verspoor, M. H., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231.
- Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*, 96(4), 576-598.
- Yang, W. & Sun, (2015). Dynamic development of complexity, accuracy and fluency in multilingual learners' L1, L2 and L3 writing. *Theory and Practice in Language Studies*, (5)2, 298-308.
- Zouhir, A. (2013). Language situation and conflict in Morocco. In O. Orié and K. Sanders (Eds.), *Selected Proceedings of the 43rd Annual Conference on African Linguistics* (271-277). Somerville, MA: Cascadilla Proceedings Project. www.lingref.com.