# Comparing Learners' General Proficiency Levels with Their Writing Productive Ability: How Correlated are They?

Bouchaib Benzehaf [a] *  iD

[a] *Chouaib Doukkali University, Faculty of Letters and Humanities, El Jadida, Morocco*

**Abstract**

Since complexity, accuracy and fluency (CAF) have been suggested as measures of language development (Larsen-Freeman, 1978), a heated debate has surrounded the issue of whether they can sufficiently capture the multi-dimensional facets of language proficiency and be reliable indices of language proficiency levels. Contributing to this debate, the present study investigates the correlation between L3 English proficiency and CAF measures in writing. It recruits 88 semester-one 2nd year Baccalaureate students who were divided into nine groups of general proficiency using the 9-point Stanine scale based on their scores in a general English proficiency test. The scores for the proficiency test were obtained based on the holistic scoring method. We, then, assigned the participants a writing task which we evaluated in terms of CAF. We used the number of dependent clauses per T-unit to measure complexity, the ratio of error-free T-units to the total number of T-units to measure accuracy and number of words per minute to measure fluency. The results showed a strong correlation between proficiency levels and CAF measures, thereby proving that CAF measures serve as a framework suitable for benchmarking language proficiency development. In light of the results, some implications are made for future research targeting language development.

*Keywords:* Language proficiency; CAF measures; writing.

## 1. Introduction

### 1.1. Proficiency models

Proficiency, which seems to be the major aim of language learning, defies attempts of definition since several questions arise as to what components need to be taken into consideration in determining a person's proficiency level as well as how competent an individual needs to be to pass as proficient. It is different things for different scholars. For some, it is restricted to a person's knowledge of grammar and lexis (Harley, Allen,

Cummins & Swain, 1990); for others, it is the knowledge of linguistic and sociolinguistic conventions (Widdowson, 1983).

One influential model of proficiency was suggested by Hymes (1972) and is referred to as "communicative competence". This model underscores the person's ability to use language to convey and interpret meaning. Canale and Swain (1980) further developed this model by dividing it into four components: grammatical competence (which refers to the learner's knowledge of vocabulary, phonology and rules of the language), discourse competence (which relates to the learner's ability to connect utterances into a meaningful whole), sociolinguistic competence (which relates to the learner's ability to use language in appropriate situations), and strategic competence (which relates to the learner's ability to make up for breakdowns in communication).

Canale and Swain's (1980) model of communicative competence had dominated the field of second and foreign language acquisition and language testing for more than a decade. Later, Bachman (1990) came with a model of "communicative language ability" which was further developed by Bachman and Palmer (1996). This model comprises two broad areas, language knowledge and strategic competence. Language knowledge consists of two main components: organizational knowledge, i.e. control over formal language structures, and pragmatic knowledge, i.e. functional and sociolinguistic knowledge. On the other hand, strategic competence refers to a set of metacognitive functions which enable language users to be involved in goal setting, planning, and assessment of communicative sources. While drawing insight from the above-mentioned models of proficiency, the present study tries to bring in other aspects of proficiency which have proved efficient in testing language proficiency development, namely, complexity, accuracy and fluency (CAF).

### 1.2. CAF components as an index of development

In concurrence with the above-suggested models of proficiency, researchers have also been concerned with ways of assessing language proficiency. Different measures to assess and determine learners' proficiency levels have been suggested in this regard. One such measure of learners' language growth, which was borrowed from the field of first language (L1) acquisition, has been based on length of certain structures (Norris & Ortega, 2009). However, this measure proved to be ineffective as rote-learned formulaic chunks gave false impressions of progress (Myles, 2012), hence, the quest for a more systematic measurement of development that could yield a more precise identification of the learners' proficiency levels and a description of their performance was initiated.

The efforts crowned with the emergence of an Index of Development operationalized as measures of Complexity, Accuracy, and Fluency (CAF). Unlike the length-based measures, CAF measures were believed to capture the pivotal aspects of L2 performance (see Ellis, 2003 & Skehan, 1998). An additional benefit of CAF is that each construct encompasses various, multi-faceted traits which can be assessed by a number of measures. CAF measures have increasingly gained significance and have

figured as major research variables in research targeting second and foreign language acquisition. Researchers have started to use them to describe oral and written performance and to measure progress in language learning.

Researchers have also used CAF to examine the development of L2 production from the perspective of its relationship to L2 proficiency. For instance, Pennington and So (1993) compared university students' writing process with their overall linguistic product and reported positive correlations between the two. Higher proficiency level learners wrote better than lower proficiency level learners. Type of task was also found to interact with complexity and accuracy as challenging tasks induced learners to produce more complex language, while simple tasks induced them to increase their accuracy. Preparation time led to an increase in complexity but not in accuracy, thereby demonstrating a trade-off between these two constructs.

Synthesising research findings of 25 studies which investigated the syntactic complexity measures and their relationship to L2 proficiency, Ortega (2003) found that this relationship varied systematically across studies. Such a variance hinged on type of language involved (2nd or foreign) and whether proficiency was defined by programme level or by holistic rating. The researcher also reported that longitudinal evidence was limited and suggested that an observation period of roughly one year of college-level instruction be a requirement for substantial changes in the syntactic complexity of L2 writing.

Larsen-Freeman (2006) studied the development of L2 complexity, accuracy and fluency in the spoken and written production of five Chinese (higher-) intermediate learners of English. The main (quantitative) finding was that every CAF domain improved at the group level indicating that they increase according to the proficiency level. Similarly, after studying L2 Korean learners' writing, Seo (2009) found that the higher the learner's proficiency level is, the higher the grammatical accuracy levels are. Fluency and complexity also increased as reflected by the number of words, clauses, and morphemes per sentence. Contrary to these results, Seo and Eo (2011) found that as proficiency increased, accuracy declined due to an increase in the proportion of syntactic errors in connective endings.

Similarly, Bulté and Housen (2015) studied the development of English L2 writing proficiency of adult ESL learners as evaluated by means of objective measures targeting different components of lexical and syntactic complexity. They also compared the scores on these measures with more holistic and subjective ratings of learner overall writing quality. One of the main results obtained was that some measures, (e.g., subordination ratios and lexical richness measures), can adequately and validly capture development in L2 writing.

In the context of Morocco, Zyad, Rguibi and Bouziane (2016) investigated the relationship between objective linguistic and lexical measures, particularly CAF components, and holistic ratings in the writing of freshmen. The main finding obtained showed a significant relationship between objective measures and holistic ratings. Also, writing quality was significantly predicted by the interaction of mean

length of T-unit, number of clauses per T-unit, number of error-free T-units and lexical diversity.

## 1.3. Aim and research questions

Since CAF measures were suggested as measures of language development, a vigorous debate has ensued regarding their nature and their ability to capture the multi-dimensional facets of language proficiency and be reliable indices of language proficiency levels. The present study aims to contribute to the knowledge and research surrounding CAF measures and their ability to correlate with overall language proficiency. In particular, it tries to probe into the ability of L3 writing CAF components to describe the learner's holistically scored overall linguistic proficiency. The following research questions guide this study:

1. To what extent does students' proficiency based on Stanines relate to their writing ability as expressed in terms of CAF?
2. To what extent do CAF measures correlate with each other in students' writing?

## 2. Theoretical underpinnings

The study is conducted within the framework of CAF measures. Though researchers agree on the validity and usefulness of these constructs, they do not agree about their definition and operationalization. One main reason why there is lack of consensus regarding their definition is that each construct encompasses various, multi-faceted traits, thus allowing for different operationalisations.

### 2.1. Complexity

Despite the interest it has generated, no consensus exists in the second or foreign language literature on the definition of complexity, and no consistency regarding its operationalization. A wealth of complexity measures is available in the acquisition literature ranging across words/T-unit, clauses/sentence, number of subordinate clauses, word types/word token, number of passive forms and number of relative clauses, etc. for Ellis and Barkhuizen (2005, p. 139), complexity is "elaborated language". For Norris and Ortega (2009), different complexity measures exist in the literature, but the most common are global complexity measures which target the length of the unit (such as words per T-units). These measures capture complexity in a general sense since any type of embedding will increase this measurement. In her study, Larsen-Freeman (2006) used one measure for syntactic complexity which is the average number of clauses per T-unit.

Using T-units as a method of analysis, two measures of complexity have been most frequently used by researchers. The first is the mean length of T-unit (e.g. Henry, 1996; Ishikawa, 1995; Larsen-Freeman, 1978, 1983), which is the average across all T-units in a text; and the second is clauses per T-unit (e.g. Bardovi-Harlig & Bofman, 1989; Flahive & Snow, 1980; Hirano, 1991), which is the number of dependent clauses

per T-unit (subordination). It is noteworthy that a T-unit is defined by Hunt as one main clause with all subordinate clauses attached to it. After surveying 27 studies, Ortega (2003) found that 25 studies employed mean length of T-unit (MLTU). Some researchers, however, have used this measure for fluency (Larsen-Freeman, 2006; Wolfe-Quintero, Inagaki, & Kim, 1998). Such a tendency has been vehemently criticised by Norris and Ortega (2009) who state that fluency should rather be measured by pause information.

## 2.2. Accuracy

Of the three CAF measures, accuracy seems to be the most easily defined construct. According to Housen and Kuiken (2009), it is "error-free" speech. However, it also poses several challenges. For instance, against which standards should we measure it? Besides, longitudinal assessment of accuracy creates challenges as learners attempt new lexical items and grammatical forms. Accuracy can also be measured specifically (accuracy of certain forms) or generally (overall number of errors or error-free units).

Ellis and Barkhuzen (2005) advocate use of general measures of accuracy, such as percentage of error-free clauses or number of errors per 100 words when data is collected from loosely structured tasks. In fact, the use of specific measures might induce participants to avoid the constructions that are targeted by the measures. Skehan and Foster (1999) agree that global measures of accuracy are more realistic and sensitive. These views are also empirically supported. Using both global and specific measurements of accuracy, Ahmadian and Tavakoli (2011) found that the global measure of error-free clauses produced the same information as the specific measure of verb forms while Michel, Kuiken, and Vedder (2007) reported that global measures were more informative. It is noteworthy that this finding applies also to complexity and fluency. Michel, Kuiken, and Vedder (2007) investigated learners' task-based L2 performance by means of global CAF measures and specific measures and found global measures to be more informative.

## 2.3. Fluency

It refers to "the production of language in real time without undue pausing or hesitation" (Ellis & Barkhuizen 2005, p. 139). Fluency has also been correlated with learner's ability to use the language with a high number of words and without extensive pauses and/or corrections within a time limit. It refers to the "delivery of speech" (Schmidt, 1992, p. 358), and it has been discussed in terms of two broad categories, temporal and hesitation. Skehan (2009b) has categorised it as repair (revisions/self-corrections), speed, breakdown in writing process, and automatization. It is obvious that most of these measures fit in with speaking but not with writing, the reason why fluency has not received much attention in writing research.

One way temporal variables have been operationalized in writing is word count. An extensive review of 18 studies by Wolfe–Quintero, Inagaki and Kim (1998)

demonstrated that 11 studies found a significant relationship between word count and writing development whereas 7 studies showed only slight increase in word number. However, word count can be rendered more reliable if it is done within a time limit. In this context, Kennedy and Thorp (2002) investigated linguistic responses to an academic writing task and reported significant differences between essays at levels 4, 6 and 8 with level 4 writers hardly meeting the word limit. Similarly, Cumming, Kantor, Baba, Erdosy, Eouanzoui, and James (2005) also reported statistically significant differences between essays at level 3 and 4 and between levels 3 and 5, though not for levels 4 and 5 in their study of differences in writing discourse in independent and integrated prototype tasks for next generation TOEFL. Other commonly adopted measures besides word count include the number of sentences and T-units produced within a certain time constraint (Godfrey, Treacy & Tarone 2014; Wu & Ortega, 2013).

In addition to studying each CAF construct separately, researchers also studied the interaction and interdependency among CAF constructs. In this regard, two hypotheses have been created: the Trade-off Hypothesis and the Cognition Hypothesis. The former makes the claim that CAF dimensions are interdependent such that they compete with each other for attention (Skehan, 1996, 2009; Skehan & Foster, 2001). On the other hand, the Cognition Hypothesis states that particularly more cognitively and functionally demanding tasks encourage the learner to produce more complex and more accurate language production (Robinson, 2001, 2005).

## 3. Method

This study follows an exploratory research design to explore the capacity of CAF measures to reflect students' proficiency levels. In light of the studies reviewed, we used the CAF triad in our measurement after having operationalized them as quantitative measures. Numerical values of the indices are, therefore, given and correlated with general proficiency. After feeding the data into SPSS, we obtained different results for these three different constructs, namely, fluency, accuracy and complexity.

### 3.1. Participants

In the present study, there are 88 participants, 37 males and 51 females. They are all high school students in second year baccalaureate, semester one. Their age ranges between 17 and 21, and they all belong to the science department. All the participants have studied English for 3 years and are in their fourth year. The reason behind collecting data from only science students is practical. They are grouped in same classes, and they constitute the type of students who can provide meaningful data, particularly data that are normally distributed and which can fit in well with stanine levels. Thus, the type of sampling we used is convenience sampling which is most common in educational settings.

## 3.2. Instruments

Data of the present study was collected by using two different sources and methods. First, we used a general proficiency test that consisted of five sub-parts: a reading text including sections for comprehension, grammar, vocabulary, functions and paragraph writing, and we scored it holistically out of 20. We designed the test according to the criteria used in national Baccalaureate exams in Morocco, but we reduced the number of questions to make the test doable within the period of one hour. The test was a nationally accredited Test of Proficiency in Morocco (TOPIK), which has won wide recognition for reliability and validity. Two high school teachers participated in scoring the test following a scoring rubric modelled on that used in baccalaureate national exams in Morocco. After scores were obtained, we converted them to Stanines.

A Stanine stands for "standard nine" and is a method of converting scores into a nine-point scale. Stanines allow us to assign to a student in the group a number which is relative to all members in that group, with number 5 being the mean. Stanines has the benefit of facilitating comparison of scores among different groups; and similar to normal distributions, it provides scores that represent a bell curve with 9 sliced-up pieces.

In addition to the proficiency test, we assigned the students a writing task with the topic of "students' dropping out of school, causes and solutions". The task was of intermediate difficulty and in line with the standards outlined in the curriculum. The participants had one hour for both planning and writing their pieces.

## 3.3. Data analysis

We used descriptive statistics to present quantitative descriptions of students' proficiency levels and generate means and standard deviations. We also used Pearson correlation coefficient which is a measure of the linear correlation between two variables. Pearson correlation coefficient has a value between +1 and −1, where 1 is total positive linear correlation, and 0 is no linear correlation, and −1 is total negative linear correlation. We used it to compute the correlation coefficient between proficiency and each CAF construct, and also to test the correlation among CAF constructs.

## 3.4. CAF analysis

We adopted Hunts' T-units to measure students' writing in terms of CAF. A T-unit is defined as the minimal terminable unit into which the sentence can be broken (Hunt, 1965). It consists of one main clause and any subordinate clauses or fragments attached to it. We attempted to operationalize CAF constructs in a way consistent with the literature. We found a multiplicity of measures; some were global while others were specific. It is worth noting in this context that Wolfe-Quintero, Inagaki and Kim, (1998) also counted over one hundred measures of accuracy, fluency, and

complexity employed in thirty-nine second language writing development studies. Based on findings of various studies regarding the informative nature of global measures, we decided to use global measures. Thus, we used one global measure for each construct. For complexity, we calculated the number of dependent clauses per T-units; for accuracy, we calculated the ratio of error-free T-units to the total number of T-units, and for fluency, we calculated the ratio of words per minute. Following is a summary of the measures adopted in the study:

Table 1. Global measures adopted in the analysis of students writing products

| Complexity | Accuracy | Fluency |
|---|---|---|
| Number of dependent clauses per T-unit | Ratio of error-free T-units to the total number of T-units | Number of words per minute |

## 4. Results

The aim of this research paper is to compare the learners' proficiency levels as scored holistically with their written productive ability as measured in terms of CAF. The study has been guided by the two research questions. Following are the results obtained for each research question.

### 4.1. Relationship between proficiency and written ability

To answer this research question, first we had to determine the participants' proficiency Stanines. Thus, we delivered a proficiency test which we scored holistically. We tested the assumption of normality via examination of the unstandardized residuals. Review of the S-W test for normality (SW=.976, df=88, p>0,05) and skewness (-0,25) and kurtosis (0,46) statistics suggested that normality was a reasonable assumption. The boxplot suggested a relatively normal distributional shape (with no outliers) of the residuals. The Q-Q plot and histogram suggested normality was reasonable.

After checking for normality of data, we converted the scores into Stanines by calculating the means for each Stanine level. The Stanine score serves to show the learner's ranking within the specific group. The lowest level within the group is assigned to Stanine 1 and the highest level to Stanine 9. The lowest three levels are below average, the next three levels are average while the highest three levels are above average. Table 2 below presents the results of the proficiency test based on Stanines.

Table 2. Students' Stanine proficiency levels

| Stanine level | Student number | Male | Female | Proficiency test mean | St deviation |
|---|---|---|---|---|---|
| 9 | 4 | 0 | 4 | 19.12 | .62 |
| 8 | 6 | 1 | 5 | 16.16 | 1.29 |
| 7 | 10 | 4 | 6 | 13.85 | .41 |
| 6 | 15 | 5 | 10 | 12.23 | .77 |
| 5 | 18 | 6 | 12 | 10.61 | .43 |
| 4 | 15 | 7 | 8 | 9.4 | .57 |
| 3 | 10 | 6 | 4 | 8 | .0 |
| 2 | 6 | 5 | 1 | 7 | .0 |
| 1 | 4 | 4 | 0 | 4.25 | 1.70 |

The participants were normally distributed across the Stanine levels, with the highest Stanine score being 19.12 which was scored by 4 female students while the lowest Stanine score was 4.25 obtained by 4 male participants. The majority of student participants came in the middle Stanine levels, namely 4, 5 and 6 scoring between 9.4 and 12.23. Thirty female students and eighteen male students were located in these levels. We also observed that girls obtained higher Stanine scores than males.

Following the proficiency test which enabled us to identify student participants' stanine scores, we administered a writing task. After scoring the writing task in terms of complexity, accuracy and fluency, we calculated the means for CAF scores for each Stanine level. Complexity was computed by the number of dependent clauses per T-unit; accuracy was calculated by the ratio of error-free T-units to the total number of T-units, and fluency was assessed by the ratio of words per minute. The results are presented in table 3 below:

Table 3. CAF measures in written production by proficiency

| Stanine level | Student number | Complexity | | Accuracy | | Fluency | |
|---|---|---|---|---|---|---|---|
| | | Mean | St dev | Mean | St dev | Mean | St dev |
| 9 | 4 | 0.53 | 0.040 | 0.48 | 0.018 | 2.91 | 0.18 |
| 8 | 6 | 0.52 | 0.044 | 0.40 | 0.020 | 2.75 | 0.042 |
| 7 | 10 | 0.50 | 0.052 | 0.35 | 0.038 | 2.69 | 0.069 |
| 6 | 15 | 0.44 | 0.047 | 0.35 | 0.30 | 2.51 | 0.116 |
| 5 | 18 | 0.44 | 0.051 | 0.31 | 0.25 | 2.47 | 0.039 |
| 4 | 15 | 0.41 | 0.061 | 0.32 | 0.045 | 2.25 | 0.024 |
| 3 | 10 | 0.32 | 0.061 | 0.26 | 0.042 | 1.86 | 0.034 |
| 2 | 6 | 0.23 | 0.077 | 0.16 | 0.034 | 1.41 | .051 |
| 1 | 4 | 0.20 | 0.081 | 0.12 | 0.10 | 1.16 | .033 |

The table shows that the higher the proficiency Stanine is, the higher the mean scores get for each construct. For complexity, learners at level one scored 0.20 while

learners at level nine scored 0.53. Learners at level 5, which is the average score, produced 0.44 subordinate clauses per T-unit. Thus, it is more likely that the more proficient learners use more subordinate clauses in their writing. These results indicate that a positive correlation exists between complexity and proficiency.

Similarly, students in higher proficiency stanines produced more accurate language than their counterparts in lower proficiency levels. For instance, student participants at stanine 9 scored 0.48 indicating an average of 48% error-free T-unit rate against the total number of T-units in their writing. In contrast, students at proficiency stanine level 1 had only 0.12 error-free T-units. This finding suggests the existence of a positive correlation of accuracy with proficiency. Mean scores for fluency, in turn, showed a positive correlation with proficiency. The higher the proficiency stanine is, the greater students' fluency becomes. Students at stanine level nine produced an average of 2.91 words per minute while students at level one wrote no more than an average of 1.16 words per minute.

To check the significance of the apparent correlations, we conducted a statistical analysis of correlations between proficiency and each construct of the CAF triad using the Pearson correlation coefficient. Table 4 below presents the results.

Table 4. Correlation between language proficiency and written production

|  |  | General proficiency | Complexity | Accuracy | Fluency |
|---|---|---|---|---|---|
|  | Pearson Correlation | 1 | .763** | .818** | .874** |
| General proficiency | Sig. (2-tailed) |  | .000 | .000 | .000 |
|  | N | 88 | 88 | 88 | 88 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4 above shows that learners' proficiency levels strongly correlate with complexity (r =0.763, p < 0.01), with accuracy (r =0.818, p < 0.01), and with fluency (r =0.874, p < 0.01). Findings indicate that learners' writing CAF components improve along with their general proficiency levels, with fluency having the strongest relationship with proficiency.

## 4.2. Correlation of CAF in writing ability

We observed that all CAF constructs increased with each proficiency stanine level. To check for the significance of that relationship, we used Pearson correlation coefficient in order to test the correlation among CAF constructs in students' written production. Table 5 below presents the results obtained.

Table 5. Correlations among CAF in writing ability

|  |  | Complexity | Accuracy | Fluency |
|---|---|---|---|---|
| complexity | Pearson Correlation | 1 | .693** | .854** |
|  | Sig. (2-tailed) | ---- | .000 | .000 |
|  | N | 88 | 88 | 88 |
| accuracy | Pearson Correlation | .693** | 1 | .825** |
|  | Sig. (2-tailed) | .000 | ---- | .000 |
|  | N | 88 | 88 | 88 |
| fluency | Pearson Correlation | .854** | .825** | 1 |
|  | Sig. (2-tailed) | .000 | .000 | ----- |
|  | N | 88 | 88 | 88 |

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5 above shows that CAF constructs are strongly correlated with each other in students' writing. Particularly between complexity and fluency the correlation coefficient ranks highest (r =0.854, p < 0.01), while it is lowest between complexity and accuracy but still significant (r =0.693, p < 0.01). Correlation between fluency and accuracy was also very strong, with a significant coefficient (r =0.825, p < 0.01). The high correlations obtained among complexity, accuracy, and fluency suggest that they interact positively with each other and develop alongside each other.

## 5. Discussion

This study had two aims: to investigate the relationship between general language proficiency development (as expressed through grammar, vocabulary, functions and reading, and scored holistically) on one part, and writing ability (as scored objectively using CAF components) on another part. The second aim was to investigate the correlation among CAF components. The results pointed to a positive correlation between proficiency and writing CAF components, and also among CAF components. Following is a discussion of these two main findings.

### 5.1. Relationship between proficiency and written ability

The results obtained show that proficiency and written production are strongly correlated with each other indicating that writing CAF components all increase with the development of proficiency. The higher the proficiency stanine level is, the higher complexity, accuracy and fluency levels are in students' written productions. This suggests that as learners improve their proficiency, they also improve their writing in terms of complexity, accuracy and fluency. This finding supports the suggestion that writing ability reflects the development of learner's language system (Verspoor & Smiskova, 2012). Thus, CAF components can be valid constructs for measuring general language proficiency.

The findings obtained are consistent with the literature as Seo (2009) also reported that the higher the learner's proficiency level is, the higher the accuracy, fluency and

complexity levels in L2 Korean learners' writing become. Hence, we align with Verspoor, Schmid, and Xu (2012, p. 239) who suggest that "one useful way to measure general proficiency in a second language (L2) is to assess writing samples". Ellis (2003) and Skehan (1998) also state that CAF measures can capture the pivotal aspects of L2 performance. One way we can explain this is that writing involves the active use of all facets of language, including lexis, grammar, as well as other levels of language features.

A side issue of the findings centres on the gender dimension. It was observed that most females were placed in upper stanines while most males were placed in lower stanines. For instance, stanine nine was exclusively female and stanine one was exclusively male. This supports the claim that education is increasingly being feminized with girls not only accounting for the large majority of the student population, but also performing higher than male students. Possible interpretations of female advantages at school go primarily to their good behaviours, positive attitudes to school work and attentiveness in classrooms. The literature abounds with studies that document female advantages at school. For instance, Kenney-Benson, Pomerantz, Ryan, & Patrick (2006) have drawn attention to the fact that girls are surpassing boys in school grades even in stereotypically masculine subject areas like maths and science. Interestingly, these researchers went beyond documenting girls' academic excellence to investigate causes for it, with findings showing that girls were more likely than boys to hold mastery over performance goals and to refrain from disruptive classroom behaviour. Similarly, Geisler and Pardiwalla (2010) documented girls' academic advantages in all courses and in all stages of education, reporting higher dropout rates, higher levels of truancy, and greater discipline problems among boys than among girls, which could account for boys' lagging performance.

### 5.2. Correlation of CAF in writing ability

The main (quantitative) finding obtained with regard to the second research question is that every CAF domain improved at the group level based on stanines. CAF components for writing ability were also highly correlated with each other suggesting that they are interconnected. Particularly, fluency is the most strongly correlated with proficiency. This finding is consistent with the one obtained by Larsen-Freeman (2006) who reported an increase in every CAF domain as proficiency improved. It is also partly consistent with Robinson's (2001, 2005) cognition hypothesis which states that it is not only possible but also natural that complexity and accuracy receive concurrent attention from the learner. However, contrary to Robinson who states that fluency develops separately, this study found that fluency increased in tandem with complexity and accuracy.

The findings of this study are different from the findings obtained in some previous studies regarding the correlation of CAF constructs. Some researchers found that CAF components do not develop concurrently and that students tend to overlook one area while concentrating on another (Skehan, 1996, 2009). Benevento and Storch

(2011) also reported significant improvements over time in language complexity and discourse, but not in accuracy. Probably one of the reasons for such inconsistency of results is the type of task and the proficiency levels of the learners. Besides, the different measures adopted in the different studies are also likely to affect the results. Whether it is a bliss or otherwise, different measures have been adopted in the assessment of CAF constructs in the literature, and that is due to the fact that each construct encompasses various, multi-faceted traits which make the construct amenable to different measures. Accordingly, different researchers used different measures, thereby making synthesis of findings as well as comparison of results across studies a complex issue. Besides, different proficiency levels of the participants may also affect the findings obtained. It is noteworthy that the proficiency levels of the participants in this study range from pre-intermediate to intermediate.

## 6. Conclusion and implications

This study examined the relationship between L2 proficiency and English written production. We used a general proficiency test that consisted of a reading text including sections for comprehension, grammar, vocabulary, functions and paragraph writing. We designed the test according to the criteria used in national Baccalaureate exams in Morocco after reducing the number of questions to make the test doable within the period of one hour. After scoring the test holistically, we converted the scores to Stanines. We also assigned the students a writing task with the topic of "students' dropping out of school, causes and solutions".

Results of the correlation analysis revealed a strong correlation between learners' proficiency levels and their written productions. All CAF constructs increased in written production as proficiency levels increased, suggesting that as the learners improve their proficiency, they are more likely to produce more complex, more accurate and more fluent output. This finding suggests that writing can capture the pivotal aspects of learners' inter-language as it is a skill that involves the active use of all facets of language, including lexis, grammar, and other discourse levels of language features.

One implication of these findings is that writing CAF components can serve as an effective framework within which to gauge learners' inter-language system and its progress. The fact that CAF can be operationalized as a set of quantitative measures providing numerical values that are more objective than holistic scoring allows them to provide better perceptibility of development. Another implication follows from the new gender gap in favour of females as suggested by the proficiency test scores. These new gender differences call for the adoption of best classroom management practices that maximise all students' achievements, males and females. Educational reforms should now give greater priority to boys' educational difficulties.

Lastly, one of the limitations of this study is that the findings have been obtained from the perspective of group means. Given that ample individual variation exists between students, we suggest that case studies should be conducted which probe into

individual differences. Combining the quantitative approach with a qualitative approach is also bound to enrich the data by presenting information regarding how learners' inter-language is being developed.

## References

Ahmadian, M. J., & Tavakoli, M. (2011). The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*, 35-59.

Bachman, L.F. (1990). *Fundamental Considerations in Language Testing*. Oxford: OUP.

Bachman, L.F. & Palmer, A.S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: OUP.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition, 11*, 17-34.

Benevento, C., Storch, N. (2011). Investigating writing development in secondary school learners of French. *Assessing Writing, 16*(2), 97-110.

Bulté, B., & Housen, A. (2015). Evaluating short-term changes in L2 complexity development. *Círculo de Lingüística Aplicada a la Comunicación, 63*, 42-76.

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

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K. & James, M. (2005). Differences in writing discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5-43.

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.

Flahive, D. E., & Snow, B. G. (1980). Measures of syntactic complexity in evaluating ESL compositions. In J. W. Oller, & K. Perkins (Eds.), *Research in language testing* (pp. 171–176). Rowley, MA: Newbury House.

Geisler, G. & Pardiwalla, M. (2010). Socialization patterns and boys' underperformance in Seychellois schools. *Journal Statistique Africain, 11*, 62-84.

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.

Harley, B., Allen, P., Cummins, J., & Swain, M. (1990). *The development of second language proficiency*. Cambridge: Cambridge University Press

Henry, K. (1996). Early L2 writing development: A study of autobiographical essays by university-level students of Russian. *The Modern Language Journal, 80*, 309-26.

Hirano, K. (1991). The effect of audience on the efficacy of objective measures of EFL proficiency in Japanese university students. *Annual Review of English Language Education in Japan, 2*, 21-30.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquistion. *Applied Linguistics, 30*(4), 461-473.

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Champaign, IL: National Council of Teachers of English.

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

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*, 51–70.

Kennedy, C., & Thorp, D. (2002). A corpus-based investigation of linguistic responses to an IELTS academic writing task: University of Birmingham.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M., & Patrick, H. (2006). Sex differences in math performance: The role of children's approach to schoolwork. *Developmental Psychology, 42*, 11–26.

Larsen-Freeman, D. (1978). An ESL Index of development. *TESOL Quarterly, 12*, 439-48.

Larsen-Freeman, D. (1983). Assessing global second language proficiency. In H. W. Seliger & M. H. Long (Eds), *Classroom Oriented Research in Second Language Acquisition* (pp. 287-304). Rowley, MA: Newbury House.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *IRAL*, 241-259.

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*(4), 555-578.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492–518.

Pennington, M. C., & So, S. (1993). Comparing writing process and product across two languages: A study of six Singaporean university student writers. *Journal of Second Language Writing, 2*, 41–63.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*(1), 27–57.

Robinson, P. (2005). Cognitive complexity and task sequencing: A review of studies in a Componential Framework for second language task design. *International Review of Applied Linguistics in Language Teaching, 43*(1), 1–33.

Seo, S. J. (2009). Study on the interlanguage development of Korean language learners by syntactic proficiency assessment: Based on analyzing syntactic features of writings. (Unpublished master's thesis). Yonsei University, Seoul, South Korea.

Seo, S. J., & Eo, J. H. (2011). Study on the accuracy variation of connective endings by Korean proficiency level. *Journal of Korean Language Education, 22*(1), 123–143.

Schmidt, R. (1992.) Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition, 14*, 357-385.

Skehan, P. (1996). Second language acquisition research and task-based instruction. In J. Willis & 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***,** 510-532.

Skehan P. (2009b). Lexical performance by native and non-native speakers on language-learning tasks. In Richards B., Daller H., Malvern D.D., Meara P. (Eds.), Vocabulary studies in first and second language acquisition: The interface between theory and application (pp. 107–24). London: Palgrave Macmillan.

Skehan, P. & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93-120.

Skehan, P. & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183-205). Cambridge: Cambridge University Press.

Verspoor, M. H., Schmid, M. S., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing, 21*(3), 239-263.

Verspoor, M. H., & Smiskova, H. (2012). Foreign language writing development from a dynamic usage based perspective. In R. M. Manchón (Ed.), *L2 writing development: Multiple perspectives* (pp. 17–46). Boston/Berlin: Walter de Gruyter.

Widdowson, H. G. (1983). *Learning Purpose and Language Use*. Oxford: Oxford University Press.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy & Complexity*. Honolulu, HI: Second Language Teaching & Curriculum Center, University of Hawaii at Manoa.

Wu, S.-L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals, 46*(4), 680-704.

Zyad, H., Rguibi, S., & Bouziane, A. (2016). The relationship between L2 writing quality and objective measures of linguistic and lexical complexity. Paper presented at the 4th International conference on "Cultures and Languages in Contact" organised by the Faculty of Letters and Humanities, Chouaib Doukkali University, El Jadida, 14-15 December, 2016.