AN OVERVIEW OF THE AGE FACTOR AND ITS PEDAGOGICAL IMPLICATIONS FOR VOCABULARY ACQUISITION

María del Pilar Expósito Morillas
TFM supervisor: Cristóbal Lozano Pozo
Máster de Secundaria, Granada 2011
University of Granada

INDEX

AN OVERVIEW OF THE AGE FACTOR AND ITS PEDAGOGICAL IMPLICATIONS FOR VOCABULARY ACQUISITION

Abstract

This research deals with the age effects on the acquisition of vocabulary as a second language in an instructed setting. The literature review on the age factor in reference to vocabulary acquisition shows an advantage of the late starters over the early starters in instructed settings. However, early starters surpass late starters in naturalistic settings in the long run when the basic conditions are fulfilled. Under the light of these assumptions, this study tries to show how late starters outperform early starters in instructed settings. The study made in this assignment is an adapted replication from the one conducted by Miralpeix in 2006. Measures such as Type-Token ratio or lexical density have been obtained via VocabProfile (Nation, 1995) and several comparisons have been done between late and early starters. The study differs from the one of Miralpeix in the number of activities that were given to the students. Moreover, new measures were obtained such as tokens per family and types per family. Findings suggest that there is an advantage of late starters over early starters in measures such as type token ratio although this fact must be taken cautiously because this study does not focus on ultimate attainment, a state in which early starters are usually better than late starters, but on rate. To finish with, some pedagogical implications in relation to age effects are exposed.

1 Introduction

Second language learning is a very complex process which involves several factors. Some of the factors that deeply influence this process are the **individual differences** in learners. There are some individual differences such as motivation, personality, aptitude, cognitive style and age among others which have an effect on language learning.

To begin with, **motivation** can become a key factor to attain native-like levels of linguistic competence given that it is the learner who is responsible for the level of effort in his/her learning. In general, motivation encloses a need to be achieved, the wish to reach an objective and the expectation for potential outcomes or rewards (Oxford and Ehrman, 1993; Dörnyei, 2001) With respect to motivation in second language learning, two factors are present: the communicative needs of the learner and the attitudes with respect to the

language community (Lightbown & Spada, 2006). In instructed settings, learner's motivation becomes different with age, so the teacher has to adjust his/her teaching to their learners' interests and goals. Furthermore, there are several kinds of motivation: integrative and instrumental motivation. Integrative motivation has to do with emotional factors: the learner wants to study a language for the sake of learning because it is a rewarding activity for him/her. However, instrumental motivation has to do with learning a language for its practical use such as travelling abroad, getting a job or passing an exam. Neither integrative motivation nor instrumental motivation is better than the other: it depends on the learner.

With respect to **personality**, some researchers have posed the question whether having certain personality traits can positively influence second language learning, but it is very difficult to test and demonstrate this fact in empirical studies. The main hypothesis with respect to personality is that extroverted people will achieve better levels in language learning given that one of the main reasons for being successful in the language learning process is producing great amounts of language and supposedly, extroverted people would do that. But producing is not only necessarily knowing the L2: an introverted person can be producing language without speaking correctly and therefore, he/she can become successful too and this is what research has proved in many cases. Probably, the most interesting aspect about personality with respect to language learning is **inhibition**. Risk taking is a basic element in language learning and inhibition dampens it in many cases. Children are supposed to have an advantage in this sense because they use to be less inhibited and they are great risk takers. **Anxiety** is another personality factor that influences language learning. Anxiety can interfere in the learning process but also, anxiety is not always negative: being anxious before an exam can lead the student to have the perfect combination of motivation and focus to succeed (Lightbown & Spada, 2006).

In relation to **aptitude**, it is defined as a special predisposition to learn a language. Moreover, John Carroll (1991) identified aptitude as an ability to learn quickly. According to these two definitions, a learner who has a better aptitude towards languages would be more successful than others, but this is not always true because slower learners can be also successful in the long run. There are several test such as Modern Language Aptitude Test (MLAT) whose authors are Carroll and Sapon (1967), and Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966) which are based in the following points: the capacity to identify and memorize new sounds, the ability to understand different functions of the words within the sentences, the faculty of being able to formulate grammatical rules from language models and

the capability to remember new words. Some of these measures are predictive of success (Lightbown & Spada, 2006)

Cognitive or **learning styles** are supposed to have an influence on language learning. Learning styles are the way by which a student prefers to absorb, process and retain new information and abilities (Reid, 1995 as cited in Lightbown & Spada, 2006) There are people who are very successful in their academic performance but then, it is very difficult for them to learn a language. In the following figure, the main cognitive styles are summarized:

Cognitive styles	
Field-dependent	Field independent
Global	Particular
Holistic	Analytic
Deductive	Inductive
Focus on meaning	Focus on form

Table 1: Cognitive styles (Saville-Troike, 2006)

Field dependent/field independent learning styles are the ones which are supposed to have a greater influence on language learning. Those people who are field-dependent are skilful in reading social signs, which is very useful for an effective communication. On the other hand, those who are field-independent have an established frame from their inner self without taking into account the context and they are more autonomous. Field-dependent students take advantage from the contextualized interaction whereas field-independent students are more analytic and they can make the most of formal instruction (Saville-Troike, 2006)

Finally, **age** is one of the most influential factors and it has been thoroughly studied in second language learning and this is one of the reasons for me to conduct this study. Furthermore, vocabulary has been seen out of the scope of age effects and this was another reason to address my study in this direction.

2 Theoretical background

There are several sections within the theoretical background. In first place, there is a section devoted to the importance of studying the age factor followed by some biological explanations. After that, there are several subsections dedicated to the critical period hypothesis, explaining what it is and its relation with the different aspects of language. With

respect to the critical period hypothesis, there is a section with possible evidences of the inexistence of this period. Moreover, it is important to highlight the difference between rate and ultimate attainment; therefore, there is a section dealing with this distinction. Furthermore, naturalistic and instructed settings are described given that this difference is relevant with respect to the age factor and some general characteristics of young and old learners. More specifically, the relation of age to second language vocabulary acquisition is put forward together with the section 'Vocabulary: an abandoned area in age-related effects in second language learning' and the relevance of studying the influence of age in vocabulary acquisition. To finish with, there is a section devoted to some general conclusions on age and the main aims of this paper.

2.1 Why study the age factor?

The age factor has been studied for over fifty years and in that time, there have been many questions answered and still there are many to solve. These studies help us to understand the language capacity that humans have and furthermore, the results of investigations are very useful for educational purposes.

When talking about the age factor, everybody has an opinion on it. Most teachers and parents agree in the fact that the younger, the better for the students. It is believed that children will learn, in a more efficient way, how to pronounce and a wider vocabulary if they are exposed to the language sooner. This is one of the reasons why I decided to present this final assignment on age: to prove if popular beliefs are true or not with respect to vocabulary acquisition. Furthermore, the study of the age factor is relevant for educational policies: Is the younger, the better in instructed settings? Are instructed settings as prepared as naturalistic settings to introduce language learning from the very beginning? These questions will be dealt with throughout this assignment.

2.2 Biological explanations

There are several hypotheses that explain the difficulties that a person can experience in learning an L2 which are related to age, such as loss of plasticity in the brain (lateralization) or the influence of universal grammar with respect to language learning. A summary of some other hypotheses is also presented in figure 1 at the end of this section.

Scientific studies have proved that there could be certain locations in the brain that are specialized in language functions. The first studies were the ones by Broca and Wernicke. Broca observed in 1861 that there is an area in the left frontal lobe that is in charge of speaking. This was proved in patients with injuries in this area. On the other hand, Wernicke found an area next to the part of the cortex whose function is to process audio stimuli. Later on, it has been proved that there are more areas involved in language processing and production all over the brain, but the more active areas related to language are typically situated in the left hemisphere. This fact is known as **lateralization**: the brain progressively specializes in certain functions and it has less plasticity. Lateralization is directly related to age given that the brain loses plasticity as the child gets older. As Saville-Troike cites in her book, 'age of acquisition influences brain organization for many second language learners' (71:2006). It is supposed that acquiring a language later in life, leads the student to have a more active right hemisphere. As Cook suggested:

> the variation in right hemisphere involvement may be due to the lack of a single route to L2 knowledge: second language may be learnt by many means rather than the single means found in L1 acquisition and consequently, may have a greater apparent hemispheric spread (1992:572)

Chomsky in 1965 proposed his hypothesis on **universal grammar**. UG is described as an innate faculty whose function is to acquire language. UG can be an explanation for that knowledge the learners have that cannot have been acquired from input. But, is this faculty available in L2 learners? If so, is it something that lasts forever? The first question can be answered by White:

> i The phenomenon in question is underdetermined by the L2 input. That is, it must not be something that could have been acquired by simple observation of the L2 input, as an effect of input frequency, or on the basis of instruction, analogical reasoning, etc.
>
> ii The phenomenon in question works differently in the L1 and the L2. If L2 learners show evidence of subtle and abstract knowledge, we want to exclude the possibility that such knowledge is obtained solely via the L1 grammar. (1989b, 1990 as cited in White, 2005)

Therefore it seems that UG has a role in language learning but, when does it finish? Bley-Vroman (1989) posited the **Fundamental Difference Hypothesis** in which it is stated that adults have no access to UG whereas children have. As adults do not have access to it, they go for problem solving skills to make sense of the grammatical structures of the L2 input.

The following figure summarizes the two hypotheses presented above and some other biological explanations related to age:

Hypothesis	Description
Loss of neural plasticity in the brain	After the closure of the critical period (CP), the neural substrate responsible for language learning is not fully available due to loss of organizational plasticity and lateralization (Penfield and Roberts, 1959)
Loss of (access to) the language learning faculty	A strong version of this hypothesis is that Universal Grammar (UG) is no longer available on closure of the CP. A weak version is that UG is mentally represented by no longer accessible. See Bley-Vroman's (1989) Fundamental Difference Hypothesis.
Maladaptive gain of processing capacity with maturation	The greater processing capacity of adults enables them to extract more from the input and thus they are faced with the problem of analyzing everything at once, whereas children's limited processing capacity means they extract less from the input but can handle it.
Use it and then lose it	The language learning faculty has served its purpose once a language has been learned; evolution has ensured that it is dismantled once it is no longer needed as keeping it would incur costs.
Use it or lose it	This draws on 'a mental muscle metaphor': that is, if the language learning faculty is not used, it will atrophy but if it is used it is maintained. Bever (1981) suggests that acquisition requires that perception and production systems need to work together but once one stops learning a language this ceases.
Learning inhibits learning	Connectionist theories see learning as a matter of the strengthening of neural connections but once the connections have been firmly established they are difficult to undo. Thus the ability to learn may change over time as a function of previous learning.
Other hypotheses	Other variables that may account for age-related difficulty in L2 learning are availability of input, social-psychological factors such as learner attitudes and motivation, and availability of instruction.

Figure 1: Explanations of age related difficulties in L2 learning (based on Birdsong 1999a: 2-9 as cited in Ellis, 2008)

2.3 Critical period hypothesis (CPH)

A **critical period** is a limited expand of time within which an occurrence can happen in order to make a transition from one point to another with a final result. As Singleton states: 'critical periods are characterized as periods with well-defined and predictable termini' (2007:1). Another definition for the critical period is the one by Harley & Wang (1997:20-21):

‘(the CPH is) a period of time during the life cycle when there is greater sensitivity to certain types of environmental stimuli than at other times’. With respect to innate behaviours, Guasti (2002:20) wrote the following: ‘Innate behaviours are often distinguished by the existence of *critical periods* during which the ability to acquire the competence reaches its peak; thereafter, the ability to acquire that competence declines’

In nature, there are several examples of critical periods such as the behaviour showed by ducklings related by De Villiers:

> This following behavior only occurs within a certain time period after hatching, after which point the ducklings develop a fear of strange objects and retreat instead of following. Within these time limits is the critical period for the following behavior. (De Villiers & De Villiers 1978:210)

With this, De Villiers refers to the period of time within which the ducklings follow the first moving object they see. Furthermore, there are examples of critical periods in humans, concretely for binocular vision: the period for developing it is thought to be from three and eight months.

With respect to first language acquisition, it is very difficult for researchers to prove with subjects if there is a critical period for first language acquisition given that it would involve depriving children of language exposure and it is not ethical. Thus, the only children they could investigate were the so called **‘feral’ children**. Genie was one of these feral children who were studied in relation to the CP. She was found when she was thirteen and in inhuman conditions and isolation. Furthermore, she was deprived from any linguistic stimuli since she was two years old. Although great efforts were made to teach her to speak, she was able to progress but not in a normal way. Something important to highlight in this point is that Genie had great progress in vocabulary which contrasts with her low capacity with respect to syntax. This fact could mean that the age factor has not had any effect on vocabulary acquisition and that the starting age is not relevant with respect to lexicon. However, linguists must be cautious with investigations on feral children because there could be other factors such as horrible living conditions that could have influenced language learning.

In reference to the **critical period hypothesis in L2 acquisition**, the neurologist Wilder Penfield and Lamar Roberts were the ones who introduced the idea of a critical period without mentioning it. They linked it to brain lateralization with the capacity to acquire languages up to puberty, where the brain loses its ability to acquire a language. Later on, **Lenneberg** in 1967

coined the term Critical Period hypothesis stating that it is a limited expanse of time where language acquisition is possible. He expressed so through this statement:

> Many animal forms traverse periods of peculiar sensitivities… or learning potentials. Insofar as we have made such a claim for language acquisition, we have postulated nothing that would be extraordinary in the realm of animal behaviour' (Lenneberg, 1967)

There are two different versions of this period: the **strong and the weak version**. The former posits that after puberty, language acquisition is almost impossible whereas the latter highlights that although language acquisition is more difficult after puberty, it is still possible to acquire it.

Regarding strong and weak versions of the critical period, there is a related distinction to it: **critical** and **sensitive** periods. A **critical period** would involve a sudden clear cut from which it is not possible to progress in language acquisition whereas a **sensitive period** suggests that after that expand of time, language acquisition is more difficult but not impossible.

Moreover, Johnson and Newport (1989) established another distinction with reference to the critical period into two hypotheses: the *exercise hypothesis* and the *maturational state hypothesis.* The **exercise hypothesis** speculates that all human beings have an innate and powerful language learning capacity that should be exercised in order not to lose it with time or maturation. If this quality is exercised, it becomes available and integral for second language acquisition. With respect to the **maturational state hypothesis**, these authors believe that this powerful capacity that humans have when they are children is lost given to maturational effects. Consequently, the main difference between these two hypotheses is that with the exercise hypothesis, adults have the possibility to keep this capacity for language learning even after childhood and therefore, they have the same opportunities for learning a language.

The critical period hypothesis has several **components** that can be distinguished: an onset, a terminus, an intrinsic component, an extrinsic component and a system which is affected by stimulation. These terms are going to be mainly explained under the light of Lenneberg. To begin with, the **onset** of the CPH is the point when it starts. Lenneberg stated that that there is a specific maturational schedule which consist on a series of capacities that arise between the second and the third year of life. For Lenneberg, children are not capable of producing language given to maturational constraints such as the brain immaturity. However, in later research it was proved that children are capable of perceiving speech by the age of six months (Ruben, 1997 as cited in Singleton, 2005). Moreover, some research done in the 70s

proved that phonemic discrimination is possible for children who are only a few weeks old. Consequently, maybe Lenneberg was cautious in his position.

With respect to the **terminus**, Lenneberg established the end of the critical period towards the age of thirteen or the beginning of puberty. By this point, humans suffer a sudden decline in their capacities for acquiring a language which takes place just as the same time as the process of lateralization ends. As it has been previously summarized in figure 2, Pinker (1994) attributes this end to the fact that children have neural structures that are taken apart given that keeping these structures would suppose metabolic costs:

> Language-acquisition circuitry is not needed once it has been used; it should be dismantled if keeping it around incurs any costs. And it probably does incur costs. Metabolically, the brain… consumes a fifth of the body's oxygen and similarly large portion of its calories (1994:294-295)

The conclusion of the critical period has to do also with universal grammar and, concretely, to the impossibility of accessing it once this period has finished. With respect to that, there is a summary in figure 4 at the end of this section that establishes the termini for the different aspects of language. This is directly related to the following point: the intrinsic component.

With respect to the **intrinsic and the extrinsic component**, Lenneberg stated that there are innate mechanisms that the child possesses for language learning. This statement is contrary to behaviouristic proposals that would attribute language learning to a stimulus-response process. From this perspective, the question would be if adult language learners are able to access these mechanisms. In reference to the extrinsic component, Lenneberg did not pay much attention to it given that probably he was following the fashion against behaviourism.

Regarding the **affected system**, Lenneberg was not precise in describing what was affected by the critical period: phonology, lexicon or syntax among others.

New insights about the critical period and its onset and terminus were brought by Birdsong in 2001. He posed that there is a quick onset followed by an area of stability of peak sensitivity and that there is a plodding tendency which leads to the end of this period of sensitivity.

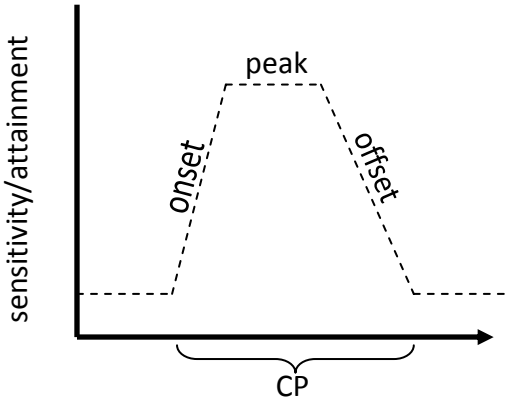


Figure 1: CP development (Birdsong, 2002)

As it was mentioned previously, Lenneberg did not account for the critical period in relation with different aspects of language such as syntax, morphology or phonology. In the next section, age-related effects linked to these aspects are going to be explained.

2.3.1 CPH and phonology

Phonology is thought to be the aspect of language in which age effects are mostly recognized. Even for those who are not used to distinguish different accents, even an inexpert ear can detect foreign accents. According to Scovel (1988, 2000) pronunciation is the sole part of language with physical properties that are determined by neuromuscular programming. As Ortega cites: 'one-third of the human brain's cortex is dedicated to controlling motor skills in the lower face, lips, tongue and throat, all involved in the production of skills.' (22:2009). Flege et al. (1999) also stated that factors such as the amount of education in L2 or the use of the second language had a bigger influence in aspects such as morphosyntax, but phonology seemed to be tied to physical factors. According to Flege, the stabilization of phonetic categories is completed around the age of five to seven. After that point, new sounds are filtered by these representations and therefore, it becomes very difficult to distinguish L2 sounds and to produce them. Hence, as Flege stated: '(foreign accents may arise) not because one has lost the ability to learn to pronounce, but because one has learned to pronounce the L1 so well' (Flege, 1999:25)

Of course, there are cases of exceptional learners such as the case of Julie, which is explained in section 2.4, that almost attain native-like levels, but it seems that the effects of age are partly overcame because of two factors: they were highly motivated towards the target of achieving native accent and furthermore, they had an L2 instruction of quality. This is also highlighted by Nikolov's study: 'these successful language learners want to sound like natives, they share intrinsic motivation in the target language which is often part of their

profession, or they are integratively motivated' (2000a:122 as cited in Johnstone, 2002). From this, it can be concluded that native-like levels can be acquired but they are most of the times exceptional cases.

2.3.2 CPH and morphology and syntax

Much of the issues that have to do with morphology and syntax are related to UG and the role that it has on these areas of language. Scovel states with respect to universal grammar and the use children made of it that it 'tunes itself in to the grammar of the particular language to which it is exposed' (Scovel, 2000). Moreover, Lakshamanan in 1995 highlighted that the grammar that children have when learning a language is influenced by UG (Johnstone, 2002) with this, the author does not mean that second language acquisition follows the same processes as in first language acquisition because different cognitive processes can arise from UG. The most precautious with the idea of different processing in adults and children is Singleton: 'There is no strong evidence that post-pubertal second language learners do not have available to them the capacities which Chomskyans think of as deriving from UG' (Singleton 1995).

2.3.3 CPH and lexicon

Researchers are increasingly recognizing the interaction that takes place between lexis and grammar. However, little attention has been paid to lexicon in relation with age. Probably, one of the reasons why a critical period has been neglected for this aspect of language is because humans keep on learning new vocabulary throughout all their lives even in their first language. However, first language vocabulary acquisition is different from second language vocabulary acquisition. There is evidence (Arenberg, 1983; Hussian, 1981: 6ff as cited in Singleton 1995) proving that younger subject tended to be better in remembering new lexicon than older subjects. As it is not easy to delimit a CPH in relation to vocabulary, it will be studied in depth in later sections.

To sum up, the following table by Singleton (2005) summarises the main proposals for the offset of different components affected by the Critical Period:

Summary of a range of proposals for CP termini	
Penfield and Roberts (1959)	Offset: age 9
Lenneberg (1967)	Onset: age 2 Offset: puberty
Molfese (1977)	Offset for phonology: age 1
Seliger (1978)	Offset for phonology: puberty
Diller (1981)	Offset for phonology: age 6-8
Scovel (1988)	Offset for phonology: age 12
Johnson and Newport (1989)	Offset of phase 1: age 7 Offset of phase 2: puberty
Long (1990)	Offset of phase 1: age 7 Offset of phase 2 for phonetics/phonology: age 12 Offset of phase 2 for morphosyntax: age 15
Ruben (1997)	Onset for phonology: 6th month of foetal life Offset for phonology: age 1 Offset for syntax: 4th year of life Offset for semantics 15th/16th year of life
Hyltenstam and Abrahamsson (2003)	Offset: shortly after birth

Table 2: Summary of a range of proposals for CP termini. (Singleton, 2005)

2.4 Against the Critical Period Hypothesis

What about those learners who seem to have obviated the critical period? There are exceptional cases of successful late L2 learners, such as the case of Julie (Ioup et al., 1994 as cited in Ortega, 2009), an English speaker who moved to Egypt when she was 21 because of her marriage to an Egyptian. Although she never received instruction in Arabic as L2, she was capable of learning the language in a naturalistic way and in some cases, people confused her with a native speaker of the language. She was tested through several means such as distinguishing Egyptian accents with very accurate answers but still, she was not considered as native-like in some other tests. However, even the main researcher (Ioup, 2005) acknowledged that there is greater evidence supporting age-related effects on second language learning. Nonetheless, the evidence presented by current research is against the CPH. Instead of determining a clear cut point where achieving native-like levels is almost impossible, research talks about a gradual decline in acquiring this ability. Further studies will be presented in the literature review section.

2.5 Rate and ultimate attainment in second language acquisition

In relation with the CPH, some of the generalizations from the works made in the 1970s were the following:

1. Adults proceed faster than children in the early stages of syntactic and morphological development (this is possible when time and exposure are held constant).
2. Within children, older children acquire faster than young children with the same conditions as the point above.
3. Those acquirers that were naturally exposed to second language when they were children achieve higher second language proficiency levels than those beginning as adults (Krashen et al, 1979/1982, reprint:161 as cited in Muñoz, 2006)

These three points are the basis for the very important distinction between *ultimate attainment* and *rate.* ***Rate*** refers on how fast the learners are in order to acquire a second language and concretely, in morphosyntactic aspects. With respect to this feature, late starters are faster than the younger ones; however, younger starters show a higher level of ***ultimate attainment,*** that is, the final stage that the learners reach in their language learning. So that, adults and older children are quicker in their learning pace and this could be attributed to the instruments and tests that have been used in the experiments: they could be more cognitively demanding and they could involve metalinguistic skills and therefore, adults have an advantage in it because they use these mechanisms in order to learn a range of aspects in second language learning (Ortega, 17:2009). Moreover, implicit learning is slower than explicit learning and therefore, the benefit that young learners have is not perceptible until they have been exposed to the second language for many hours. Several studies (Oyama, 1976; Patkowski, 1980; Long, 1990; Aoyama et al., 2008; as cited in Ortega, 2009) have proved that the initial advantage the older learners possess when learning languages disappears through time and eventually, younger learners catch up older learners. However, recent studies on age have shown that older learners keep their advantage over younger learners even after five years of instruction: this means that where language learning is only available through instruction, older learners are not caught up by the younger ones. This fact is due to the **amount of exposure** to the language: in five years, a student can be exposed to an average time of 540 hours in comparison to 7,000 hours of L2 exposure in L2 environments. As a result, Singleton (2003) that in L2 instructed settings more than 5 years were necessary to find enduring differences between subjects with different starting ages.

With respect to ***ultimate attainment***, this is an important term related to the end state of L2 acquisition. Regarding the studies that have been made in this line, it is essential to consider the native-like features because it is the main frame of reference to consider how successful learners can be in learning a second language. Studies around this idea suggested that those learners who began learning a language before puberty tend to have closer intuitions to those of native-speakers. With respect to ultimate attainment, Krashen, Long and Scarcella (1979/1982 as cited in Miralpeix, 2008) make a distinction between formal and informal learning contexts, referring the later to situations in which the learners are being informally instructed in naturalistic settings.

2.6 Naturalistic vs. instructed settings

With respect to the age factor, it is necessary to make a distinction between naturalistic vs. instructed settings. To begin with, there are different terms with respect to the settings in where language learning takes place.

Regarding the first occasion when the learner has his/her first contact with language, it is possible to distinguish between **age of onset (AoO)** and **age of exposure (AoE).** Age of onset refers to the age when the learner is in contact for the first time with the target language in naturalistic settings. Age of exposure refers to the first time when the learner is exposed to the target language for the first time in instructed settings. From this, also two other concepts derives, which are **age of arrival (AoA), length of residence (LoR)** and **length of exposure (LoE)**. Age of arrival has to do with the chronological age the learner has when he/she first arrives to a country where he/she is exposed naturalistically to a foreign language. Related to this, the term length of residence arises and it has to do with the time during which the learner has been exposed to the target language in the country where the language is spoken as the mother tongue. In reference to the length of exposure, it refers to the time during which the learner is in contact with the target language.

So, with respect to the terms explained above, an example of a **naturalistic setting** would be a family moving to a foreign country and the children being exposed to naturalistic language through interaction with friends, with everyday life situations or at school through other subjects which are taught in the target language. **Instructed settings** are those which most students are used to, such as learning the second language at school but the family has not moved to another country. With respect to naturalistic settings, Singleton (2000, as cited by Johnstone, 2002) highlighted that there is 'evidence of more rapid initial learning on the

part of adult and adolescent subjects, but also of younger beginners catching up with and beginning to overtake older beginners after about 12 months of exposure' (Singleton, 2000:22). In reference to instructed settings, Singleton stated the following:

> there is a consisting finding that learners exposed to a second language at primary and who then at secondary level are mixed in with later beginners do not maintain an advantage for more than a modest period over these latter' (Singleton, 2000:22)

This quotation is related to some of the issues found in the study: it is not surprising to find that late starters can surpass early starters in instructed settings.

The main differences between **naturalistic vs. instructed settings** are the following:

i. First, the difference has to do with **input**. In instructed settings, input is very poor because learners are only exposed to it within the class and in many cases they are not even exposed to it there because the teacher is not capable of dealing with the language due to the lack of confidence or shortage of knowledge.
ii. Second, the exposure to the language is time-bound especially when the teacher uses the **mother tongue in the classroom**: one of the very few opportunities the learners have to be exposed to the language.

Torras and Celaya discussed in their book the following:

> The problem one comes across in formal contexts is that the advantage in ultimate attainment of younger learners that seems to exist in naturalistic context cannot always be tested empirically in instructional settings […] so there is a need for studies measuring the long-term effects of an early introduction to a FL (2001: 105)

Nonetheless, in formal settings these long-terms effect are not considered as 'ultimate attainment'

With respect to the essential conditions in learning, Muñoz (2006) states that the concept of an end-state in language acquisition is not possible if in a formal context conditions such as the amount of exposure and the quality of input fail to be sufficient.

2.7 Characteristics of young and old learners

In Johnstone (2002) it is possible to find a very important clarification about such terms:

> There is of course no such thing as '*the*' younger learner or '*the*' older learner, since there are many stages in the life-long process between 'young' and 'old', at any of which there is considerable variation among individuals (11:2002)

Therefore, he poses the following **advantages of younger learners over older learners**:

- They tend to acquire the basis of the sound system easily.
- Their anxiety with respect to languages is lower and therefore, their affective filter is up and they are able to absorb a bigger amount of language.
- With respect to time, it is easy to conclude that they are going to have more time available than older learners because they start before them.
- With respect to biological features, starting languages at an early age allows the learner to make links within languages and this brings a huge benefit to children's language awareness.
- Languages are also about different cultures and exposing children to languages from very early ages leads them to be more intercultural conscious and this is a positive fact for their educational development.

With respect to older learners, they have some advantages that the younger learners do not possess, but once the younger ones grow up, they become also available for them, so these are some of the **advantages of older learners**:

- With respect to vocabulary acquisition, they are capable of transferring concepts of the world that the learners possess in their first language to the second language and it helps greatly in the aspect mentioned above.
- Regarding the discourse of conversation, adults may be more capable of negotiating meaning and therefore, they may obtain more feedback from the speakers or the teachers.
- Given to the experience they have acquired in learning, older learners possess a broader assortment of learning strategies.
- Older learners may have an established target beforehand and they may address their efforts towards it.

This is a short summary made by Johnstone (2002) with respect to some of the advantages that younger and older learners have. As it has been previously said, the characteristics that the older learners possess are not exclusive of them: they become available for younger learners once they get older.

2.8 The relation of age to second language vocabulary acquisition

The best introduction for this section is clearly summarized by a quotation of Ellis:

> SLA refers to all the aspects of language that the language learner needs to master. However, the focus has been on how L2 learners acquire grammatical sub-systems […]. Research has tended to ignore other levels of language. A little is known about L2 phonology, but almost nothing about the acquisition of lexis (1985:5)

Furthermore, Singleton pointed out: 'the age factor, as it relates to second language lexical acquisition, is not a matter that receives a great deal of attention' (1995:10). With these quotations, it is highlighted that most of the studies have been on grammar but there is little research on lexis, probably because it has been considered out of reach of the mechanisms of language acquisition and having more to do with memory affairs, and especially in the field of age: it seems that age effects do not affect vocabulary learning, but research does not point this out so. In the literature review section, this matter will be dealt with.

Still, since the 80s more research on vocabulary and second language acquisition has been done referring to early and late vocabulary learning, differences regarding learning styles depending on their age, word associations, meaning production, transfer and representation and access of words (Miralpeix, 23:2008). In relation to the study of the effects of second language acquisition, most of the efforts have been addressed to the study of phonology and morphosyntax probably because their effects are more visible and easier to study than those related to vocabulary. Some authors such as Long (1990: 272) also argue that 'there appears to have been little or no published work on ultimate attainment in the area of lexis and collocation'. Moreover, Harley and Wang (1997:24) taking into account the hypothesis stated by Lenneberg concluded that he 'seems to have viewed vocabulary learning as exempt from maturational constraints, or at least, that his primary concern was with syntax and phonology'

Moreover, vocabulary has been an abandoned area in age studies as it is developed in the next section.

2.8.1 Vocabulary: an abandoned area in age-related effects in second language learning.

Why is vocabulary not as central as syntax or phonology in age-related studies in second language learning? There are some answers for this question. In first place, the main objective of many studies in relation to age has to do with finding a critical period in first or second language acquisition. Muñoz (2006b as cited in Miralpeix, 2008) stated that the idea of a

critical period is very closely related to innatist ideas. Those ideas emerge from Chomsky's Universal Grammar whose main focus is on grammar and therefore, vocabulary does not have a place in there.

Moreover, Bialystok (1994) proposed another model of language organization which consisted on a language centre which is a similar idea to Chomsky's Universal Grammar, an area in charge of representing meaning which is similar to a place for storage of semantic knowledge and the last components are the Language-Specific details (LSD) which work providing the lexicon with two components: the pragmatic rules and the grammatical surface features. Once again, vocabulary is seen as exempt from maturational constraints.

One of the strongest reasons for why vocabulary has not been deeply studied in relation to age effects is because it is a life-long process. Even in our first language, humans can learn vocabulary throughout their lives. However, grammar is learnt at once: once it is learnt, the process ends. The following quotation clearly represents this fact:

> Lexical competence simply never approaches this kind of completeness [that takes place in grammar]. The learning of new vocabulary is clearly very rapid in childhood, and then slows down. But a person's vocabulary may nevertheless keep growing throughout their whole life. New meanings can be learned for old words, and new relations between words can be formed. (Stubbs, 1986:59 as cited in Miralpeix, 2008)

As it is stated in the quotation above, lexicon is a part of language that can expand continuously but syntax is considered as a set of rules.

As a summary of this section, the combination of all these factors can be the reason why the study of vocabulary and concretely of the age-related factors to vocabulary acquisition have been left apart, but there are some reasons why these studies are relevant with respect to vocabulary acquisition that will be dealt with in the next section.

2.8.2 Is it relevant to study the influence of age in vocabulary acquisition?

Sometimes, the question above has even been forgotten because it seemed that vocabulary is out of the reach of age-related effects. In this section, some of the ideas that Miralpeix exposed in her PhD (2008) are going to be summarized, mainly because such reasons are the ones that led me to address my research on age effects towards the issue of vocabulary acquisition.

There are two main reasons to study age effects on vocabulary: the first one has to do with the fact that vocabulary learning is not only explicit but also implicit. The second reason

has to do with biological factors: the areas in the brain which are in control of syntax are in different locations to those which are in charge of vocabulary acquisition. This does not mean that the area which is only affected is syntax given that, possibly, the areas that control vocabulary storage and production could possibly be affected by age-related effects too.

To begin with, it is important to make a distinction between implicit vs. explicit learning and incidental vs. intentional learning. Then, applying the definitions of implicit learning to vocabulary acquisition, for Ellis it happens 'when the meaning of a new word is acquired totally unconsciously as a result of an abstraction from repeated exposures in a range of activated contexts'. Similarly, explicit vocabulary learning implies that 'there is some benefit to vocabulary acquisition from the learner noticing novel vocabulary, selectively attending to it, and using a variety of strategies to try to infer its meaning' (1994b:219 as cited in Miralpeix, 2008) What is the difficulty on this distinction? Gass (1999) realized that it would be difficult to distinguish if the focus is on deliberate attention has really been previously premeditated.

According to the difference between *incidental* vs. *intentional* learning, Hulstijn (2001:271 as cited in Miralpeix, 2008) stated that incidental learning could be 'the learning of vocabulary as a by-product of any activity not explicitly geared to vocabulary learning' The main aim of incidental vocabulary acquisition is focusing on meaning and not on form (Huckin & Coady, 1999 as cited in Miralpeix, 2008) However, intentional vocabulary acquisition is 'any activity geared at committing lexical information to memory' (Hulstinj, 2001:271). Consequently, neither incidental nor implicit vocabulary learning are synonymous and that either intentional or explicit learning are.

Gass in 2003 proved through an investigation on focused attention and its effect on syntax and vocabulary that there were better results in syntax than in lexicon. Although it was thought that focused attention would be more useful for lexicon given that it seemed to be a more isolated aspect and less complex than grammar, the results proved that learners benefit from focused attention in syntax more than in vocabulary because a large amount of learning in vocabulary took place incidentally. With the examples given above, it is clear that vocabulary acquisition is not only a matter of explicit learning. Furthermore, if there are age effects in other areas of language, vocabulary acquisition cannot be left away of these effects.

With respect to the second argument, the fact that vocabulary and syntax acquisition are neurobiologically different and have different areas in the brain does not imply that age has no effect on vocabulary areas. It is true that the distribution of the areas involved in

language learning in early and late starters are different for syntax and very similar for lexicon, but this is not a direct proof to firmly state that lexicon is out of the reach of age effects. Just as in many age-related studies, research with hearing impaired people provides very useful findings in this field. Ledeberg and Spencer (2005) suggested a critical period in certain aspects of semantic development. Miralpeix (2008), in reference with this study, states the following:

> age of exposure seems to affect the automatic semantic processing in the long run as well as the lexical growth rate: this rate is faster for those children exposed early to the language, that is, those who have had a cochlear implant early in life. It is not clear, however, if the growth rate declines quickly after one or two years of age or if the decline is more gradual through childhood. In addition, they point out that if there is a CP for syntactic processing, it will also slow lexical growth. (35:2008)

Mayberry and Eichen also stated with respect to the matter of age of acquisition that it 'exerts one effect that reverberates throughout the processing of language structure' (1991:507 as cited in Miralpeix, 2008). According to this, difficulties in learning in other areas of language could originate from one point: lexical access. Therefore, it is not logical to separate those maturational constraints that affect other areas of language from lexicon.

2.9 Some general conclusions on age

As a summary of this section, some general conclusions exposed by Ellis (2008) are summarized in the following points:

- Adults are better in the first stages when rate is involved, especially in grammar but in the long run, the younger learners will overtake the older ones.
- Younger learners tend to be better at pronunciation than older learners. Moreover, they acquire better grammatical levels.
- There is no consensus around the matter of whether there is a critical period for learning a second language.
- Especially relevant to instructed settings is the fact that younger learners can outperform older learners providing that there is a sufficient exposure to the L2: if it is not enough, older learners will keep their advantage over the younger ones.

2.10 Aims of this paper

The main aim of this paper is to check if the general trend that is present throughout studies on age can be found in the adapted replication of the study by Miralpeix (2006). If hours of exposure are kept constant, it would be easy to think that the results on early and late starters would be very similar, but most studies show an advantage of the LS over the ES at the time the study is made but in the long run, early starters surpass late starters. This fact will be proved in the following section with several studies made on vocabulary and age-related effects.

One more aim of this study is to draw some conclusions on the effects that age can have in instructional settings and the possible applications that can be made in order to improve vocabulary learning. Moreover, some conclusion can be made on the statement 'the younger, the better' in instructional settings.

3 Literature review

In this section, a historical overview of the age factor and the critical period hypothesis is summarized. Later on, this literature review will be focused on vocabulary studies which are relevant for the aim of this paper.

3.1 Age and L2 acquisition

The greatest influence of age is on rate (a.k.a. speed) and success (a.k.a. ultimate attainment) of SLA.

In terms of **rate**, older learners are better: that is, if adults and children are exposed for the same period to the L2, the older learners are the ones who achieve higher levels of proficiency. However, this generalization should be modified in two important ways: **Snow and Hoefnagel-Höhle (1978)** have shown that the quicker learners are adolescents. They carried an experiment to study the natural acquisition of Dutch in two groups: English speaking children from 8 to 10 years old and adolescents from 12 to 15 years old. They were studied over a 10-month period in three different stages: after three months, after six months and at the end of the study. The results showed that adolescents were the best followed by the adults. Children obtained the worst result of the three of them. So that, it demonstrates that although age is a factor which improves learning capacity, performance may reach its peak during adolescence. After this period, performance declines. This study was also important to find that age was a relevant factor when making reference to morphology and syntax given that small differences were found in pronunciation tests. More conclusions towards this address are posed by some

author as **Krashen, Long, and Scarcella (1979)**. These authors based their conclusions in the experiment carried out by Snow and Hoefnagel-Höhle and stated that adults are superior to children in rate of acquisition and that older children learn more rapidly than younger children.

Furthermore, in terms of **pronunciation** age plays also an important role but in two different directions. On the one hand, Olsen and Samuel (1973) noticed that adolescents and adults speakers of American English outperformed children after 15-25 in sessions of German pronunciation. On the other hand, Cochrane (1980) researched the ability of 54 Japanese children and 24 adults to discriminate between English /r/ and /l/. In order to measure the results, the hours of naturalistic exposure were calculated in both groups. The result was that the adults needed 245 hours and the children 193 hours. In this case, the children did better than the adults because they needed less time to achieve the goal. Towards these two results, the research gives more support to Krashen, Long and Scarcella's generalization of the adults learning quicker than children. It is supposed that adults are better learners of grammar but children surpass adults in pronunciation, although the latter is not totally true because adults do better in this area in formal learning situations as it will be demonstrated with this study on vocabulary acquisition.

With respect to the critical period hypothesis, Lenneberg (1967) supports the CPH basing his arguments in **injuries to the right hemisphere**. He found cases in which children underwent surgery of the left hemisphere and they did not have speech disorders and they were able to recover language control but adults did not do so and the linguistic impairments that they suffered from were permanent. Furthermore, for him there is an onset at the age of two which coincides with the lateralization process. However, this does not demonstrate that it is easier to acquire languages before puberty.

Selinger proposed in 1978 that there are **multiple critical periods**. As the process of lateralization and localization of language functions is a gradual one that takes place during several years, different aspects of the language are influenced at different points in the process named before. Diller (1981) also supports the theory of multiple critical periods. His proposal was that young children were the only ones that could achieve authentic accent whereas adults learned cognitive aspects of L2.

Neufeld (1978) made a study with 20 adults who were English natives. They were given 18 hours of intensive training in Chinese and Japanese and they were rated from 1 as 'heavily accented' to 5 as 'unmistakably native'. The result of this study was that nine of them were rated as natives for Japanese and eight of them were rated as natives in Chinese. With this study the purpose is to demonstrate that with the adequate training, the level of proficiency in adults can be of a native. This study would not support the critical period hypothesis. As many studies, it received criticisms about the procedures used in the experiment. **Coppieters (1987)** made an experiment which had similar results to the one obtained by Neufeld but he found differences in the grammatical competence of the non-native speakers. He tested 21 speakers of French that begun to learn English as adults. There were no big differences among the speakers and the non-native speaker besides grammar.

3.2 Age and L2 vocabulary acquisition

With respect to the main findings on second language vocabulary and age, this section will begin with research on the beliefs of people on the effects of age on L2 learning. Then, some experiments on the naturalistic and instructed settings will be presented.

To begin with, if someone conducts a spontaneous survey on the street asking about who are better at languages, most of the people will answer that children are better in general. Maybe, if we ask with respect to vocabulary, many people may think the opposite: it seems that the common belief is that vocabulary is out of the reach of age effects given that it is generally thought that vocabulary has to do more with rote memory than with language acquisition mechanisms. However, in a study conducted by Burstall et al. (the NFER Evaluation Project, 1974 as cited in Miralpeix, 2008) most of the teachers that were asked about the introduction of French in primary schools in the UK when they were younger than 12 considered this measure as something positive because this 'would help pupils to acquire a wider vocabulary' and furthermore it was the time 'to get children speaking French quite naturally, assimilating new words and sounds without difficulty (1974:69-70 as cited in Miralpeix, 2008)

Parents also agree with teachers in the fact that the earlier, the better. Tragant & García (1997 as cited in Miralpeix, 2008) did interviews to those parents who had children who began their foreign language instruction at a young age. The results of their surveys were that parents thought that children were better in assimilating languages, vocabulary being the area in which they learn the most. With respect to my study, the results are not in the same line of this belief given that late starters surpass early starters in almost all aspects.

Nevertheless, these beliefs are only beliefs and empirical studies will prove them or not. For us, it is useful to distinguish between naturalistic vs. instructed settings studies.

3.3 Naturalistic settings studies

Most studies in **naturalistic settings** show **an advantage for older learners but only in the short term**. In the section above, the study conducted by Snow and Hoefnagel-Höhle has been described with respect to the results in general. With respect to vocabulary, they also conducted a study with English learners of Dutch using the Peabody Picture Vocabulary Test – PPVT – (Dunn, 1959 as cited in Miralpeix, 2008) in which once again older learners were superior to younger ones.

Swain in 1981 carried out a research in late French **immersion programmes** in Canada comparing them to younger students in early immersion programmes[1]. The results showed that the younger learners performed as well as older learners in the cloze test after 4,000 hours of exposure: older learners only needed 1,400. Harley (1986 as cited in Miralpeix, 2008) did a similar research with immersion students and found out older learners had a wider scope of **verb vocabulary** than the younger ones when the exposure was kept constant to 1,000 hours of exposure.

With respect to **vocabulary** and the possibility to be affected by a critical or sensitive period, there is a study by Spadaro (1998) who concluded that the closure time for this sensitive period is around 6. In his study in 1998 (Miralpeix, 2008) the lexical performance of four groups was analyzed: there was a control group of native speakers (N=10) and three groups of non-native speakers (N=38) who had different L1 and different ranges of age of onset of learning (0-6, 7-12 and later than 13). The three groups obtained similar results, but the one made of the younger learners were judged to be more native-like in an oral task. In Germany, there was another study conducted by Mägiste in 1987 (as cited in Miralpeix, 2008) with students from primary and secondary schools. They were asked to name pictures and numbers in German and Swedish. The results showed that elementary students used less time to acquire an elementary vocabulary in the foreign language. Conversely, if the task was more difficult (more cognitively demanding), the results of both groups were alike. She concluded that maybe there is not a clear cut point for second language learning but maybe an optimal age given that 'if the language task allows for the students' cognitive level, younger students will generally acquire that task with greater ease' (1987:56 as cited in Miralpeix, 2008)

[1] Although this is a language immersion, it does not mean that the students will have a contact with the SL outside it because it is not an immersion in naturalistic settings.

3.4 Formal settings studies

Studies in **formal settings** show a **better performance by the older learners, but only in the short run**. Crucially, in both contexts (naturalistic and formal) it seems that when time is held constant, older learners surpass younger learners: rate increases with age. There was a study by McLaughlin, Osterhout and Kim (2004 as cited by Miralpeix, 2008) in which they stated that adult L2 learning is not 'uniformly and laborious' because 'some aspects of the language are acquired with remarkable speed' (2004: 704 as cited in Miralpeix, 2008). They draw these conclusions after the first classes of adult learners in a SL. Ervin-Tripp stated that adults 'tend to pay most attention to vocabulary' (1974:123 as cited in Miralpeix, 2008). There are two more studies cited in Singleton, 1995 that are in line with this trend. The conclusion after observing classes of foreign language in primary schools was that pupils' acquisition of vocabulary is very poor.

However, there is a study by Yamada et al. (1980) with thirty Japanese students of English. They did not have previous exposure to English and they were divided in three groups with ages 7, 9 and 11. After the experiment consisting of matching the pictures with their word in two different learning sessions, they concluded that young children had better rote memory.

Nonetheless, contrary to what happens in naturalistic settings, **younger learners do not surpass older learners in the long run in instructed settings** although this cannot be totally assured given that there are no studies seeking ultimate attainment in instructed settings. In 1974, Oller and Nagato made a study with subjects from grades 7, 9 and 11. All of them were Japanese students learning English. Although there was an advantage of 6 years for the group of young learners, late starters outperformed them at the end of secondary education. Furthermore, Griffin in 1993 made a study on the long-term achievements of two groups made up of American learners of French at the end of high-school. The difference between the two groups was the age of onset: one of the groups began learning French between kindergarten and Grade 4 (ES) and the other one started between grades 5 to 8 (LS). Although the late starters received less instruction, they performed better than the early starters. In this study, vocabulary was measured through an oral storytelling task, a cloze and a composition. As Miralpeix states on this issue: 'The older the students were, after a similar amount of English language instruction, the greater the lexical complexity found in their compositions' (Lasagabaster & Doiz, 2003 as cited in Miralpeix, 2008)

The study that has been the main model for the present assignment was carried out by **Miralpeix (2006)**. In this study, the author compared two groups, the early starters whose onset age was 8 and the late starters, with an onset age of 11. Hours of instruction were kept constant (726 hours) but the age of onset was different. She used oral and written tests in order to compare two groups and the instruments *VocabProfile* by Nation (1995) and *D_Tools*, an instrument created by her and Paul Meara (2008). The results showed that **early starters did not show an advantage over late starters** in the different vocabulary measures she used in the experiment.

Therefore, I would like to conclude this section with a quotation by Singleton:

> The age factor operates in relation to second language vocabulary learning in the same way as it operates in relation to other aspects of second language learning, i.e. older beginners exhibit an initial advantage which is progressively eroded as younger beginners catch up with them and eventually overtake them. This pattern is clear in the naturalistic evidence, and is undisturbed by most of the evidence from formal instructional timescale that must be required for the eventual advantage of an early start to manifest itself under conditions of sparse exposure (Singleton, 1995:20)

This quotation is a good summary for what most of the studies have proved. Most of the results in this section have showed an advantage of the older learners over the younger ones but all the authors warn us about the fact that this is something temporal and these results can change in the long run.

4 Measures to assess this study

In order to offer a clearer explanation of the procedures of the research, I found it important to include in this section an explanation of the measures that have been undertaken here.

One of the measures that have been used in the study is **Type-Token Ratio (TTR)**. As Miralpeix defines in his PhD: 'TTR is the number of different words [types] as a ratio of the total number of running words [tokens] in a text' (57:2008). The aim of this measure is to show the possibilities that the learner has to repeat the same words. The main problem of this measure is that it is length sensitive, therefore it could be a difficulty in the study that was partially solved as it will be related in the section 'method'.

Another measure that has been used here is **Lexical Density (LD)**. Its main aim is to evaluate lexical richness. It measures the proportion of *content words* versus *function words*

(Ure, 1971 as cited in Miralpeix, 2008). It is not a good measure if the study is evaluating low-level students because they tend to use telegraphic style, but this seems not to be the case of this research.

To compensate the possible mistakes that the previous measures have presented, I used one more measure which is the ***D measure***, proposed by Malvern and Richards (1997, 2002 as cited in Miralpeix, 2008). *D* works using a process of curve fitting and it is an index to evaluate lexical diversity. It is more precise and informative than TTR because this program represents several values of TTR, concretely, 'how TTR varies over a range of token sizes for each speaker or writer' (Miralpeix, 58:2008). There are two more advantages: the length of the text has not to be standardized because *D* uses all the data available and furthermore, short texts can be analyzed: fifty tokens are enough. This is a great advantage if the researcher is planning to work with low-level learners because they are likely not to produce much.

An additional measure was obtained via *VocabProfile* which relates to **word families**. The term 'word family' has been defined by Bauer and Nation (1993). The definition they provide in their study is the following:

> a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately. So, *watch, watches, watched,* and *watching* may all be members of the same word family for a learner with a command of the inflectional suffixes of English. As a learner's knowledge of affixation develops, the size of the word family increases. The important principle behind the idea of a word family is that once the base word or even a derived word is known, the recognition of other members of the family requires little or no extra effort. (253:1993)

Other measures that have been taken into account are the **tokens**, **types** and the **tokens per type**. Although these measures sometimes are not useful because for example in many cases it does not depend on the age of the subject if they produce more or less tokens, it could be significant at some point so I decided to show some data related to these measures.

To finish with, these measures have been obtained through ***VocabProfile*** (Nation, 1995) and ***D_Tools*** (Meara & Miralpeix, 2008). In later sections, they will be explained in depth.

5 Hypotheses

In this section, some research questions and hypotheses are going to be presented. When I first decided to do this final assignment on age, I chose to address my efforts towards the study of vocabulary because from my point of view, it is the aspect in language that has been forgotten in terms of age effects. Therefore, these were my main concerns when I started this assignment.

5.1 Research questions

The main research question with respect to this study is: **does age, and concretely, age of onset really have an influence in vocabulary acquisition?** As it has been previously explained, vocabulary is also influenced by age effects although these effects cannot be spotted in many cases in instructed settings. In section 3.4, we saw that older learners acquire some aspect of language very quickly, so it leads to another question: **is the age effect in favour of the late starters or the early starters with respect to vocabulary?** Taking into account that older learners are quicker in some aspects of language, maybe an early start does not have an effect on vocabulary acquisition. Some of these general questions will be refined in the hypotheses.

5.2 Hypotheses

These are the hypotheses for my study:

H_1: Late starters will perform better than early starters in TTR measures in the non-standardized free-writing activity. TTR shows wider vocabulary knowledge, which is why late starters are supposed to outperform early starters in this measure.

H_1: AoO and TTR: late starters > early starters

H_2: Early starters will perform better than late starters in lexical density measures in the non-standardized free-writing activity. Miralpeix in his paper on *Age and Vocabulary Acquisition in English as a Foreign* Language highlights the fact that the writing style of the younger learners could influence lexical density. Such measure is supposed to be in favour of the early starters because the measure could be affected by length, in this case, short texts with telegraphic style that early starters are more likely to produce.

H_2: AoO and Lexical Density: late starters < early starters

H_3: In the standardized free-writing activity, late starters will outperform early starters more significantly than in the non-standardized free-writing activity. Once length is standardized, it is possible to observe a better performance of late starters over the early starters with the same number of tokens.

H_3: AoO and all lexical measures in standardized writing: late starters < early starters

H_4: The older the age of testing is, the more correct answers the students will get in the controlled writing activity. This could happen because of a higher cognitive development as years go by. It is more likely to happen in controlled writing because they have limited correct answers for the gaps and therefore, more developed cognitive skills must be used in order to achieve better results.

H_4: Positive correlation between AoT ↔ Controlled Writing

H_5: In general and with respect to all the measures studied in the different activities, late starters will perform better than early starters except for those measures affected by text length. As it has been previously explained in the L2 literature review, older learners have an advantage over younger learners in instructed settings and therefore, they are expected to outperform them. Those measures that can be affected by length are more likely to be in favour of the early starters given to the fact that they are expected to write in telegraphic style and with fewer tokens.

H_5: AoO and all lexical measures not affected by length: late starters > early starters

6 Method

6.1 Variables

The variables that appear in this study are the following:

- Independent variables:
    - Age of Onset (AoO): Age at which the subjects started learning their second language. As the participants are investigated within a learning context, the

appropriate term is AoO because it does not happen in a naturalistic context. This variable has two levels (which correspond to two groups): early starters and late starters which are used in order to perform all the T-Tests. In the correlations, this independent continues to be the independent one given that it is the only measure together with age of testing that depends on itself.

  - Age of Testing (AoT): Age at which learners were tested.

- Dependent variables:
  - Tokens: Number of running words in the text.
  - Types: Different word classes in the text.
  - Type-Token Ratio (TTR): As it has been explained in section number 3, it is the figure obtained from the division of the number of types by the number of tokens.
  - Tokens per type: As its name indicates, it is the figure obtained after dividing the total number of tokens by the number of types.
  - Lexical Density (LD): proportion of content words versus function words.
  - *D:* Similar to TTR but not affected by length.
  - Families: As previously explained, a word family is a group of words organized around a base form with all its derived forms.
  - Tokens per family: Number of words which pertain to a determined word family.
  - Types per family: Number of different types for each family.
  - Correct answers: Number of right answers in the gap-filling activity.

6.2 Participants

The following figure presents a summary of the participants of the study:

	Length of Exposure (LoE)	**Age of Onset (AoO)**	**Age of Testing (AoT)**	**Marks (in English)**
Early Starters (N=23)	9.57	6.28	15.51	7.52
Late Starters (N=21)	9.90	8.21	17.88	7.90

Table 3: Number of participants and their means of length of exposure, age of onset, age of testing and marks.

The participants for this study are students from IES Fray Luis de Granada (Granada, Spain) and IES San Juan de la Cruz (Úbeda, Jaén). The reason for having subjects from different centres is the difficulty to find subjects with the same length of exposure. There are in total 44 students. They are divided into two groups: **Late starters** (N=21), students from 2nd Bachillerato and with a mean length of exposure of 9.90 years at the time of the study. Their mean of age at testing was 17.88. The second group is the **Early starters** group (N=23), students from 3rd and 4th of Compulsory Secondary Education (CSE) with a mean of exposure of 9.57. Their mean of age at testing was 15.51.

After applying the T-Test to check if the difference with respect to length of exposure between early starters and late starters was significant or not, the results showed that it was not (p=0.123, t=1.576). This group is of different educational levels given the impossibility to find more subjects for my study, because if I chose only those from 3rd of CSE, the number of subjects would have been very reduced (N=13) and my aim was to find two groups with a similar number of subjects.

Furthermore, I tried to keep constant their level of English and the main measure I could find was their final mark for the previous year. Late starters have an average mark of 7.90 and Early starters have a mean of 7.57. After applying a T-Test to check the observed difference between the two groups, the result was that it was not significant (p=0.474, t=0.723).

6.3 Materials

In order to collect the data, an adaptation from the study of Miralpeix (2006) has been done. There were three questions in the materials: a free writing activity, a controlled activity and a semi-controlled activity.

The first one was a **free writing activity** in which they had to write about themselves: their hobbies, their family or whatever that was related to them. By these means, they could express themselves in similar terms but at the same time in different directions. They were asked to write about 100 words although they had the freedom to continue if they wanted to.

The second activity was a controlled activity from the study of Miralpeix, 2006. The activity consists of a gap-fill exercise about Little Red Riding Hood. It has thirty gaps in it distributed in a short introduction and three more paragraphs.

The last activity was a story-telling activity taken also from Miralpeix, 2006. The students were presented with six frames constituting a story. They had a frame and three lines to fill in for each of the cartoons. Although there were three lines, they could use more or less lines if they wanted to and therefore, the number of tokens was not controlled, something done on purpose to observe if there was a difference between the late and the early starters.

All these instruments are available in the Appendixes. The main difference of these instruments with those of Miralpeix is that I implemented them in the written form whereas she made the free composition and story-telling orally. The lack of time and the lack of resources and judges (I am the only one) led me to do it in the written form.

6.4 Data analysis

Once all the materials were collected, I used ***VocabProfile*** [2](Nation, 1995) in order to analyze the free writing and the semi-controlled writing. Moreover, in order to measure D, the program ***D_tools*** *v.2*.0[3] has been used. Furthermore, *Microsoft© Excel* together with the plug-in ***EZAnalyze***©[4] have been used in order to analyze the results **with both descriptive and inferential statistics**. More details on the use of both programs in the section *procedure.*

6.5 Procedure

In this section, I will present the whole procedure, from data collection, through data coding and analysis to the statistical analysis.

6.5.1 Data collection

The first step in this study was to **collect the samples** from the students. In order to do that, I asked for permission in my centre and in the one in Ubeda to carry out the study. Before implementing the tests, I asked the students about the first year in which they started studying English as their second language. After I checked that I had most of the subjects in the classrooms, I administered the test without notifying them about this exercise given that if they had known about the test, they could have studied something related to it and therefore, results would not have been accurate for the study. Furthermore, I presented this exercise as a test in order to obtain the best results from them and this was the best way to motivate them.

[2] This instrument is available online at http://www.lextutor.ca/vp/eng/
[3] Software available at Paul Meara's webpage *Lognostics*: http://www.lognostics.co.uk/tools/index.htm
[4] This plug-in can be freely downloaded from http://www.ezanalyze.com/download/

The day of the test, they were separated to avoid cheating and their level of anxiety was reduced by telling them that although it was an exam and it was important, the final mark would influence their mark only in a positive way. That was completely true because the exams were checked afterwards and the marks were given to their teachers to count for their final evaluation. So, they were more relaxed and their level of anxiety would not highly influence them. The tests were carried out in different hours and different days but the students who did the exercise after the first group did not know about the test given that those students did not have a relationship and they were not informed. The span of time allotted was one hour. Many students handed it in before, none of them needed more time. Previous to the test, a **background questionnaire** was given to the students in order to collect data about private lessons, their age of onset and their date of birth among others (see appendix 1). The complete test can be found in the Appendix 2. Both, the exercises in the texts and the complete questionnaire, were written in Spanish in order to avoid any misleading instructions.

6.5.2 Data manipulation/transformation

After all the tests were collected, a selection was made to choose the subjects that were of an approximate length of exposure and suitable for the study by obtaining the information from the previous test.

Once the tests were selected, all the free-compositions and the story-telling were **computerized** in order to use the instruments above related. Furthermore, all the free-composition exercises were saved in individual **plain text** (*.txt) files because it was necessary to process each essay via *D_tools*.

Next step was copying and pasting one by one the **free-writing activities** of the students first and then the semi-controlled activities into the *Web VP v 3 Classic* (part of the *VocabProfile* tools). Every time a piece of material was inserted, it was monitored to correct mistakes such as *childrens* into *children*; those words which did not correspond to English words were deleted together with Spanish names. Abbreviations such as *gonna* were changed into *going to*. Moreover, different spellings were changed consistently into one form, such as *program* and *programme.* This stage was completely necessary given that if words such as Spanish names were left in the text, the program would understood that it is another type of an English word and therefore, the analysis would have been misleading. The *gap-fill* exercise was corrected to check the number of correct answers.

6.5.3 Data coding.

Once all the data was obtained from the *Web VP v 3 Classic,* all the figures were introduced into an excel sheet in which there were two different groups: early starters and late starters. More information was collected from the questionnaire (see appendix 1) and introduced in this chart, but information such as private lessons turned into not being necessary and relevant for the study. Their date of birth was used to calculate their exact age on the day of testing (16th of May, 2011). Additionally, their sex was also introduced in the chart, but this data was not finally analyzed.

6.5.4 Data analysis

Another step in the study was to take all the essays and analyze them in *D_Tools* (Meara & Miralpeix, 2008). In order to do so, all the texts from the students were taken individually by their *.txt file and put into the programme. Previously, those texts were similarly adjusted to the measures exposed above. Firstly, there was an attempt to analyze the semi-controlled activity via *D_Tools* too, but some texts were so short that they did not even reach the minimum tokens to be analyzed. So that, facing the fact that there were going to be some gaps in the data, it was finally decided not to measure *D* for this activity.

6.5.5 Data standardization

Furthermore, another step was introduced in the study. As length in the free-writing activity was different for each of the students, the tokens were standardized to fifty to check if length had some influence in the results that were obtained in the free-writing activity.

6.5.6 Statistical analysis

Once all the data was consciously organized and prepared into Excel sheets and performed descriptive statistics (mean). The plug-in *EZAnalyze* was used to perform inferential statistics (***T-Tests*** on independent samples on the data, as well as **correlations** by using the *Pearson* correlation test). **Alpha** was set at 0.05. After that, correlations between the age of onset and different values were obtained. The results are in the following section.

7 Results

In this section, there are going to appear several letters which stand for the following terms:

- F: Free-writing activity (unstandardized)
- FS: Standardized Free-writing activity

- SC: Semi-controlled writing activity

Moreover, these letters will appear together with the different measures that have been obtained corresponding to the figures obtained for each activity.

7.1 T-tests: Free writing and age of onset

1. UNSTANDARDIZED RESULTS (Hypothesis 1 and 2). The following charts summarize the results in the different parts of the test that was provided to the students. In figure 5, the results for the free-writing activity are reported (Complete results on appendix 3):

	Tokens	Types	TTR[5]	TPT[6]	LD[7]	D	Families	TPF[8]	Types PF[9]
ES N=23	126.26	66.26	0.53	1.89	0.47	46.07	54.61	2.18	1.11
LS N=21	109.62	66.24	0.61	1.66	0.44	54.96	56.10	1.87	1.11

Table 4: Mean scores for the free-writing activity

In this charts, the figures for tokens, types, type-token ratio, tokens per type, lexical density, D, families, tokens per family and types per family are compared. These figures correspond to the **free-writing activity** without having standardized the number of tokens. In order to make the results clearer, the following figures will represent the results graphically:

[5] Type-token ratio
[6] Tokens per type
[7] Lexical Density
[8] Tokens per family
[9] Types per family

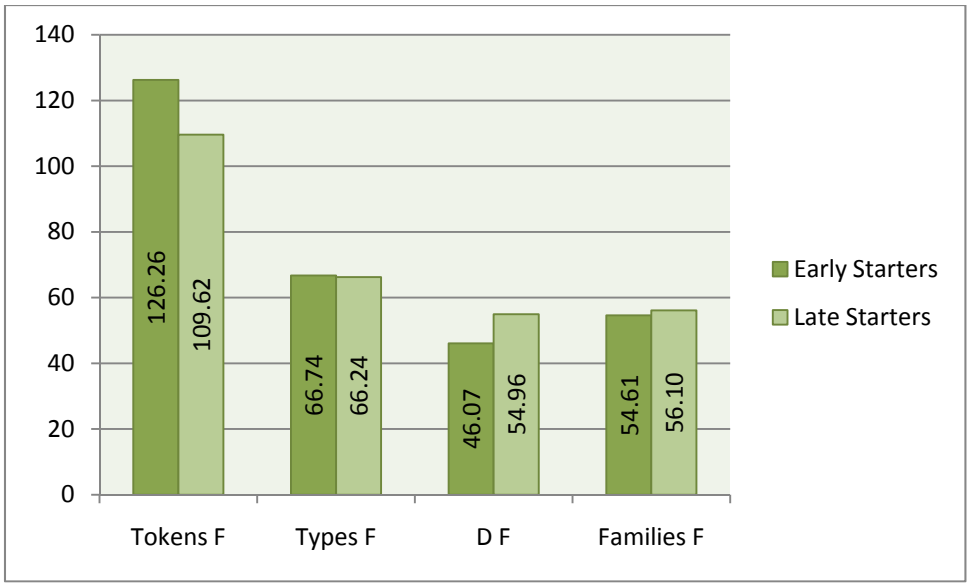


Figure 2: Means of tokens, types, D and families in Free writing

Figure 6 and 7 are organized as presented because the scale fits better to the figures presented in them: for tokens, types, D and families, values ranging from 0 to 140 are more suitable whereas for TTR, tokens per type, lexical density, tokens per family and types per family values range from 0 to 2.5 and results are more clearly shown.

Surprisingly, there is a higher number of **tokens** for the early starters (126.26) as it can be seen in figure 6 but there is not a significant difference (p=0.123, t=1.576). There is also a little difference between the number of **types**, once again in favour of the younger ones (66.74), but it is not significant either (p=0.923, t=0.097). With respect to ***D***, the difference is in favour of the late starters (54.96) and although it is not significant (p=0.075, t=1.824) but it approaches signification. With respect to the number of **families**, late starters (56.10) performed better than early starters (54.61), but once again the difference is not significant (p=0.716 , t=0.366).

The following figure represents the rest of the measures for the free-writing activity:

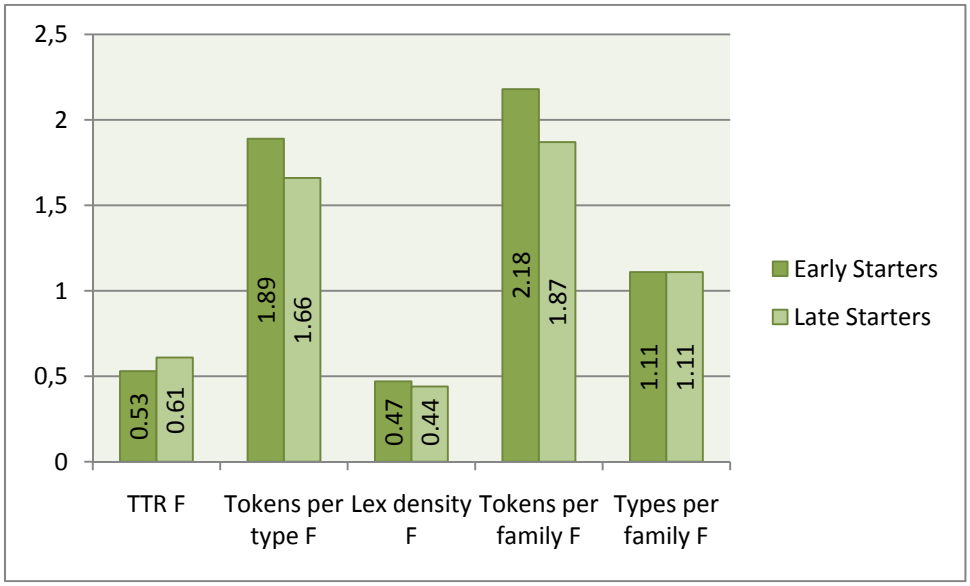


Figure 3: Means of TTR, tokens per type, lexical density, tokens per family and types per family in free writing

In figure 7, it is possible to find some differences. When discussing **TTR**, late starters (0.61) appeared to be better than younger ones (0.53) with a highly significant difference ($p < 0.001$, t=4.792). This would confirm **hypothesis 1.** With respect to **tokens per type**, there is a higher number for the early starters (1.89) ($p < 0.001$, t=4.566). in reference to **lexical density**, numbers are in favour of the early starters (0.47) with a significant difference (p=0.019, t=2.432). This result would be in favour of **hypothesis 2**. In reference to **tokens per family**, early starters (2.18) write a higher number of this value being the difference very significant ($p < 0.001$, t=5.029). With respect to the **types per family**, there is no difference at all.

2. STANDARDIZED RESULTS (Hypothesis 3). Tokens in the free-writing activity were standardized, that is, the number of tokens was 50 for all the texts and these were the results:

	Tokens[10]	Types	TTR[11]	TPT[12]	LD[13]	D	Families	TPF[14]	Types PF[15]
ES N=23	50	32.05	0.64	1.58	0.47	37.99	27.26	1.79	1.11
LS N=21	50	34.61	0.69	1.45	0.44	47.65	29.9	1.63	1.15

[10] There is the same number of tokens in both groups because they are standardized.
[11] Type-token ratio
[12] Tokens per type
[13] Lexical Density
[14] Tokens per family
[15] Types per family

Figure 4 Standardized free-writing activity

A more visual report of this data can be found in the following figures:

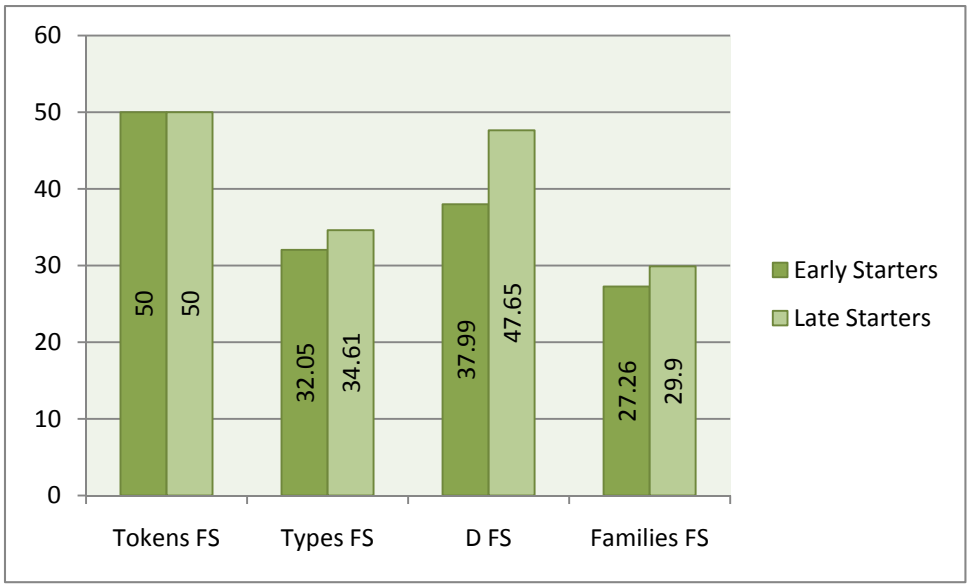


Figure 5: Means of tokens, types, D and families in the standardized free-writing activity

In this figure, there is no difference in the number of **tokens** given that they have been standardized to avoid differing length effects. With respect to the number of **types**, late starters (34.61) wrote a higher number of different types with respect to early starters (32.05) with a significant difference (p= 0.024, t=2.344). With respect to the measure ***D***, it is possible to find again a difference between the two groups in favour of the late starters (47.65), but this time the difference is significant (p=0.031, t=2.237). In reference to the amount of **families**, late starters (29.9) overpass early starters (27.26) and the difference is significant (p=0.008, t=2.789).

Figure 6 summarizes the rest of the measures:

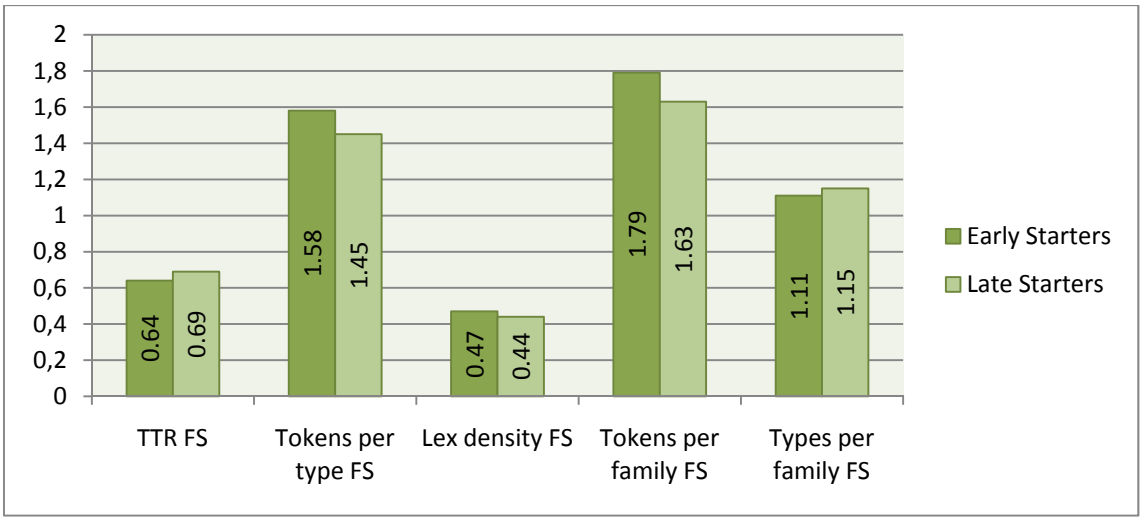


Figure 6: Means of TTR, tokens per type, lexical density, tokens per family and types per family in the standardized free-writing activity

With respect to **TTR**, late starters (0.69) surpass early starters (0.64) with a significant difference (p=0.025, t=2.324). In reference to the number of **tokens per type**, early starters (1.58) wrote a higher number of them with a significant difference too (p=0.027, t=2.295). **Lexical density** is in favour of the early starters (0.47) but the difference is not significant (p=0.140, t=1.505). In reference to **tokens per family**, the group of the early starters (1.79) wrote a higher number of them with a significant difference (p=0.020, t=2.421). However, when referring to **types per family**, late starters (1.15) outperform early starters but the difference is not significant (p=0.299, t=1.053).

In general, the results in almost all the measures are in favour of the **late starters**, which confirms **hypothesis 3**.

7.2 T-tests: Controlled writing and age of onset (Hypothesis 4)

With respect to the **controlled writing activity**, the results were the following:

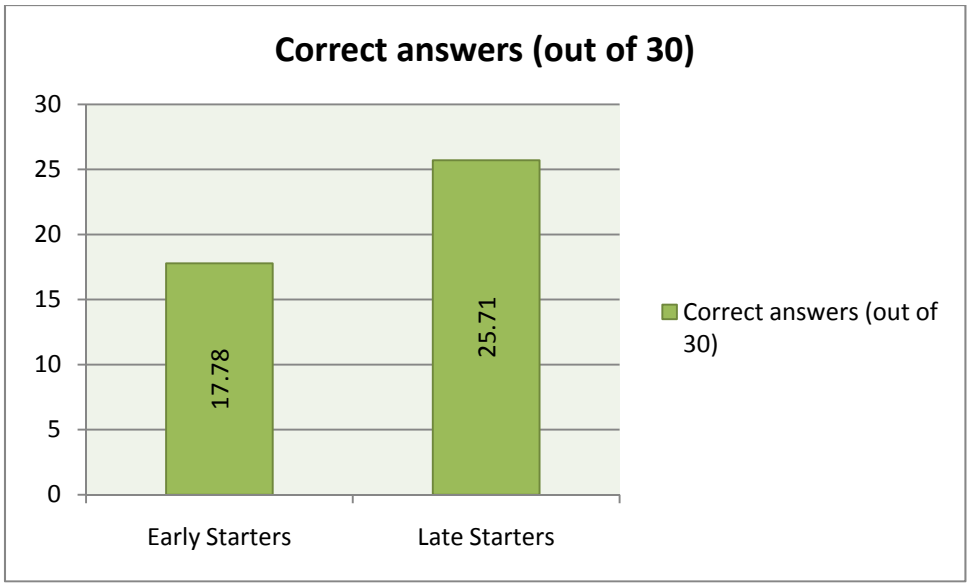


Figure 7: Means of correct answers out of 30 in the controlled writing activity

As it is possible to observe in figure 7, late starters significantly outperformed early starters (p=0.001, t=3.681) with a mean of 25.71 correct answers. Early starters achieved a mean of 17.78 correct answers. This confirms what was stated in **hypothesis 4.**

7.3 T-tests: Semi-controlled writing and age of onset

In reference to the **semi controlled activity**, results were as follow:

	Tokens	Types	TTR[16]	TPT[17]	LD[18]	Families	TPF[19]	Types PF[20]
ES N=23	72.65	38.65	0.55	1.86	0.47	34.26	2.03	1.08
LS N=21	98.05	48.81	0.51	2.00	0.43	42.95	2.23	1.11

Figure 8 Semi-controlled activity

In figure 8, the figures for tokens, types, type-token ratio, tokens per type, lexical density, families, tokens per family and types per family are compared. In order to make the results clearer, the following figures will represent the results graphically:

[16] Type-token ratio
[17] Tokens per type
[18] Lexical Density
[19] Tokens per family
[20] Types per family

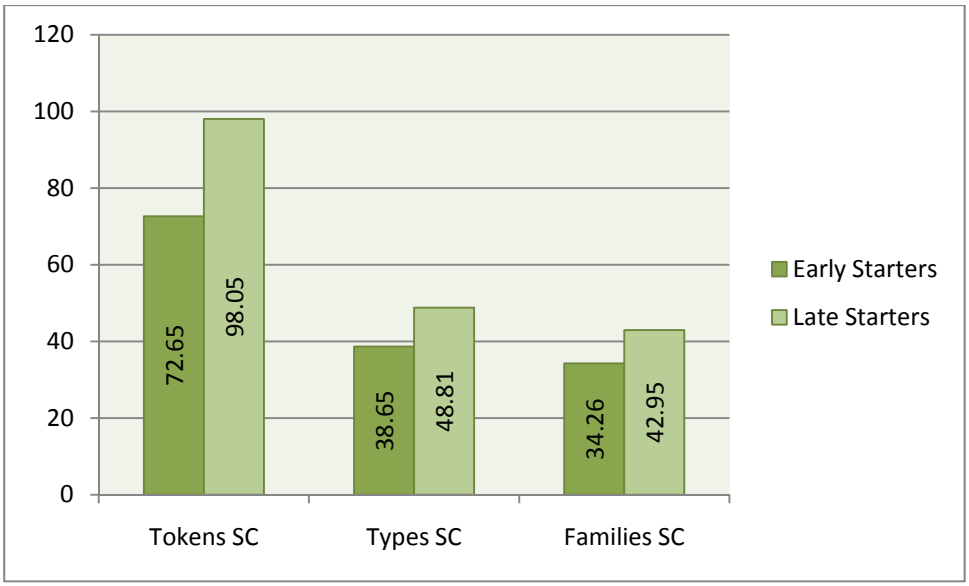


Figure 9: Means of tokens, types and families in semi-controlled writing

In figure 9, it is possible to observe that late starters outperform early starters in all measures. In particular, late starters (98.05) write a higher number of **tokens** with a significant difference (p=0.005, t=2.984). Therefore, late starters outperformed early starters (72.65). With respect to the number of **types** that both groups wrote, there is also a significant difference in favour of the late starters (48.81) (p=0.015, t=2.533). In reference to the number of different **families**, once again late starters (42.95) outperform early starters (34.26) with a significant difference (p=0.019, t=2.434).

The following figure represents the rest of the measures for the semi-controlled activity:

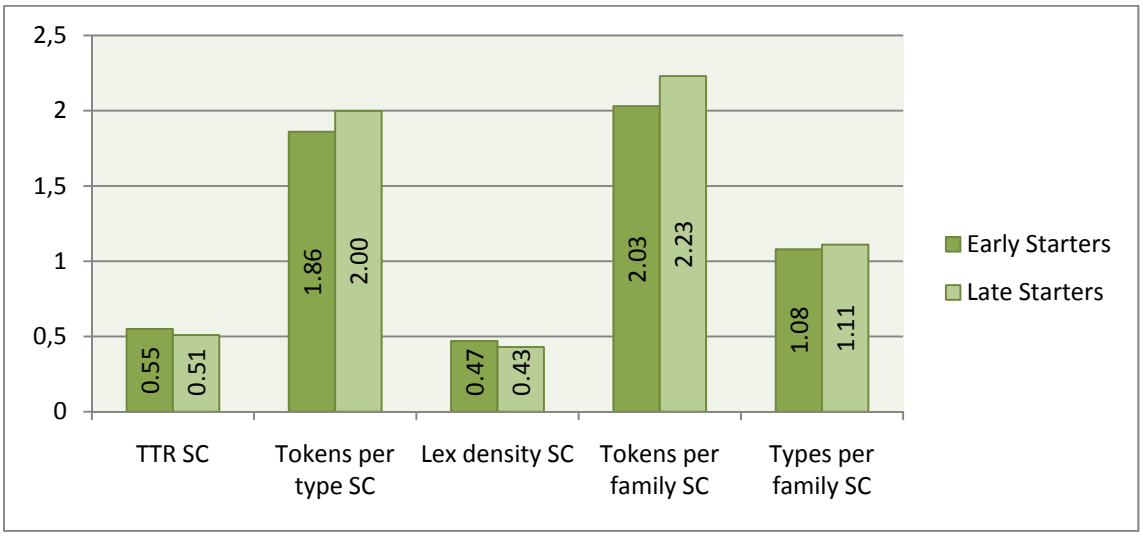


Figure 10: Means of TTR, tokens per type, lexical density, tokens per family and types per family in semi-controlled writing

With respect to the **TTR**, early starters (0.55) outperform late starters (0.51) with no significant difference but remarkable (p=0.072, t=1.845). With respect to **tokens per type**, late starters (2.00) wrote more tokens for each type. The difference once again is not significant but remarkable (p=0.072, t=1.846). In **lexical density**, early starters (0.47) significantly outperform late starters (0.43) (p=0.002, t=3.246). With respect to **tokens per family**, late starters (2.23) wrote a higher number of them with a significant difference (p=0.030, t=2.253). In the case of **types per family**, late starters (1.11) outperform early starters (1.08) in that figure. The difference is not significant but almost (p=0.051, t=2.006)

Several **correlations** were made to illustrate in a more detailed way the relationship **between AoO and several dependent variables measuring lexical attainment in L2**. Here are some of the most interesting ones:

7.4 Correlations: Free-writing activity

The first interesting correlation that it is possible to find related to the AoO is the following:

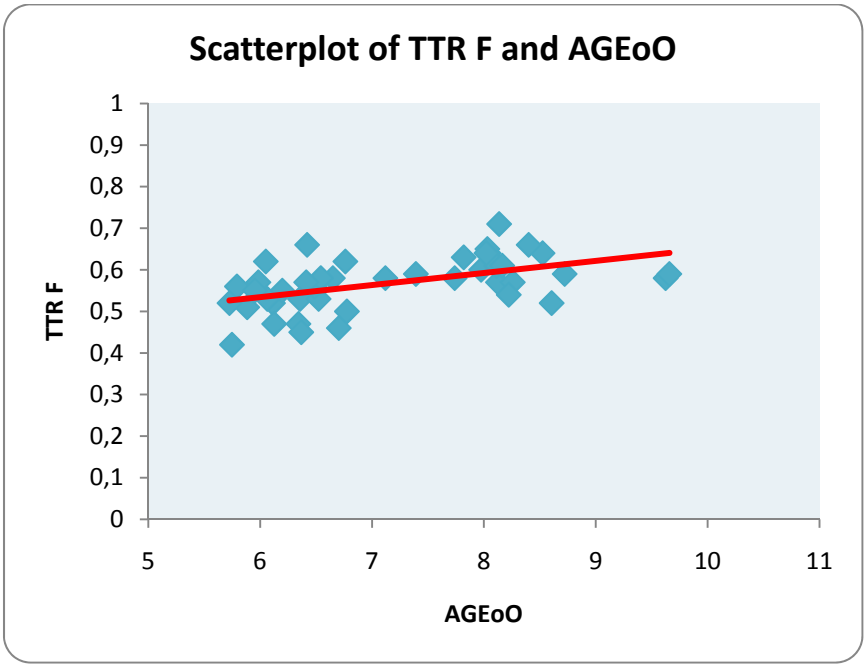


Figure 11 Scatterplot of TTR in free writing and AoO

In this figure, there is a **positive correlation** between the **AoO** and **TTR** in **free writing**. The present correlation is **significant** ($p < 0.01$, r=0.517). Dots are in two main blocks: between ages 6 and 7 and between the ages 8 and 9. These blocks correspond to 1st and 3rd of primary education respectively and their ages vary according to their date of birth which has been previously calculated via Excel.

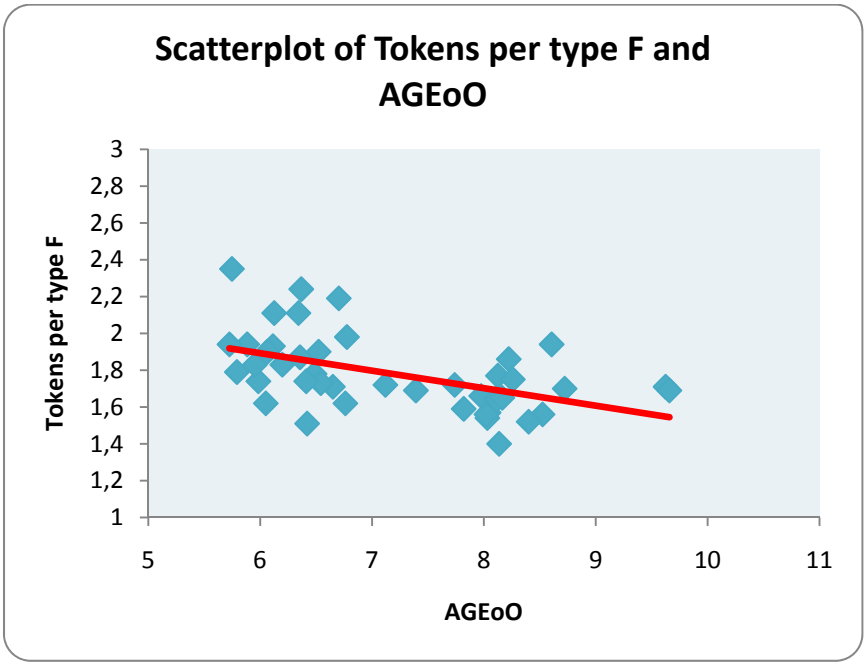


Figure 12 Scatterplot of tokens per type in free writing and AoO

With respect to this **correlation** presented in figure 16, the **age of onset** and **tokens per type** have been evaluated. This negative correlation is **highly significant** ($p < 0.01$, r=-0.512) and what it shows is that the older the student started learning English, the fewer tokens s/he writes per type given that s/he is able of writing more types.

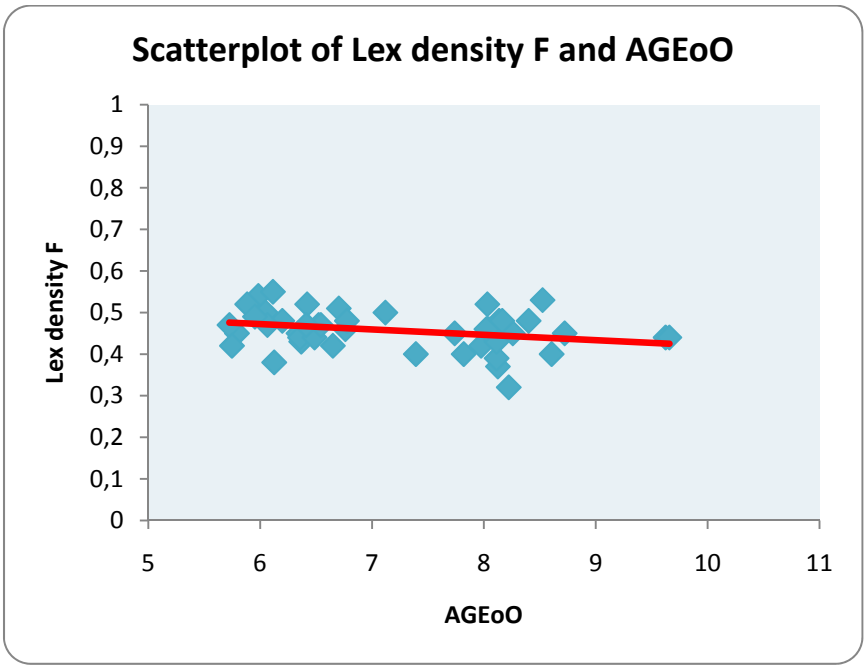


Figure 13 Scatterplot of lexical density in free writing and AoO

In this **negative correlation**, it is possible to observe how the value of the measure **lexical density** decreases as the **age of onset** increases. This correlation is statistically **significant** (p=0.050, r=-0.298).

The rest of the values and correlations of free writing activity are in the following chart:

Measure	Tokens F and AoO	Types F and AoO	TTR F and AoO	TPT[21] F and AoO	LD[22] F and AoO	D[23] F and AoO	Fam[24] F and AoO	Tok PF[25] F and AoO	TyPF[26] F and AoO
p value	0.308	0.767	< 0.01*	< 0.01*	0.050	0.092	0.539	< 0.01*	0.100*

* = significant

Figure 14 *p* value in correlations for each of the measures in free writing

There is one more significant correlation that has been illustrated below that is the **amount of tokens per family**. The older the learner started learning English, the fewer tokens per family they write as can be seen in figure 15:

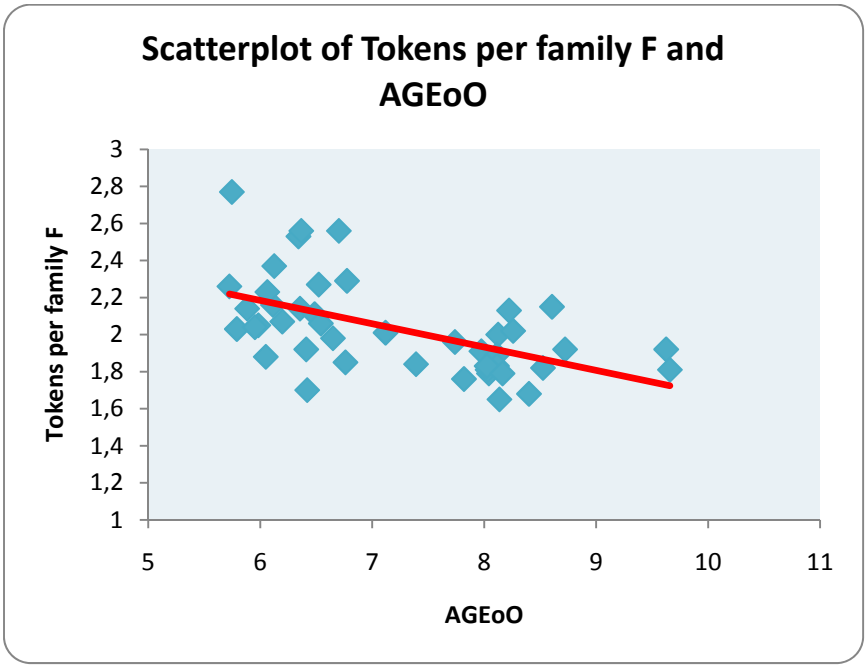


Figure 15: Scatterplot of tokens per family and age of onset in free, non-standardized writing.

The correlation is statistically significant (p < 0.01, r=-0.550)

[21] Tokens per type
[22] Lexical Density
[23] D value
[24] Families
[25] Tokens per family
[26] Types per family

The correlation between tokens and AoO is not significant just as the number of types, the families or the types per family they write. D value is not significant but remarkable. Besides that, values such as TTR, tokens per type, lexical density or tokens per family are significant just as the data shows.

7.5 Correlation: Controlled writing activity

The following figure represents the **correlation** between the **age of onset** and the **correct answers** the students obtained in the **controlled writing activity**. This correlation is statistically **significant** (p=0.010, r=0.496). The bigger is the value of the age of onset, the more correct answers they achieve. Of course, as it can be observed, there are exceptions, but in this case it is the general rule.

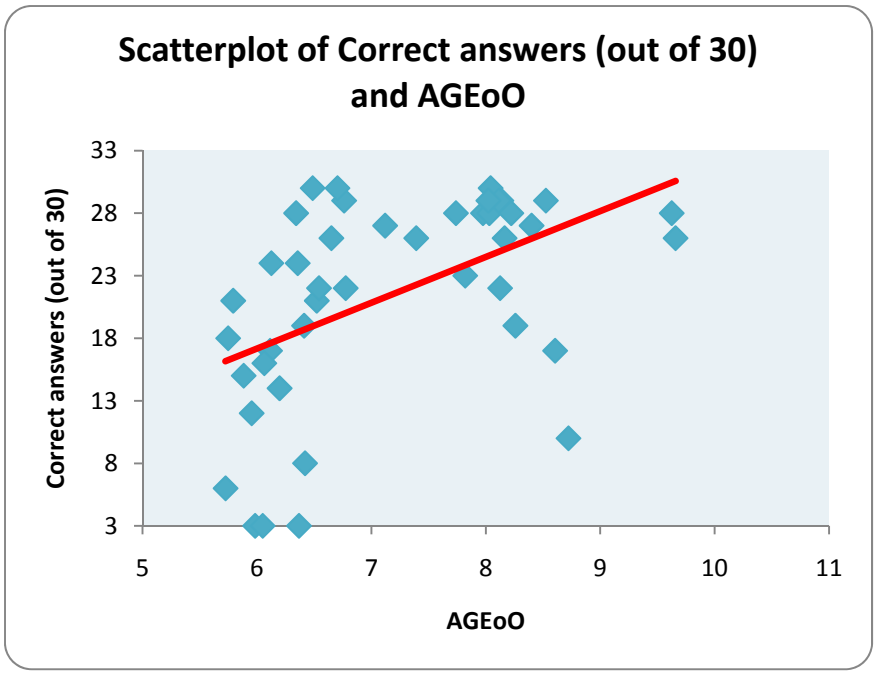


Figure 16 Scatterplot of the correlation between correct answers and AoO

In figure 17, the figures show a **positive correlation** between the **age of testing** and **the number of correct answers** the students get. The correlation is significant (p=0.021, r=0.346)

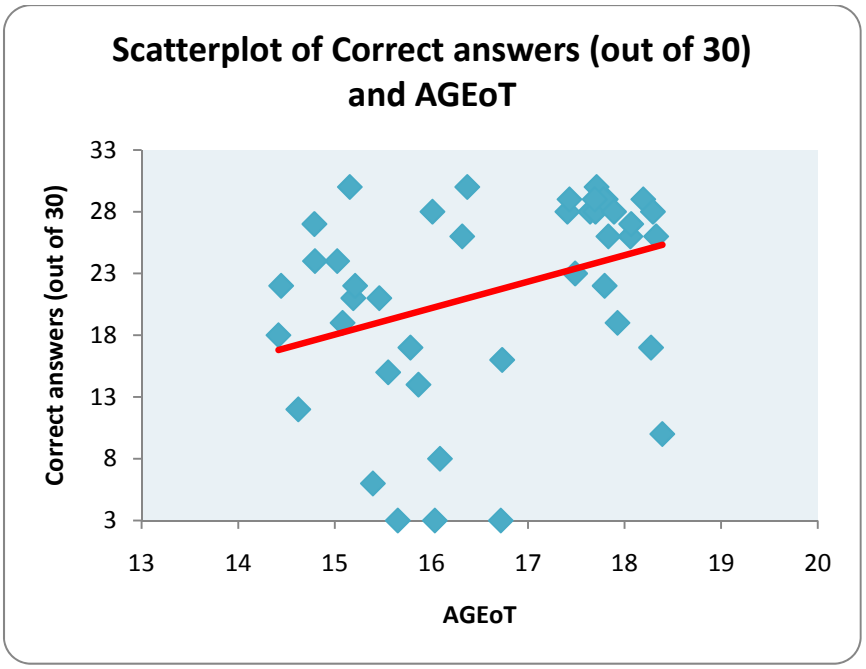


Figure 17: Scatterplot of correct answers and age of testing

This correlation would confirm **hypothesis 4.**

7.6 Correlations: Semi-controlled writing activity

In this section, some of the most interesting correlations with respect to the semi-controlled writing activity are presented.

In this **correlation**, it is possible to observe how the **number of tokens** increases as the **age of onset** rises. This correlation is statistically **significant** (p=0.016, r=0.362) and it is relevant for the study as it will be analyzed in the discussion section.

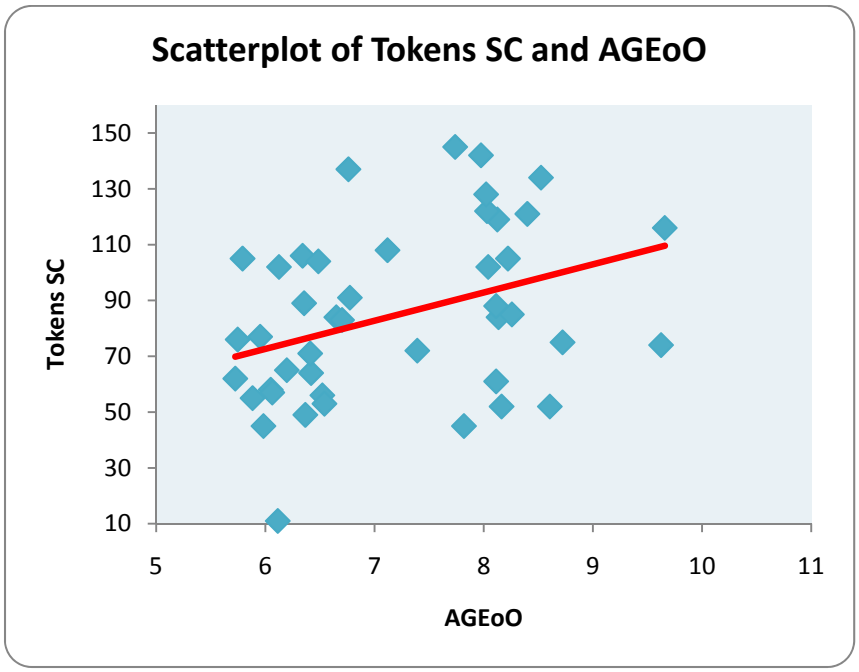


Figure 18 Scatterplot of the number of tokens in semi-controlled writing and AoO

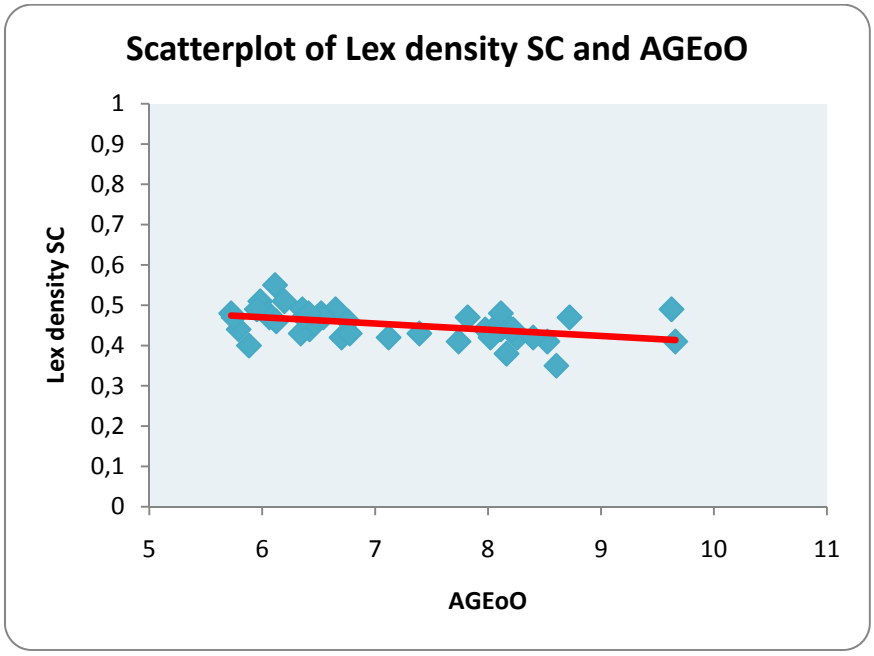


Figure 19 Scatterplot of the correlation between the lexical density value in semi-controlled activity with AoO

The data of this **significant negative correlation** in figure 18 (p=0.002, r=-0.455) is related to the one in the free-writing activity. In both cases, as the **age of onset** increases, the value of the **lexical density** decreases. It will be analyzed in the discussion section.

The rest of the values and correlations of the semi-controlled writing activity are in the following chart:

Measure	Tokens SC and AoO	Types SC and AoO	TTR SC and AoO	TPT[27] SC and AoO	LD[28] SC and AoO	Fam[29] SC and AoO	Tok PF[30] SC and AoO	Ty PF[31] SC and AoO
p value	0.016*	0.033*	0.115	0.123	0.002*	0.035*	0.080	0.109

* = significant

Figure 20 p value in correlations for each of the measures in semi-controlled writing

The statistically significant measures in this section are the three exposed above (tokens and lexical density) and also types and families. The *p* value for TTR, tokens per type, tokens per family and types per family are close to be remarkable, but out of the boundaries of being significant.

[27] Tokens per type
[28] Lexical Density
[29] Families
[30] Tokens per family
[31] Types per family

8 Discussion

One of the aims of this study was to confirm empirically whether age had an effect on vocabulary acquisition in the acquisition of L2 English by native speakers of Spanish in an instructed setting. The bulk of the results indicate that, **when length of exposure is kept constant, learners who were exposed to L2 English at a later age (late starters) significantly outperform early starters in several measures of lexical richness**.

There is a limitation in this study, though: that these results cannot be compared after several years of exposure, when the effects of age seem to be more remarkable, but at least, the results in this study are going to be discussed.

To begin with, the number of **tokens** in free writing turned out to be not significant in the study. Although late starters were expected to write a higher number of tokens, early starters outperformed them. It must be remembered that although the number of tokens required in the activity was one hundred, learners had the freedom to do some more or some less. Hence, it is possible to understand the variety of results with respect to the number of tokens. One of the main reasons for the early starters to write more tokens than late starters is because late starters are students from 2nd Bachillerato and they are trained for the Selectivity exam to not write more than the tokens required or they could be penalized for it. However, once the learners had the freedom to describe the story with the number of tokens they wanted to, late starters significantly outperformed early starters, which is an indicator of wider vocabulary resources. The number of tokens in free writing is taken into account although it does not give relevant information in this activity because it is relevant in the semi-controlled activity and therefore, I wanted to highlight this difference.

With respect to the number of **types**, it is significant to highlight that although early starters wrote a higher amount of tokens and therefore it is easier to find more different types on it, late starters were very close to the mean of types in the free writing activity. Moreover, once length was standardized, older learners wrote significantly more types of words in fifty tokens; therefore, **late starters outperform younger learners showing once again that the vocabulary (types) they have is wider than that of the early starters**. Additionally, the number of types was also higher for the late starters in the semi-controlled activity, a result that was somehow expected given that the number of tokens was higher too, but once again, another reinforcing result for the better performance of the older learners in vocabulary turned out.

In reference to the measure **TTR**, **older learners surpass younger learners significantly in free writing**, something that also occurs when length was standardized. As it was explained in section 4, this measure indicates the possibility to repeat the same word and therefore, if a learner is able of repeating less often the same words, it is **because the size of their lexical repertoire is wider**. However, results in the semi-controlled activity showed an outperformance of the early starters over the late starters. This could be due to the fact that as it was previously explained, TTR is a measure which is length-sensitive and hence, as late starters wrote more words in the semi-controlled writing activity, the figure of TTR has been affected, which does not directly imply that early learners are better in this task.

The measure **tokens per type** works inversely: a higher figure means a smaller size of lexicon because if more tokens are written for each type, it means that there are less different types and therefore, a narrower vocabulary extension. Therefore, in light of the results, late starters are better than early starters because the younger ones have a higher amount of tokens per type in free writing, which amounts to saying that **late starters are lexically more varied than early starters** in the standardize free writing activity but not in the semi-controlled activity. This last result could have also been affected by length measures.

***D* value** is **also in favour of the late starters over the early starters**. This result is significant given that it is not supposed to be affected by length factors. In the free writing activity, the difference is not significant but it is remarkable in the favour of late starters. Although it is not affected by length factors, once the free-writing activity was standardized, a higher significant value was achieved in favour of the late starters. This value could not be measured in the semi-controlled activity because the number of tokens for some of the participants in the group of the early starters was so little that it could not have been analyzed by *D_Tools* and therefore, this value was not measured for this activity.

With respect to the **number of families**, it is a measure that could show a predisposition to enlarge vocabulary given that when a student knows the stem that is the main element of a family to derive it into other words, s/he is supposed to have an inclination to easily learn the words that are related to that stem. Moreover, knowing a large number of families involves knowing a larger number of words and therefore, having a wider vocabulary. Results in this study indicate that **late starters outperform early starters** (suggesting that **late starters have a wider vocabulary than early starters**), both in free writing when the length of the text is standardized and in the semi-controlled writing activity too with a significant difference. However, although late starters also surpass early starters in the free writing

activity, the difference turned out to be not significant. This result could have been also affected by length but in any case, although the difference is insignificant, it is once again in favour of the late starters.

In reference to the measure **tokens per family**, it works in the same way as tokens per type: the more tokens per family, the narrower the scope of the learner's vocabulary. Once again, **the difference is in favour of the late starters** who write fewer tokens per family than early starters in the free writing activity, where the difference is very significant, suggesting that **late starters show a wider scope of vocabulary than early starters**. In the standardized free writing results, differences were not very big, but once again, very significant differences appeared in the semi-controlled activity.

Regarding **types per family**, differences were not significant at all in any of the data collected in free writing (standardized and non standardized) or in semi-controlled writing. Therefore, **no relevant conclusions can be drawn** from that.

With respect to **lexical density**, **results are in favour of the early starters** in free writing, being the significant difference; furthermore, the difference was also significant in the semi-controlled writing activity. However, it is not significant when the text is standardized. Maybe, it is wrong to conclude that younger learners are better than older learners in this measure because, as it has been previously explained, **LD can be sensitive to telegraphic writing and therefore, the figures can be misleading**.

Concerning the **controlled writing activity**, **late starters significantly outperformed early starters**. Although it can be attributed to the superiority in terms of vocabulary that older learners seem to show, it can be also linked to maturational and cognitive constraints. For this reason, I give more importance to the other measures presented in the previous lines.

The **correlations** showed in the results section are a graphic representation of what has been discussed throughout this section. The correlation is *positive* when a higher value of the measure coincides with older students such as the case of TTR in free writing, correct answers in controlled writing or number of tokens in semi-controlled writing. The correlation is *negative* when a lower value of the measure corresponds to older learners. That is the case of the number of tokens per type and lexical density in free writing and the value of lexical density in semi-controlled writing. Overall, correlations indicate that **lexical richness goes up as age of onset to the L2 increases**, i.e., the older you are exposed to the L2, the more lexically varied you can get (at least in the short-term during intermediate stages).

As a **conclusion** to this section, **most of the measures show a better performance from the side of the late starters who also happen to be the older learners**. It is true that the cognitive maturity of the students could have influence the results; therefore, the influence of this chronological age can be stronger than an early start in language learning. Some measures are in favour of the early starters but, after analysing the data, it seems that those figures have been affected by **length factors** given that when the text has been standardized, figures continue to be on the side of late starters besides lexical density, a measure that could have been affected by the telegraphic writing style of the younger learners. Nonetheless, it is very important not to forget that this study has been conducted in an instructional setting and in a determined point in time and probably these subjects will not be ever tested again to check if early starters surpass late starters **in the long run**.

8.1.1 Limitations to this study

There are several limitations to this study. Some of them have been dealt with in order to avoid them, but there are others that cannot be dealt with at all.

The first limitation which was found was the **number of subjects**. Although at first around 60 people were tested, many data was taken out given to the length of exposure they had. Therefore, from a wider sample of subjects, the study only has 44 cases. Moreover, their length of exposure is not totally equal but approximate though, crucially the difference was not significant. This was one of the reasons to leave out so many students.

Another limitation in this study is the number of researchers in a project of this size. Given to that fact, I could not apply spoken tasks because more than one judge would have been needed and hours of transcription too. Therefore, although the written form provides a high amount of data, having oral data could have been interesting too.

Lack of time to carry out some activities is also a limitation to the study. The subjects that were available were tied to a course schedule and therefore, I could not make an individual interview to implement oral instruments as it has been explained above. For this reason, I am very thankful for the facilities I got in both centres. Moreover, getting ready a study of these dimensions requires a lot of time and great efforts were made to conduct this kind of study in such a short period of time.

In conclusion, for the reasons explained above, time and facilities have been the factors that have hampered the development of this study. However, relevant results have

been obtained besides the difficulties showed and obstacles have been solved by different means and thanks to the help of the institutions and supervisors.

9 Implications for vocabulary second language teaching

Under the light of these results with respect to second language vocabulary acquisition, is the earlier, the better? As it has been previously explained, it seems that age effects are more visible in areas such as pronunciation and sometimes in grammar: it seems that vocabulary is to some extent less affected than the other areas, but research in the long term has proved that it is not totally true. In order to make improvements to the lexicon, the points explained by Zurawsky (2006) can be applied to vocabulary teaching:

i. In first place, the **amount of input** is essential. The mechanisms to absorb vocabulary are **implicit** but also **explicit**; therefore, it is essential to provide the students with great amounts of input because although at first it could seem that they are not acquiring any vocabulary, innate mechanisms are taking part and in the long run they can turn up. However, explicit mechanisms are also useful but for older learners. Consequently, vocabulary teaching should be taught implicitly but also explicitly because learners of all ages can benefit for each kind of teaching depending on their cognitive development.

ii. Furthermore, older learners acquire vocabulary at a **faster pace**; therefore, if a younger learners is not as quick as an older one, it does not mean that vocabulary learning is for later stages: therefore vocabulary must not be forgotten in language teaching and being relegated to memory issues: vocabulary acquisition requires a gradual process just as other aspects of language affected by age.

iii. Teachers should not expect a spectacular performance of those who started early: it must be reminded that if the **basic conditions** are not fulfilled such as the following:

   - Sufficient amount of input: it must be taken into account that input is not enough for language learning in many cases. The total hours to which a student is exposed is around 110 hours per year, a very poor figure taking into account that if it were naturalistic exposure, the number of hours would be 1460 being conservative and taking a figure of 4 hours per day.
   - Teacher's knowledge and its transmission to the students: With this MA, there is a starting point to prepare language teachers for the classrooms. Teaching a language is not something that is innate in people, we must learn how to teach because it is not just knowing the language or about it, but also knowing how

to teach it. Moreover, with respect to the age factor, the teacher has to be aware that not all stages of learning are the same, and therefore, s/he has to adjust her/his teaching to their students.

If the above basic conditions are not fulfilled, late starters are going to surpass early starters anyway, no matter if ES had a longer exposure to the target language.

What about **immersion programmes**? Older students require more than immersion to achieve a good level in their vocabulary: they benefit from direct instruction and explicit teaching. So that, the teacher has to adequate his/her way of teaching in order to cater for differences. Furthermore, immersion is beneficial, but may not be enough given that once the student is out of school, the input s/he receives is almost none. Vocabulary acquisition can somehow benefit from this programmes in the sense that students are exposed to different vocabulary depending on the subject they are taught and therefore they can widen their number of types of words and also of families.

Stay-abroad programmes are highly beneficial for the acquisition of vocabulary. Nowadays, these programmes are more abundant in universities although little by little, they are being introduced in high schools. The possibility that these programmes offer is to learn vocabulary in a natural setting, something that is absolutely essential to learn well is the semantic form and function of lexicon. Moreover, curiously enough, the students who benefit from these programmes are usually older learners, probably given to the independence and maturity of the person to live on his/her own in a foreign country. Furthermore, these learners who up to now have not been early starters seem to have benefited from these programs for what we know in our everyday experience, especially in vocabulary. Therefore, policymakers should think about the relevance of these programmes in language education.

As a summary for this section, Johnstone (2002) suggests the following measures:

- It has been shown that an early start is not enough in many cases; therefore, lifelong language learning should be ensured by the Governments.
- Teachers should be well trained. Basic intuitions are not enough when teaching languages, consequently, teacher training should be of a regular basis in education.
- The design of adequate materials according to the age of the student.
- Educational systems should be also revised from time to time, therefore, an adequate system of self-evaluation is necessary.

To conclude, **policy makers** can benefit from research and studies on the age factor in relation to vocabulary acquisition and be aware that vocabulary is part of language and as such, it is not out of the reach of age-related effects and therefore it must not be taken as part of memory affairs but of language acquisition and implement measures and mechanisms to improve it.

10 Conclusion

This research has dealt with the age factor in its different dimensions such as the biological explanations, the critical period hypothesis, rate and ultimate attainment, naturalistic and instructed settings and especially about age and its relation to second language vocabulary acquisition in a formal context. As it has been explained, this area has been abandoned for a long time as though age did not have any effects on vocabulary acquisition, but later studies have shown the contrary. Although studies in instructed settings can be misleading in the sense that results are in favour of the older learners and teachers, parents and policy makers have to take into account that vocabulary is another part of language acquisition just as is pronunciation or syntax. The problem is that effects are not so clear because instruments to measure vocabulary are not easy to implement and not so visible at first, but in the long run, those who started earlier and are under adequate conditions are capable of surpassing late starters.

With respect to the study, the hypotheses that were previously made were confirmed because previous studies have backed up those hypotheses and they were easy to predict. Many difficulties were faced during the time when this research was being conducted, but finally results were rewarding and the study was satisfactory. The main results of this study were that late starters outperformed early starters as it was expected given that they are students in an instructed setting. Another obstacle is that it will be impossible to replicate this study with the same subjects, to test them again with the possibility of obtaining similar results, something that would be very useful in order to confirm the current research.

To conclude, it can be highlighted that although there are always factors that hamper language acquisition, success in language learning is always possible although it can be more or less difficult depending on the conditions and the individual differences of each learner although older learners are less likely to be successful. This must not be an obstacle for

language acquisition, but rather a motivation to achieve our final goal in language learning: learning a language to communicate. Finally, I am pleased to contribute to second language research (although in a minor scale) as a future teacher to be able to spot the difficulties and advantages that students are going to have and to help them in the light of research and future experience.

11 References

Bauer, L. & Nation, P. (1993). Word families. *International Journal of Lexicography* 6, 4: 253-279.

Bialystok, E. (1994). Representation and ways of knowing. Three issues in second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549-569). London: Academic Press.

Birdsong, D (2002). Interpreting age effects in second language acquisition. University of Texas: Unpublished manuscript. Available online: http://ccwf.cc.utexas.edu/~birdsong/publications/birdsong_1-4-02_interpr_age.pdf

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41–68). Cambridge: Cambridge University Press.

Cenoz, J. (2002). Age differences in foreign language learning. *ITL Review of Applied Linguistics*, 135-136, 125-142.

Cook, V. (1992). Evidence for multicompetence. *Language Learning*, *42* (4): 557–91.

De Villiers, Jill and Peter De Villiers (1978). *Language Acquisition.* Cambridge, MA: Harvard University Press.

Diller, K. (1981). 'Natural methods' of foreign language teaching: Can they exist? What criteria must they meet? *Native Language and Foreign Language Acquisition*, HarrisWinnitz (ed.), 75–91. New York: The New York Academy of Sciences.

Dornyei, Z. (2001). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics*, *21*:43–59.

Ellis, R. (1985). *Understanding second language acquisition* (1995 edition). Oxford: Oxford University Press.

Ellis, R. (2008). *The Study of Second Language Acquisition* (2nd ed.). Oxford: Oxford University Press.

Flege, J.; Yeni-Komshian, G.; & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language, 41,* 78-104.

Gass, S. M. (1999). Incidental vocabulary learning. Studies in Second Language Acquisition, 21(2), 319-333.

Genesee, F. (1978). Is there an optimal age for starting second language instruction? *McGill Journal of Education*, 13(2), 145-154.

Griffin, G. G. (1993). *The relationship between starting age and second language learning.* Unpublished Master Thesis, Faculty of the Dominican College, San Rafael, CA.

Harley, B., & Wang, W. (1997). The Critical Period Hypothesis: where are we now? In A. de Groot & J. Kroll (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.

Johnstone, R. (2002). *Addressing 'The age factor': Some implications for languages policy*. Strasbourg: Council of Europe. Retrieved from www.coe.int/t/dg4/linguistic/source/JohnstoneEN.pdf

Lederberg, A. R., & Spencer, P. (2005). Critical periods in the acquisition of lexical skills. Evidence from deaf individuals. In P. Fletcher & J. F. Miller (Eds.), *Developmental theory and language disorders* (pp. 121-145). Amsterdam: John Benjamins.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.

Lightbown, p. M., & Spada, N. (2006). *How Languages are Learned (3rd edition)*. Oxford: Oxford University Press.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, *12*, 251-285.

Miralpeix, I. (2006). Age and Vocabulary Acquisition in English as a Foreign language (EFL). In C. Muñoz (Ed.), *Age and the rate of foreign language learning* (pp. 89-106). Clevedon: Multilingual Matters.

Miralpeix, I. (2008). *The influence of age on vocabulary acquisition in English as a foreign language.* Universitat de Barcelona, Barcelona.

Muñoz, C. (Ed.). (2006). *Age and the rate of foreign language learning*. Clevedon: Multilingual Matters.

Oller, J., & Nagato, N. (1974). The long term effects of FLES: An experiment. *The Modern Language Journal, 58*(1-2), 15-19.

Ortega, L. (2009). *Understanding Second Language Acquisition*. Hodder Education.

Oxford, R. L. & Ehrman, M. (1993). Second language research on individual differences. *Annual Review of Applied Linguistics, 13*:188–205.

Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York: Harcourt Brace Jovanovich.

Poynton, T. (2007) *EZAnalyze. http://www.ezanalyze.com/index.htm*

Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.

Scovel, T. (1988). *A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Language*. Rowley, MA: Newbury House.

Scovel, T. (2000). A critical review of the Critical Period Hypothesis. *Annual Review of Applied Linguistics* 20: 213–223.

Singleton, D., & Lengyel, Z. (1995). *The age factor in second language acquisition*.

Clevedon: Multilingual Matters.

Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics*, *43*, 269-285.

Singleton, D. (2007). The critical period hypothesis: some problems: *Interlingüística*, 17, 48-56

Torras, M. R., & Celaya, M. L. (2001). Age-related differences in the development of written production. An empirical study of EFL school learners. *International Journal of English Studies, 1*(2), 103-126.

White, L. (2005). On the Nature of Interlanguage Representation: Universal Grammar in the Second Language. In C. Doughty & M. Long (eds.) *The Handbook of Second Language Acquisition.* Blackwell Publishing, 2005. Blackwell Reference Online. 25 March 2011 http://www.blackwellreference.com/subscriber/tocnode?id=g9781405132817_toclevel_ss1-3

Zurawski, C. (2006). Foreign language instruction: Implementing the best teaching methods. *Research Points*, 4(1), 1-4. Downloadable http://www.aera.net

APPENDIXES

APPENDIX 1: BACKGROUND QUESTIONNAIRE

Cuestionario previo

Nombre y apellidos:…….

Fecha de nacimiento:……

Lengua materna:……..

Lengua materna de tu padre……. Lengua materna de tu madre……..

Lengua que hablas en casa……

Sexo: □ Mujer □ Hombre

Curso y grupo en el que estás……

Nota en inglés el curso pasado…….. Nota en inglés en el segundo trimestre……..

¿En qué curso empezaste a aprender inglés?.....

¿Has estado apuntado a clases particulares de inglés? (profesor particular, academia…)…….

Si has respondido sí, ¿En qué curso/s?..... ¿Cuántas horas a la semana?.....

¿Has realizado alguna estancia larga fuera de España? (más de un mes) .....

Si has respondido sí ¿dónde?.....

APPENDIX 2: WRITING ACTIVITIES

Writing exam

Nombre…….. Curso……

1. Escribe una redacción en inglés sobre ti mismo. Puedes hablar sobre lo que quieras: tu aspecto, tus hobbies, tus mascotas, tu familia o cualquier cosa que se te ocurra. Debes escribir alrededor de 100 palabras.

2. Completa en inglés el siguiente texto sobre 'Caperucita Roja'

A

Hello. I am 'Caperucita Roja'. My favourite colour is (1) _____
I am ten years old. I live with my parents.
I have got no (2)_____or sisters.
I have got a grandmother.
She is (3)_____years old.

B

Once upon a time, there was a little girl. She was a very good girl. Her grandmother loved the girl (4)_____lot and she often gave her presents. One day Little Red Riding Hood's (5)_____made her a beautiful red cape and hood. The little (6)_____liked it a lot and she always wore (7)_____cape and hood.

Little Red Riding Hood lived with her mother and father (8)_____a village near a large forest. Her mother worked at home and her (9)_____in the forest. He was a woodcutter.

Little red Riding Hood's grandmother (10) _____ a kilometer away. She lived alone in a little (11)_____inside the forest. Little Red Riding Hood loved her grandmother and every (12) _____she went to visit her.

C

One day her mother called her (13)_____said, "Little Red Riding Hood, there (14)_____a cake and a bottle of milk in this basket. I (15)_____like you to take it to your grandmother. She (16)_____ill this morning and she needs to eat."

Before the (17) _____left, her mother told her "Be careful and (18)_____stop to play in the forest. And walk carefully. Don't run or you (19)_____break the bottle. Then you will have no milk for your (20)_____." Little Red Riding Hood took the basket. Then she (21)_____goodbye to her mother and started to walk to her grandmother's (22)_____.

Ten minutes later, Little Red Riding Hood (23)_____a wolf. She had never seen a wolf before and she (24)_____not know he was a bad creature. She thought he (25) _____ a large dog and she was not afraid of (26)_____.

"Good morning, Little Red Riding Hood. Where are you going so early (27)_____the morning?" asked the wolf. "To my grandmother's house," she answered.

Then the wolf asked her, "And what have (28)_____got in your basket"?. Little Red Riding Hood said, "A cake and some (29)_____. My grandmother is ill and this basket is (30)_____her."

Cuenta en inglés lo que está sucediendo en estas viñetas. Puede haber varias respuestas correctas.

4

5

6

APPENDIX 3: RAW DATA

ID	GROUP	NAME	COURSE	DoB	AGEoO	AGEoT	LoE	MARK	L1	ML1	FL1	L1 at home	SEX
1	Late starters	DCU	2ºBc	07/02/1993	8,61	18,28	10	7	SP	SP	SP	SP	F
2	Late starters	RCP	2ºBc	06/03/1993	8,53	18,19	10	10	SP	SP	SP	SP	F
3	Late starters	HCHH	2ºBc	30/07/1993	8,13	17,79	10	8	SP	SP	SP	SP	F
7	Late starters	AML	2ºBc	26/07/1993	8,14	17,81	10	9	SP	SP	SP	SP	M
8	Late starters	CMS	2ºBc	12/06/1993	8,26	17,93	10	6	SP	SP	SP	SP	F
9	Late starters	JMA	2ºBc	30/01/1993	9,63	18,29	9	7	SP	SP	SP	SP	M
10	Late starters	PRH	2ºBc	24/09/1993	7,98	17,64	8	9	SP	SP	SP	SP	M
11	Late starters	CRG	2ºBc	04/08/1993	8,11	17,78	10	8	SP	SP	SP	SP	F
12	Late starters	JGA	2ºBc	24/04/1993	7,39	18,06	11	8	SP	SP	SP	SP	M
13	Late starters	LCA	2ºBc	31/08/1993	8,04	17,71	10	9	SP	SP	SP	SP	F
14	Late starters	MTS	2ºBc	20/11/1993	7,82	17,49	10	8	SP	SP	SP	SP	F
15	Late starters	CMR	2ºBc	04/09/1993	8,03	17,70	10	9	SP	SP	SP	SP	F
16	Late starters	FMCE	2ºBc	04/08/1993	8,11	17,78	10	7	SP	SP	SP	SP	M
17	Late starters	MGC	2ºBc	21/04/1993	8,40	18,07	10	6	SP	SP	SP	SP	F
18	Late starters	JHF	2ºBc	07/09/1993	8,02	17,69	10	9	SP	SP	SP	SP	M
19	Late starters	ART	2ºBc	16/07/1993	8,16	17,83	10	9	SP	SP	SP	SP	F
20	Late starters	SGP	2ºBc	18/01/1993	9,66	18,33	9	8	SP	SP	SP	SP	F
21	Late starters	JMLH	2ºBc	25/12/1992	8,72	18,39	10	5	SP	SP	SP	SP	M
22	Late starters	NLR	2ºBc	25/06/1993	8,22	17,89	10	9	SP	SP	SP	SP	F
23	Late starters	SVF	2ºBc	19/12/1993	7,74	17,41	10	8	SP	SP	SP	SP	M
24	Late starters	MJNP	2ºBc	11/12/1993	6,76	17,43	11	7	SP	SP	SP	SP	F
25	Early starters	LFS	4º	21/01/1995	6,65	16,32	10	7	SP	SP	SP	SP	F
26	Early starters	MBB	4º	21/09/1995	5,98	15,65	10	6	VL	VL	VL	SP	F
27	Early starters	RJF	4º	30/11/1995	5,79	15,46	10	8	SP	SP	SP	SP	F
28	Early starters	AAH	4º	14/04/1995	6,42	16,09	10	6	SP	SP	SP	SP	M
29	Early starters	LFL	4º	27/10/1995	5,88	15,55	10	6	SP	SP	SP	SP	F
30	Early starters	MJMO	4º	02/01/1995	6,70	16,37	10	10	SP	SP	SP	SP	F
31	Early starters	PEL	4º	04/08/1995	6,11	15,78	10	5	SP	SP	SP	SP	M
32	Early starters	SZ	4º	12/05/1995	6,34	16,01	10	10	AREG	AREG	AREG	SP/AREG	F
33	Early starters	AMG	4º	27/08/1994	6,05	16,72	11	2	SP	SP	SP	SP	M
34	Early starters	GRM	4º	03/05/1995	6,37	16,04	10	5	SP	SP	SP	SP	M
35	Early starters	AAV	4º	04/07/1995	6,20	15,87	10	6	SP	SP	SP	SP	M
36	Early starters	JAFP	4º	24/12/1995	5,73	15,39	10	6	SP	SP	SP	SP	M
37	Early starters	AMF	4º	22/08/1994	6,06	16,73	11	7	SP	SP	SP	SP	F
38	Early starters	DVC	3º	02/08/1996	7,12	14,79	8	9	SP	SP	SP	SP	M
39	Early starters	EAGO	3º	07/05/1996	6,36	15,03	9	10	SP/IT	SP/IT	SP	SP/IT	M
40	Early starters	IMC	3º	07/03/1996	6,52	15,19	9	7	SP	SP	SP	SP	F
41	Early starters	RGA	3º	06/12/1996	6,78	14,44	8	9	SP	SP	SP	SP	F
42	Early starters	PFA	3º	20/03/1996	6,49	15,16	9	10	SP	SP	SP	SP	F
43	Early starters	IRA	3º	16/12/1996	5,75	14,42	9	9	SP	SP	SP	SP	M
44	Early starters	MDB	3º	02/10/1996	5,95	14,62	9	9	SP	SP	SP	SP	F
45	Early starters	CDP	3º	29/02/1996	6,54	15,21	9	9	SP	SP	SP	SP	F
46	Early starters	ACR	3º	30/07/1996	6,13	14,79	9	10	SP	SP	SP	SP	F
47	Early starters	NLS	3º	17/04/1996	6,41	15,08	9	7	SP	SP	SP	SP	F

			Free writing						
PRIVATE LESSONS	TIME ABROAD	WHERE	Tokens F	Types F	TTR F	Tokens per type F	Lex density F	D F	Families F
N	N		124	64	0,52	1,94	0,4	41,15	55
Y	N		147	94	0,64	1,56	0,53	56,43	71
Y	N		159	90	0,57	1,77	0,37	56,52	78
Y	N		101	72	0,71	1,4	0,48	98,15	57
N	N		103	60	0,57	1,75	0,45	36,86	51
Y	N		106	62	0,58	1,71	0,44	45,52	52
Y	N		148	89	0,6	1,66	0,42	88,45	75
N	N		117	71	0,61	1,65	0,39	62,39	61
Y	N		91	54	0,59	1,69	0,4	35,48	49
N	N		102	65	0,64	1,57	0,46	62,27	56
N	N		92	58	0,63	1,59	0,4	50,42	51
Y	N		83	54	0,65	1,54	0,52	51,21	43
Y	N		90	55	0,61	1,64	0,43	40,6	47
N	N		94	62	0,66	1,52	0,48	61,28	53
N	N		100	64	0,64	1,56	0,46	57,87	52
N	N		109	66	0,61	1,65	0,48	53,61	57
N	N		88	52	0,59	1,69	0,44	41,42	47
N	N		107	63	0,59	1,7	0,45	46,83	53
N	N		121	65	0,54	1,86	0,32	47,24	56
Y	N		131	76	0,58	1,72	0,45	65,67	67
N	N		89	55	0,62	1,62	0,46	54,77	47
N	N		113	66	0,58	1,71	0,42	56,16	55
S	N		87	50	0,57	1,74	0,54	34,54	40
N	Y		161	90	0,56	1,79	0,45	67,73	76
Y	N		95	63	0,66	1,51	0,52	57,5	50
Y	N		105	54	0,51	1,94	0,52	33,12	49
Y	N		160	73	0,46	2,19	0,51	37,25	59
Y	N		108	56	0,52	1,93	0,55	29,91	44
N	Y	EGYPT	272	129	0,47	2,11	0,45	48,57	101
N	N		84	52	0,62	1,62	0,5	51,33	43
Y	N		101	45	0,45	2,24	0,43	26,67	39
N	N		95	52	0,55	1,83	0,48	37,79	44
Y	N		62	32	0,52	1,94	0,47	20,83	27
Y	N		163	86	0,53	1,9	0,47	56,32	71
Y	N		146	85	0,58	1,72	0,5	75,67	68
N	N		148	79	0,53	1,87	0,44	40,81	64
Y	N		116	61	0,53	1,9	0,47	40,76	48
Y	N		117	59	0,5	1,98	0,48	40,83	48
N	N		158	89	0,56	1,78	0,44	87,39	71
N	N		120	51	0,42	2,35	0,42	25,92	40
N	N		99	54	0,55	1,83	0,49	35,93	46
Y	N		116	67	0,58	1,73	0,47	52,05	50
Y	N		177	84	0,47	2,11	0,38	63,89	73
N	N		101	58	0,57	1,74	0,47	38,65	50

		Controlled writing					Semi-controlled writin
Tokens per family F	Types per family F	Correct answers (out of 30)	Tokens SC	Types SC	TTR SC	Tokens per type SC	Lex density SC
2,15	1,05	17	52	29	0,56	1,79	0,35
1,82	1,08	29	134	69	0,51	1,94	0,41
2	1,12	22	119	60	0,5	1,98	0,45
1,65	1,14	29	84	46	0,55	1,83	0,46
2,02	1,14	19	85	32	0,38	2,66	0,42
1,92	1,1	28	74	43	0,58	1,72	0,49
1,91	1,13	28	142	68	0,48	2,09	0,44
1,89	1,13	29	88	45	0,51	1,96	0,44
1,84	1,08	26	72	41	0,57	1,76	0,43
1,79	1,13	30	102	49	0,48	2,08	0,44
1,76	1,1	23	45	27	0,6	1,67	0,47
1,81	1,14	28	122	63	0,52	1,94	0,43
1,83	1,11	29	61	33	0,54	1,85	0,48
1,68	1,08	27	121	56	0,46	2,16	0,42
1,83	1,13	29	128	56	0,44	2,29	0,42
1,79	1,04	26	52	26	0,5	2	0,38
1,81	1,04	26	116	51	0,44	2,27	0,41
1,92	1,09	10	75	38	0,51	1,97	0,47
2,13	1,13	28	105	57	0,54	1,84	0,44
1,96	1,13	28	145	73	0,5	1,99	0,41
1,85	1,13	29	137	63	0,46	2,17	0,46
1,98	1,13	26	84	46	0,55	1,83	0,49
2,05	1,13	3	45	19	0,42	2,37	0,51
2,03	1,12	21	105	48	0,46	2,19	0,44
1,7	1,08	8	64	35	0,55	1,83	0,44
2,14	1,1	15	55	32	0,58	1,72	0,4
2,56	1,14	30	83	42	0,51	1,98	0,42
2,16	1,05	17	11	10	0,91	1,1	0,55
2,53	1,15	28	106	58	0,55	1,83	0,43
1,88	1,14	3	58	29	0,5	2	0,48
2,56	1,13	3	49	30	0,61	1,63	0,47
2,07	1,09	14	65	39	0,6	1,67	0,51
2,26	1,15	6	62	31	0,5	2	0,48
2,23	1,14	16	57	29	0,51	1,97	0,47
2,01	1,13	27	108	54	0,5	2	0,42
2,14	1,09	24	89	44	0,49	2,02	0,49
2,27	1,15	21	56	32	0,57	1,75	0,48
2,29	1,08	22	91	40	0,44	2,27	0,43
2,11	1,14	30	104	61	0,59	1,7	0,47
2,77	1,13	18	76	39	0,51	1,95	0,47
2,04	1,09	12	77	44	0,57	1,75	0,49
2,06	1,1	22	53	33	0,62	1,61	0,47
2,37	1,1	24	102	53	0,52	1,92	0,46
1,92	1,06	19	71	41	0,58	1,73	0,48

g							Standardize length	
Families SC	Tokens per family SC	Types per family SC	Tokens FS	Types FS	TTR FS	Tokens per type FS	Lex density FS	D FS
25	2,08	1,16	50	28	0,56	1,79	0,56	24,6
61	2,16	1,11	50	35	0,7	1,43	0,5	53,36
52	2,29	1,15	50	37	0,74	1,35	0,46	52,08
39	2,1	1,13	50	42	0,84	1,19	0,44	57,88
28	3,04	1,14	50	32	0,64	1,56	0,42	37,99
37	1,86	1,11	50	31	0,62	1,61	0,35	43,52
60	2,3	1,08	50	36	0,71	1,42	0,41	45,44
38	2,29	1,16	50	38	0,76	1,32	0,38	41,5
32	2,19	1,22	50	32	0,64	1,56	0,36	31,12
43	2,35	1,12	50	35	0,7	1,43	0,48	54,2
24	1,83	1,08	50	36	0,72	1,39	0,38	48,27
57	2,09	1,07	50	37	0,74	1,35	0,52	61,92
28	2,11	1,11	50	35	0,7	1,43	0,4	57,09
48	2,46	1,15	50	37	0,74	1,35	0,48	57,81
51	2,47	1,06	50	33	0,66	1,52	0,48	37,85
23	2,22	1,09	50	34	0,68	1,47	0,54	49,42
50	2,32	1,02	50	33	0,66	1,52	0,4	41,18
34	2,15	1,09	50	31	0,62	1,61	0,46	23,46
52	2	1,08	50	35	0,7	1,43	0,34	44,98
64	2,19	1,11	50	36	0,72	1,39	0,48	83,55
56	2,39	1,07	50	36	0,72	1,39	0,48	53,36
39	1,92	1,05	50	35	0,7	1,43	0,4	60,72
16	2,81	1,19	50	32	0,64	1,56	0,54	33,84
42	2,33	1,07	50	36	0,72	1,39	0,46	53,36
28	2,29	1,25	50	35	0,7	1,43	0,44	50,37
30	1,83	1,07	50	33	0,66	1,52	0,48	30,04
37	2,16	1,11	50	32	0,64	1,56	0,52	33,31
9	1,11	1	50	31	0,62	1,61	0,54	25,17
50	2,1	1,14	50	40	0,8	1,25	0,44	72,33
28	2,07	1,04	50	34	0,68	1,47	0,5	45,71
28	1,75	1,07	50	32	0,64	1,56	0,46	32,97
37	1,76	1,05	50	31	0,62	1,61	0,44	33,7
28	2,11	1,04	50	26	0,52	1,92	0,48	14,49
27	2,11	1,07	50	35	0,7	1,43	0,5	44,02
48	2,21	1,1	50	33	0,66	1,52	0,56	43,54
41	2,12	1,02	50	36	0,72	1,39	0,46	57,74
30	1,7	1	50	34	0,68	1,47	0,48	40,92
33	2,64	1,09	50	23	0,46	2,17	0,44	18,39
53	1,91	1,11	50	34	0,68	1,47	0,38	49,66
34	2,06	1,06	50	30	0,6	1,67	0,44	22,88
40	1,83	1,05	50	36	0,72	1,39	0,52	43,19
28	1,79	1,07	50	26	0,52	1,92	0,38	21,47
45	2,18	1,09	50	25	0,5	2	0,42	25,33
37	1,86	1,08	50	31	0,62	1,61	0,52	20,66

Families FS	Tokens per family FS	Types per family FS
26	1,85	1
28	1,57	1,04
30	1,53	1,1
33	1,42	1,18
27	1,85	1,19
26	1,88	1,15
32	1,59	1,13
32	1,53	1,16
28	1,75	1,11
30	1,6	1,1
33	1,52	1,09
33	1,48	1,09
33	1,52	1,06
32	1,53	1,13
28	1,71	1,11
29	1,62	1,97
30	1,6	1,03
25	1,88	1,12
30	1,63	1,13
31	1,61	1,16
32	1,53	1,09
30	1,57	1,07
25	1,84	1,12
29	1,66	1,17
30	1,57	1,1
31	1,61	1,06
27	1,78	1,11
25	1,84	1,08
32	1,47	1,16
31	1,58	1,06
28	1,71	1,07
25	1,88	1,12
20	2,4	1,2
30	1,6	1,1
26	1,77	1,12
29	1,62	1,14
30	1,6	1,1
20	2,45	1,1
30	1,67	1,13
25	1,8	1,04
32	1,53	1,09
23	2,17	1,13
23	2,17	1,09
26	1,81	1,08

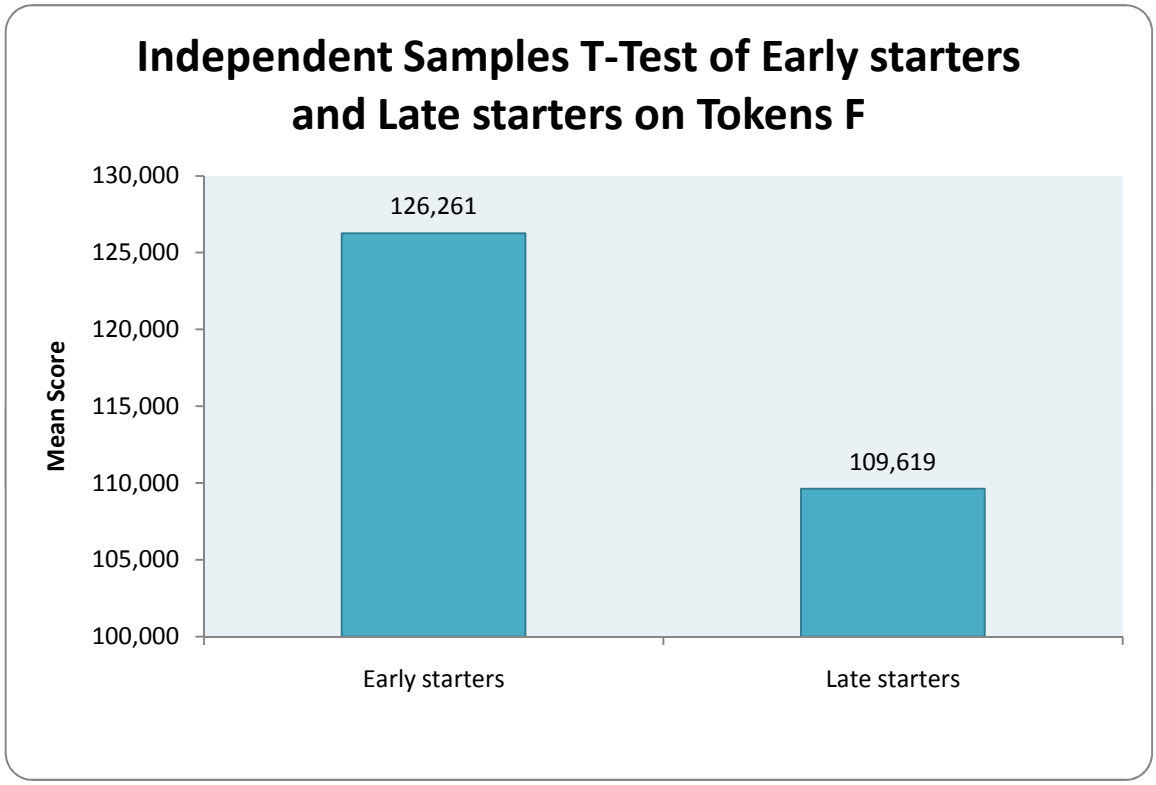
APPENDIX 4: T-TESTS AND CORRELATIONS

T-TESTS: FREE-WRITING ACTIVITY

EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens F

GROUP	Early starters	Late starters
Mean:	126,261	109,619
Std. Dev:	43,748	21,602
N:	23	21
Mean Difference:	16,642	
T-Score:	1,576	
Eta Squared:	,053	
P:	,123	

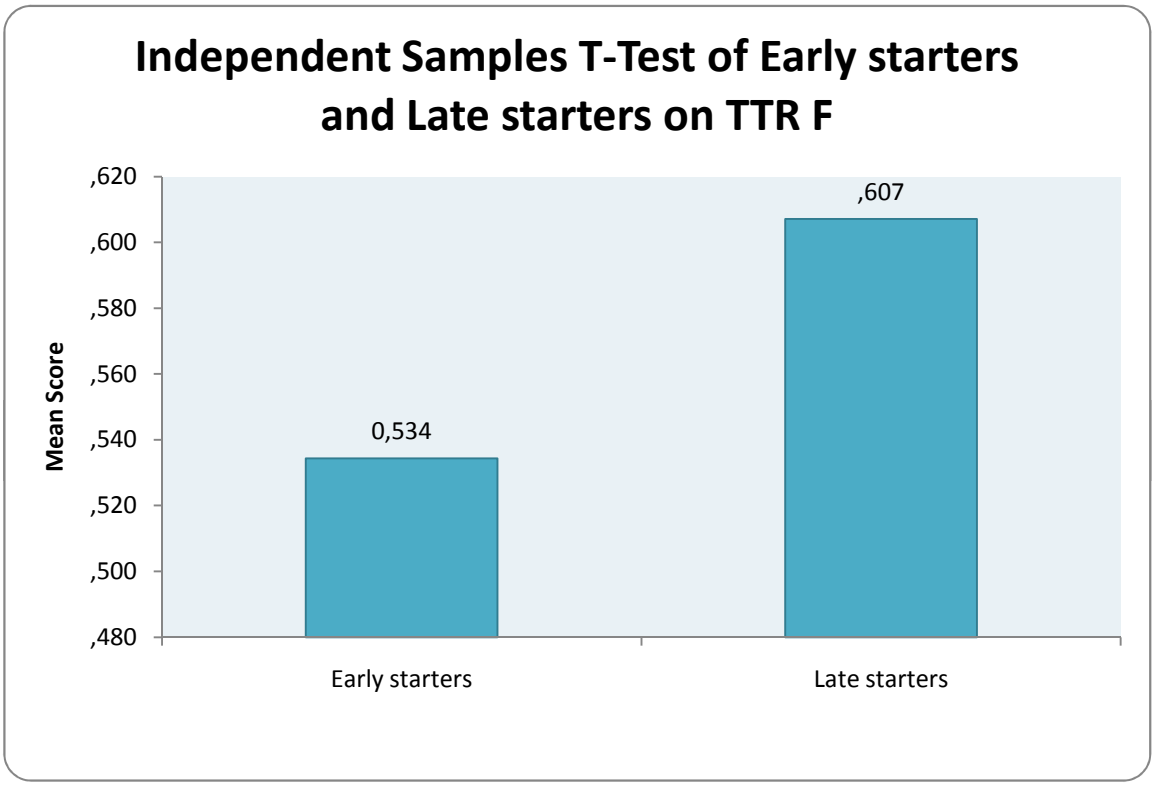
The observed difference between the group means is not significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on TTR F

GROUP	Early starters	Late starters
Mean:	,534	,607
Std. Dev:	,056	,043
N:	23	21
Mean Difference:	,073	
T-Score:	4,792	
Eta Squared:	,343	
P:	,000	

The observed difference between the group means is significant

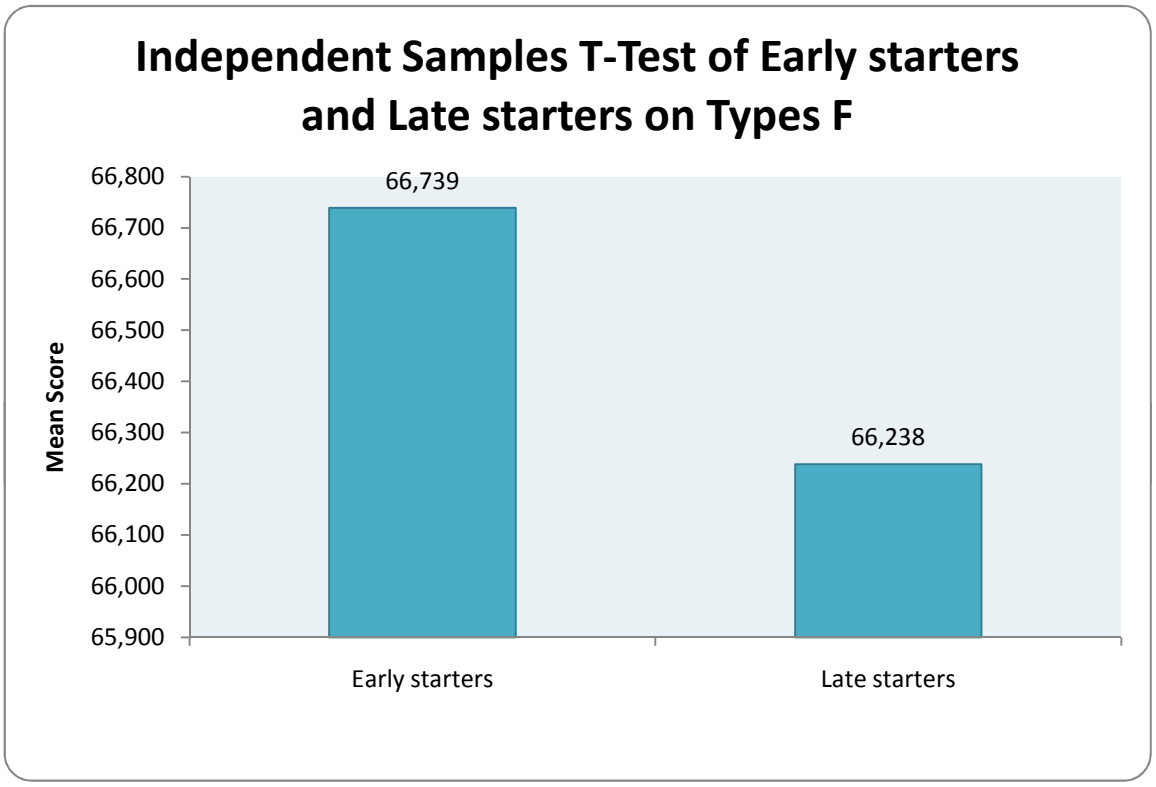


EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Types F

GROUP	Early starters	Late starters
Mean:	66,739	66,238
Std. Dev:	20,638	12,095
N:	23	21

Mean Difference:	,501
T-Score:	,097
Eta Squared:	,000
P:	,923

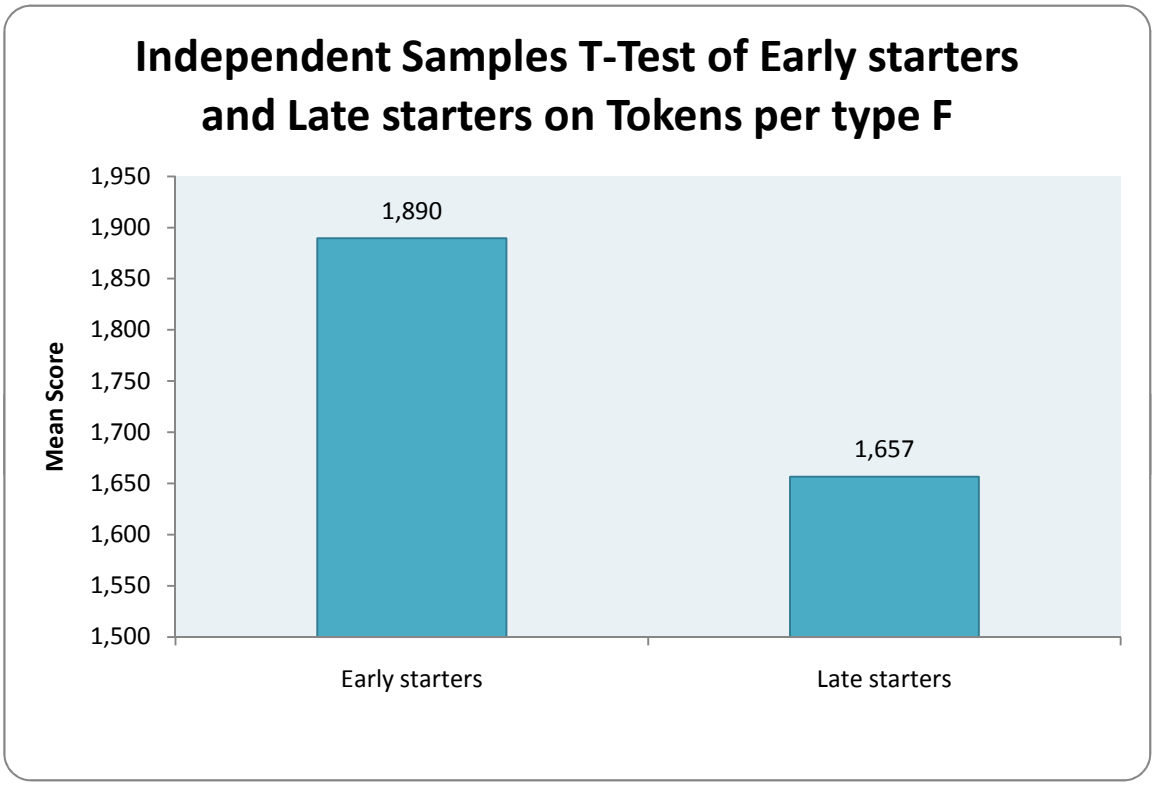
The observed difference between the group means is not significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens per type F

GROUP	Early starters	Late starters
Mean:	1,890	1,657
Std. Dev:	,204	,119
N:	23	21
Mean Difference:	,233	
T-Score:	4,566	
Eta Squared:	,322	
P:	,000	

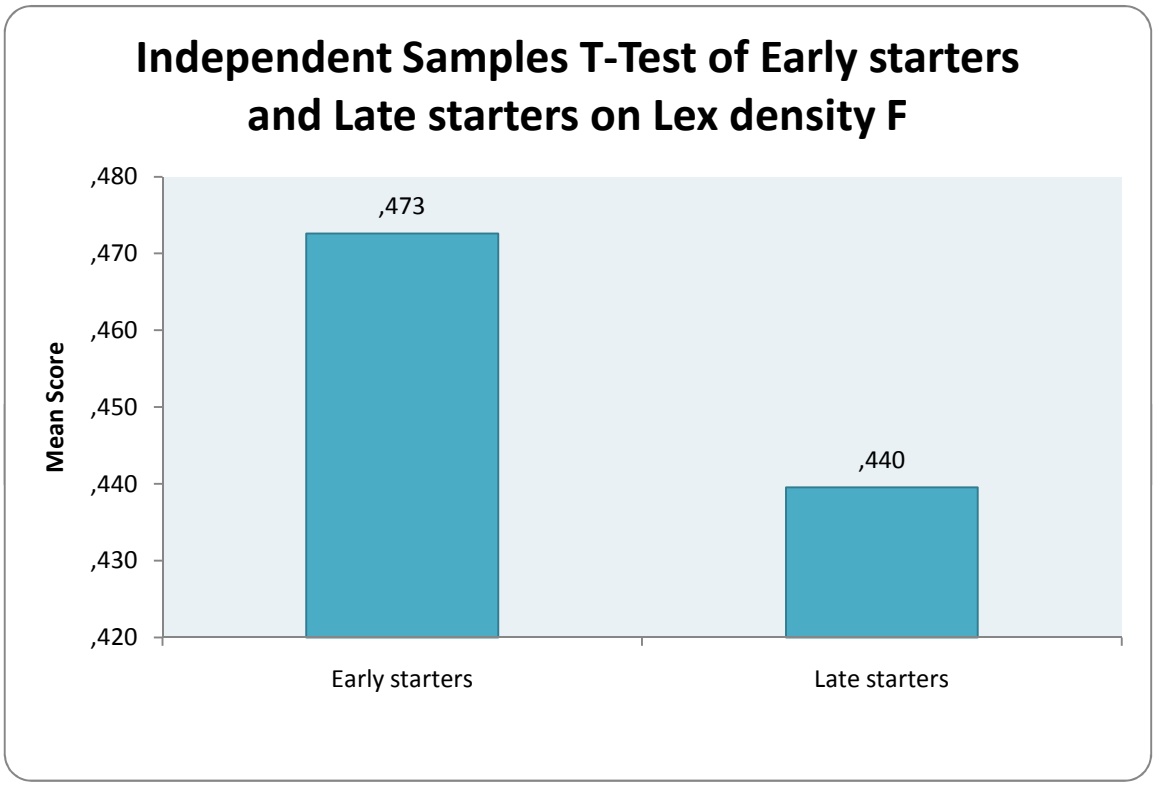
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Lex density F

GROUP	Early starters	Late starters
Mean:	,473	,440
Std. Dev:	,041	,049
N:	23	21
Mean Difference:	,033	
T-Score:	2,432	
Eta Squared:	,118	
P:	,019	

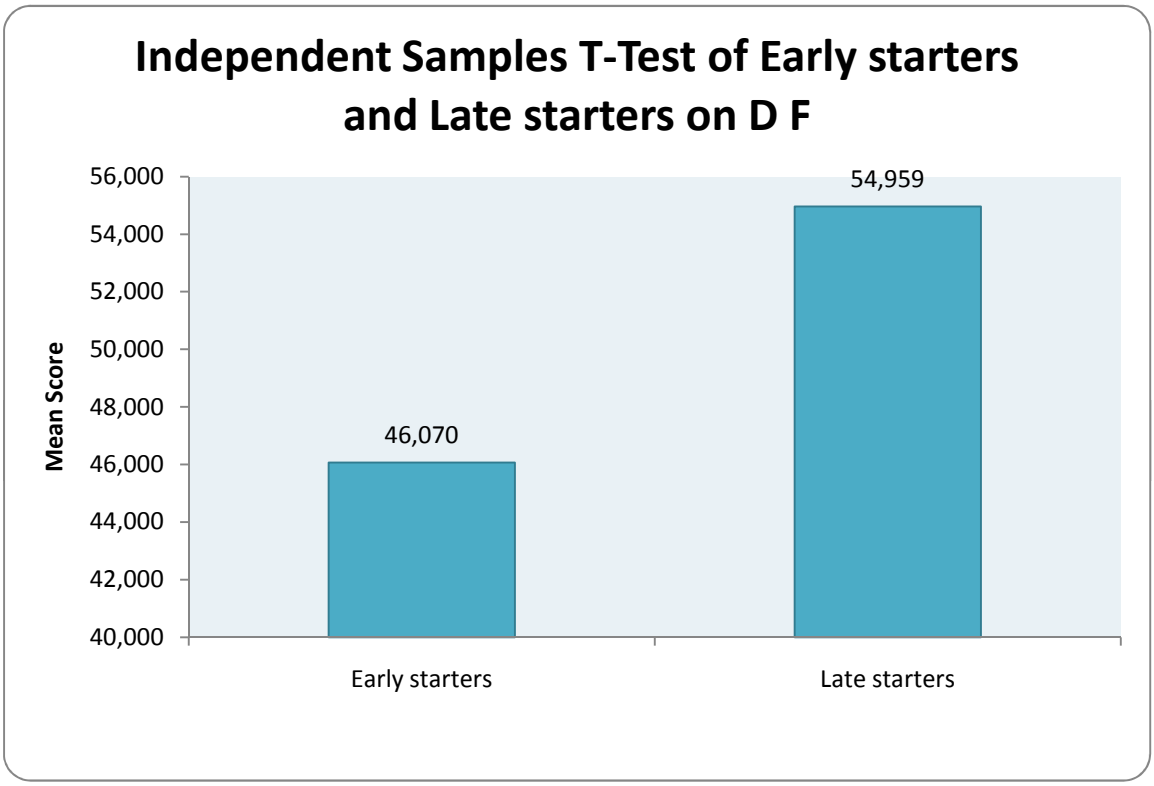
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on D F

GROUP	Early starters	Late starters
Mean:	46,070	54,959
Std. Dev:	16,714	15,495
N:	23	21
Mean Difference:	8,889	
T-Score:	1,824	
Eta Squared:	,070	
P:	,075	

The observed difference between the group means is not significant

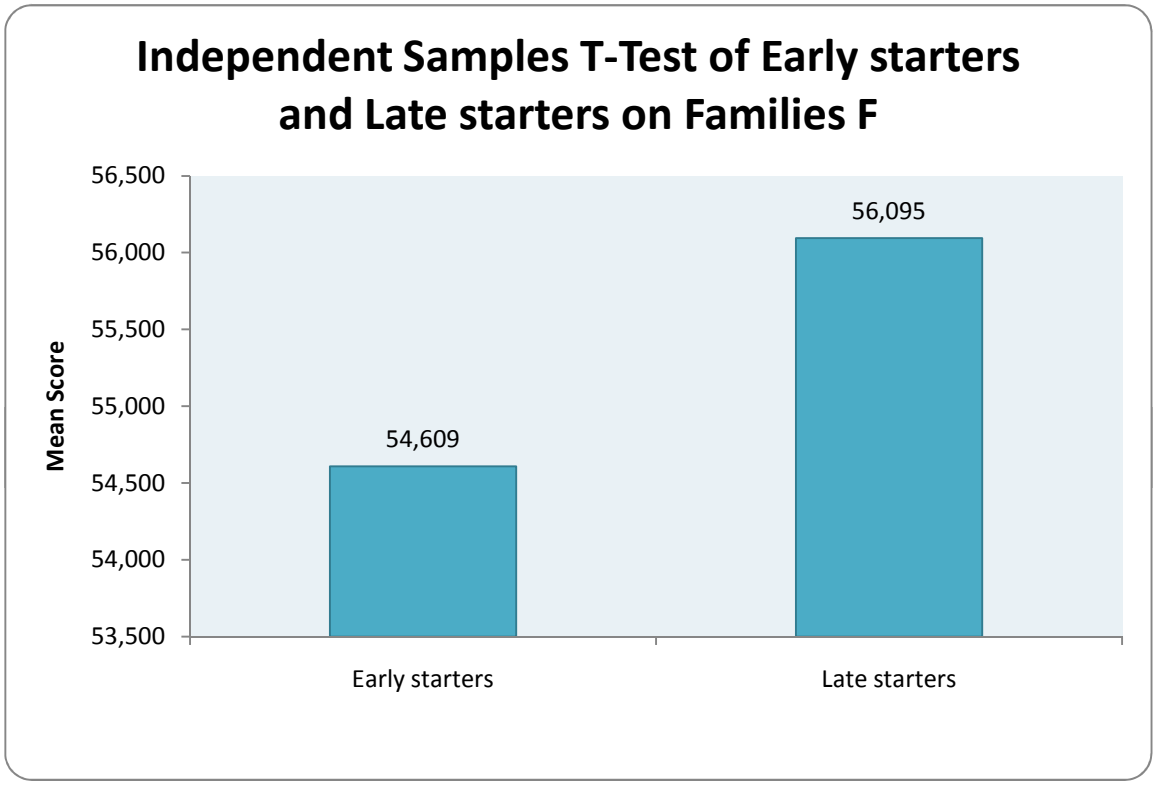


EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Families F

GROUP	Early starters	Late starters
Mean:	54,609	56,095
Std. Dev:	16,267	9,433
N:	23	21

Mean Difference:	1,487
T-Score:	,366
Eta Squared:	,003
P:	,716

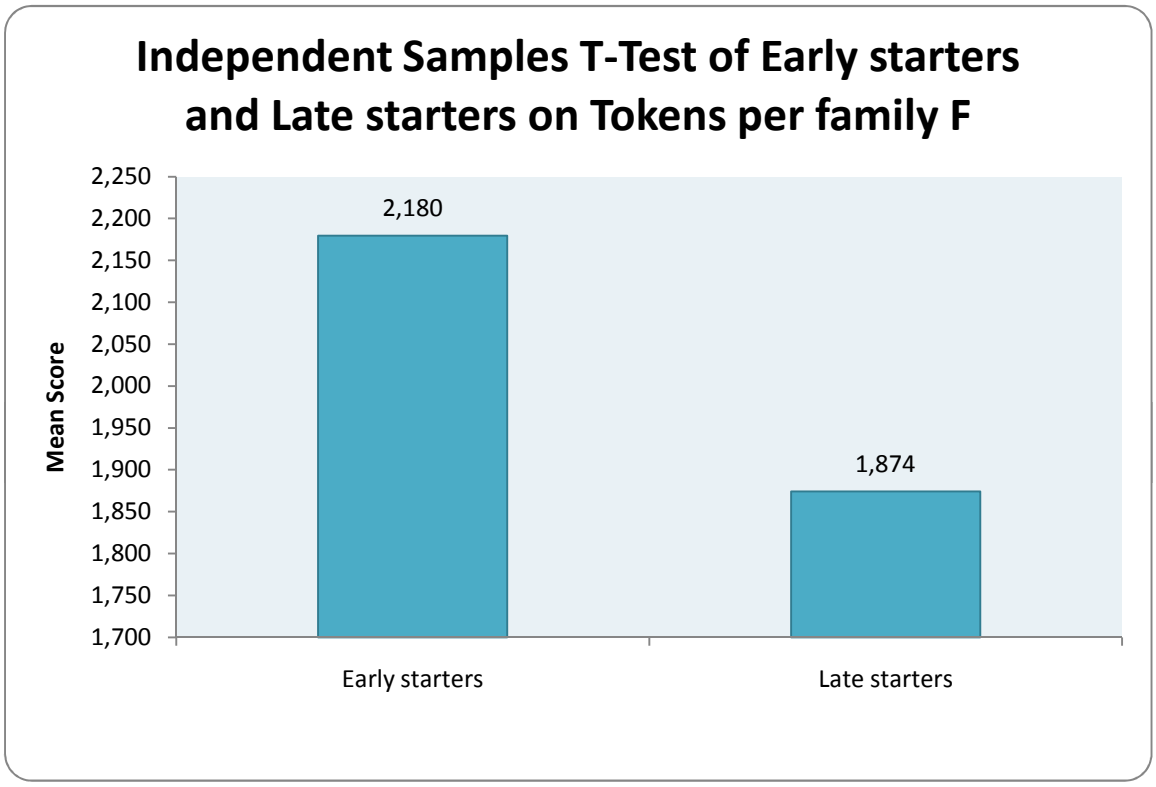
The observed difference between the group means is not significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens per family F

GROUP	Early starters	Late starters
Mean:	2,180	1,874
Std. Dev:	,250	,127
N:	23	21
Mean Difference:	,305	
T-Score:	5,029	
Eta Squared:	,365	
P:	,000	

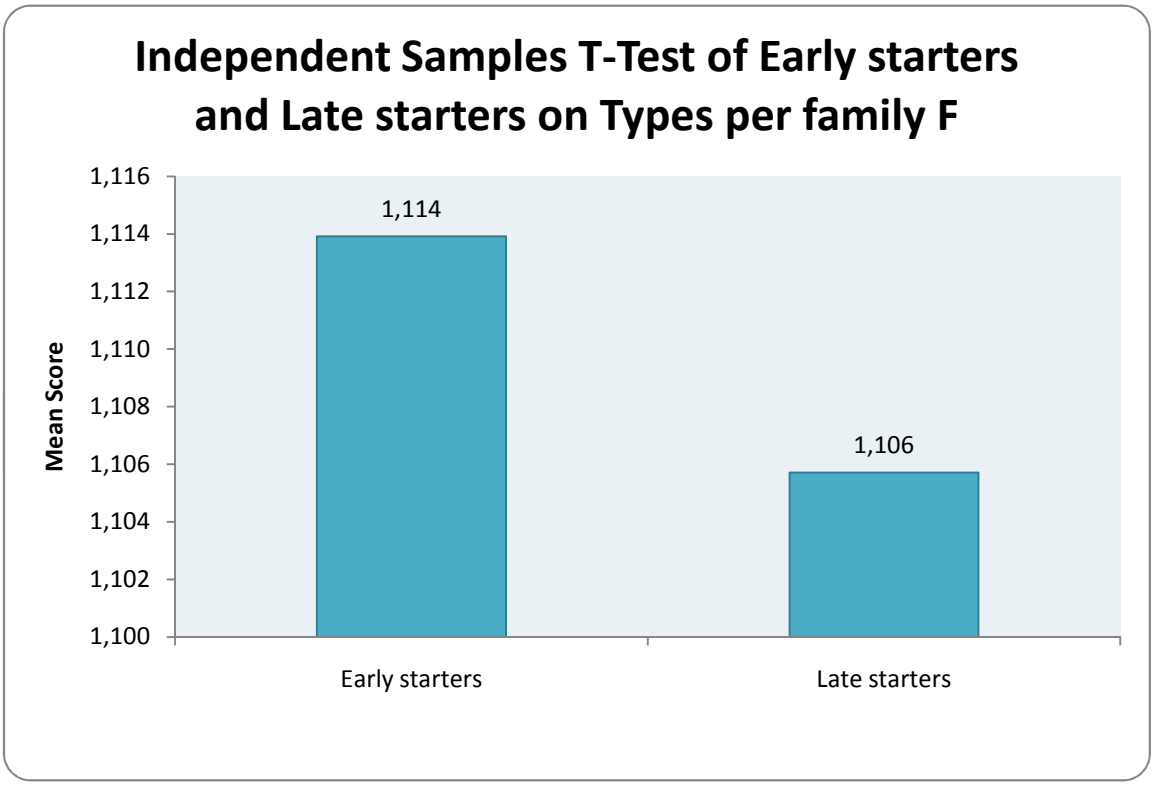
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Types per family F

GROUP	Early starters	Late starters
Mean:	1,114	1,106
Std. Dev:	,030	,033
N:	23	21
Mean Difference:	,008	
T-Score:	,865	
Eta Squared:	,017	
P:	,392	

The observed difference between the group means is not significant

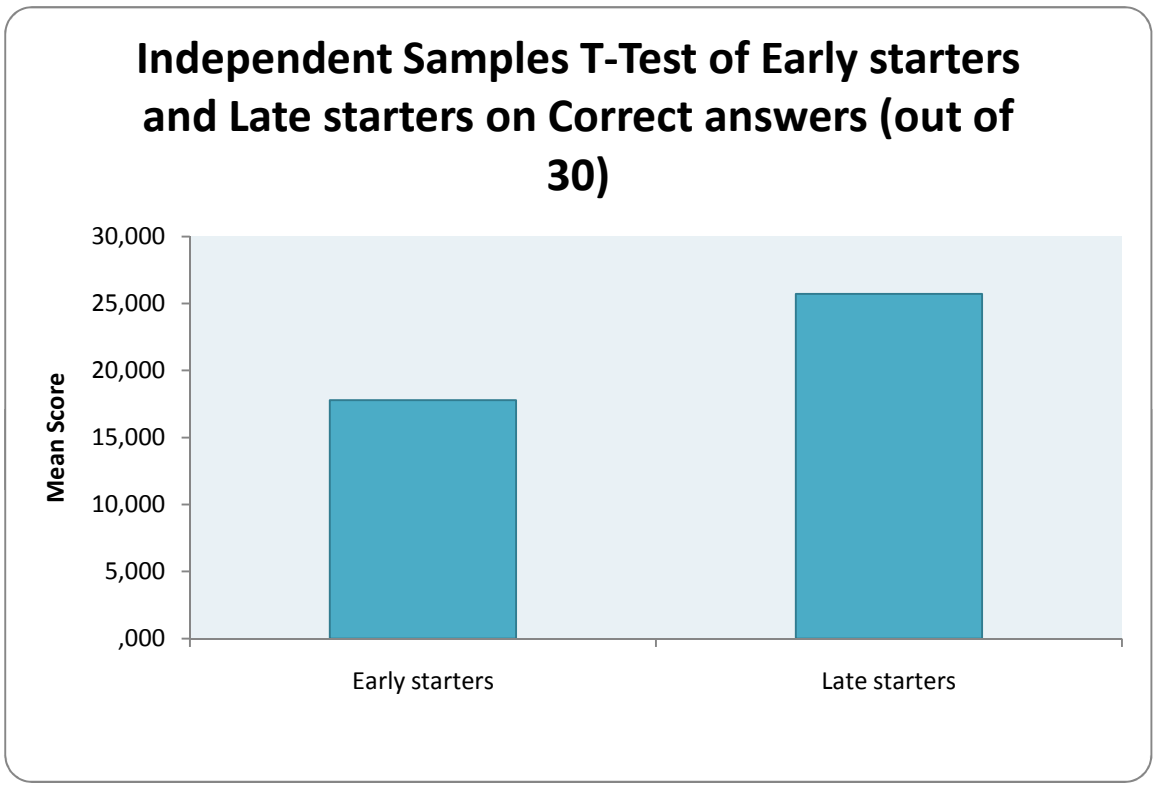


T-TEST: CONTROLLED WRITING ACTIVITY

EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Correct answers (out of 30)

GROUP	Early starters	Late starters
Mean:	17,783	25,714
Std. Dev:	8,634	5,001
N:	23	21
Mean Difference:	7,932	
T-Score:	3,681	
Eta Squared:	,235	
P:	,001	

The observed difference between the group means is significant

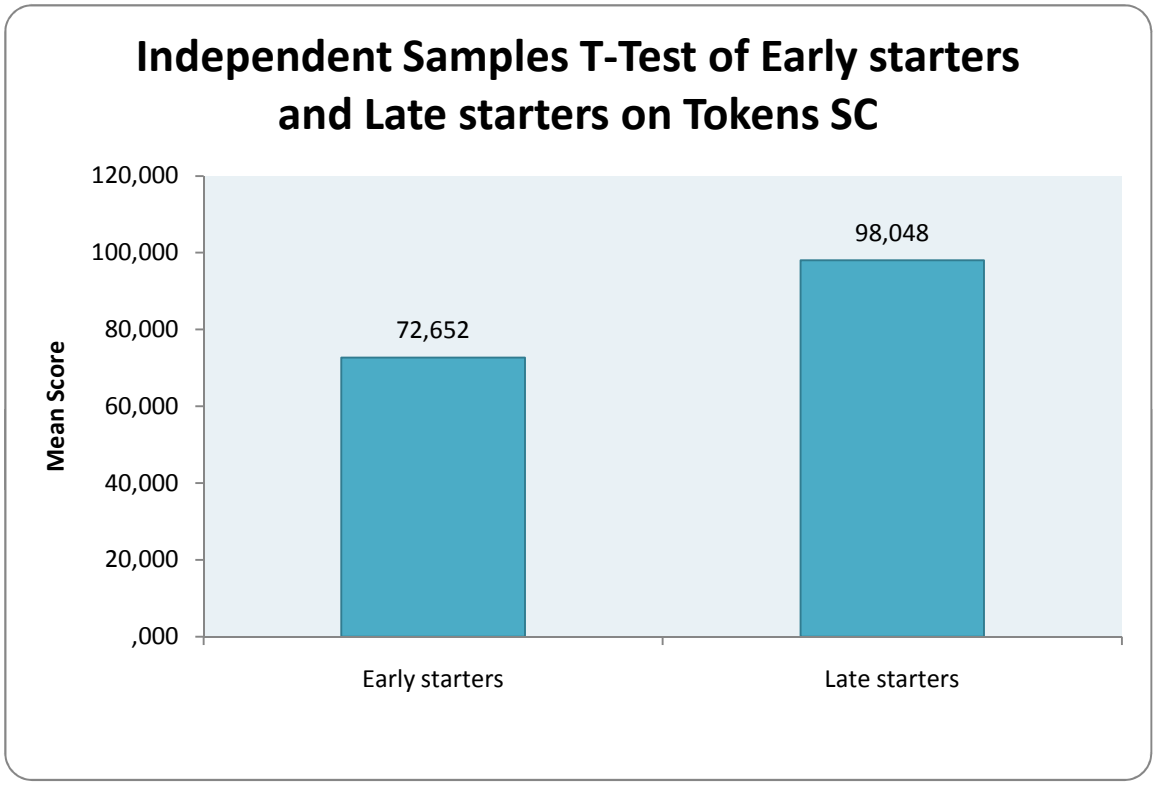


T-TESTS: SEMI-CONTROLLED WRITING ACTIVITY

EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens SC

GROUP	Early starters	Late starters
Mean:	72,652	98,048
Std. Dev:	24,244	31,979
N:	23	21
Mean Difference:	25,395	
T-Score:	2,984	
Eta Squared:	,168	
P:	,005	

The observed difference between the group means is significant

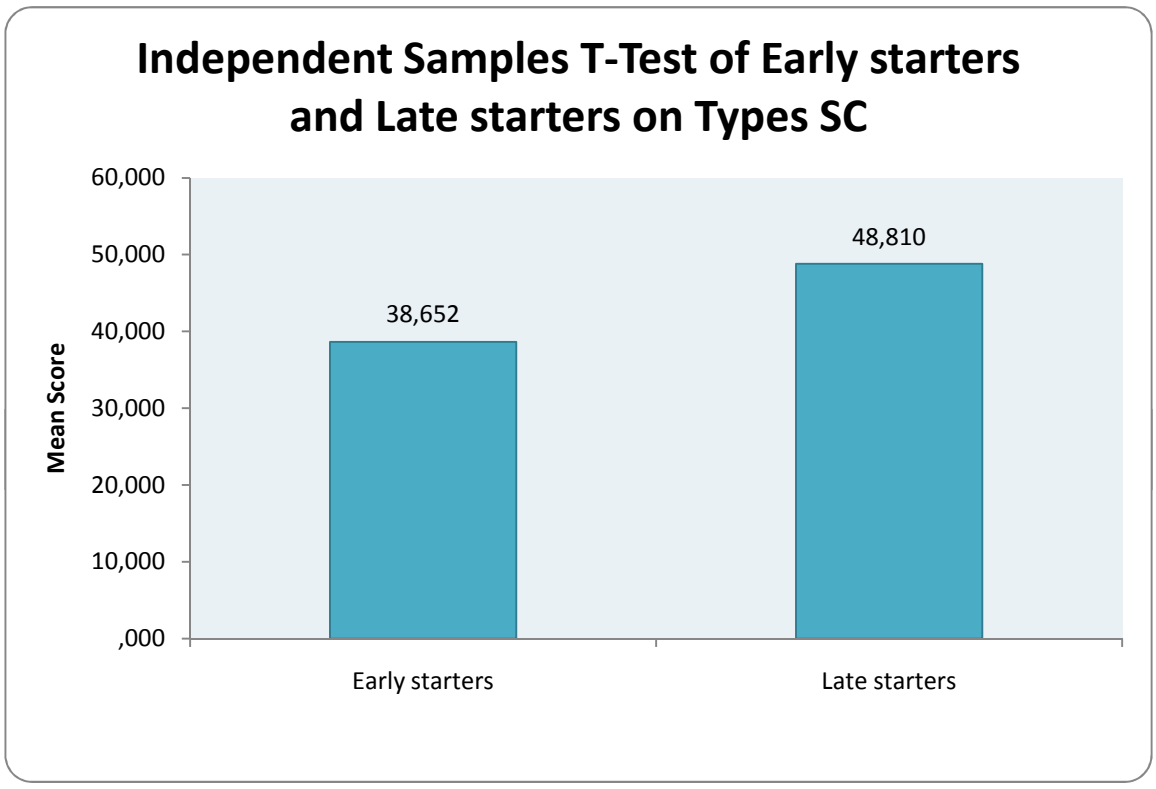


EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Types SC

GROUP	Early starters	Late starters
Mean:	38,652	48,810
Std. Dev:	12,048	14,525
N:	23	21

Mean Difference:	10,157
T-Score:	2,533
Eta Squared:	,127
P:	,015

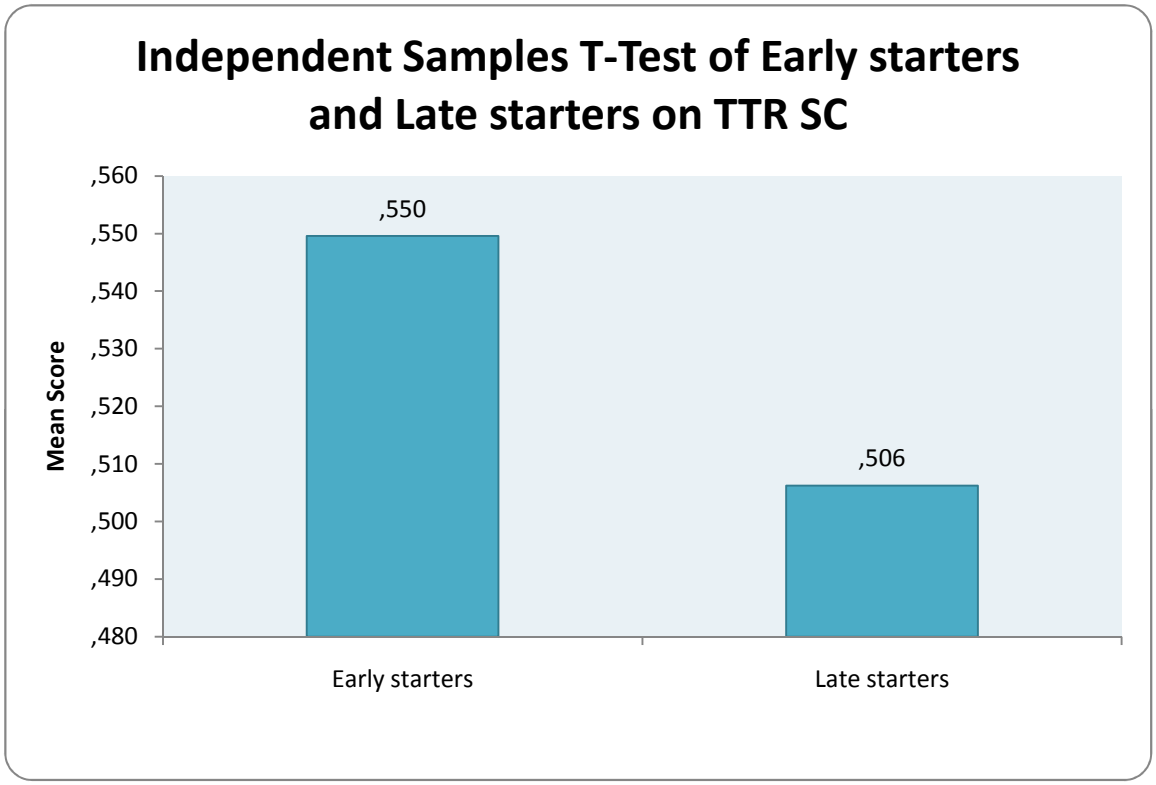
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on TTR SC

GROUP	Early starters	Late starters
Mean:	,550	,506
Std. Dev:	,095	,053
N:	23	21
Mean Difference:	,043	
T-Score:	1,845	
Eta Squared:	,072	
P:	,072	

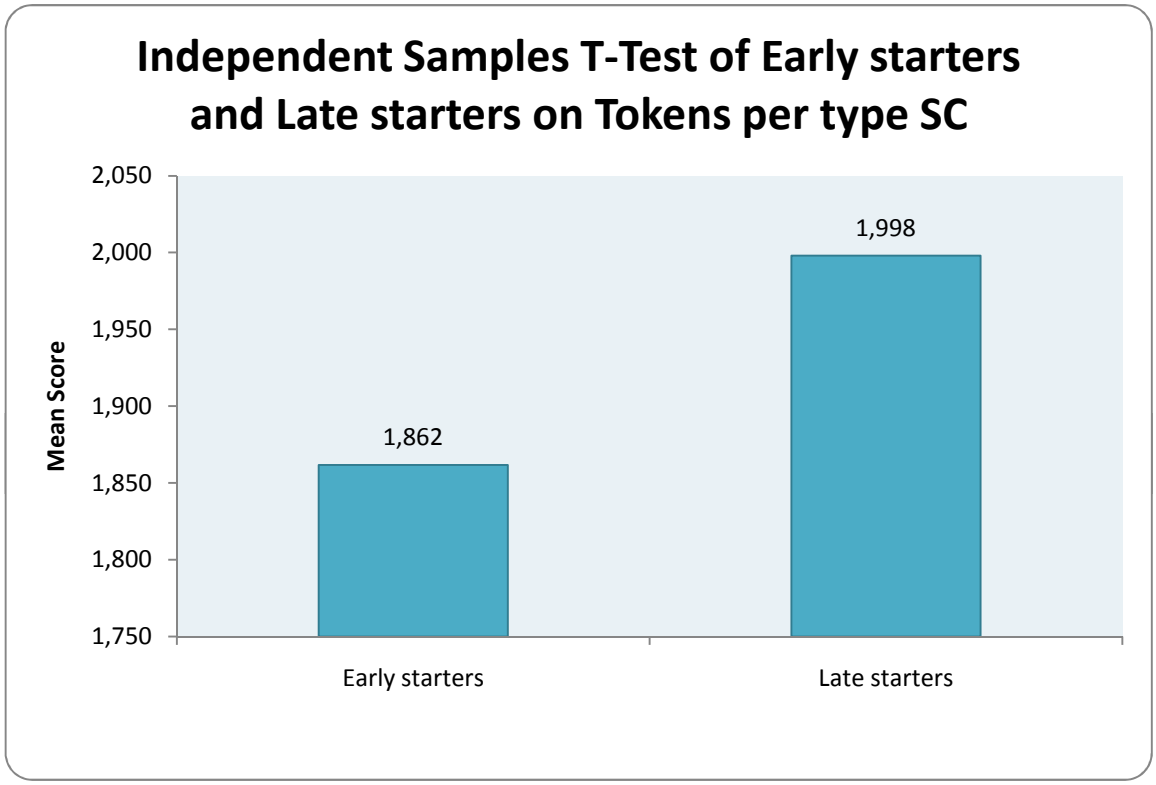
The observed difference between the group means is not significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens per type SC

GROUP	Early starters	Late starters
Mean:	1,862	1,998
Std. Dev:	,259	,228
N:	23	21
Mean Difference:	,136	
T-Score:	1,846	
Eta Squared:	,072	
P:	,072	

The observed difference between the group means is not significant

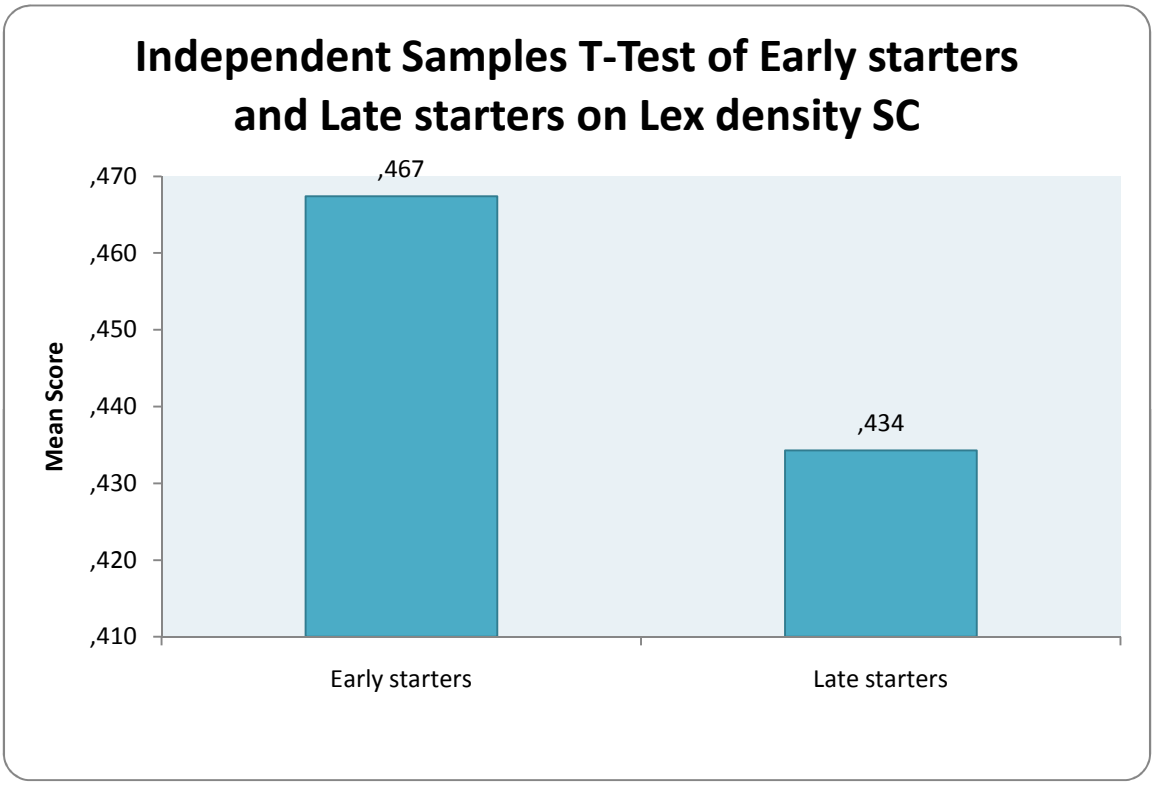


EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Lex density SC

GROUP	Early starters	Late starters
Mean:	,467	,434
Std. Dev:	,034	,033
N:	23	21

Mean Difference:	,033
T-Score:	3,246
Eta Squared:	,193
P:	,002

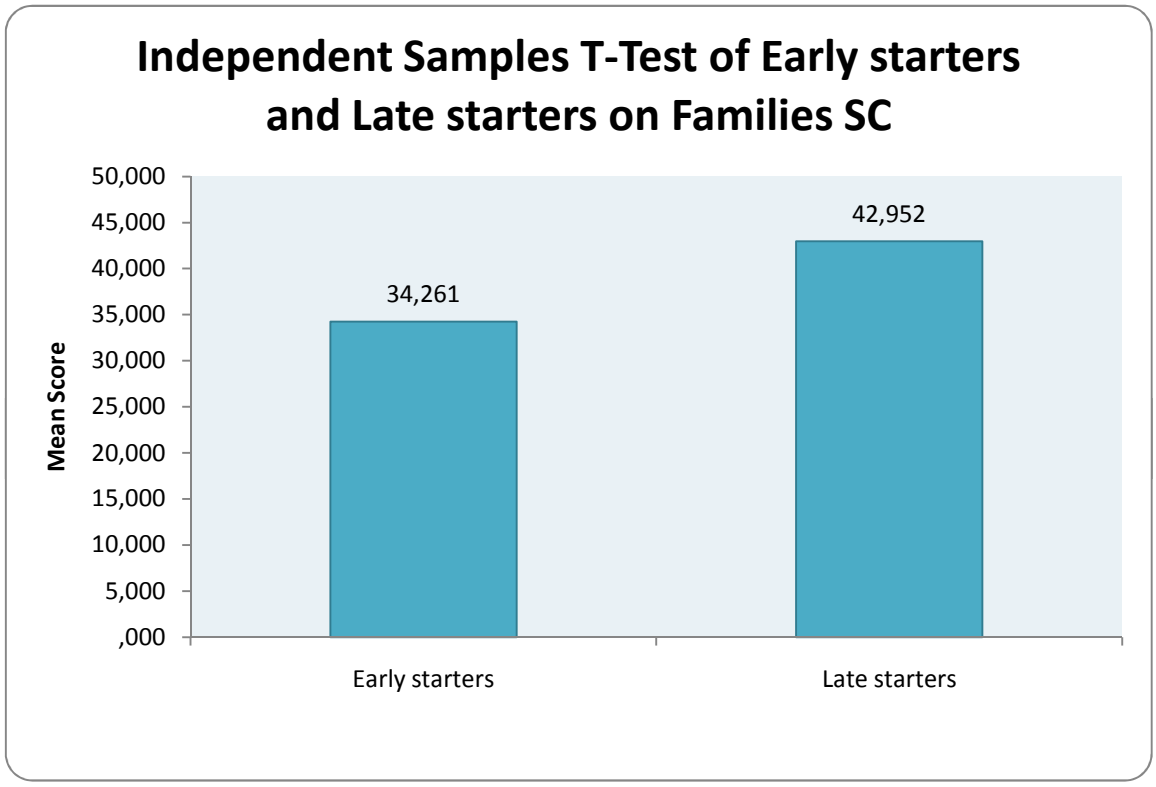
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Families SC

GROUP	Early starters	Late starters
Mean:	34,261	42,952
Std. Dev:	10,323	13,291
N:	23	21
Mean Difference:	8,692	
T-Score:	2,434	
Eta Squared:	,119	
P:	,019	

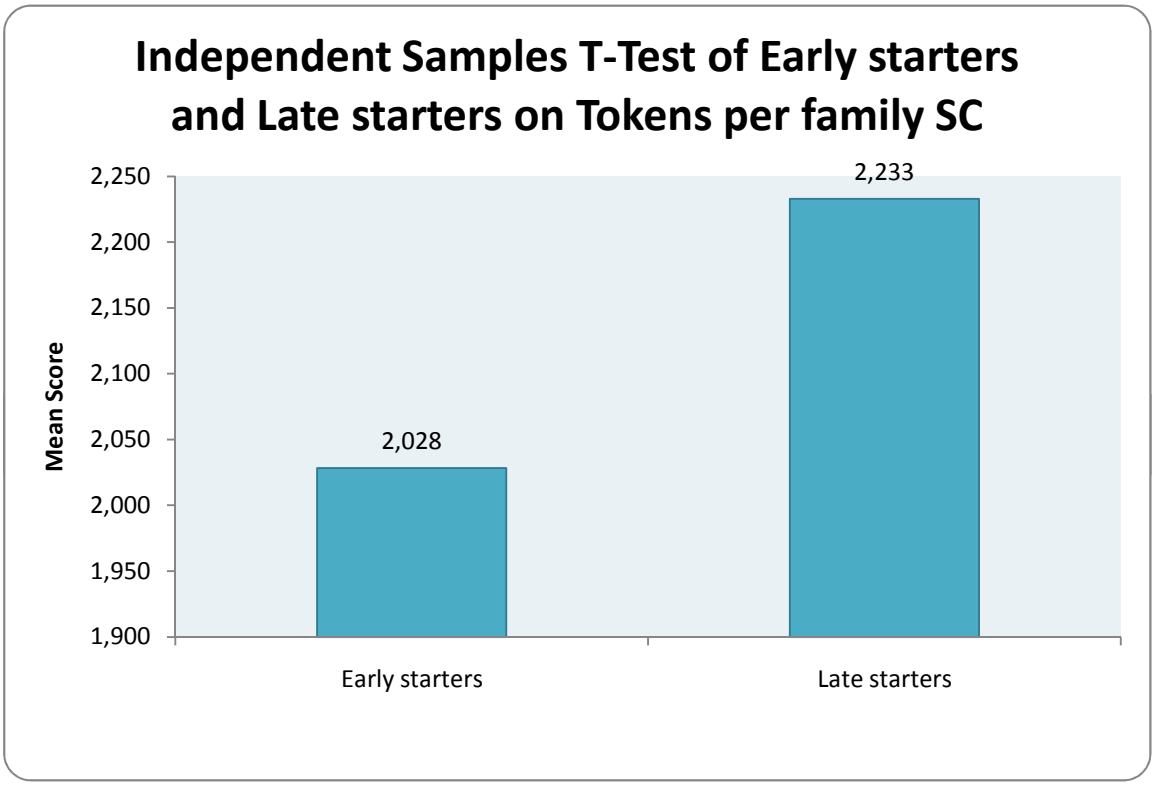
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens per family SC

GROUP	Early starters	Late starters
Mean:	2,028	2,233
Std. Dev:	,340	,251
N:	23	21
Mean Difference:	,205	
T-Score:	2,253	
Eta Squared:	,103	
P:	,030	

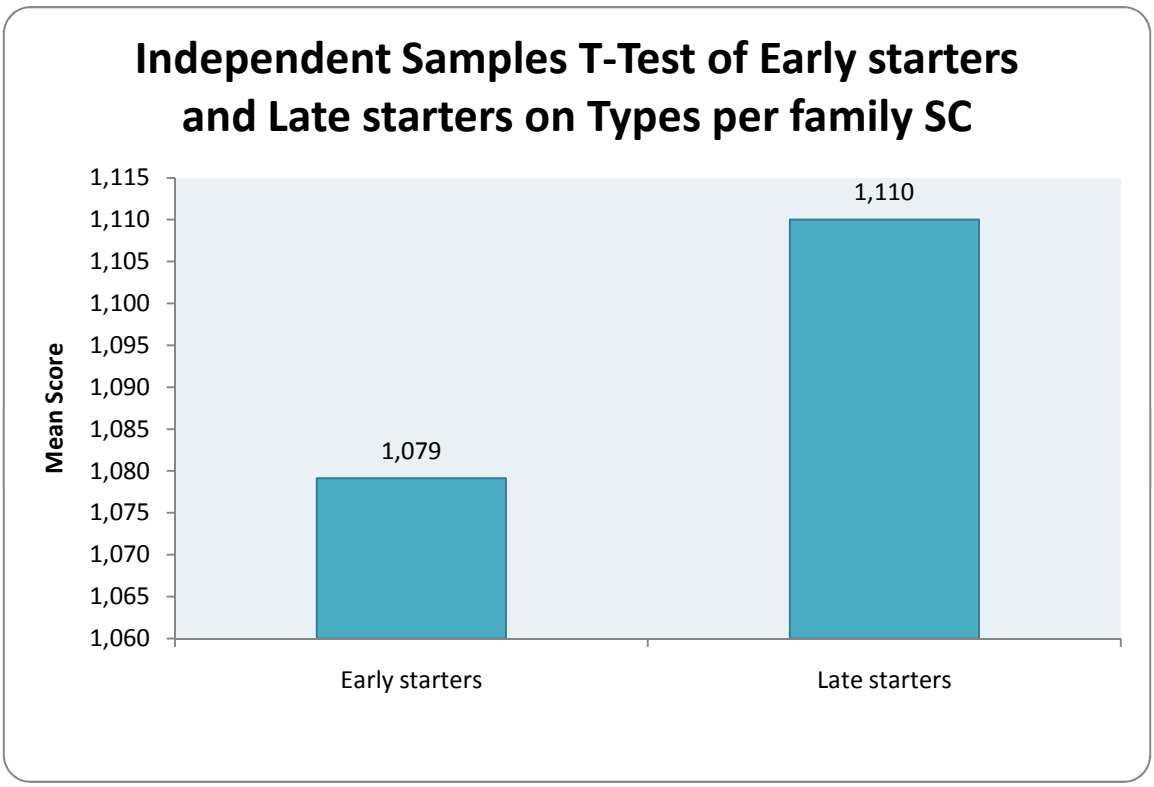
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Types per family SC

GROUP	Early starters	Late starters
Mean:	1,079	1,110
Std. Dev:	,056	,044
N:	23	21
Mean Difference:	,031	
T-Score:	2,006	
Eta Squared:	,084	
P:	,051	

The observed difference between the group means is not significant

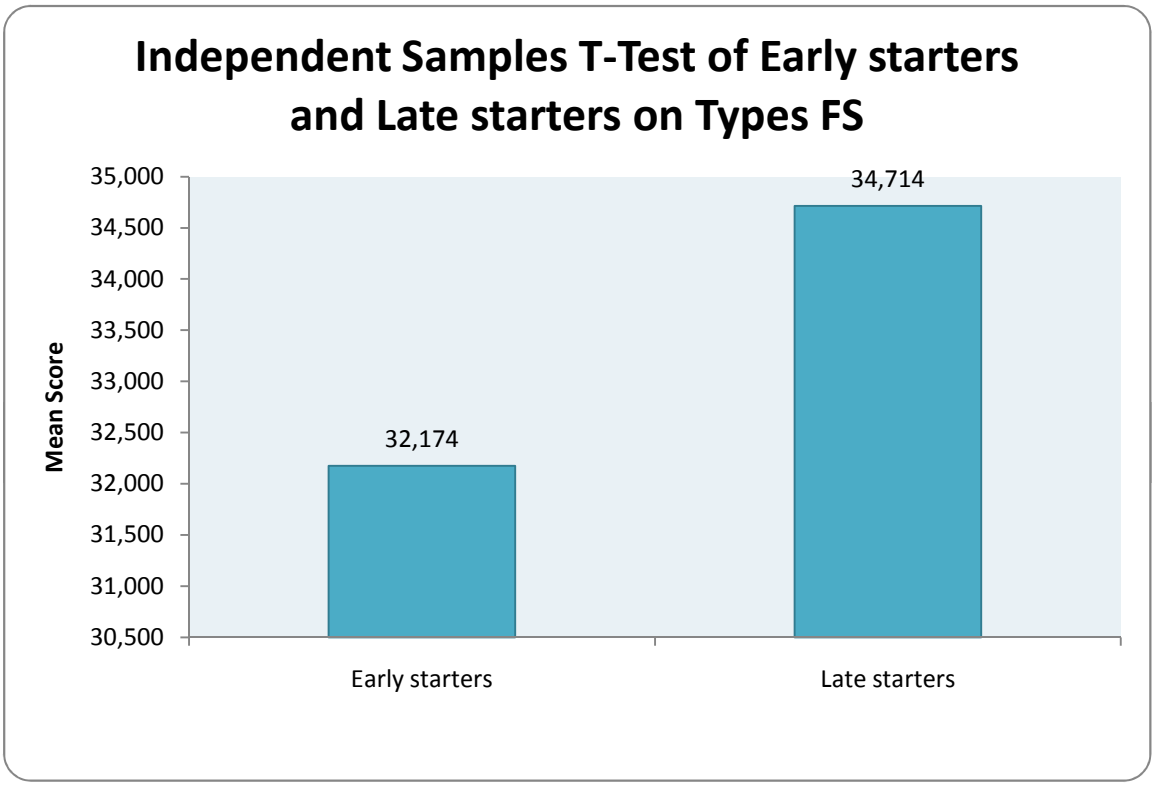


T-TESTS: FREE-WRITING ACTIVITY (STANDARDIZED)

EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Types FS

GROUP	Early starters	Late starters
Mean:	32,174	34,714
Std. Dev:	4,053	3,002
N:	23	21
Mean Difference:	2,540	
T-Score:	2,344	
Eta Squared:	,111	
P:	,024	

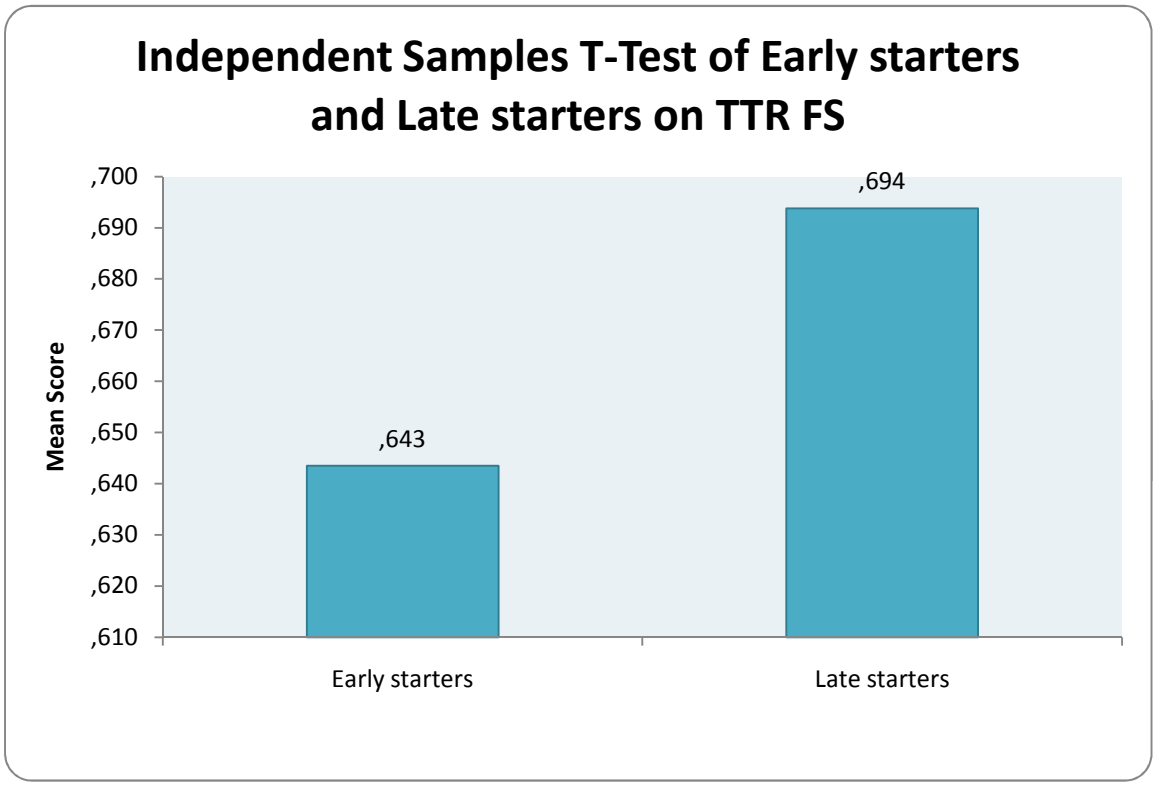
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on TTR FS

GROUP	Early starters	Late starters
Mean:	,643	,694
Std. Dev:	,081	,060
N:	23	21
Mean Difference:	,050	
T-Score:	2,324	
Eta Squared:	,109	
P:	,025	

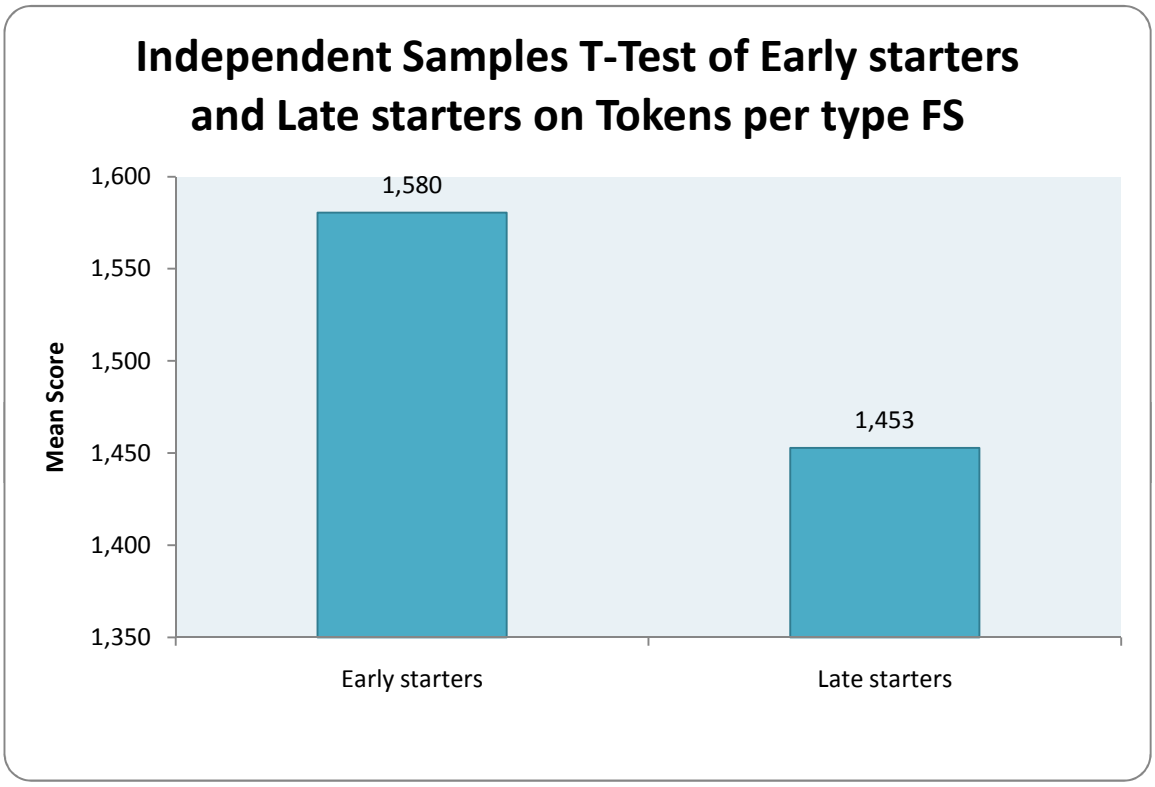
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens per type FS

GROUP	Early starters	Late starters
Mean:	1,580	1,453
Std. Dev:	,223	,128
N:	23	21
Mean Difference:	,128	
T-Score:	2,295	
Eta Squared:	,107	
P:	,027	

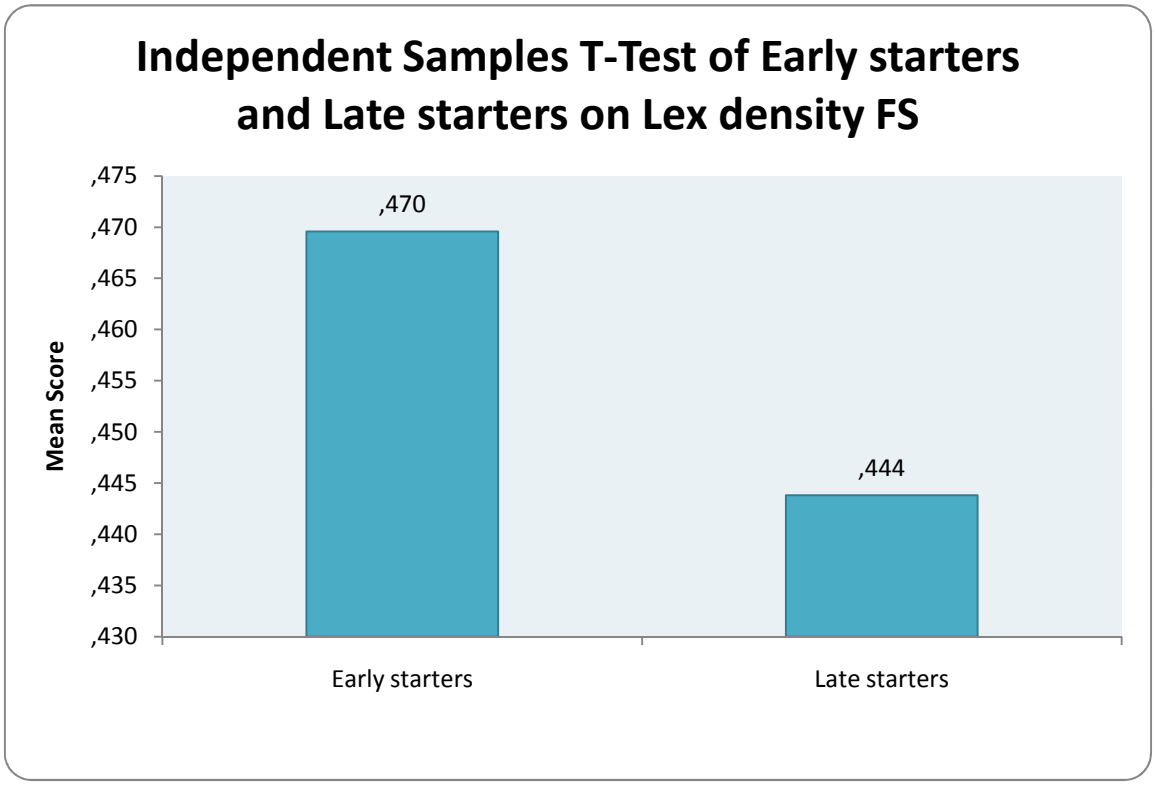
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Lex density FS

GROUP	Early starters	Late starters
Mean:	,470	,444
Std. Dev:	,050	,063
N:	23	21
Mean Difference:	,026	
T-Score:	1,505	
Eta Squared:	,049	
P:	,140	

The observed difference between the group means is not significant

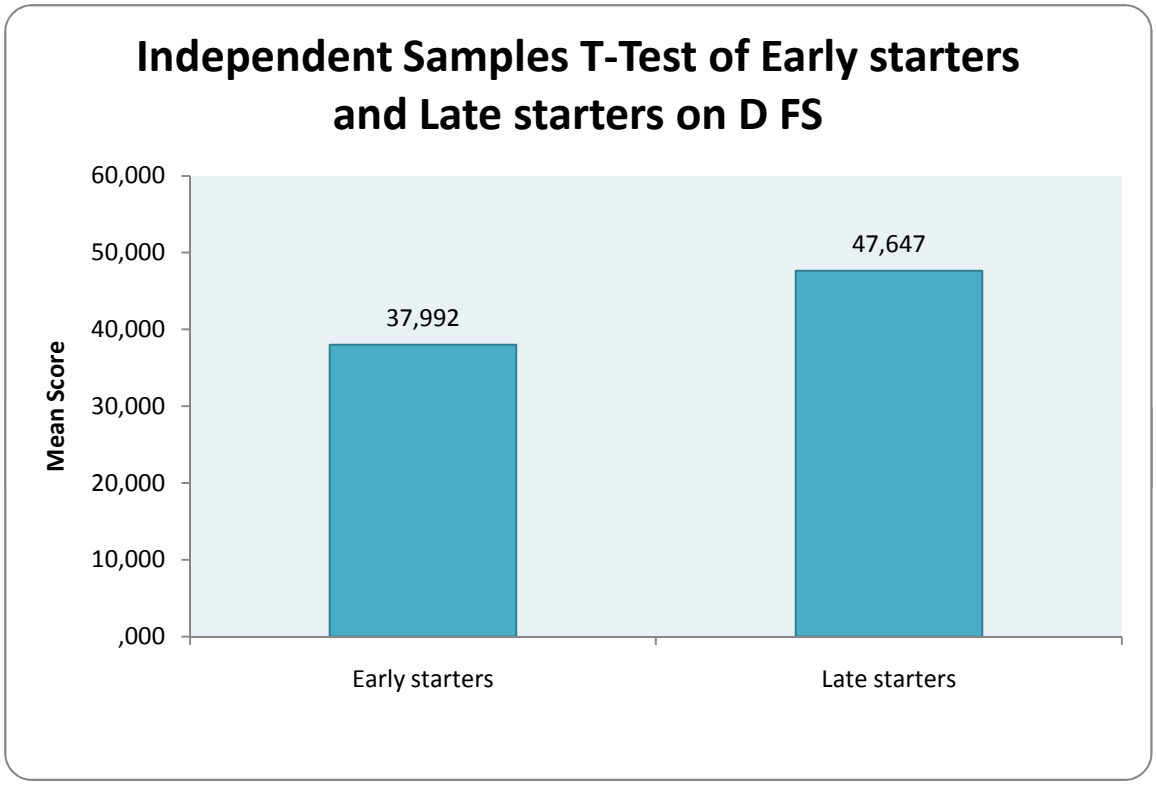


EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on D FS

GROUP	Early starters	Late starters
Mean:	37,992	47,647
Std. Dev:	15,049	13,432
N:	23	21

Mean Difference:	9,655
T-Score:	2,237
Eta Squared:	,102
P:	,031

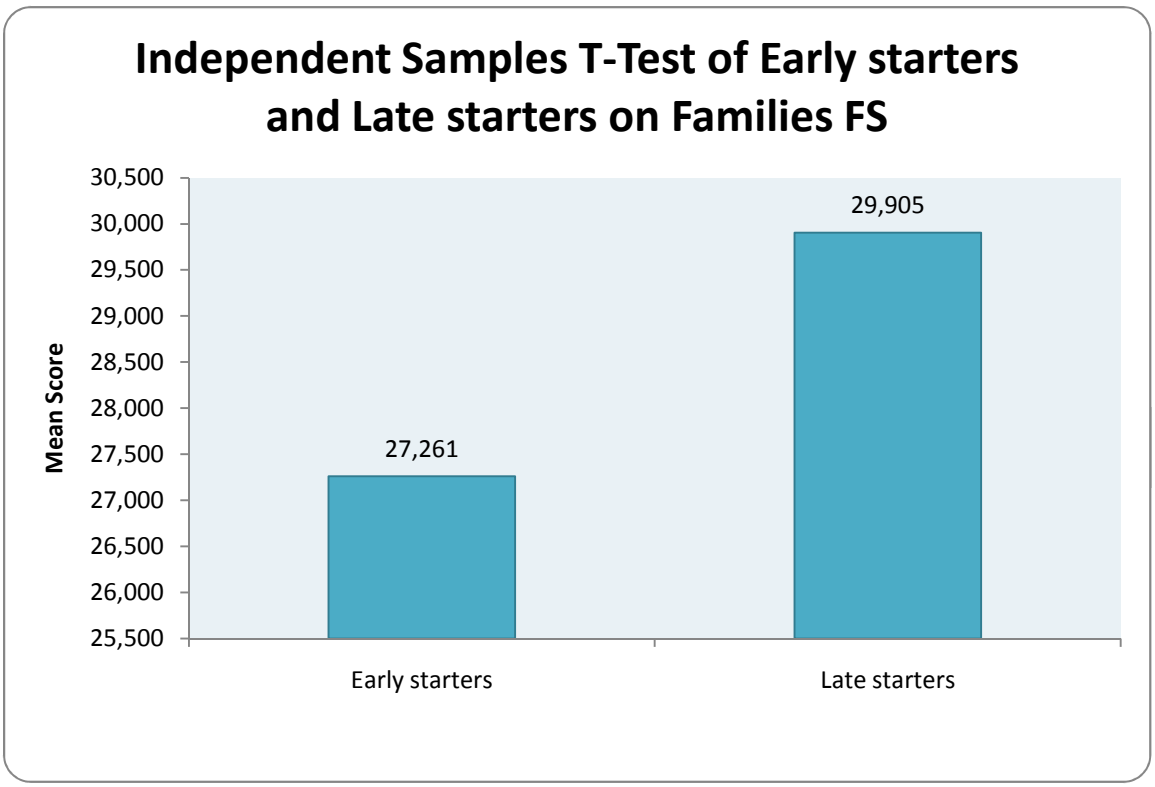
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Families FS

GROUP	Early starters	Late starters
Mean:	27,261	29,905
Std. Dev:	3,583	2,567
N:	23	21
Mean Difference:	2,644	
T-Score:	2,789	
Eta Squared:	,150	
P:	,008	

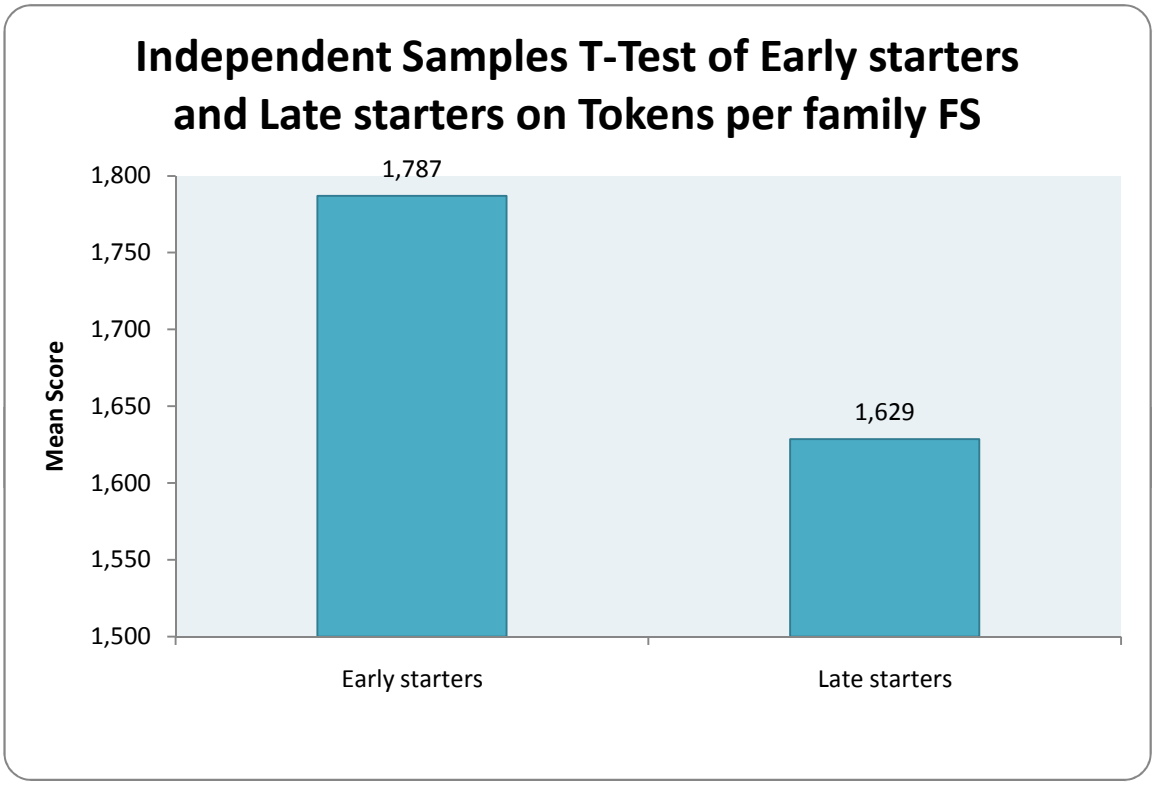
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Tokens per family FS

GROUP	Early starters	Late starters
Mean:	1,787	1,629
Std. Dev:	,269	,138
N:	23	21
Mean Difference:	,158	
T-Score:	2,421	
Eta Squared:	,118	
P:	,020	

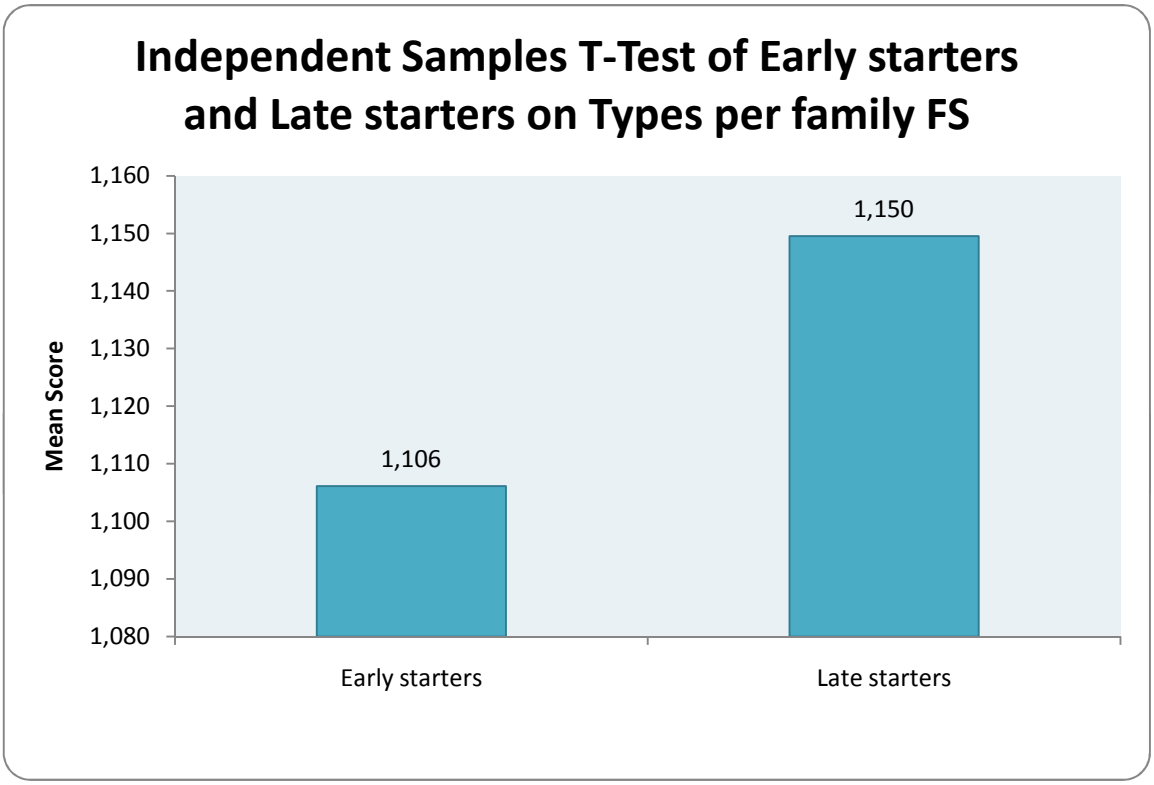
The observed difference between the group means is significant



EZAnalyze Results Report - Independent T-Test of group Early starters and Late starters on Types per family FS

GROUP	Early starters	Late starters
Mean:	1,106	1,150
Std. Dev:	,038	,194
N:	23	21
Mean Difference:	,043	
T-Score:	1,053	
Eta Squared:	,025	
P:	,299	

The observed difference between the group means is not significant

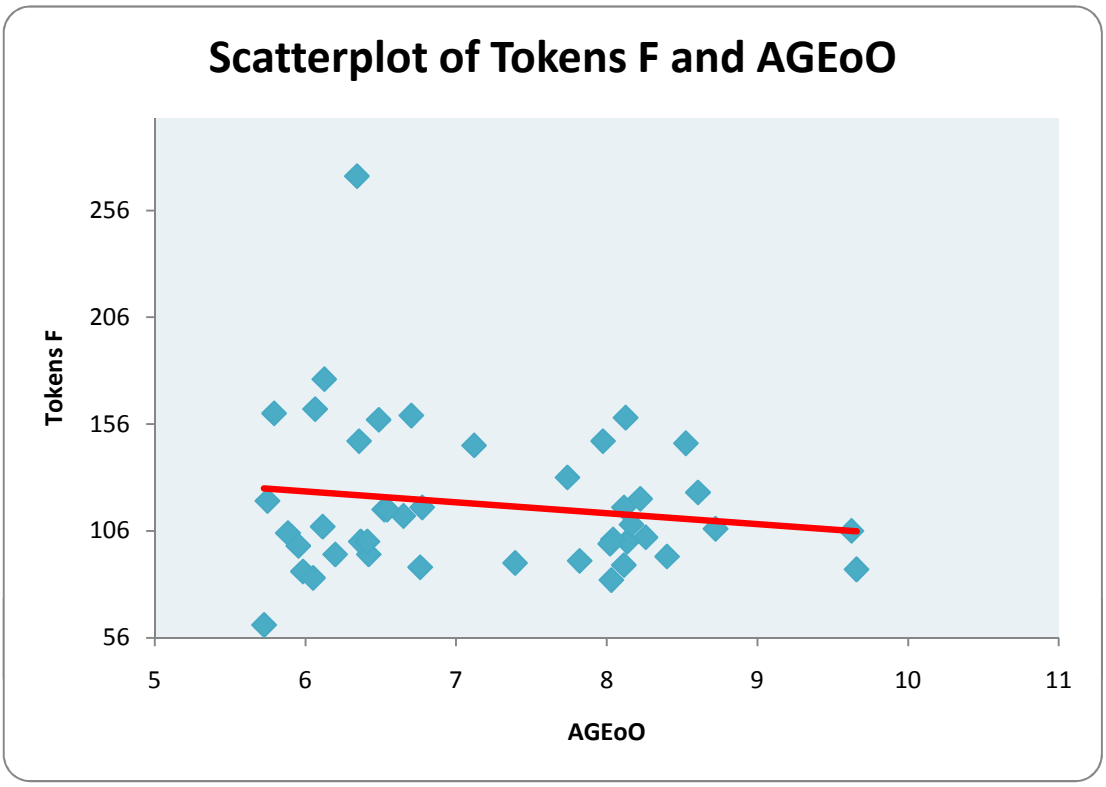


CORRELATIONS: AGE OF ONSET

EZAnalyze Results Report - Correlation of Tokens F with AGEoO

Pearson Correlation	-,157
N	44,000
P	,308

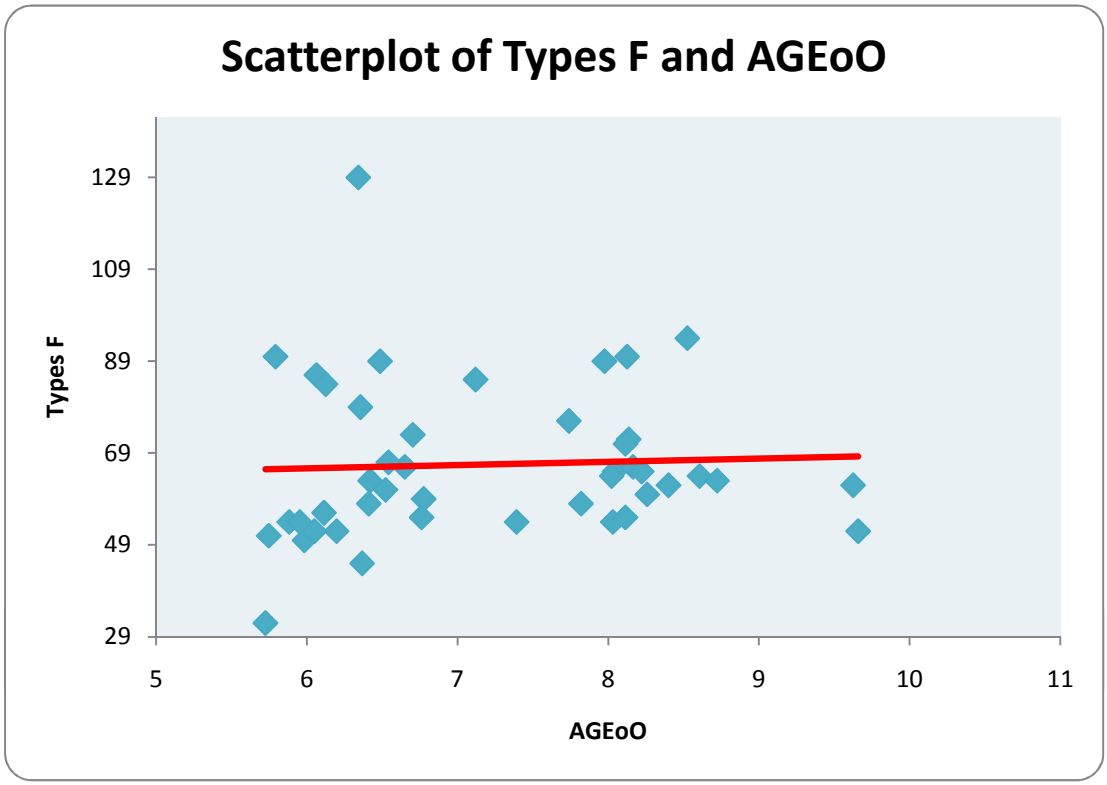
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Types F with AGEoO

Pearson Correlation	,046
N	44,000
P	,767

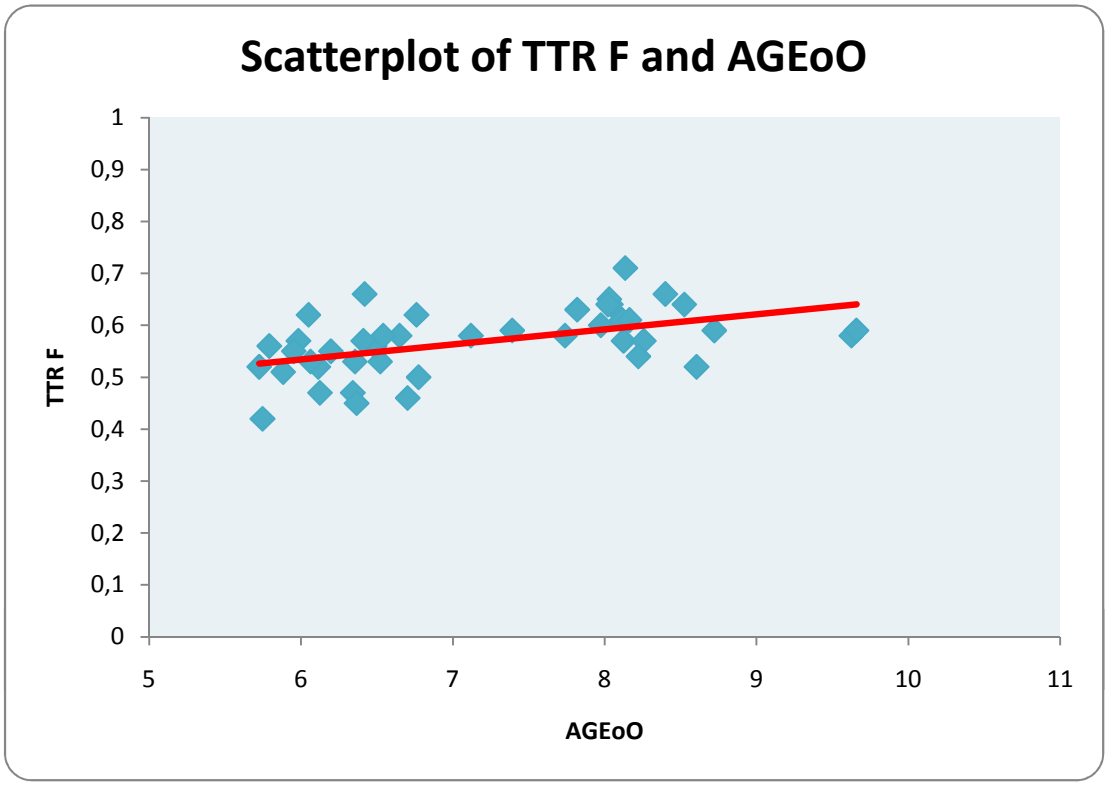
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of TTR F with AGEoO

Pearson Correlation	,517
N	44,000
P	,000

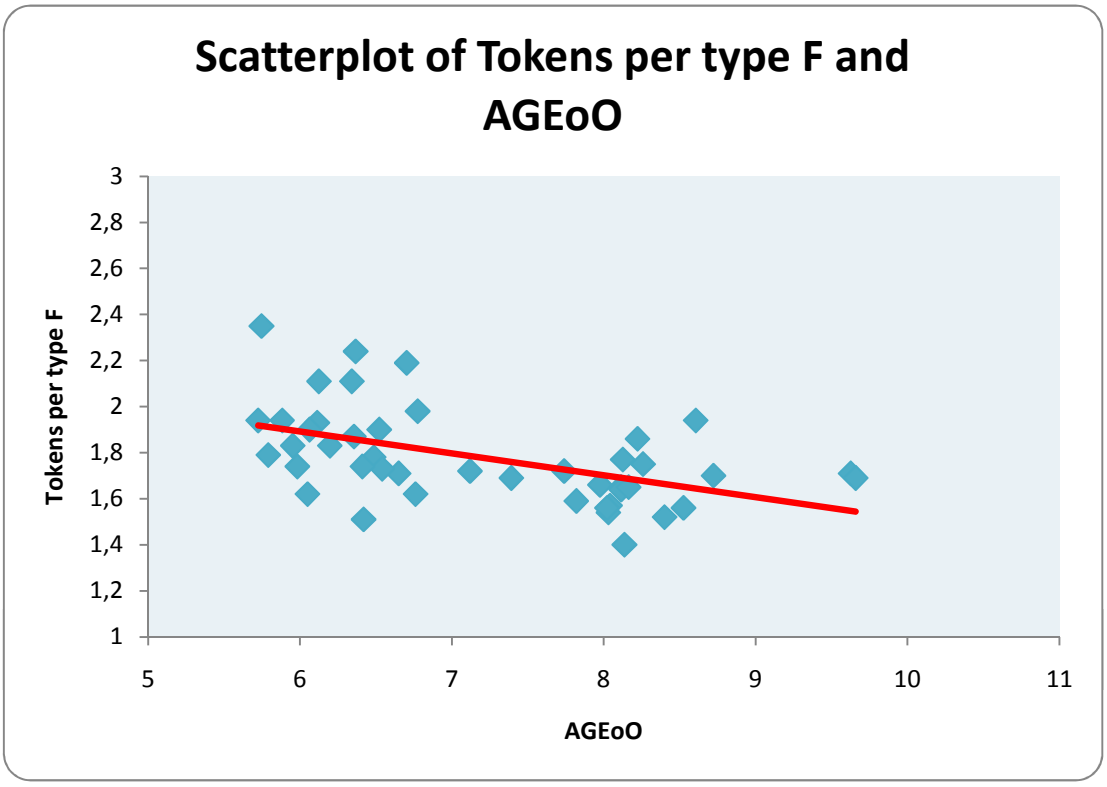
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Tokens per type F with AGEoO

Pearson Correlation	-,512
N	44,000
P	,000

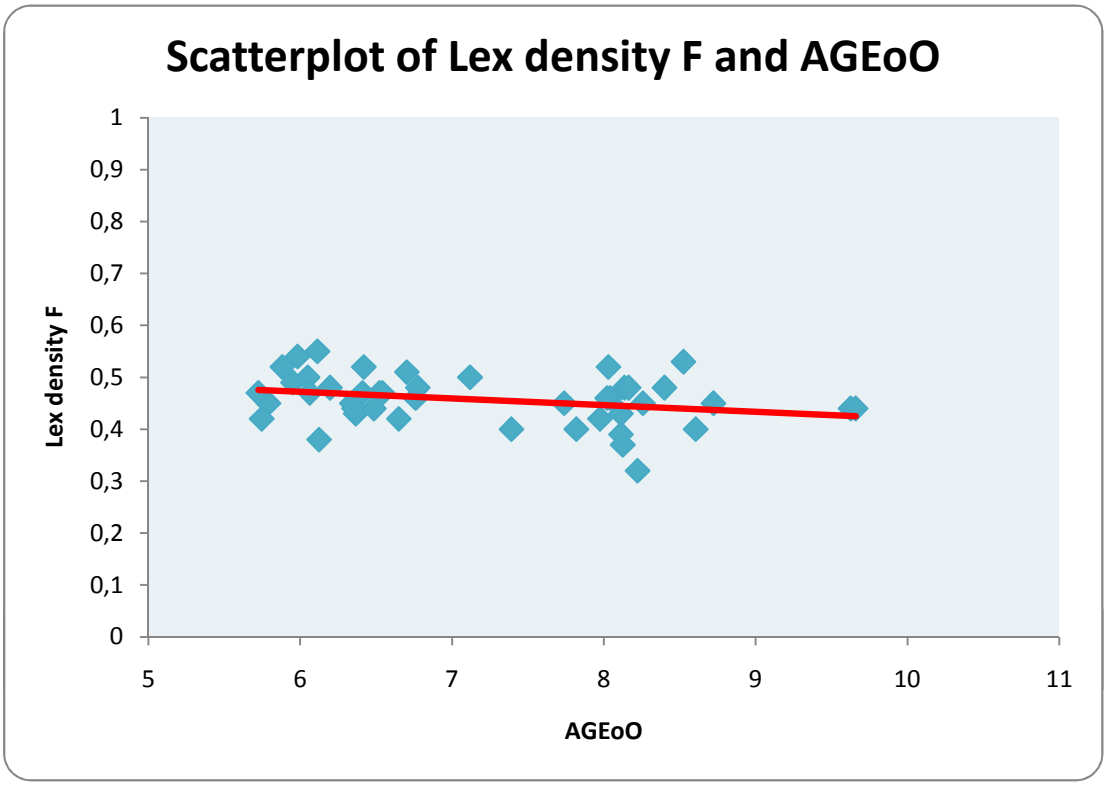
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Lex density F with AGEoO

Pearson Correlation	-,298
N	44,000
P	,050

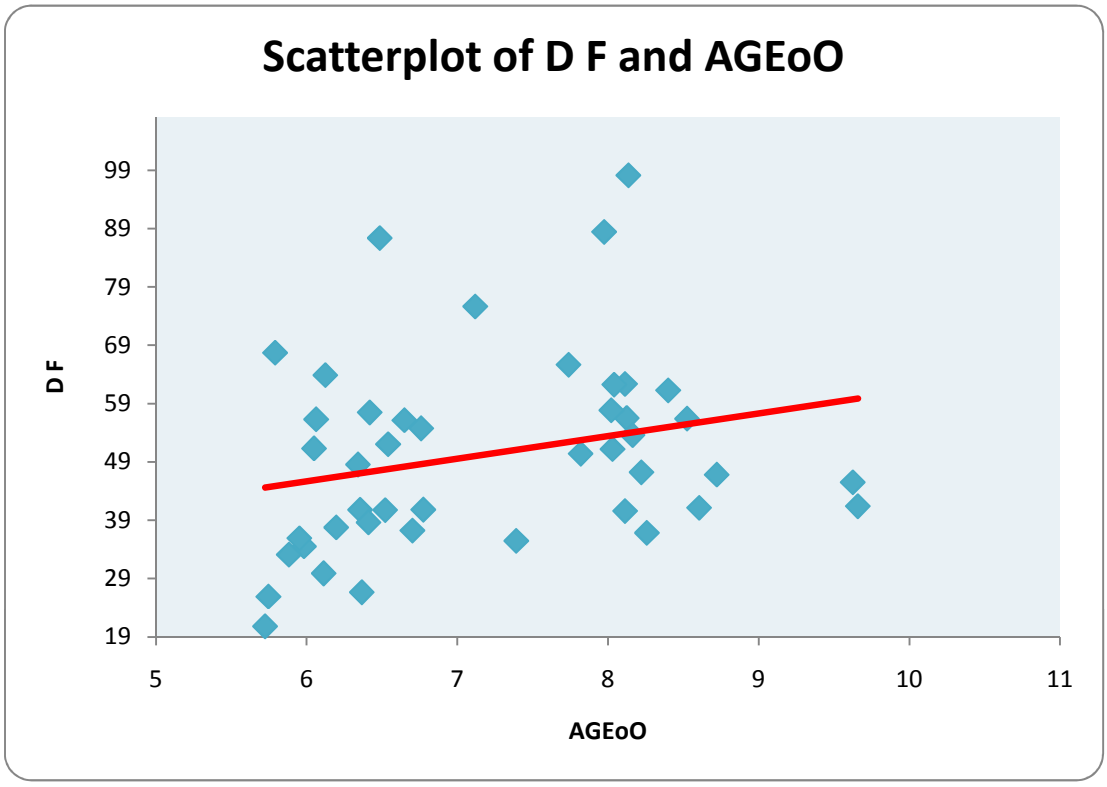
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of D F with AGEoO

Pearson Correlation	,257
N	44,000
P	,092

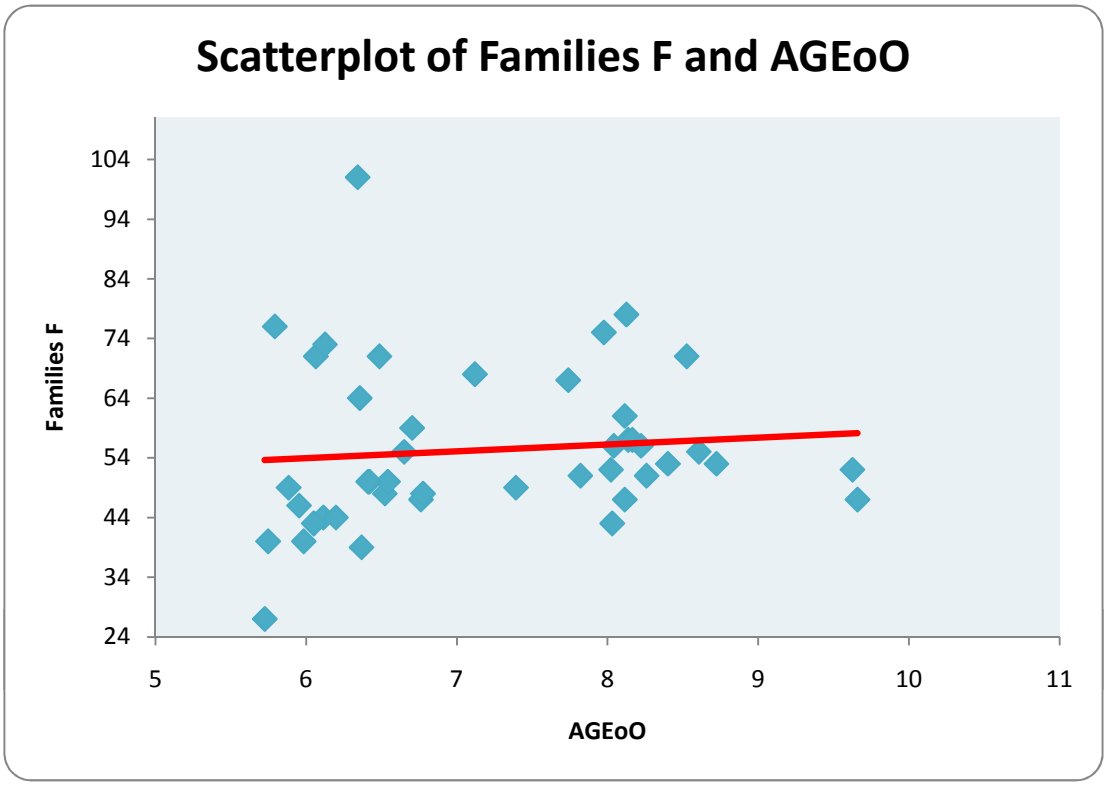
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Families F with AGEoO

Pearson Correlation	,095
N	44,000
P	,539

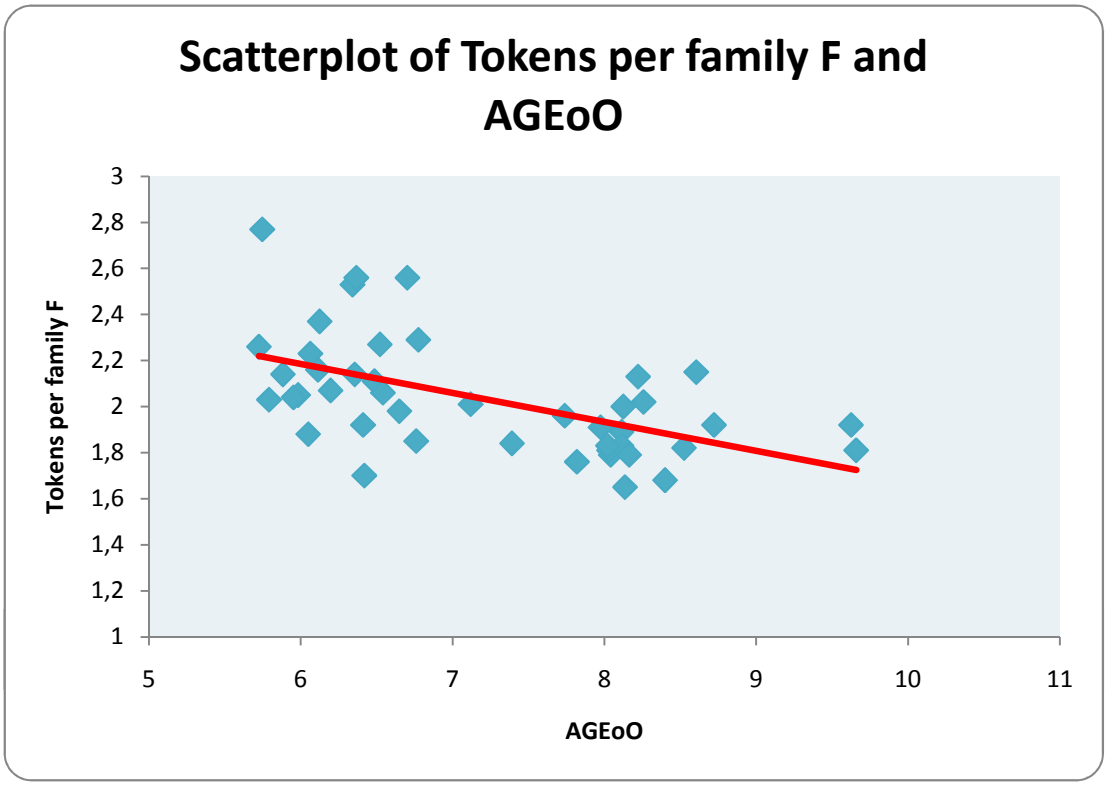
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Tokens per family F with AGEoO

Pearson Correlation	-,550
N	44,000
P	,000

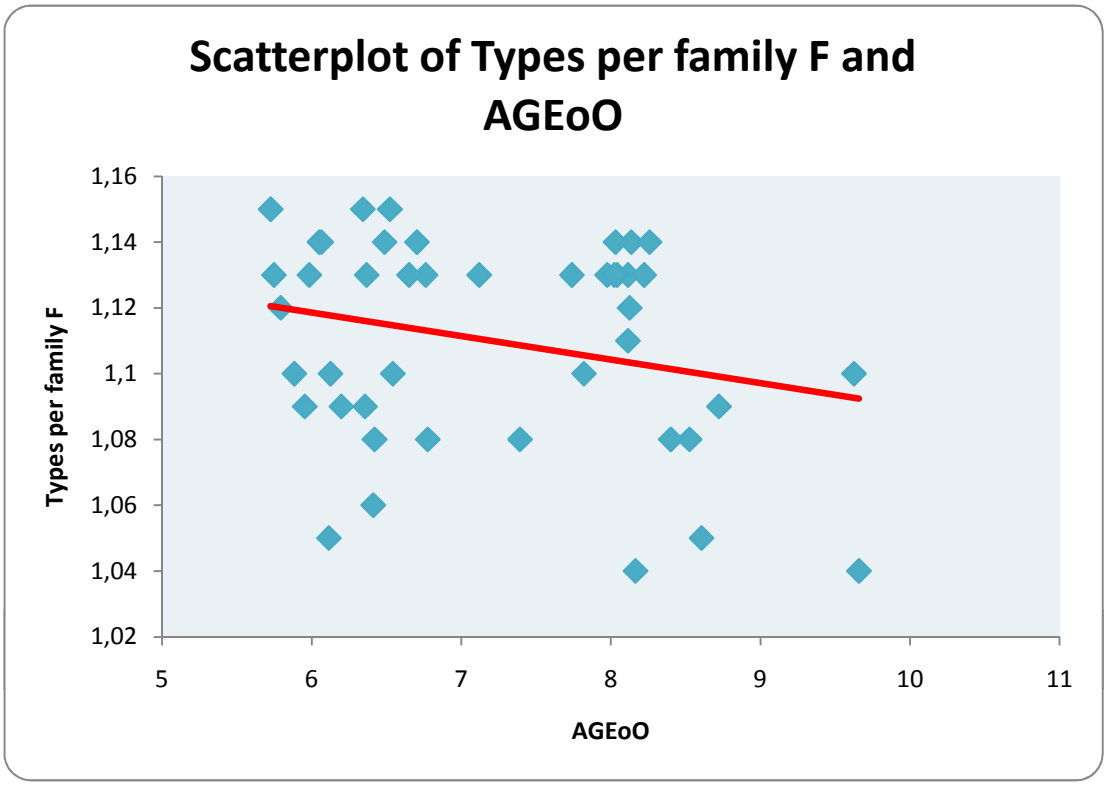
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Types per family F with AGEoO

Pearson Correlation	-,251
N	44,000
P	,100

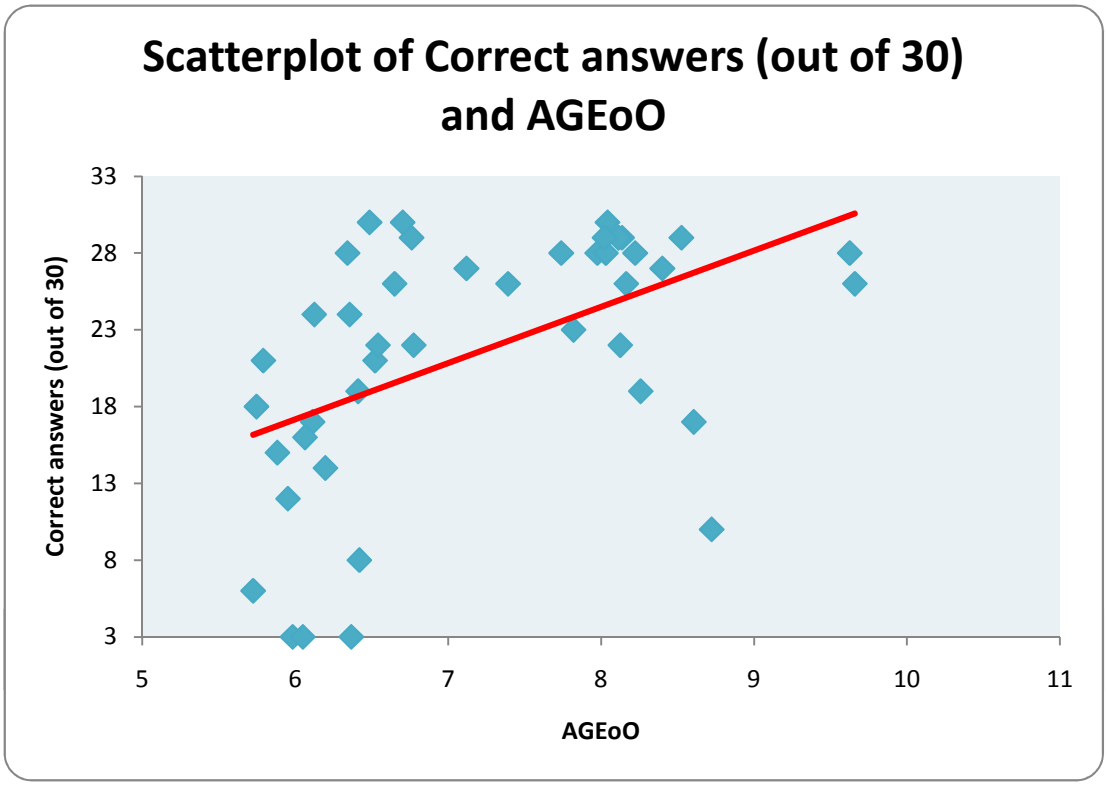
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Correct answers (out of 30) with AGEoO

Pearson Correlation	,496
N	44,000
P	,001

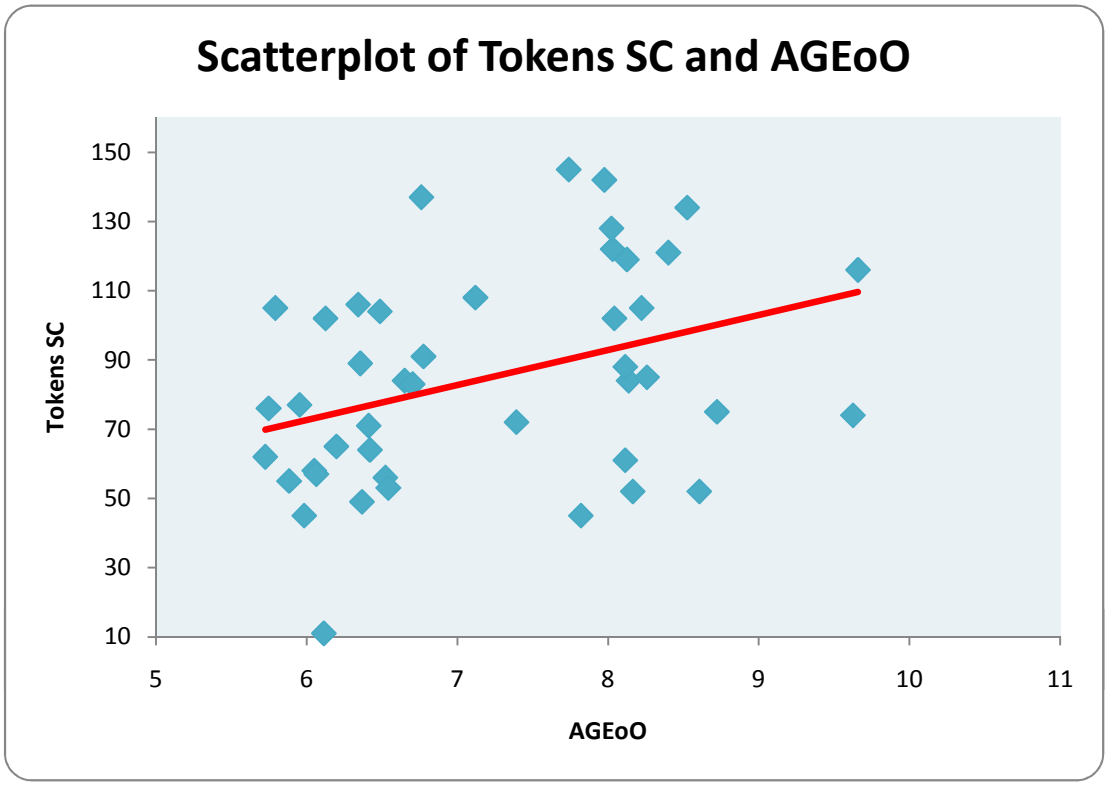
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Tokens SC with AGEoO

Pearson Correlation	,362
N	44,000
P	,016

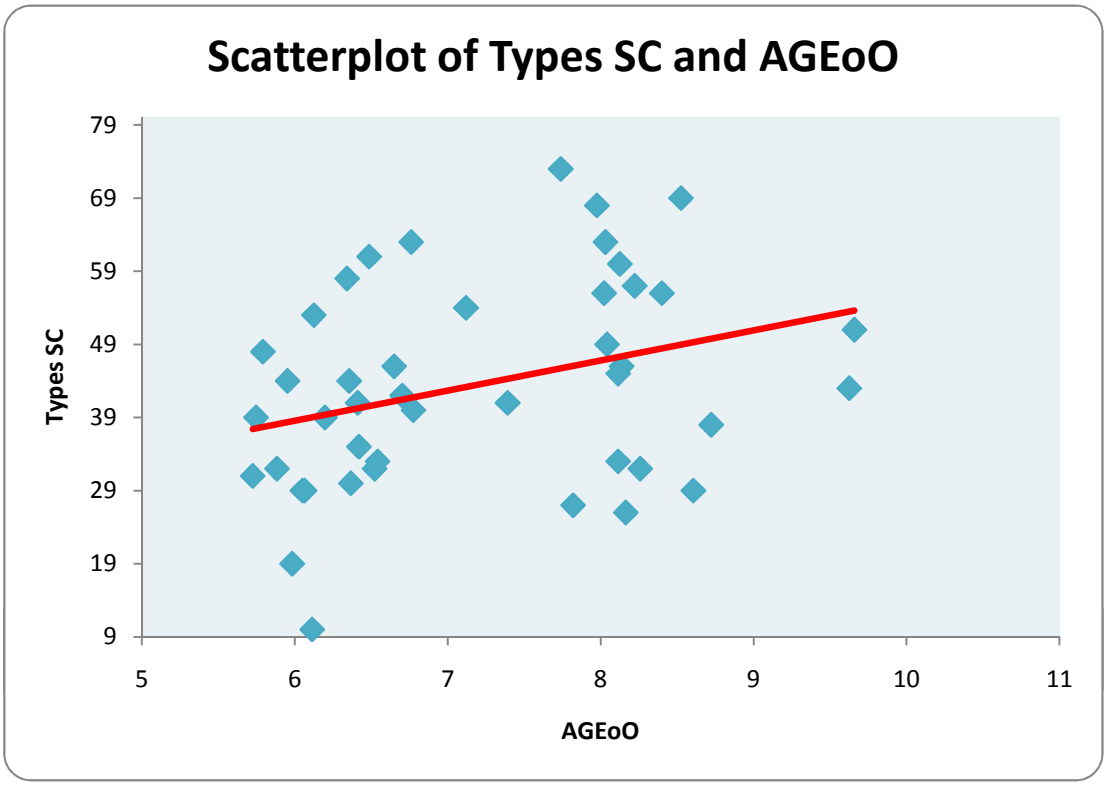
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Types SC with AGEoO

Pearson Correlation	,321
N	44,000
P	,033

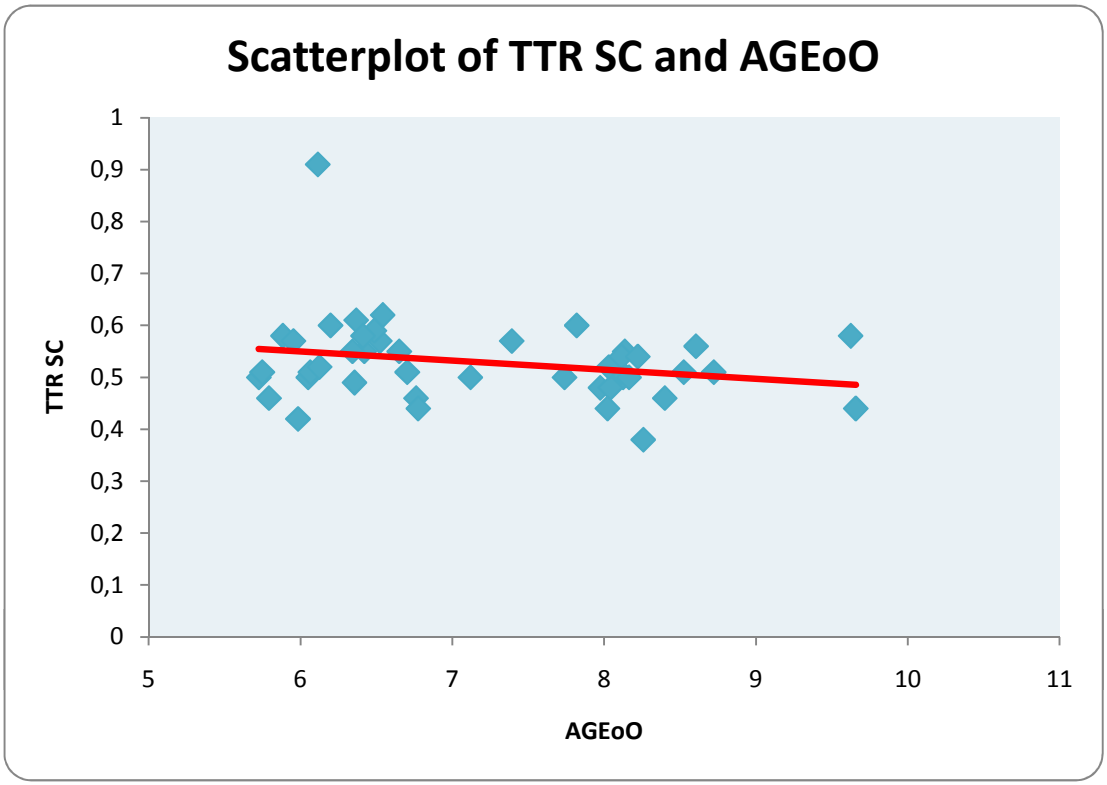
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of TTR SC with AGEoO

Pearson Correlation	-,241
N	44,000
P	,115

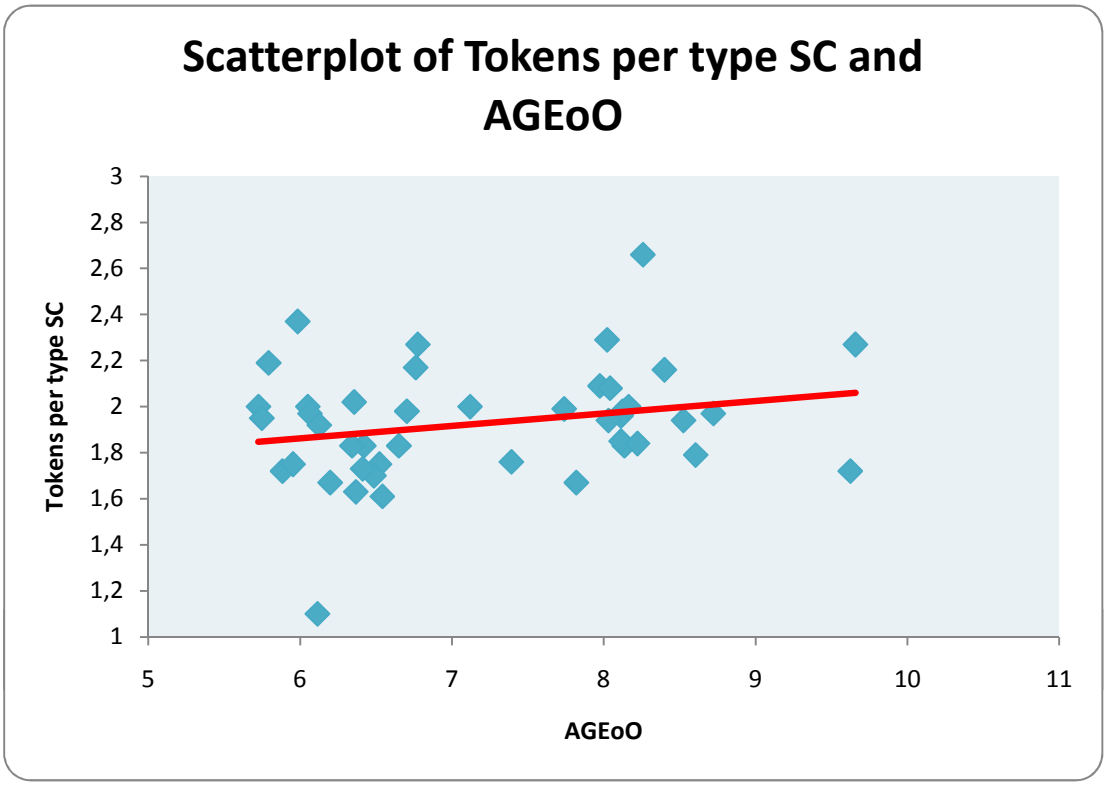
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Tokens per type SC with AGEoO

Pearson Correlation	,236
N	44,000
P	,123

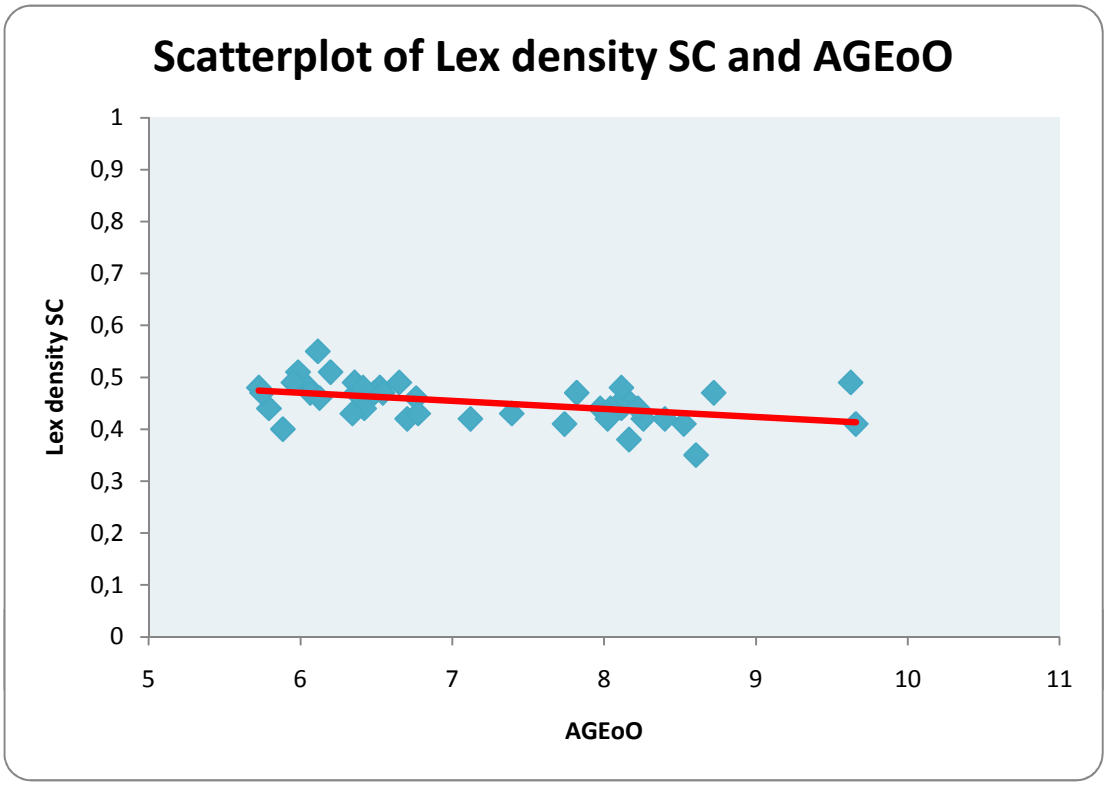
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Lex density SC with AGEoO

Pearson Correlation	-,455
N	44,000
P	,002

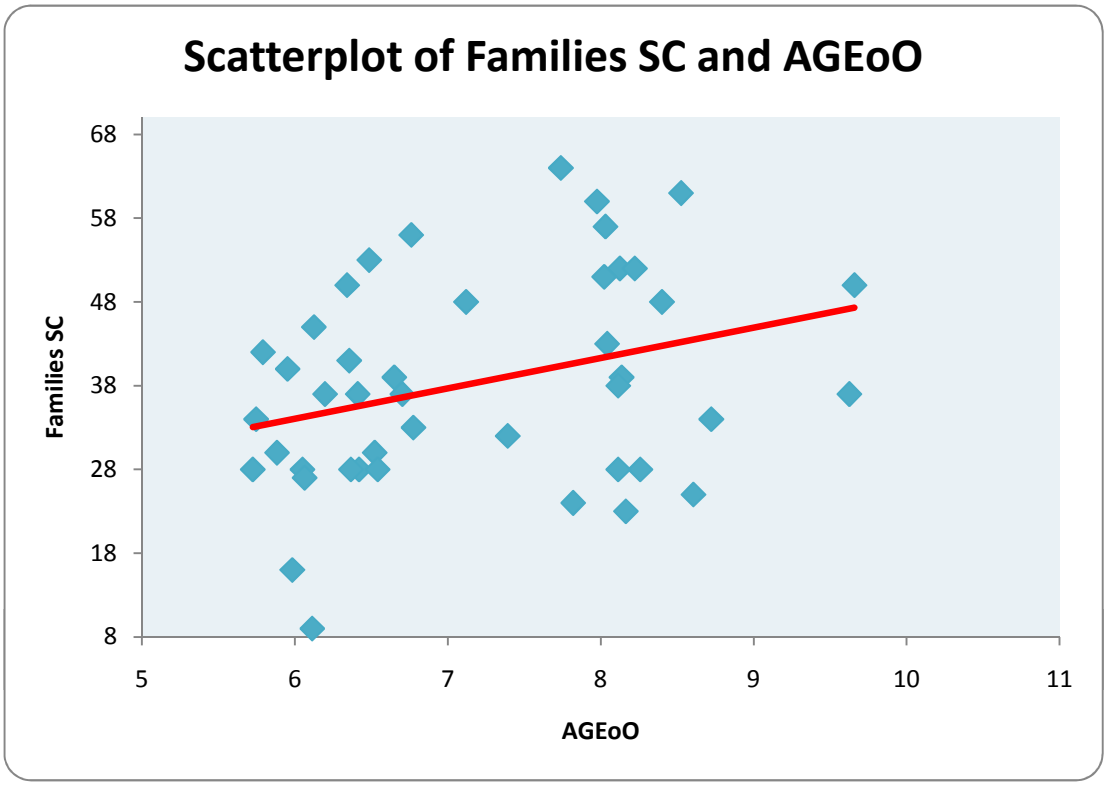
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Families SC with AGEoO

Pearson Correlation	,319
N	44,000
P	,035

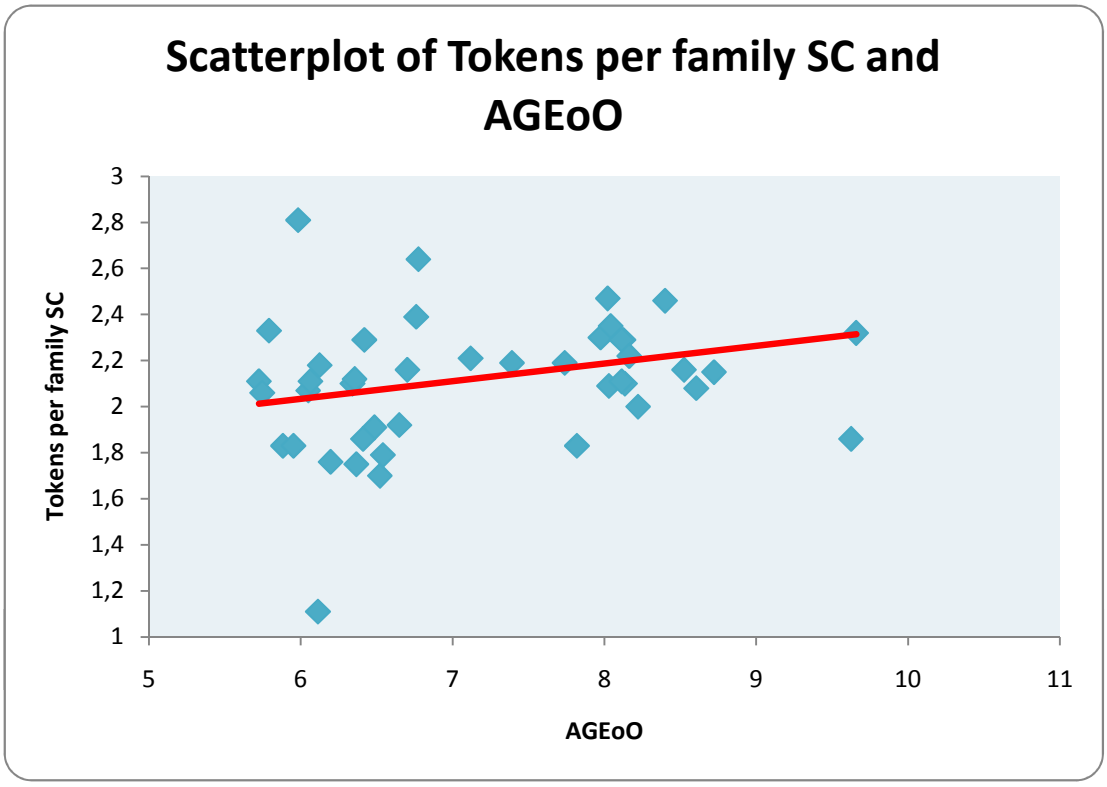
The observed correlation is statistically significant



EZAnalyze Results Report - Correlation of Tokens per family SC with AGEoO

Pearson Correlation	,267
N	44,000
P	,080

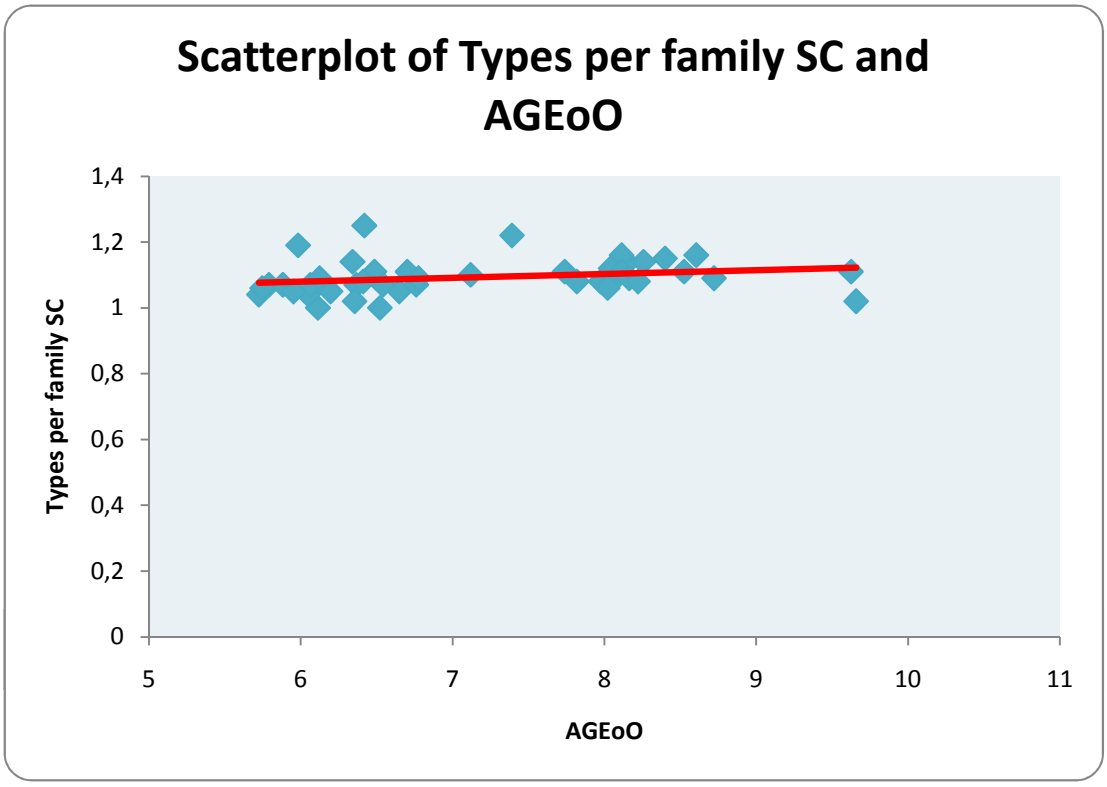
The observed correlation is not statistically significant



EZAnalyze Results Report - Correlation of Types per family SC with AGEoO

Pearson Correlation	,245
N	44,000
P	,109

The observed correlation is not statistically significant

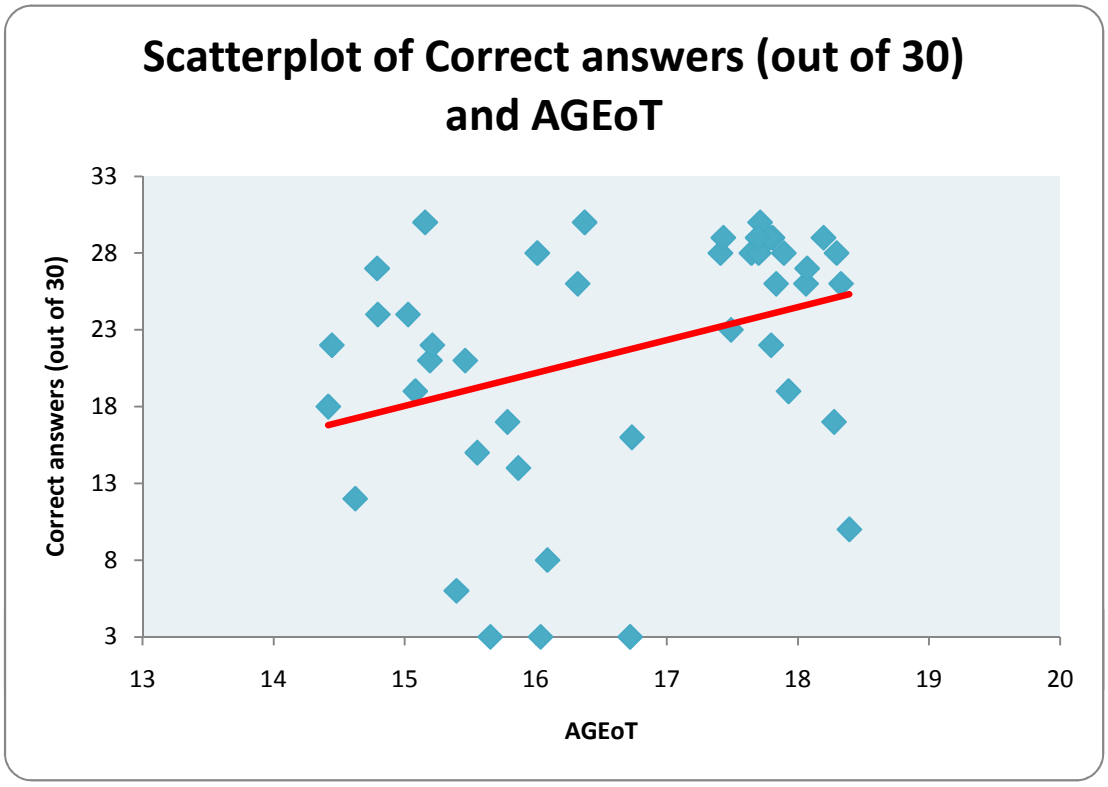


CORRELATION: AGE OF TESTING

EZAnalyze Results Report - Correlation of Correct answers (out of 30) with AGEoT

Pearson Correlation	,346
N	44,000
P	,021

The observed correlation is statistically significant



APPENDIX 5: STUDENTS' PRODUCTION

1

I'm a girl and I'm 18 years old. Physically, I have short dark hair, and I am tall. I am a little thin.

I think that I'm very extroverted, and I like to listen others. Sometimes I have bad character, and I am very nervous. I like so much the sports. I do gymnastic rhythmic since I was 6 years old. I like go away with my friends, and we go to cinema or to the shopping.

I love to go to the party, and we go at the disco. I haven't got any brothers and sisters, and I love so much my parents. They understand me. In my house, we don't have any pet, but I would like.

2

I am a tall, thin girl. I'm brown-haired boy everybody says I'm ginger. I usually wear glasses when I'm at the high school but I wear contact lenses when I'm out with my friends.

I like bright colours, shorts and boots and I hate baggy clothes because they never suit me.

I also enjoy playing the guitar, my acoustic guitar and I love singing anywhere, anytime. I like listening to music too: rock, hard rock, metal, pop, grunge… and I like adventurous and mysterious books and romance comedy and horror movies.

I am realistic and a bit ambitious because all the things I want to become in my life. I am very sociable with any people.

I just want to be studying English translation in a few months time and that's why I'm here. This is all I ask for.

3

I'm Helena, and I'm 17 years old I live in the centre of Granada which is a good place because is quiet. I study in the high school Fray Luis de Granada but today finish my studies here because I have to start study for selectivity. I'm very nervous. I think about what career I will want to do, but I'm not sure. Actually, I would like to think only about the next summer. However, if I want to achieve, I will have to study harder. My parents are always to help me, they are called Mª Carmen and Jose Carlos. It is said that I look like at my mother and I think so too.

My father is more different than me, but I love them so much.

On balance, I like my life although it is sometimes hard but it is normal. I could like to do something that I will like in the future. Despite this, I will try to enjoy nowadays.

7

Well, my name is Alberto, as you can see. I'm a tall guy, blond and thin, who lives in Granada, Spain. My birthday has just passed, it was two weeks ago, so I'm eighteen years old.

Related to my family, I live with my mother, and I also have an older brother, called Armando. We have a beautiful cat, too, whose name is Misi.

Talking about hobbies, I love sport. I couldn't live without it; it's the center of my leisure activities. I practice athletism, known also as "Track and Field", and I am a short distance-runner, I'm a sprinter.

8

My name is Carolina. I'm seventeen years old. I live in Cúllar Vega which is a village of Granada. I like with my parents and my two sisters. My parents work in a shop and my sisters are studying.

I study in a high school. This year I finish and the next year I will go to the university I don't know. I will study yet but I like teachers. I'm a girl tall and thin. I have got long and blond hair. I like reading a book and write histories. I like going shopping and going to the cinema too. And stay with my friend.

9

Hello, I'm Jesus I live in Granada. In a village, my village is called Cúllar Vega. Cullar Vega is near to the town (Granada). My school is in Granada. I come to school by car or by bus. I study in Granada because on my village there are no high schools.

I am blond; I have a straight nose, big ears and blue eyes. I think that I am intelligent because I pass all exams. I am generous and kind and I am very extroversive.

I want to be oboist, a famous oboist. I love classic music, and I should play in famous orchestras. I usually play football with my friends. I like football.

10

At first, I have to say that I'm not use to speak about myself.

I'm going to start describing me. I'm tall, strong and thin or at least (that is what people said to me). I have got dark, straight and short hair and a face square-shaped. I look like athletic because I'm very sportive I love playing rugby (I have been in the university Rugby club for two years) and swimming.

But I love books and computers too. I have got three or four each month.

I haven't got hundreds of friends, because actually, I am shy, buy my friends are time ?, and when I am with them I'm a funny and friendly person.

I could continue speaking about me, but I'm afraid, that I would pass the limit of the number o words, so, this is how I am.

11

My name is Carmen, I'm seventeen and I love acting in theatres, I do it since I was six years old and for me is one of the best things in the life, when I act, I can't explain how I feel but is a sensation very nice. I'm in a theatre academy the Granada's theatre school. I go to class the Fridays from four o'clock to half past five, but I would like that it was more time. The eighteenth of July we will act in Caja Granada's theatre, we will play Coma Ético that is a play which has been made by us. I am little nervous but I'm sure that we will do it very good.

12

I am Jesús Gómez Arias. I am eighteen years old and I was born in Úbeda. My birthday is the 24th April.

I am blond, tall and I am not fat. I have blue eyes but at night it is green.

I am very nice and I always a laugh in my fat. I like a lot of the sports and especially the football. I play football too well and I am playing in Úbeda Viva. The last weekend we won Real Jaén.

To sum up I have a sister, Inmaculada, and my father and my mother. I love them.

13

My name is Luna. I'm seventeen years old and I live in Úbeda. I have blue eyes and brunet and long hair. I'm short. I am shy quiet good person, friendly and people say that I am intelligent. I like listening to music and going out with my friend I also love travelling I would like to be doctor so I'm studying hard for get go to the university.

I live with my parents and my sister, who is fifteen and we are very close. I too have a brother but he lives in Toledo and he is twenty-three years old.

14

My name is Miriam Torrecillas Sánchez. I am 17 years old and I am student.

I live in Úbeda in a house with my mother and my father and I have got two brother, but they don't live with us because their works. Although they usually come to visit us. I like all kind of music and love meeting with my friends. I have people who lay and ?.

My boyfriend is Alvaro and we have know each other since for years. I am student in San Juan de la Cruz, but next year I want go to Jaén to study a career.

15

My name is Carmen. I am seventeen years old. I think that I am cheerful, happy, nice and friendly. I love making laugh to everybody. I have got short hair and fair. I'm tall. I am studying in the high school "San Juan de la Cruz", the bachiller sciences and technology. My favorite subjects are Biology and Math. I would like to study Biotechnology in Seville. My hobbies are painting - I love painting and I got to classes since I was eight-, listening to music, meeting with my friends…

16

I'm Fco Manuel. I am 17 years old. I live in Úbeda with my parents and my dog in a flat. I haven't got brothers or sisters. I study in "San Juan de la Cruz" high school and next year I spend to go to the university. I would like to study medicine in Granada.

Some of my hobbies are play football, watch TV… but I like most is go out with my friends! They are the best!

Every weekend I watch my football team, Villareal C.F. on TV and sometimes I go to the stadium with my father!

17

My name is María, but my friends call me Mari. I am eighteen years old. My date of birthday is 21 of April, 1993. My favourite colour is yellow because I think it is cheerful. I have got no dogs or cats, my mother detests the animals. My best friend is Sonia, she think that I am crazy. I like play games with Sonia, go to the cinema, go to the beach… and I don't like the people that isn't generous. Personally, I prefer sleep that go to the school, but I study for became police-woman.

18

My name's Jesús and I am 17 years old. I live with my mother, my father and my brother. I've got brown eyes and dark hair. I have got twin brother and we usually go out together. I like playing tennis and I go to play it when it's possible.

I am outgoing but also a bit shy in new situations. I've got a lot of friends and they are always able to help me school, I am doing well, and I want to be a doctor. I am good at English, biology and I get good marks.

19

My name is Arancha, I am seventeen years old. I live with my parents and brother, he is twelve years old. My house is very big and nice. I have got green eyes, blond and long hair. I have got a piercing in my nose. I usually wear t-shirt and jeans. I'd like listening music and reading books of fantasy and romantic. I haven't got many free time. When I have it I'd like walking for the park. I haven't got any pets, gut in the future I will have a little dog. I will have on enormous house with a swimming pool in the garden.

20

I'm eighteen years old and I'm studying in Úbeda at "San Juan de la Cruz". I have brown eyes and brown and large hair. I have a piercing too, but I haven't got a tattoo although I like it. I love listen to music in my bedroom and play with my brother with computer games, in my family is very amazing to see films together because we laugh a lot. I have a dog her name is Nika, my dog is brown with a small nose and love to play with ball.

21

I and my life.

Hello, my name is Jose Manuel Lima and I'm 18 years old. I'm tall and I have got brown hair and brown eyes. My father is a Guardia Civil and my mother is a nurse. My hobbies are watch a good film, paint play with my videogames etc.

I'm good, honest and funny and I always wear my tennis jeans and t-shirt. My best friends are Jaime, Ramón, Guille, and Manuel. They are the best people of Úbeda. Right now I'm studying in the high school "San Juan de la Cruz", and my teacher is Ramón, who are in "Saber y Ganar" the last week. Finally I think that my life will be very interesting the next year.

22

My name is Nuria and I have lived in Úbeda since I was born. Nowadays I am seventeen years old and the next month, I'll be eighteen. I have thought many times what I would do when I

was eighteen. I believed it would be fantastic because I could do everything I want, but now, I know it isn't like that, because I'll have more responsibilities. Moreover, there is a difficult exam I'll have to be accepted at university and after that, I have to study a lot. It's true that I would be older and I'll be able to do more things but while I am living at home, I'll do everything my parents say.

23

I'm going to talk about myself. My name is Seba I was born in Ubeda in 1993, the 19th of December. I'm studying the second course of the high school and I hope next year I will study at university. About my town I think it's a very nice place but it has got some negative things, for example streets should be more clean.

My family is quite big. I've got a brother and two sister.

I met my friend when I was a child at primary school. I've got some friend since a lot time so they are very good friends.

All in all I think I'm a normal person but I like my life, my school, my family and my friends so I'm happy.

24

I like a lot the animals but my parents don't want to have animals at home. Two years ago, I bought a tropical little bird and I get it at home. Later, my mother liked more the animal and we called it, Pipi. Pipi was very funny, it was like a dog: you called it and Pipi came to you.

Few days ago, I return to my house and my sister told me that the bird had left the jail and escaped.

Now, I'm trying to have other bird.

25

I'm Laura. I'm sixteen years old. I'm a tall and thin girl. I have blue eyes and my hair is brown. When I'm boring, I like listening music at the same time that I'm surfing on internet.

I haven't got any pet, but when I was little I had rabbits, cats and two dogs, but they died.

In the school, I have a lot of friends, but my best friend is only one and I pass a lot of time with her.

I have a little brother, his name is Sergio and he's two years old. He is very nervous and a bit bad, but I love him.

26

My name is Miriam. I have fifteen years old. I am short, eyes blue and hair blonde. My hobbies are dance and play tennis. Basketball.

I live with my parents and my brother. My parents have forty three years old, my father works in mercagranada and my mother works in my house.

My brother has nineteen years old. He's studying 2nd bachiller and he will like to study police. I have two birds and two animals. The two animals are black and white and the two animals are funny.

27

Hi, my name is Rocío, Rocío Jiménez Fernández. I am fifteen years old but I will be sixteen years old this year, on 30th November. I'm from Boston, Boston is a city of Massachusetts (EEUU). I was born in this city because my parents lived there for three years, my mother is a university teacher and an investigator and my father have a food factory and he is too a investigator, they studied in Granada university. When I was three months, my mother noticed that I have a problem, I couldn't hear, so, my parents gave me for go to the hospital and the doctor said that I am deaf.

Later, ten months ago, we went to Spain for do an operation, the operation was for put me an advice for hear, and I am here, writing that with two advices, one in the left ear and other in the right hear.

I am very happy with my advices, because I can hear well.

28

Hello! My name is Abel and I'm sixteen years old. I live in Granada, a village near of Granada, in Cúllar Vega. I have two brothers the bigger than me. I play football in a team, but my favorite

sport is the snowboard. The snowboard is … GENIAL , I love it. Also I play other sports like basketball, tennis, paddle…

I study in high school in Granada. I come all day to the high school on bus. I like travelling on the world. I have gone to Holland, Germany, France, Italy and other countries that right now I don't remember.

29

Hello! I am Laura Mª Fajardo León and I am fifteen years old. I am a girl, I am not tall… I have got black hair and long, my eyes are brown.

I dance every day, but the Saturdays and the Sundays, no. I dance classic, contemporary,… and I study music. I like dance.

Too, I like drawn, but I don't draw every day.

I have got two birds, the first is blue, and orange and green, the second is green, red and blue.

My family is very sympathetic. I have got two sister, they have got twenty years old and eighteen years old, and they are studying.

30

My name is Mª José, I'm sixteen years old and I study in IES Fray Luis de Granada.

I'm very tall, I'm thin. I have got brown hair, brown eyes and my hair is very curly.

I'm optimistic, cheerful and shy.

I love drawing, usually I draw strange creatures, then I pint this pictures. It's very funny.

I hate people who are … and proud, they are very stupid.

I love pets. I have got a brown, small dog called Rocky. Rocky is very cheerful and funny. I have got two cats too, called Laika and Simba. Laika and Simba are ?

About my family… I only can say that my mother is Silvia, she is very generous, friendly and cheerful. My father, José, is very cheerful and outgoing.

I only have got a sister, her name is Silvia, she is very cheerful.

My sister Silvia has got a boyfriend, her boyfriend is Javi, he is very outgoing. Javi lives in Cádiz, but he studies in Granada's University. Silvia lives in Granada, but she studies in Huelva's university.

31

Hi, my name is Pablo, I'm tall and young, I have got the hair medium-large and blond.

My favourite hobby is sing, I sing and listen Rap. I have got a one pet, is one dog, it name is Nilo. It is big, tall and strong, Nilo is very fast and furious.

My family is very big, in total I have got ten cousins. I have got other hobby: the tennis, I play tennis very well in the club of Peligros, and my third hobby is see my girlfriend and my friends in the weekend, and play football every Sunday in the morning with my friends.

32

Hello my name is Sara. I'm sixteen years old. I'm going to speak a little bit about myself. I was born in Granada but my parents are from Egypt. So I can speak Egyptian Arabic too, and I'm very proud of it. I've got three brothers and two sisters. My oldest sister is married and she has a daughter, and my oldest brother lives with me and my mother. My other three brothers live in Egypt with my grandmother and my aunt. I like so much to speak in English and my dream is to find a great job in the United States and to live there. I love go to the cinema or to watch movies at home, actually it's my favourite hobby, but I like to watch them in their original language English so I can learn much more about it. I also like to travel around the world with my family and my friends, I travelled to Switzerland, to Egypt, to Morocco, to Paris and to Bruxelles, but I want to visit much more. I think what I like more is to go out with my friends and to pass excellent holidays with my family in Egypt. Well, really I don't know what more can I say… oh yes! I love to sing! In my free time I play my favourite songs on my computer with their lyrics and sing them all the time. Where I really want to go is to Los Angeles, but unfortunately it's very expensive but I promised to myself that I'm going to save money and go there.

33

My name is Alejandro, I'm sixteen years old and the next month I have seventeen years old. I have got one sister, she is twelve years old. I'm a boy with hair black. I go to the gym two days at week. I like go out with my friend all weekends and we gone to party. I live in Huetor Vega (Granada) My hobbies are play to football, swim in summer to the beach and help my mother with the house. All summers I go to Malaga

34

I was born in Granada. When I had four years, I ? to Cúllar Vega. Cúllar Vega is a small village of Granada. I live in a small house, but two years ago, my Family ? to big house. When I had five years, I began play football in Cúllar Vega.

In this team I was very happy because we win very matches and I was happy with my football friends. But I thought that I was could play in a better team of Granada. This team was Granada 74

Granada 74 was a team with very footballers because is a big club.

In this club I been four years but this year I began to Granada C.F.

35

My name is Antonio, I'm sixteen years old. I living in Granada in Colomero Street. I'm tall and handsome. I want to study journalism in a university of Madrid.

My hobbies are the English and the music. I like all the type of music, but my prefer type of music is flamenco.

I had a fish when I was eleven years old. But my fish was died.

My eyes is brown, my hair is brown too. I want speak English very well, because English is my favourite language.

I living with my parents and my little sister.

36

Hello, my name is Juan, I'm fifteen years old. I am Spanish and I live in Granada with my family. I have three sisters, Aurora that she is twenty nine years old, Irene that she is twenty five years old and Fátima that she is nineteen years old. My father name is Juan and my mother Fátima.

I'm studying in high school Fray Luis de Granada and it is very good.

37

Hello, my name is Ana, I have sixteen years old.

I have blonder hair and brown eyes, I am not very tall. I like play tennis and football, but I love parties, I go to parties with my best friends, her name is Sara. Now I am going to speak about my family.

My father is a good man, he loves his job and sport. He likes play football on Sunday with his friends. My mum is more boring, she is shy, but she has much friends, she love the holidays, she likes go to the beach. I have only one sister, her name is Lucia, she is very different than me. She listen music all the time, she likes heavy and rock and roll, I think that this type of music isn't normal because I don't understand. I have too a pet, it is a dog, his name is Baloo, he has brown hair and he is small, he has only two years old.

38

I am David Vargas and I'm fourteen years old. I am tall and I've brown eyes and brown hair. I like watch and play football (my favourite team is Real Madrid), basketball and tennis. I like play computer games too.

I've had a lot of pets, like turtles, fish, dogs or rabbits. Exactly, I've had a white rabbit until a few months ago. I also had a dog, but only for three days.

I don't have any brothers or sisters, but I've much cousins and uncles. The name of my mother is Puri, she's forty eight years old and she likes fruit and vegetables. My dad's name is Jose Manuel, he also likes watch football and basketball.

One of my hobbies are the computers: I like linux, freeware, and open source software. I hate Microsoft, and I belong to Anonymous, I've taken part in it since 2010.

39

I was born in Granada, the 7 of May of 1996. Today I am 15 years old. I am very tall, handsome, intelligent and strong. In the institute my marks are good. I love basketball, play videogames and walk in the environment. I haven't got pets, but before I had a hamster called Tutankamon and my grandmother's dog called Pato. I love this pets.

My birthday was many years ago, and it was fantastic, I went with my family to a restaurant, after I went with my cousins to the bowling,, and at the night with almost all my family, we had the dinner and at the end I saw my presents.

The day before yesterday I went with my parents to the "Palacio Municipal de Deportes" to see the Harlem Globetrotters. They are fantastic player and the show was very fantastic with impossible slam dunk and alley up and it was very funny, too.

40

My name is Irene. I'm 15 years old. I was born in Madrid but I live in Granada since eleven years. My favourite sport is tennis, I play tennis at the weekend. I have got one pet, a rabbit but I don't like. I have got one sister, she is ten years old and her name is Nuria. The last summer I stayed in New York, Boston and Canada, I stayed there three weeks. I went with my parents and my sister and friends of my parents. This summer, I'm going to go to England four weeks, with my cousin and if I ? all my exams I going to go to Galicia with my parents and my friends.

41

My name is Alba, I'm 14 years old. I'm from Granada (Spain) I have got one brother, he's name is Mario, he is got 17 years old, he is from Granada too. My mother is Yolanda and, my father is Mario. I have got brown and curly hair, and brown eyes. I'm friendly, I'm generous. My favourite hobby is listen to music, and play the guitar. The music I listen is all type of music, but I like very much the roc music, and pub music. Also have got a dog, she's name is Gala. I love the animals, so , when I was 9 years old, I want to be vet, but know, I think it better.

42

Hey! My name is Paula and I'm fifteen years old.

I study in "IES Fray Luis de Granada" in the class 3rd A. If you go there and you want to meet me, you have to know that I'm the shortest of the class, my hair is brown and curly, my eyes are brown too and I usually wear a lot of jewellery; with this information you'll recognize me easily, I think.

I have an eldest brother called David, he's twenty three years old. My parents are Javier and Gloria, they're both English teachers. I have a dog too, her name is Branie and it's a Shi-tzu. She's lovely.

My hobbies are go swimming, reading and listening to music. I'm almost always listening to music.

I love travelling too.

I've been in Italy, France, Scotland, Germany and England (for 5 or 6 times) but I want to go to the USA! Maybe next time. So this is all about me.

43

My name is Nacho, I have got 14 years old. I have two brothers, Javi is in 3rd ESO and Rafael is in 1st Bach.

The hobbies: my best hobby is the basketball, I like play but I don't like see. The other hobby is ski, I go to the mountain in Granada (The Sierra Nevada) with my best cousin. I don't like the TV because in my house don't have TV, but I like the Play Station and Wii. In my house have a TV for play to Wii o Play Station 3.

I don't have dog because my dog Luna dead the past year.

In the future I'm going to the university because I want be doctor.

My mother is doctor and my father is architect

44

Hi my name is María. My favorite hobby is dancing, but also I love listen to music and talking with my friends. I'm 14 years old but second of October is my birthday.

I sometimes read a books. My favorite book is "Tres metros sobre el cielo" the history is very beautiful.

I like also go to cinema with my family or friends, stay in the park in the summer with my friends, listening the birds or talking is fantastic.

Sometimes go shopping with my friends or my mum. I love go shopping! I hate study and do my homework. This is my hobby.

45

My name is Carmen. I'm fifteen years old. I have one sister, her name is Sara. My father's name is Jose Miguel and my mother's name is Carmen as me.

I have one dog, it name is Badu. I like it! But it isn't live with me, it lie with my grandparents.

I live in Maracena, a little village near Granada. I like painting, travelling, chatting on Internet, going out with my friends, playing volley…

My favourite subjects are Biology, French and Science.

I have travelled Paris, London, Scotland, Germany, Italy. I'm tall, and thin. My hair is blond and straight. My eyes are green and small.

This summer I'm going to the beach. The name of the beach is Almuñécar.

46

Hello, my name is Alicia. I'm 14 years old. I've born in Granada, I have got a brother, he is 19 years old, but now he isn't in Granada, he is studying in Ceuta, also I have got a sister, she is 16 years old but in November she is going to have 17.

With her and my parents, we live in Granada, on a flat. I live in the 11th floor. I don't have any pet.

My uncle live in the same flat, so when I'm boring I go to her house and I play with my cousins. They are younger than me. I can say that they're babies. So sometimes I work like a nanny.

At summer we go to the beach, we have a cottage there.

But this summer I'll go to England for three weeks. I'm going to live with a host family, but it isn't new for me, because last summer I travelled there for two weeks. But this time I'm going to go alone, without friends.

47

Hello my name is Natalia, I have 15 years old, I have brown and curly hair, and I'm very tall. My hobby is sing and all about music. My favourite type of music is rock and pop.

I have four dogs. My favourite animal is tiger. My favourite team of music is Mc Fly and I love the music of Bon Jovi.

My best friends is Alba, she is very kind, very clever and very friendly.

I'd like will working something about English or with animal or music. I don't travel so much, but I will travel in the future to America and Asia.