Context-Dependent and Context-Independent Task Development in Examining Verb Complementation Patterns in English

Serap ATASEVER BELLİ[1] & Sibel SÖĞÜT[2]

Ph.D., Erzurum Technical University, Erzurum, TURKEY
ataseverbelli@gmail.com
ORCID: 0000-0003-3605-4943
[2]Ph.D., Sinop University, Sinop, TURKEY
siblsogut@gmail.com
ORCID: 0000-0002-3395-7445

Abstract: The aims of the present study are: a) to examine the methodological issues on the development of context-dependent and context-independent task development process in investigating the cognitive verb complementation patterns in English; b) to design and administer task items to analyze the learners' recognition and production levels in the use of verb complementation patterns and their related verb senses. The paper adopted an exploratory sequential mixed methods design and followed a three-phase procedure covering the creation of task items and the development of instruments, administration of context-dependent and context-independent tasks to 200 Turkish EFL learners and conducting the reliability and validity measurements based on the interpretation of the results. Based on the validity and reliability analyses results, the tasks designed for factive and non-factive cognitive verb complementation analysis were found to be acceptable. Thus, this study offers valid and reliable context-dependent and context-independent tasks to examine the learners' competence at the recognition level and their performance at the production level in the use of cognitive verb complementation patterns and their related verb senses in the English language. These tasks provide a novel application by shedding light on the developmental relationship between the learners' state of knowledge regarding verb complementation and their vocabulary knowledge and address both the learners' comprehension and production of cognitive verbs. In this sense, a number of implications are suggested for pedagogical purposes and further inquiries.

Anahtar Sözcükler: Test geliştirme, eylem yanulamlaması, bağlam içi ve bağlam dışı testler

Türkçe Başlık: İngilizcede Eylem Yanulamlama Görünümlerinin İncelenmesi için Bağlam İçi ve Bağlam Dışı Testlerin Geliştirilmesi

Özet: Bu çalışmanın amaçları; a) İngilizcede biliş eylemlerinin yanulamlama görünümlerinin incelenmesinde kullanılacak bağlam içi ve bağlam dışı test geliştirme süreçlerine yönelik yöntemsel aşamaların açımlanması; b) biliş eylemlerinin yanulamlama görünümlerinin yapısal ve anlamsal özelliklerinin öğrenciler tarafından edim ve kullanım düzlemlerinde bilgilerini ölçecek testlerin geliştirilip uygulanmasıdır. Çalışmada keşfedici sıralı karma desen kullanılarak bağlam içi ve bağlam dışı test maddelerinin oluşturulması, testlerin anadili Türkçe olan İngilizceyi yabancı dil olarak öğrenen 200 öğrenciye uygulanması, elde edilen bulgular ışığında testlerin geçerlik ve güvenirlik hesaplamalarının yapılması olmak üzere 3 basamağı kapsayan bir süreç izlenmiştir. Çalışmanın sonucunda, olgusal ve olgudışı biliş eylemlerinin yanulamlama görünümlerini incelemek için hazırlanan test maddelerinin geçerlik ve güvenirlik değerlerinin kabul edilebilir ve geçerli olduğu bulgulanmıştır. Bu bağlamda bu çalışma; öğrencilerin, İngilizcede biliş eylemlerinin yanulamlama görünümlerini eylemlerin sözdizimsel ve anlambilimsel özelliklerini göz önünde bulundurarak tanıma ve üretim düzlemlerinde incelemek amacıyla geçerli ve güvenilir bağlam bağımlı ve bağlam bağımsız testlerin geliştirilmesi açısından alanyazına özgün bir katkı sağlaması bakımından önemlidir. Bu çalışma, biliş eylemlerine özgü yanulamlama yapılarının farklı sözcük düzeylerindeki öğrenciler tarafından edinim ve üretim durumlarını ele alarak aradil çalışmalarına yönelik alanyazında bir eksikliği gidermektedir.

1. Introduction

In the process of language production, language learners make a variety of conceptual and linguistic decisions (Levelt & Maassen, 1979). In the generation of a language item, the learners decide on content selection, which is about "what should be expressed, given the speaker's intentions and given mutual knowledge in the speaker-listener context" (Levelt & Maassen, 1979, p. 221). They also decide on linguistics elements, which are classified as functional and positional decisions (Garrett, 1975). Functional decisions involve the lexical choices, and the words are "selected for the expression of thoughts by process of lexical search", whereas positional decisions are about "the realization of appropriate syntactic frames and grammatical formatives" (Levelt & Maassen, 1979, p. 221). In order to have a comprehensive overview of the learners' recognition and production of a language item, both functional and positional decisions should be taken into consideration.

Receptive knowledge of a language refers to recognizing the word in written and spoken forms, knowing general and specific meanings in context, knowing word parts, and the positive and negative connotations (Gass & Selinker, 2008, pp. 451-452). Productive knowledge refers to knowing the accurate pronunciation and spelling, precise word meaning in context, context-specific features, and precise context of use (Gass & Selinker, 2008, p. 452). Considering the receptive and productive knowledge of language, the knowledge especially tied to verbs in the language in terms of syntactic and semantic features is particularly important and essential (Wilson & Garnsey, 2009, p. 369) in comprehending the meaning conveyed through the verb and the patterns used in sentences because sentence grammar is considerably verb grammar (Hubbard, 1994, p. 69). In the literature, the issue of the syntactic and semantic relationships of the required constituents within the environment the verb occurs is dealt with under the term 'verb complementation'. As for semantic features, a majority of English verbs realizes a number of senses in various contexts, and in terms of syntactic features, they appear with various complementation patterns in different syntactic structures. Complementation patterns, i.e., complements, are defined as the elements that the verb takes to form a grammatical sentence, and they consist of phrases and clauses (Herbst, Heath, Roe & Götz, 2004, p. xxiii). Phrases are comprised of noun phrases (NP), adjective phrases (ADJP), prepositional phrases (PrepN), whereas clauses include ing-clauses (V-ing), to-infinitive (to-INF), that-clauses (that-CL), and wh-clauses (wh-CL) (ibid.). In this respect, the knowledge of verb complementation features is helpful for language users as it facilitates both language comprehension and production (Hare, McRae & Elman, 2003, p. 281). Hence, in this study, productive and receptive knowledge of language users regarding verb senses associated with various complementation patterns has a prominent influence in effective use of the target language, and this issue has been examined and highlighted. Based on the aforementioned points, the type of tasks used in examining the comprehension and production of verb complementation has a great impact on the results revealed in analyzing learner language. The production task involved producing a string of words for which the meaning is already known, while in the comprehension task, students assign meaning to a string of sounds or words (Anderson, 1976). In this respect, two different grammars, production and comprehension grammar (Naiman, 1974) have been suggested.

In the relevant literature, there is a controversy about whether production and recognition tasks are reliable measures of learners' performance or whether these tasks provide a comprehensive picture by complementing each other. Studies investigating the relationship between learners' performance on grammaticality judgment tasks and production data are quite inconclusive in their findings (Leow, 1996). Some studies suggest that the

grammaticality judgment task may be relatively more reliable for performances on written production tasks when compared to oral production tasks (Leow, 1996, p. 129), and "they are indeed reflective of patterns of second-language use" (Gass, 1994, p. 320). On the other hand, as there is a considerable inconsistency found in his subjects' grammaticality judgments, "learners' judgments can be inconsistent, and, therefore, unreliable, when they are unsure" (Ellis, 1991, p. 181). As Milton (2009) states, language knowledge is not something like "a directly accessible quality and we rely on learners to display their knowledge in some way so it can be measured (p. 6)". In other words, learners' explicit knowledge cannot be measured by means of a single test but will require multiple instruments to demonstrate concurrent validity (Ellis, 2004, p. 267). Through only recognition tasks such as grammaticality judgment tasks, the larger part of the difficulties to be experienced most probably by language learners in their production is left unexamined since such 'elicitation tasks impose a high degree of control over the participants' output production' (Chan, 2004, p. 68). That is to say, it is difficult to highlight problems encountered by language users, and thus productive tasks are also needed for further evidence for the maximum performance in the target language and for better understanding of in/correct uses in language production. Thus, this study takes a relatively different approach by addressing verb complementation in terms of the use of tasks measuring learners' performances through sentence production and sentence completion tasks and examining their competence by administering a grammaticality judgment task and a fill-in-the-blanks task and using item analyses.

Verb complementation patterns and their related verb senses have been investigated through the use of either context-dependent or context-independent tasks in the literature. These data collection tools are grammaticality judgment tasks, sentence completion and production tasks, translation tasks, substitution tasks, multiple-choice tests, fill-in-the-blanks tasks, and picture description tasks. In these tasks, speakers of a language are presented with a set of linguistic stimuli to which they must react (Tremblay, 2005). Scholars argue that there are significant differences between the verb subcategorization frequencies revealed from experimental methods and corpus methods (Roland & Jurafsky, 1998). One way of eliminating the possible effect(s) of the type of task on learner production is to utilize not a specific type of task but a variety of data collection sources such as Sentence Production and Sentence Completion Tasks (Schwarte, 1982) to highlight the overall picture through both context-dependent and context-independent tasks. Therefore, the learners' reception and production of verb complementation patterns and verb senses by triangulating the data are essential. By combining both context-dependent and context-independent tasks to analyze learners' recognition and production of the verb complementation, this study presents a comprehensive and novel investigation of the task development procedure, which may also be adapted to examine different language items in English. Figure 1 below depicts the tasks developed within the scope of this study.
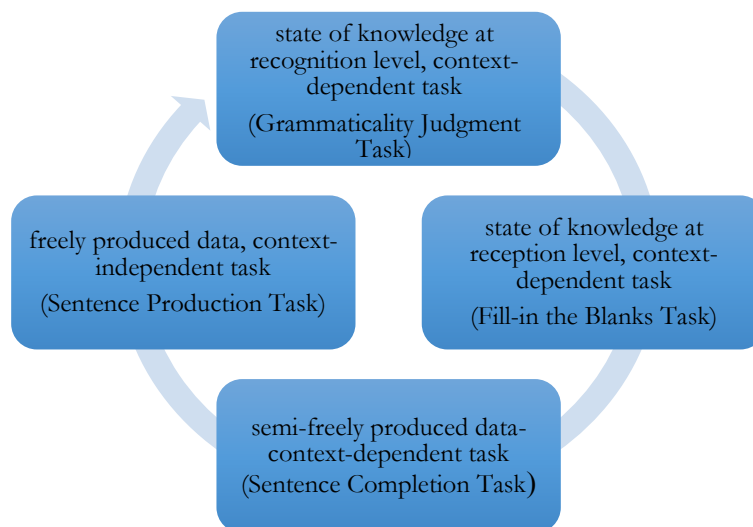
Figure 1. Task types

It is necessary to specify that the context refers to the "accompanying text, the wording that came before and after whatever was under attention" (Halliday, 1999, p. 3) within the domain of this study. The context has three components: the underlying social activity, the persons or "voices" involved in that activity, and the particular functions accorded to the text within it (Halliday, 1999). These components were specified in the context-dependent tasks, whereas they were not provided in the context-independent tasks. The students were expected to identify the processes in which the texts are located and to pay attention to the surrounding written discourse through these tasks. In language test development, the process is not linear or fixed but iterative and dynamic (Fulcher & Davidson, 2007, p. 89). This process requires reaching "the optimal design to release a working, crafted product, suitable for its intended purpose" (Fulcher & Davidson, 2007, p. 89). Hence, this research would be informative by providing a novel contribution to the literature with the development of context-dependent and independent tasks, which are used to examine the learners' competence and performance of verb complementation patterns and their related verb senses.

2. Method
2.1. Research Design

The current study adopts an exploratory sequential mixed methods design. Accordingly, researchers first collect the data, analyze the results, develop an instrument, and then administer it to a sample of a population (Creswell & Creswell, 2018, p. 306). As there are no adequate data collection tools to investigate learners' verb complementation patterns in both receptive and productive levels in the relevant literature, the researchers developed them in the present study. In this sense, they employed a three-phase procedure. The first phase is the exploratory step, which consists of the identification of the salient verb complementation patterns and their related verb senses, analysis of verb occurrences, concordance lines, and wider contexts of cognitive verbs in the Corpus of Contemporary American English (COCA). The second phase includes the development of data collection instruments and conducting the reliability and validity measurements. The third phase

consists of administering the tasks to the learners and gaining insight into their competence and performance in verb complementation use.

2.2. Participants

The participants of the study consist of 200 Turkish EFL learners majoring in the English Language Teaching Program in their first, second, third, and fourth year at a state university in Turkey. Data collection tools consisted of four types of tasks that were developed within the scope of the study. These tools along with the task development procedure and the data analysis procedure were explained in detail in the following sections.

2.3. Task Development Procedure

The tasks were developed at a number of phases, and these phases were explained in detail in the following order.

2.3.1. *The selection of cognitive verbs, their complementation patterns, and senses*

Cognitive verbs, also labeled as verbs of belief (Papafragou, Cassidy & Gleitman, 2007), opinion verbs (Klotz, 2007), private verbs (Quirk, Greenbaum, Leech, & Svartvi, 1985), and mental verbs (Biber, Johansson, Leech & Conrad, 1999), refer to the verbs which communicate information about mental states and actions such as emotion, desire, thought and perception (Owen Van Horne & Lin, 2011, p. 2). They are universal, and they occur in every language (Givón, 1973, p. 891). Regarding syntactic and semantic properties and cognitive demands, they are highly complex lexical items (Cappelli, 2008, p. 538). Syntactically, they occur in multiple environments appearing with various types of complementation patterns such as sentential complementation patterns, noun phrases, prepositional phrases etc. (Naigles, 2000; Nixon, 2005). At the semantic level, they are abstract, and they possess a variety of verb senses (Naigles, 2000; Stojičić, 2008, p. 27). In cognitive terms, these verbs express certainty about the proposition at varying levels (e.g., *know* with a high degree of certainty) (Moore, Bryant & Furrow, 1989; p. 168; Wellman & Estes, 1987, p. 152). Considering these complexities, they pose problems for both native and non-native speakers of English (Nixon, 2005; Owen Van Horne & Lin, 2011). Therefore, contextual information and syntactic information are important as they may provide language users with the discovery of verb senses (Papafragou et al., 2007).

Nine cognitive verbs were chosen in the study consisting of five factive verbs and four non-factive cognitive verbs. These verbs include *know, understand, regret, forget, remember, think, believe, assume,* and *suppose.* They were selected based on the high-frequency criterion. They are high-frequency verbs in English as it has been evidenced in prior studies (Shatz et al., 1983; Kidd et al., 2010) and in frequency lists of corpora such as Longman Spoken and Written English Corpus (Biber et al., 1999) and in COCA covering over one billion words. They are claimed to be problematic in terms of syntactic, semantic, and cognitive reasons. As for syntactic concerns, they occur with various complementation patterns, from clausal complementation patterns, inflectional phrases, verb phrases to noun phrases (Nixon, 2005). In semantic concerns, they express complex, abstract, and unobservable concepts (Papafragou et al., 2007, p. 126). In this sense, they are polysemous and ambiguous as they have a number of verb senses appearing in many syntactic structures (Booth & Hall, 1995; Booth, Hall, Robison & Kim, 1997; Naigles, 2000; Stojičić, 2008; Verdaguer, 2010). They are hard to acquire for second/foreign language learners and even for native speakers' meanings (Barak et al., 2012; Naigles, 2000; Papafragou et al., 2007). Regarding cognitive concerns, cognitive

verbs have a factivity (factive vs. non-factive) dimension and express varying levels of certainty about the proposition (i.e., *know* with a high degree of certainty and *think* with less certainty) (Moore et al., 1989, p. 168; Wellman & Estes, 1987, p. 152), which make them cognitively complex and demanding (Nixon, 2005, p. 20).

Prior to the development of task items, the verb complementation patterns and senses of eleven cognitive verbs (*think, believe, assume, suppose, know, understand, regret, forget* and *remember*) were checked and determined based on the source of Valency Dictionary of English (Herbst et al., 2004) and Verb Net (Schuler, 2005). Valency Dictionary of English is a comprehensive dictionary of English verbs, nouns, and adjectives providing extensive information about verb complementation. It is based on the Bank of English Corpus covering 320 million words and reflecting authentic present-day English (Herbst et al., 2004). A verb entry possesses four components which comprise the complementation patterns given for the verb, example sentences given for the valency patterns, information on the meaning, and idiomatic phrasal verbs (ibid.). As depicted in figure 2, the formal description of the following verb information was considered while selecting the patterns and senses for the tasks within the scope of this research.

+ (that)-CL (>30%)

C I *think* you have the right to use the road and you don't need to pay for it. – I *think* that's a good point. – I don't *think* that that is entirely true. – One would have *thought* that they would have got something which was probably a little but more compatible. - It is *thought* that the balance of power will be held by the Liberal Party.

+wh-CL/wh

A I've never thought why I like something or why I don't like something. -I was just *thinking* how awful it must have been.

C I can't *think* why.

+about Np/V-ing/about wh-CL/wh to-INF (frequent)

A I was just *thinking* about that. – I never really *think* about creativity. – you don't *think* about how much you are actually paying for the things you buy. – Just *think* about where you were and where you are now. – The Churches now have to *think* about how to create a new basis for support.

B I wouldn' even *think* about lying to you.

+of Np/V-ing /of wh-CL/ wh-to- INF (frequent)

A He *thought* of how he'd kissed her.

B I've searched everybody that I can *think* of. Can't *think* of a concrete example. – I can't really *think* of a a way round it. You know. – How can you *think* of winning a national championship when you do that sort of thing?- We had to *think* of how to sort this out.

Figure 2. Verb entry information for the verb think in the Valency Dictionary of English (Herbst et al., 2004, p. 869)

Apart from the Valency Dictionary of English, VerbNet (Schuler, 2005) is another source of online verb lexicon that was checked as reference data in this research. This lexical database was intentionally chosen since it is said to be the largest online verb lexicon currently available for English, and it covers a very comprehensive account of the thematic roles, selectional restrictions, and syntactic and semantic frames. By checking these sources, the basic verb senses and complementation patterns were identified for each cognitive verb. During this identification, the 'frequency information' criterion was taken into consideration. Frequency is a crucial criterion in the selection of verb complementation patterns and the verb senses since words, meanings, and the patterns that are most frequent in the language are the ones that are or need to be taught to and used by foreign language learners (Herbst et al., 2004). Hence, the frequency of the patterns was taken into account in task development.

2.3.2.   Tasks focusing on the learners' competence at the recognition level

Two types of tasks, i.e., Grammaticality Judgment Task (GJT) and Fill-in-the-Blanks Task (FBT) were developed to measure learners' competence of verb complementation patterns and senses at recognition level. During the context-dependent task development procedure, the following steps were followed.

Each cognitive verb was searched in the COCA to gather extracts from wider contexts that represent the use of the selected verb complementation patterns and senses from various text types, i.e., registers consisting of magazines, newspapers, fiction, spoken language academic texts to form the task items. COCA was intentionally used for the formation of task items and checking the different uses of cognitive verbs in a variety of contexts since the main aim is to include the language that reflects the naturally occurring language as it is used by English native speakers rather than giving sentences that the researcher makes up. Among other present-day English language corpora such as BBC or MICASE, COCA was specifically chosen as it is the only large and freely available corpus of English and the most widely-used data, which consists of one billion of words in spoken and written language equally divided among the text types including fiction, newspaper, magazine, blogs, spoken language, TV and movie subtitles, web pages and academic texts (http://corpus.byu.edu/coca). Moreover, it includes samples of the English language dating back to 1990 up to 2020. In the present research, specifically, the keyword, i.e., each cognitive verb was entered into the 'List' search line column in COCA. Texts, in which the cognitive verbs were used, were extracted from different types of registers along with their wider contexts. The contexts were examined in terms of sense and complementation patterns of verbs. Next, the task items were formed for both GJT and FBT. In this procedure, in the GJT, which was intended to tap modalities other than production, a counter-balanced set of grammatical and ungrammatical sentences were formed, and a total of 20 task items were developed for the measurement of the knowledge of non-factive cognitive verb complementation and 20 items were developed for the assessment of the knowledge of factive verb complementation. In this task, participants were asked to read each sentence and to decide on the grammaticality of the sentences by evaluating the appropriateness of the sentences in terms of whether the sentence sounds good, bad, or not sure. As for FBT, in each task item, an excerpt from COCA was formed by taking out the cognitive verb, and a number of verbs, including the cognitive verbs under investigation within the scope of this research were given at the very beginning of the task in a box for the subjects to choose the appropriate verb among others and to fill the gap in the item. Participants were asked to fill in the gap with the expected cognitive verb based on context. In total, ten task items were formed for the exploration of the learners' state of knowledge regarding non-factive

cognitive verb complementation and ten items for the factive cognitive verb complementation knowledge.

2.3.3. Tasks focusing on the learners' performance at the production level

Two types of tasks, i.e., Sentence Completion (SCT) and Sentence Production (SPT) Tasks, were developed to identify language users' preferences of verb complementation patterns and verb senses at production level.

SCT is a written elicited production task designed to test learners' productive knowledge of verb complementation. In SCT, an extract from a wider context retrieved from COCA was formed into a task item, and the part following the cognitive verb in this sentence fragment with contextual information was left blank for the language user to complete the sentence with their own words. The deleted words were placed by a blank, and the learners were constrained to provide responses by taking the contextual clues into consideration. The frequent verb complementation patterns used with the cognitive verbs and their related verb senses were detected first in the Valency Dictionary of English, and then extracts from different texts were chosen using COCA. In total, 20 items were developed for the measurement of learners' receptive and productive knowledge of factive cognitive verbs and 20 for the non-factive verbs. Through this design, the present study enables us to provide some understanding in how contextual processes and constraints operate in language processing. As for SPT, it is an uncontrolled task designed to measure the complementation patterns that language learners used with the cognitive verbs under investigation and the verb senses they know at the production level. As exemplified below, participants were asked to freely create two sentences for each cognitive verb and write down the meaning of the verb in each sentence they formed. In this task, participants were given the freedom to choose whichever verb complementation pattern and verb sense they knew to find out their tendencies. Eighteen sentences and eighteen verb senses were expected to be written for nine verbs (five factive verbs and four non-factive verbs) in total.

2.4. Data Analysis

In this study, each task was evaluated based on 100 points for both factive and non-factive verbs. In these tasks, a point of zero was allocated to the incorrect response(s). Incorrect responses include the choice of irrelevant complementation patterns and the use of verb senses that are not among the meanings of the verb, and the use of nouns rather than verbs (e.g., my assumptions were correct). In addition to this scoring, item analyses, item format, and content analyses were conducted for the tasks developed in the study.

2.4.1. Item format analysis

Item format was analyzed to find out whether the task items were properly written and how well task items were formatted. The guidelines designed by Brown (1996, p. 50-51) were followed, and each item in the guideline was checked. Feedback sessions were carried out through face-to-face and online consultations with two native English speakers and one non-native English language instructor. Based on these sessions, necessary changes were made in terms of clarification and elaboration of the instructions, omission of any biased contexts, avoidance of clues, clarification of unfamiliar terms, organization and order of the test items, preparation of a key for each task, standardizing the blanks, specifying the scoring procedure and giving personal IDs for each learner to make their responses anonymous.

2.4.2. Item content analysis

After developing the task items, the content validity of the tasks was checked by consulting opinions of two English native speakers consisting of one instructor in the School of Foreign Languages at a state university in Turkey and one practicing teacher of English in Japan, and one instructor at English Language Teaching Program. During this procedure, they were asked to analyze each item and to give feedback in terms of the overall organization of tasks, the wording of items, the instructions, contexts, and the use of the verbs. In this process, interaction with one English native speaker from Japan was carried out through e-mails, whereas the one with the other two experts was realized through face-to-face sessions. Based on their feedback, necessary changes and omissions were made on the tasks. Specifically, the following decisions were made:

- The instructions in some tasks were revised and elaborated.
- Some contexts in the FBT and SCT were found to give little indication of possible answers and draws on background knowledge that subjects might not have. Thus, such contexts were simplified to provide contextual clues in the surrounding contexts. In addition, sophisticated words in such complex contexts were omitted.
- Some of the task items had extra information in parenthesis to make context revealing for the participants, whereas some of them did not have information. Since extra information in some items was found to be unnecessary, these parts were excluded, and contexts were standardized.
- For some of the items where conversation between two speakers occurred, as speakers' names or identities were not present in the contexts where they occurred in COCA, speakers were named as "A" and "B" while in many task items, speakers' actual names were extracted from COCA as they were. To standardize this in all items, speakers were provided with names rather than named as A or B.
- Some words were found to be difficult for students to know or remember. Therefore, they were changed with their synonyms which were expected to be easier for them to know/remember. Also, some words/patterns were found to be inaccurate and inappropriate in the context where they occurred. Thus, instead of these words/patterns, accurate versions were written (e.g., instead of "slice of rice", "portion of rice" was written).

After making editions based on the expert feedback, a pilot study of the tasks was conducted to test the difficulty level of sentences, the length and complexity of statements and the familiarity of vocabulary (Henning, 1987, p. 10). The tasks were administered to 46 undergraduate EFL learners (1st, 2nd, 3rd, and 4th year) majoring in the English Language Teaching Program. To minimize the possible effect of language exposure on the learners' responses and not to affect the learners' verb use, each task was administered every other week in the order of SPT-SCT-FBT and GJT, respectively.

2.4.3. Item analysis

Item analyses were carried out through a pilot study based on the answers gathered from language users to decide whether items were developed correctly or not and to increase the reliability and validity of the tasks. For this purpose, item facility and item discrimination indices were calculated.

Item Facility (IF) Analysis: IF, which is also called item difficulty, is defined as the proportion of test-takers who answer an item correctly (Fulcher & Davidson, 2007). It is generally assumed that items should not be too easy or too difficult for the population for whom the test has been designed. In order to calculate the IF index, the number of task-takers who correctly answered a particular item is added, and then the sum is divided by the total number of task-takers (Brown, 1996). An IF index ranges from .00 to 1.00 for different items (ibid, p. 65). According to Brown (1996), an ideal item has an average IF value of .50, but items barely have such an index. In this regard, items that fall within the range of around 0.3 to 0.7 are regarded as acceptable (Brown, 1996; Henning, 1987, p. 50). Thus, items that fell within the allowable IF range were accepted in this study.

Item Discrimination (ID) Analysis: ID analysis was conducted to reveal the degree to which an item separates the task-takers who performed well from those who performed poorly (Brown, 1996). The following stages were pursued in calculating ID indices.

- Task-takers in the top and bottom groups on the whole test were identified.
- IF for upper and lower groups was calculated separately for each item.
- The IF for the lower group was subtracted from the IF for the upper group.

ID indices range between +1.00 and -1.00 (Brown, 1996, p. 68). Ideal items are the ones that have the highest available ID (Brown, 1996). So, among items that have acceptable IF value, only the ones that have the highest ID indexes were selected for retention in revised version of tests. In the evaluation of ID indices of items in the current study, Ebel and Frisbie's (1991, p. 232) ID range guidelines were followed. Based on the ID analysis of items, necessary changes were made in the tasks.

2.4.4. *Reliability and validity*

Based on the pilot study, the reliability of four types of tasks was examined, and the agreement between the raters was calculated. The reliability analysis was conducted through Kuder Richardson Formula 20 (KR-20) for estimating the internal consistency of the data collection tools, i.e., the tasks. KR-20 formula is the most commonly used and reported formula by researchers, and it is accepted as the most accurate estimate of reliability (Brown, 1996). KR-20 value shows to what extent the results are consistent and how well the tasks measure what they aim to measure. It is a method of reliability that is suitable only for the tests or tasks with dichotomous items, which means that the answer of an item is either right or wrong (usually scored as either zero (0.00) or one (1.00)) (Bademci, 2011). In other words, A value of high degree demonstrates the high reliability of a task, which means that the closer the value is to 1, the more reliable the test is. The ranges of internal consistency coefficients proposed by Salvucci, Walter, Conley, Fink and Saba (1997, p. 350) were used in the study, and the reliability coefficients obtained from the pilot study analysis were interpreted.

Inter-rater Reliability Analyses. Inter-rater reliability was calculated to explore to what extent an agreement was reached between two raters in scoring the tasks. Interrater reliability analysis was explicitly performed for two kinds of tasks: sentence completion and production tasks. These context independent tasks included many sentences created by the task-takers and required the researchers to decide on the correctness and appropriateness of verb complementation patterns and senses produced by students. In the first phase, two researchers independently analyzed 30 % of the writings and scored the items. Then, the raters compared their analyses and scores and discussed the points they disagreed with, and came to a conclusion at the end. In this procedure, Cohen's Kappa statistics were used to

measure the agreement between raters. The benchmarks showing the strength of agreement in Kappa Statistics suggested by Landis and Koch (1977, p. 165) were adopted in this research.

3. Findings

The context-dependent and context-independent tasks were administered to the participants, and the responses of participants to the task items were analyzed to determine the reliability and validity of tasks. Based on the item analyses, the findings obtained from the item facility and item discrimination analyses are demonstrated in Appendix D.

The reliability of the four types of tasks was calculated through the Kuder-Richardson-20 Formula. Accordingly, as far as the task items designed to explore the use of non-factive verb complementation use, it was revealed that KR20 for Sentence Completion Task is .84 whereas it is .64 for Fill-in the Blanks Task. As for the tasks designed to measure factive verb complementation use, it was revealed that GJT and FBT tasks had high reliability values (i.e. 0,87 and 0,80 respectively) and SCT had moderate degree of reliability (0,78).

Inter-rater reliability was calculated to find the agreement between two raters in scoring the tasks. Considering the tasks designed for non-factive cognitive verbs Cohen's Kappa for Sentence Production Task was between 1.0 and .64 indicating an acceptable (i.e. substantial and almost perfect) agreement (Landis & Koch, 1977, p. 165). Accordingly, there was almost perfect agreement between the raters for thirteen items and substantial agreement for three items. As far as the agreement values SCT is concerned, Kappa value was between 1.0 and .80 indicating almost perfect agreement between the raters. Specifically, there was almost perfect agreement between the raters for nineteen items and substantial agreement for one item (See Appendix A for the agreement values for each item in SPT and SCT designed for non-factive and factive verbs). Considering the tasks designed for factive cognitive verb complementation analysis, there was substantial agreement between two raters for four items in SCT, that were, Item 1 ($\varkappa$=.77), Item 2 ($\varkappa$=.80), Item 4 ($\varkappa$=.75), and Item 6 ($\varkappa$=.77) and almost perfect agreement for the rest of items. In addition, there was substantial agreement between two raters for nine items in SPT that were, Item 1 ($\varkappa$=.70), Item 2 ($\varkappa$=.77), Item 3 ($\varkappa$=.77), Item 4 ($\varkappa$=.77), Item 6 ($\varkappa$=.75), Item 9 ($\varkappa$=.75), Item 13 ($\varkappa$=.64), Item 14 ($\varkappa$=.64), and Item 18 ($\varkappa$=.77) and almost perfect agreement for the rest of items.

In the second phase, an American native speaker of English analyzed 20 % of the data. The native speaker was informed that tasks were administered to a group of L2 learners and she was asked to carefully read the sentences produced by the task-takers in SCT and SPT and to judge the sentences in terms of acceptability of verb complementation patterns used and the cognitive verb meanings. In this procedure, the native speaker was asked to put plus (+) for the well-formed sentences and verb meanings and minus (-) for the ill-formed sentences and verb meanings. During data analysis procedure, 'acceptability' term was used as "a concept that belongs to the study of performance, whereas grammaticalness belongs to the study of competence" and "grammaticalness is only one of the many factors that interact to determine acceptability" (Chomsky, 1965, p. 11). For the cases on which no compromise was reached between the native speaker and the researcher, COCA was checked as a reference resource. For example, there was no compromise between the native speaker and the researcher in the use of assume in imperative form, the use of assume was checked in COCA and it was revealed that there were such uses in various registers in English. Thus, such uses were accepted in language users' responses.

Apart from that, the native speaker rated the language users' responses to task items. The agreement between the native speaker and one of the researchers was calculated through Cohen's Kappa. Considering the tasks designed to explore non-factive cognitive verb complementation use, in SPT, there was almost perfect agreement between the raters for eleven items in terms of correctness and appropriateness of verb complementation and sense use and substantial agreement for five items (See Appendix B). As for SCT, in terms of pattern acceptability, there was almost perfect agreement for seventeen items and substantial agreement for the rest of items, i.e. three items. In terms of meaning acceptability, there was almost perfect agreement between the raters for seventeen items and substantial agreement for three items. (See Appendix B). As for the productive tasks of factive cognitive verbs, it was found out that there was almost perfect agreement between the two raters for all items in specifically SCT considering acceptability of patterns and senses produced by the students as the Kappa values of all these items were over .81 on the basis of Landis and Koch's (1977) model. As for SPT, there was almost perfect agreement between the raters for nine items and substantial agreement for one item in terms of pattern and meaning acceptability. The Kappa Statistics of the items in production tasks were depicted in Appendix C.

4. Discussion and Conclusion

With a purpose to develop tasks to figure out the English language learners' receptive and productive levels in the use of cognitive verb complementation patterns and their related verb senses, the current study adopted an exploratory sequential mixed methods design. Firstly, the verb complementation patterns and related verb senses of cognitive verbs (*think, believe, assume, suppose, know, understand, regret, forget* and *remember*) were checked and determined based on the source of Valency Dictionary of English (Herbst et al., 2004) and Verb Net (Schuler, 2005). Based on these reference data, selectional restrictions and syntactic and semantic frames of the cognitive verbs were identified based on the frequency criterion and the most salient verb senses and complementation patterns were identified for each verb. Prior to constructing the items in the tasks, extracts from COCA were analyzed through the search of cognitive verbs from a variety of text types consisting of magazines, newspapers, fiction, spoken language and academic texts to form the task items. GJT and FBT were developed to measure learners' competence of verb complementation patterns and senses at recognition level whereas SPT and SCT were developed to decipher the learners' performances at production level. To sum up, two types of context-dependent and two types of context-independent tasks were developed within the frame of the present study. The items were developed by taking item format guidelines into consideration. Depending on the expert opinions, the items were revised and the instructions were re-organized. Based on pilot study results regarding item analyses, reliability analyses of tests and inter-rater reliability analyses, all tasks were revealed to be reliable and the majority of items were found to be acceptable and the rest of the items that were poor were omitted and new items were developed. Item format analysis and item content analysis were conducted through feedback sessions and necessary changes were made on the tasks. Additionally, item facility and item discrimination analyses were administered for the task items. Agreement values for each task were calculated and pattern and meaning acceptability of the learner responses to task items were identified.

The tasks developed within the scope of this study are expected to contribute to the field of foreign language teaching by addressing the English language learners' competence and performance levels through context dependent and context independent task items. Thus, the current study provides a unique pedagogical application for both language teachers and

learners and fills a gap in the literature by combining recognition and production levels. In other words, the core contribution of this research is that the context-dependent and context-independent tasks bring to light the language users' recognition and production of verb complementation in the target language. Another significant methodological contribution of the current study is that it provides an insight into the use of corpus for pedagogical purposes. Specifically, as an implication, English language teachers and researchers may make use of this method to prepare teaching and testing materials by corpus tools. In this regard, as more publications that contain ready-made exercises based on authentic speech and writing from different text types and language varieties and focused on language items that are of central importance and/or troublesome for learners are needed (Römer, 2010, p. 30), the current study fills a gap in corpora use for pedagogical purposes in English language teaching. Additionally, this study is likely to promote task creation through corpora use and foster the creation of testing and teaching materials through different reference corpora such as Michigan Corpus of Academic Spoken English and British National Corpus. Another pedagogical contribution of this approach would be to promote using multiple tasks and tools to decrease the influence of the type of the task used to assess and evaluate different language skills. In the relevant literature, scholars argue that Grammaticality Judgment Tasks rests on the distinction between competence and performance: internalized mental knowledge and actual language use (Han, 2010, p. 177). They may not provide a sensitive evaluation (about whether or not a sentence is grammatical) of a learner's developing linguistic abilities, rather, they elicit a response indicating the learner's belief about the L2 grammar (Munnich, Flynn & Martohardjono, 1994, p. 229). For this reason, as a pedagogical implication, the use of different tasks to decipher the learners' actual performance on the use of a specific language item is essential and the current study provides a comprehensive overview through the development of context-dependent and context-independent task development process in investigating the cognitive verb complementation patterns in English.

Considering that the scope of this study is limited to the analysis of nine cognitive verbs and their verb complementation patterns used by 200 participants with the aim of highlighting the receptive and productive knowledge of Turkish learners of English, as an implication, these tasks could be administered to wide numbers of participants or language learners from different native language backgrounds other than Turkish to make generalizations regarding other ESL (English as a second language) or EFL (English as a foreign language) contexts. One further implication is that these tasks could be used to explore the use of verb complementation by language users from different educational backgrounds such as learners at different proficiency levels or primary/secondary schools. To provide further evidence for the findings obtained from these tasks, it would be ideal to conduct think-aloud protocols with these subjects to gain more insights regarding their knowledge of verb complementation and thoughts and intuitions about their choices. Moreover, as the current study is limited to the development of written tasks, it would also be ideal to collect data through spoken medium other than think-aloud protocols such as role-plays or spontaneous conversations in addition to these tasks. The scope of the future studies could also be extended to the analysis of other classes of verbs or different parts of speech other than verbs.

Acknowledgements

Note on Ethical Issues

The authors confirm that ethical approval was obtained from Anadolu University (Approval Date: 26/02 /2018).

References

Anderson, J. I. (1976). *The acquisition of English sentential complementation by adult native speakers of Puerto Rican Spanish* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

Bademci, V. (2011). Kuder-Richardson 20, Cronbach'in alfası, Hoyt'un varyans analizi, genellenirlik kuramı ve ölçüm güvenirliği üzerine bir çalışma. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, 17*, 173-193.

Barak L., Fazly, A., & Stevenson, S. (2012). Modeling the acquisition of mental state verbs. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2012)* (pp. 1-10).

Berndt, R. S., Haendiges, A. N., Burton, M. W., & Mitchum, C. C. (2002). Grammatical class and imageability in aphasic word production: Their effects are independent. *Journal of Neurolinguistics, 15*(3-5), 353-371.

Biber, D., Johansson, S., Leech, G., & Conrad, S. (1999). *Longman grammar of spoken and written English.* U.K.: Pearson Education.

Booth, J., & Hall, W. S. (1995). Development of understanding of the polysemous meanings of the mental state verb know. *Cognitive Development, 10*, 529-549.

Booth, J. R., Hall, W. S., Robison, G. C., & Kim, S. Y. (1997). Acquisition of the mental state verb know by 2- to 5-year-old children. *Journal of Psycholinguistic Research, 26*(6), 581-603.

Brown, J. D. (1996). *Testing in language programs.* New Jersey: McGraw-Hill College.

Cappelli, G. (2008). Antonymy and verbs of cognitive attitude: When know is the opposite of think and believe. In M. Bertuccelli Papi, A. Bertacca & S. Bruti (Eds.), *Threads in the complex fabric of language* (pp. 529-546). Pisa, Italy: Felici Editore.

Chan, A. Y. W. (2004). Syntactic transfer: Evidence from the interlanguage of Hong Kong Chinese ESL learners. *The Modern Language Journal, 88*(1), 56-74.

Chomsky, N. (1965). *Aspect of the theory of syntax.* Cambridge: The M.I.T. Press.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: SAGE Publications.

COCA. Corpus of Contemporary American English. Retrieved November 12, 2019, from https://corpus.byu.edu/coca/

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice-Hall.

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition, 13,* 161-186.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language learning, 54*(2), 227-275.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment.* New York: Routledge.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *The psychology of learning and motivation advance in research and theory.* (Vol. 9, pp. 133-177). New York, U.S.: Academic Press.

Gass, S. M. (1994). The reliability of second language grammaticality judgments. In E. E. Tarone, S. M. Gass & A. D. Cohen (Eds.), *Research methodology in second-language acquisition.* (pp. 303-322). New York: Lawrence Erlbaum.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course.* (Third Edition). New York: Routledge.

Givón, T. (1973). The time-axis phenomenon. *Language, 49*(4), 890-925.

Halliday, M. A. (1999). The notion of "context" in language education. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (Vol. 169), (pp. 1-24). Amsterdam-Philadelphia: John Benjamins Publishing.

Han, Y. (2000). Grammaticality judgment tests: How reliable and valid are they? *Applied Language Learning, 11*(1), 177-204.

Hare, M., McRae, K., & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language, 48*(2), 281-303.

Henning, G. (1987). *A guide to language testing: Development, evaluation, research.* New York: Newberry House Publishers.

Herbst, T., Heath, D., Roe, I., F., & Götz, D. (2004). *A valency dictionary of English.* Berlin: Mouton de Gruyter.

Hubbard, P. (1994). Non-transformational theories of grammar: Implications for language teaching. In T. Odlin (Ed.), *Perspectives on pedagogical grammar* (pp. 49-71). New York: Cambridge University Press.

Kellerman, E. (1986). An eye for an eye: Crosslinguistic constraints in the development of the L2 lexicon. In E. Kellerman & M. Sharwood Smith (Eds.), *Crosslinguistic influence in second language acquisition.* (pp. 35-48). New York: Pergamon.

Kempen, G. (1977). Conceptualizing and formulating in sentence production. In S. Rosenberg (Ed.), *Sentence production: Developments in research and theory* (pp. 259-274). Hillsdale, NJ: Erlbaum.

Kidd, E., Lieven, E., & Tomasello, M. (2006). Examining the role of lexical frequency in the acquisition and processing of sentential complements. *Cognitive Development, 21*, 93-107.

Klotz, M. (2007). Valency rules? The case of verbs with propositional complements. In T. Herbst, & K. Götz-Votteler, (Eds.), *Valency: Theoretical, descriptive, and cognitive issues* (pp. 117-129). Berlin: Mouton de Gruyter.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159-174.

Levelt, W., & Maasen, B. (1979). Lexical search and order of mention in sentence production. In W. Klein & W.J.M. Levelt (Eds.), *Crossing the boundaries of linguistics.* Studies presented to Manfred Bierwisch. (pp. 221-252). Dordrecht: Reidel.

Leow, R. P. (1996). Grammaticality judgment tasks and second-language development. In J. E. Alatis, C. A. Straehle, M. Ronkin, & B. Gallenberger (Eds.), *Linguistics, language acquisition, and language variation: Current trends and future prospects* (pp. 126-139). Washington, DC: Georgetown University Press.

Milton, J. (2009). *Measuring second language vocabulary acquisition.* Bristol, UK: Multilingual Matters.

Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60*, 167-71.

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and Grammaticality Judgment Tasks: What they measure and how they relate to each other. In E. E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 227-243). Hillside, NJ: Lawrence Erlbaum.

Naigles, L. R. (2000). Manipulating the input: Studies in mental verb acquisition. In B. Landau, J. Jonides, E. Newport, & J. Sabini (Eds.), *Perception, cognition, and language: Essays in honor of Henry and Lila Gleitman* (pp. 245-274). Cambridge, MA: MIT Press.

Naiman, N. (1974). The use of elicited imitation in second language acquisition research. *Working Papers in Bilingualism, 2,* 1-37.

Nixon, S. M. (2005). Mental state verb production and sentential complements in four-year-old children. *First Language, 25*(1), 19-37.

Owen Van Horne, A. J., & Lin, S. (2011). Cognitive state verbs and complement clauses in children with SLI and their typically developing peers. *Clinical Linguistics and Phonetics, Early Online, 25*(10), 1-18.

Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105,* 125-165.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* London: Longman.

Roland, D., & Jurafsky, D. (1998, August). How verb subcategorization frequencies are affected by corpus choice. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 2* (pp. 1122-1128).

Römer, U. (2010). Using general and specialized corpora in English language teaching: Past, present and future. In M. C. Campoy-Cubillo, B. Bellés Fortuño, & M. L. Gea-Valor (Eds.), *Corpus-based approaches to English language teaching* (pp. 18-38). London: Continuum.

Salvucci, S., Walter, E., Conley, V., Fink, S., & Saba, M. (1997). *Measurement error studies at the National Center for Education Statistics (NCES).* National Center for Education Statistics.

Schuler, K. K. (2005). *VerbNet: A broad coverage, comprehensive verb lexicon.* Unpublished PhD Dissertation: Pennsylvania: The University of Pennsylvania.

Schwarte, B. (1982). The acquisition of English sentential complementation by adult speakers of Finnish. *Jyvaskylti Cross-Language Studies, 8,* 1-161.

Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14,* 301-321.

Stojičić, V. (2008). Functional labeling with polysemy verbs. *Linguistics and Literature, 6*(1), 27-34.

Tremblay, A. (2005). Theoretical and methodological perspectives on the use of grammatically judgement tasks in linguistic theory. *Second Language Studies, 24*(1), 129-167.

Wellman, H. M., & Estes, D. (1987). Children's early use of mental verbs and what they mean. *Discourse Processes, 10,* 141-156.

Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language, 60,* 368-392.

Verdaguer, I. (2010). A corpus-based contrastive study of mental verbs. *Language, translation, reception. To honour Julio-César Santoyo.* Universidad de León.

Appendix A. *Agreement Values for Sentence Production Task items-Non-factive cognitive verbs Agreement Values for Sentence Completion Task items--Non-factive cognitive verbs*

Sentence Production Task																	
Verb	**Think**				**Believe**				**Assume**				**Suppose**			
Item ID	Sentence	Verb Meaning	Sentence	Verb Meaning	Sentence	Verb Meaning	Sentence	Verb Meaning	Sentence	Verb Meaning	Sentence	Verb Meaning	Sentence	Verb Meaning	Sentence	Verb Meaning
Agreement Value (Cohen's Kappa)	1	1	.75	1	1	.64	1	.92	1	.76	.84	.83	1	1	.91	.91

Sentence Completion Task																				
Item ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
A. Value	1	1	1	.92	.84	.91	1	.89	1	1	.92	1	.92	1	1	.83	.91	1	.80	1

Agreement Values for SCT and FPT Items (between researcher and Turkish researcher)- Factive Cognitive Verbs

Test type	Cohen's kappa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SCT	K	.77	.80	.88	.75	.88	.77	.89	1.0	.89	1.0	1.0	.88	1.0	1.0	.88	1.0	.86	.83	1.0	1.0
FPT	Cohen's kappa	1	2	3	4	5	6	7	8	9	10										
pattern	K	.70	.77	1.0	.81	.75	1.0	.64	1.0	1.0	1.0										
Sense	K	.77	.77	.75	.83	1.0	.84	.64	.90	.77	.89										

Appendix B. *Agreement Values (between the researcher and English Native Speaker) for Sentence Production Task Items-non-factive cognitive verbs*

Sentence Production Task																
Verb	**Think**				**Believe**				**Assume**				**Suppose**			
Item ID Sentence Number & Verb Meaning (VM)	1	VM	2	VM	3	VM	4	VM	5	VM	6	VM	7	VM	8	VM
Agreement Value (Cohen's Kappa)	1	1	.65	1	1	1	.65	1	.65	.65	.86	.77	1	1	.89	.87

Agreement Values (between the researcher and English Native Speaker) for Sentence Completion Task Items-non-factive cognitive verbs

Sentence Completion Task-Pattern Acceptability																				
Item ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
A. Value	1	1	1	.62	1	.72	1	1	1	1	.78	1	.90	1	1	.87	1	1	.87	.87
Sentence Completion Task-Meaning Acceptability																				
Item ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
A. Value	.88	.94	.89	.89	.88	.84	.81	1	.77	.78	.83	.94	.85	.86	1	.87	.80	.90	.90	.91

Appendix C. *Interrater Reliability Analysis of Production Tests (between the researcher and English Native Speaker) -Factive Cognitive Verbs*

SCT Item	Cohen's kappa	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
Pattern Accept.	K	1	.96	1	1	1	1	.89	1	1	1	1	1	1	.84	1	1	1	.96	1	1
Sense Accept.	K	1	.83	1	1	1	1	.96	.95	1	1	1	1	.95	.84	1	1	.94	1	1	1

FPT Item	Cohen's kappa	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10
Pattern Accept.	K	.65	1	1	1	1	1	.84	.89	1	1
Sense Accept.	K	1	1	1	1	1	.95	1	1	.65	1

Appendix D. *Item Analyses in the Tasks Developed for Examining the Factive and Non-Factive Cognitive Verbs*

Non-factive		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
SCT	IF p	0,49	0,38	0,73	0,63	0,48	0,80	0,19	0,84	0,68	0,85	0,63	0,45	0,57	0,63	0,85	0,70	0,66	0,74	0,70	0,64
	ID	0,57	0,40	0,44	0,67	0,51	0,44	0,36	0,34	0,51	0,26	0,59	0,53	0,53	0,65	0,40	0,61	0,69	0,65	0,63	0,67
GJT		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
	IF p	0,56	0,41	0,66	0,25	0,65	0,57	0,54	0,42	0,62	0,58	0,48	0,65	0,81	0,33	0,26	0,62	0,25	0,47	0,70	0,49
	ID	-0,04	0,46	0,34	0,36	0,44	0,56	0,20	0,40	0,38	0,32	0,44	0,44	0,14	0,40	0,30	0,34	0,32	0,40	0,62	0,46
FBT		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10										
	IF	0,64	0,53	0,80	0,49	0,81	0,84	0,45	0,58	0,43	0,59										
	ID	0,44	0,61	0,46	0,55	0,40	0,40	0,61	0,38	0,61	0,67										
Fact.	IF p	.51	.78	.69	.63	.63	.66	.66	.72	.90	.81	.81	.69	.57	.60	.60	.75	.69	.87	.54	.63
SCT	ID	.61	.30	.72	.72	.72	.72	.72	.72	.30	.51	.51	.51	.82	.51	.72	.41	.61	.30	.72	.30
GJT		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
	IF p	.52	.91	.55	.50	.50	.61	.44	.47	.69	.72	.47	.61	.55	.36	.75	.44	.13	.41	.50	.52
	ID	.73	.10	.31	.73	.84	.73	.31	.84	.10	.10	.84	.63	.63	.31	-.21	.42	.31	.73	.42	.73
FBT		Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10										
	IF p	.62	.45	.64	.62	.62	.62	.45	.48	.16	.56										
	ID	.56	.40	.24	.72	.64	.56	.40	.72	.48	.56										