

Construction Grammar Meets Language Testing and Evaluation: Redefining & Ensuring Construct Validity

Tan Arda Gedik¹

¹ FAU Erlangen Nürnberg, Germany

Contact:	Tan Arda Gedik		tan.gedik@fau.de
----------	----------------	---	--

ABSTRACT

Theoretical linguistics and teaching are concurrent, that is, neither can exist without the other. In other words, decisions made in (foreign) language teaching will inevitably affect how language testing and evaluation as a concept is designed and administered. For example, if a curriculum states that the English as a foreign language teachers need to make a distinction between a gerund and an infinitive in the classroom and teach it as such, then, this distinction will be tested or evaluated in one way or another. In this article, language testing and evaluation (LTE), an important component in language teaching, is combined with construction grammar (henceforth CxG), a salient theory of language with ample evidence to support its claims. Specifically, this article discusses how construct validity, test items, and rubrics can be reimagined from the perspective of usage-based construction grammar.

KEYWORDS:

construction grammar;
usage-based approaches;
construct validity;
language knowledge;
test-items

ARTICLE HISTORY:

Received 10 March 2022
Accepted 16 June 2022

Introduction

Within LTE, there are fundamental concepts that help prepare and administer a test reliability. These can be categorized as reliability and validity. Under validity, one can find construct validity, criterion validity, and content validity. Focusing on construct validity as it is the focus of the present article, it is a way of examining whether a test is testing what it claims to test. Following Brown (2000), construct validity is how we understand and define what language proficiency is. That is, the definition of how L2 speakers' language proficiency is defined is a construct. Bachman and Palmer (1996) stated "English grammar, vocabulary, reading and listening comprehension" which was "widely used at that time" in the English as a foreign LTE practice was separated. They also suggest that "language ability was viewed as a set of finite components- grammar, vocabulary, pronunciation, spelling-that were realized as four skills-- listening, speaking, reading, and writing" (Bachman & Palmer, 1996). Since then, this view seems to have caught on in many textbooks and consequently testing practices. When foreign

language textbooks tend to draw a sharp dividing line between lexis and grammar, teachers follow the same lead, although they do not necessarily always follow such practices, and to keep their tests' validity and reliability scores high, they may separate lexis from grammar in tests to follow the textbooks. Similarly, Mahlberg (2006) reports that most English language textbooks divide lexis from grammar in teaching cohesion and cohesive devices. See pages 3-4 in *Green Line 10* under *Sprachliche Mittel*, and pages 2-4 in *Look 1*, National Geographic's English textbook for young learners for such a separation.

The issue at stake is that some tests designed reflect on a decades-long linguistic tradition (see (Römer, 2017) on the speaking section of standardized tests), i.e., a construct for language proficiency based on the separation of lexis from grammar. As Shohamy (1995) states "language testing has always followed linguistic theories of the time" and as such how researchers defined language proficiency. From that perspective, the Chomskyan understanding was that speakers had a perfect state of knowing language, i.e., competence, and a state affected by extra-linguistic factors, i.e., performance (Chomsky, 1965). This was later on expanded upon by Hymes (1972) as a criticism of this paradigm and was defined as linguistic and communicative competence and linguistic and communicative performance. Since then, there have been many turns in language testing and assessment, for instance in the 1970s there was a criticism of test-items not being communicative enough (see (Shohamy, 1995, 1996) for a survey) and Chomsky and Hymes' models served as the basis for many upcoming frameworks (Oyinloye, Adeoye, Fatimayin, Osikomaiya, & Fatola, 2020). The succeeding frameworks are Canale and Swain's (1980), Bachman (1990), and Bachman and Palmer (1996). These frameworks essentially divided communicative competence into grammatical, sociolinguistic, and task or strategic competence. Bachman and Palmer's model (1996) was developed to apply these divisions in language testing. In any case, while there are ongoing discussions about whether competence and performance should be separated or not (see (Skehan, 2013) on how competence cannot predict performance), this article focuses on the separation of lexis from grammar, irrespective of such discussions.

Criticizing tests of any form today is difficult. This is because there have been many turns in LTE and consequently it can be argued that the present-day exams are a mix of all of them. Nevertheless, one assumption that seems to persist in many standardized (Römer, 2017) and non-standardized tests is the separation of lexis from grammar. Despite the fact that there have been improvements with communicative testing or task-based testing in LTE, as Römer (2017) states that "more recent models of language ability...continue this separation of lexis and syntax as distinct aspects of 'grammatical knowledge', separating these aspects of language ability from knowledge of language functions, which is subsumed under 'pragmatic knowledge'". While it is difficult to pinpoint where this separation comes from historically,

generative assumptions certainly propagated a separation with the Universal Grammar. This, as Shohamy (1996) suggests, forms the base of what language proficiency might be because “competence” refers to ‘unobserved’, underlying knowledge” (Shohamy, 1996). Thus, if applied linguists, textbook designers, and teachers do away with generative ideas of a construct for language proficiency then that means that there will be a shift in how language proficiency is defined which will change the construct. Generative ideas or approaches refer to the assumption that grammar is separated from lexis in competence and performance in the rest of the article.

Alderson and Kremmel (2013) scrutinize whether vocabulary and grammar knowledge in reading sections should be treated separately. Shiotsu and Weir (2007) also explain that “the role of vocabulary appears somewhat overstated while that of grammar [is] understated” in reading sections. Similarly, speaking sections also deal with the same issue. The rubrics of many speaking tests seem to favor a separation of grammar from lexis (Luoma, 2004; L. B. (Ed. . Taylor, 2011; Weir, Vidaković, & Galaczi, 2013). Since usage-based construction grammar redefines what linguistic knowledge is, i.e., construct of language knowledge, the redefinition of this construct will have implications for the current understanding of a construct of language knowledge in LTE.

Literature Review–Construction Grammar

For many decades, lexis and grammar have been accommodated as two separate entities. This clearly stems from the fact that for many years researchers tried to answer questions such as: What do you need to know to know a language? Many scholars answered this question differently, but they more or less agreed that lexis and grammar were acquired separately and thus should be tested separately. I will refer to these ideologies as the generative understanding of language for ease of reference. Starting in the 1980’s, there was a gradual shift in how researchers looked at language and linguistic knowledge, and this shift can be referred to as the lexicogrammatical view. There are different approaches to this lexicogrammatical view of language but perhaps the most famous one is construction grammar (Goldberg, 1995, 2006). Constructions are form-meaning pairings, in that they do not separate meaning from form.

While generativist approaches acknowledge the importance of learning or environmental factors in language acquisition, they subscribe to a Language Acquisition Device, which helps speakers become proficient in a language regardless of environmental or learning factors (Pullum & Scholz, 2002). On the contrary for CxG, there are not a set of a priori rules governing languages universally, nor is there something called the language acquisition device. With the advent of computers and corpora in the 1970s and 80s, researchers realized not only language was based on chunks, but also these chunks were highly repetitive. For cognitive linguists,

language was not a set of a priori rules but rather a dynamic web of interrelated signs, i.e., both grammar and lexis, that unfold and is learned over time through domain-general cognitive abilities (Beckner & Bybee, 2009; Tomasello, 2003). These abilities are perception, hearing, attention, memory, automation, and abstraction, to name a few (Divjak, 2019; Tomasello, 2003). Thus, cognitive linguists, or usage-based linguists formulated that language learning is based on exposure to a set of highly repetitive chunks via general-cognitive abilities. It is difficult to draw a line between the two since both fields are intertwined with one another and heavily draw on each other's practice. In this study, I use both terms interchangeably to refer to the same field. Another important aspect of CxG is that it bases its assumptions on lexicogrammar, a continuum where there is no separation of grammar from lexis (see figure 1). This can account for a wide array of linguistic phenomena, even those which were considered in between lexis and grammar by previous scholars. Many research studies suggest that speakers use fixed expressions or partially fixed expressions to communicate (Biber, 2009; Römer, Roberson, O'Donnell, & Ellis, 2014; Sinclair, 1997, 2014).

Prefixes, suffixes <i>un-, re-, de-...</i>	Words <i>cat, dog,</i> <i>bird...</i>	Fixed expressions <i>Here you are</i>	Partially fixed expressions <i>The Xer,</i> <i>The Yer</i>	Fully abstract schemas <i>Subj Verb</i> <i>Obj Oblique</i>
--	---	---	--	--

Figure 1. Assumptions on lexicogrammar

This means that any linguistic unit can be defined in terms of form-meaning pairings, regardless of how big or small the unit is. The definition that we subscribe to in this paper is the following:

"Any linguistic pattern is recognized as a construction ... even if they are fully predictable as long as they occur with sufficient frequency." (p.5)

In figure 1, it is also clear that the continuum illustrates the gradience of what can be thought of as lexis and as grammar. The gradience of the color gray represents the level of schematicity, that is, the darker it gets, the higher the schematicity is. Moreover, in CxG, unlike in generativist approach, grammar is regarded as meaningful. That is, a highly abstract schema such as the caused-motion construction can coerce with nonce verbs or verbs that cannot typically occur in such argument structure patterns, and give meaning to them. One of the leading examples for this is the following:

(1) *She sneezed the foam off the cappuccino* (Goldberg, 2006)

(1) is an unusual use of the verb *sneeze*. In a generativist approach, *sneeze* would be given two entries, one for its transitive and one for its intransitive use. However, as Goldberg explains (1995, 2006), the constructional schema for the caused-motion construction is *subject+verb+object+oblique*, and when combined with semantically coherent verbs, it enriches the verb with the meaning of 'X moves Y along Z'. This renders grammar, those highly abstract rules, meaningful. Thus, in one way, it is possible to claim that there is an interplay between grammar and lexis from a constructionist point-of-view. As such, CxG moves away from a verbocentric approach found in generativist approaches, and distributes the labor of meaning creating across different levels of constructions, for instance words and argument structure constructions.

Many researchers point out the importance of item-specific knowledge (Boas, 2003; Herbst, 2020; Perek, 2015). A good example for this is the fusion of *give* and the ditransitive construction. Herbst (2020) demonstrates that the verb *give* is the most frequently used item in the verbal slot in the construction. This might not be surprising, since the ditransitive construction itself carries the meaning of 'transfer' and *give* is a very prototypical item for such a meaning. In any case, as Herbst (2020) shows, it is possible to talk about the *give*-ditransitive construction and there is ample corpus-evidence to suggest that it might be entrenched in the minds of speakers as such. Moreover, Goldberg (2006) demonstrates that argument structure constructions are learned by generalizing over specific items and this skewed frequent occurrence of one verb helps with learning.

Then, from a usage-based constructionist perspective, one can define the construct validity for language proficiency or knowledge as follows:

In a usage-based approach, "To know a language, one must know its schemata" (Turner, 2018) on the basis of ambient language, i.e., input, and must know form-meaning pairings with their pragmatic and discourse functions. Furthermore, a speaker must have entrenched the transitional probabilities of constructions, i.e., what comes after *give birth* or whether it is *heavy rain* or *strong rain*." Knowledge of language is to be modeled as knowledge of constructions, and nothing else in addition" (Hilpert, 2014) since constructions already unify every faculty of language, i.e., form, meaning, discourse, pragmatics, frequency and so on, that was previously separated in previous linguistic traditions (J. R. Taylor, 2012). Then, knowing a language is knowing" constructions all the way down" (Goldberg, 2010).

From CxG to Language Testing and Evaluation

Assuming that teaching materials and teaching are aligned with the tenets of and evidence from usage-based CxG, construct validity within LTE will need to be rethought in line with the new construct validity proposed in this paper. This section aims to demonstrate how that might look. Clearly, the shift from previous traditions in LTE where the proficiency of a speaker

was tested based on isolated-grammar-questions to communicative contexts where linguistic units are not separated is a step-forward in the right direction in light of CxG literature. Current approaches do not employ communicative test items, but they make a distinction between lexis and grammar (Harding, 2014). Applying the principles outlined in the previous section, construct validity in LTE informed by usage-based constructionist studies need to adhere to the following principles to ensure that the construct validity is intact. Furthermore, Goldberg's explanation of speaking conventionally is also important (2019). A usage-based constructionist account of construct for language proficiency is as follows:

Language learning takes place with general domain cognitive abilities in social contexts, i.e., pattern recognition, abstraction, generalization, perception, and keeping a record of type and token frequencies to name a few. Linguistic knowledge consists of constructions, i.e., form-meaning pairings, and the mental organization of constructions arises from usage and frequency effects with mechanisms such as entrenchment and statistical preemption. Because speakers strive to speak like everyone else, conventionalized constructions are preferred over unconventionalized constructions.

Communicative test items and CxG and lexicogrammar in testing

Communicative testing has been around for several decades (Fulcher, 2000; Harding, 2014). This framework is based on testing language abilities in context (see Davies et al., 1999). This also serves as the base for the CEFR ((Ed.), 2020) and also standardized foreign language tests, such as TOEFL iBT (Services, 2021). Harding (2014) summarizes the current landscape in their article and states "there is strong evidence to suggest that the communicative approach has become the dominant paradigm in modern language testing". From this perspective, communicative language testing is compatible with CxG because from a constructionist standpoint language is learned in and through usage-events, especially in context or in communication (see (Goldberg, 2019) for a detailed discussion). As an example, the ditransitive construction characterizes a basic human act of transferring (Goldberg, 1995). However, the communicative approach in testing has received its criticism (McNamara, 1996) mainly for lacking what ability to use. The newer iterations of this approach (Byram, 1997; Leung, 2005; Roever, 2011; Timpe, 2012) include how a testee can adapt to new and challenging tasks in real-time. As an example for this, the speaking section of TOEFL iBT, whose construct has a "task-based design" (Services, 2021) and also that TOEFL tests "communicative competence—the ability to put language knowledge to use in relevant contexts" (Services, 2011). However, even then, this approach does not necessarily point at a merging of lexis and grammar. As Römer (2017) points out, speaking sections of standardized language tests make a distinction between lexis and grammar.

As several researchers have explained (Alderson & Kremmel, 2013; Shiotsu & Weir, 2007), it is redundant to separate lexis from grammar in any section of a language test. To ensure construct validity in LTE informed by usage-based CxG, this paper advocates the merging of the two in tests. In this case, not only will our test items change, but also the rubrics for writing and speaking will need to be rethought, which is discussed later. As Römer (2017) points out, the rubrics for many speaking tasks still favor the separation. Example test items presented here are not conclusive. In other words, they are open for development and would need to be adapted in accordance with the needs of a test. Nevertheless, they are presented to give the reader an idea of how such an approach to testing would manifest itself in test items.

One way to test lexicogrammar in reading sections would be cloze tests. Adhering to the rules, every 7th word is deleted in a passage. Thus, for instance, if there is an adjective phrase construction in the passage, it is possible to test both lexis and grammar at the same time. The following example can be used in both a final achievement test and a proficiency exam:

(Example 1) Select the odd one out.

The (1)_____ rabbit disappeared after a while. *The passage continues...*

(1) a) scared b) feared c) afraid d) terrified

As speakers generalize over items, the fact that A-class adjectives do not occur in adjectival positions becomes more entrenched. Thus, by forming the multiple choices with semantically related but unconventional items, lexis and grammar can be combined. As per tradition, the rest of the reading section questions would test for general understanding and inference, to name a few. By nature, reading questions require a lexicogrammatical approach, since they test whether the student understands for instance the “*unless*” construction, both the meaning it carries and the specific items that occur within another example would be the way-construction and its metaphorical extensions (Goldberg, 1995). A question that asks the comprehension of a sentence such as “she talked her way to fame” requires the understanding of the grammar and its meaning extension onto talk. The same can be said to apply to the listening section, too. By including questions such as the one outlined above in example 1, the listening section can be enriched with the testing of lexicogrammar. The rest of the listening questions, just as the reading questions do, require an understanding of lexis and grammar at the same time.

As it has been outlined, constructions themselves carry meaning and detailed grammatical information. They have discourse-specific functions (see the discourse-specific properties of the ditransitive construction (Goldberg, 1992)). They also have slot-specific requirements. For instance, the recipient in the ditransitive construction should be animate.

Thus, a sentence such as *I gave the table a flower* would not be attested in a corpus. Thus, if the teaching includes teaching constructions and consequently their lexicosemantic meaning properties, then the test items should reflect that.

A lexicogrammar section might test the following things in a multiple-choice question. The following example is intended for a proficiency exam:

(Example 2)

Laura's Evening

It was a rainy evening. I was working at the diner. Two customers sat down at one of the tables and waved their hands. I approached them and asked what they would want. They asked for some soda. So, I went back but there were no sodas left. So, I came back and (1) _____ . Five minutes later, I realized I had not seen the sodas below the counter. I quickly ran to the table, apologized for the inconvenience, and said (2) _____ .

Select only one option to complete Laura's speech in (1).

- a) I explained them the situation
- b) I explained the situation to them

Select only one option to complete Laura's speech in (2).

- a) I will bring your sodas to you, now
- b) I will give you your sodas, now
- c) I will bring you your sodas, now
- d) I will give your sodas to you, now

Certain verbs are item-specific, in that, they will only appear in a certain construction, i.e., *explain this to me* but not *explain me this* (see (Goldberg, 2019) for a detailed discussion). What is tested here is the collocational patterns of two verbs, *explain*, *give* and *bring*. In other words, the question tests item-specificity. As Gries and Stefanowitsch (2004) demonstrate, *give* and *bring* appear in the ditransitive and the to-prepositional construction more often respectively. Thus, if the proficiency levels of constructions are identified, test items such as example (2) can be of help in identifying testees proficiency levels with greater accuracy. This way, test designers also ensure a unibody construct validity in that language proficiency is speaking like the rest of the community (Goldberg, 2019).

However, including lexicogrammar is not only testing lexicosemantic properties of constructions, but also item-specificity.

(Example 3)

Put the words in order to build a sentence. There are two extra items in each question.

a) I // him // book // to // donate // a // give → _____.

In a), the testees can come up with two attested sentences (i) *I donate a book to him*, or (ii) *I give him a book*. With this test item, the teachers test item-specificity, that is the fact that *donate* can only occur with the to-dative construction among others, i.e., *explain*, *donate*, and *return* to name a few.

(Example 4)

Fill in the blank using *should* and the verb in brackets.

Scott: Wow, I find this new topic in biology about birds really cool.

Rachel: I know. We should do more research on it.

Scott: Do you have your phone and mobile data?

(1) _____? (*google*)

Rachel: Of course. Let's do it.

(1) _____?

This test item is intended for a proficiency exam and it tests the *should* construction, the transitive construction and its extensibility. That is, students will have been exposed to attested examples such as *I should look this up* or *I should search this*. By seeing *google* as a verb, students' knowledge extending the transitive construction to other verbs is tested.

One final example is the testing of an argument structure construction. In this example, the item can be embedded in a listening section, albeit it is also possible to include it in speaking or writing by enhancing the rubrics used. This test item is based on marking statements as true or false.

(Example 5)

AUDIO: ... and as such, Sally talked her way to fame.

a) Sally talked and became famous without any obstacles ____

b) Sally overcame obstacles by talking and became famous ____

This item tests both the metaphorical extension of the way construction, moving from a literal motion meaning to a metaphorical meaning of overcoming obstacles (see (Goldberg,

1995) on a detailed discussion for this), and it also tests the form of the argument structure construction, see the difference between options a & b.

Corpus-driven test items

Corpora shed light on how L2 speakers learn language and what they can do. In other words, they provide insight into proficiency levels, as well. Clearly, this is not something noteworthy. Corpus-based testing has been discussed in the literature before (Barker, 2012; Troike, 2006). Association of Language Testers in Europe (ALTE), for instance, has advocated for a shift in how testers should define construct validity in regard to using corpora in testing. As Barker (2012) outlines, corpora in LTE have been employed to determine "(a) defining user needs and test purpose, (b) designing tests, and (c) refining task rating" (p. 2). Barker (2012) also mentions the following:

More generally, learner corpora are used by test writers to explore the collocational patterning in learner or native production, so that common or less frequent patterns can be tested to distinguish between candidates at a particular proficiency level. Additionally, the most frequent errors or misuses of specific collocational pairings can be used to provide suitable *distractor* items for multiple-choice questions. Corpus evidence—whether from learner or native corpora—is used alongside experienced question writers' intuitions about what learners can be expected to know at a certain level, so is not considered to replace human question writers in the test writing process.

However, the reason why it has not been widely used so far is the mismatch between the construct validity of a test and the teaching. Unless the teaching also follows a corpus-based and thus assuming learning language in and through usage events, the test cannot reflect that assumption.

In this approach, there are two conceivable ways of approaching this issue at hand. Proficiency and final achievement tests will be discussed in turn as examples. First, for proficiency exams, test items can be designed in regard to how many times a construction occurs in a native speaker corpus. With this, a simple inverted correlation is assumed, that is, the higher the frequency count of a construction in a native speaker corpus, the more likely the testee may have learned this at lower stages in their language learning journey, i.e., the more entrenched one can expect the construction to be. This, however, is not a one-size-fits-all solution to the phenomenon at hand. This is because, as Dabrowska (2012) demonstrates, our mental constructions are slightly differently shaped than other speakers', since frequency effects tend to and will inevitably vary from one person and speech community to the other, i.e., socio-economic background, education, willingness to learn among many other factors. Weir (2005) also supports this with their *socio-cognitive framework*, in which the effects of testee demographics and how the test is administered are also mentioned. Another issue with

this approach is that high-frequency count of a construction may not necessarily entail entrenchment. This is because frequency effects, as influential as they may be, are not the only driving force in the process. There are others such as memory, repetition, salience of an item among others (Divjak, 2019).

The frequency-entrenchment paradox can be argued to not be as important for a proficiency test as a final achievement test since the tester expects testees of a proficiency exam to already know certain constructions. In any case, careful material and teaching design can overcome this and exact statistical numbers can be obtained (Madlener, 2016). The individual-grammars-paradox can also be overcome by using a set of sophisticated tools and procedures. For instance, by comparing learner and native speaker corpora for a specific construction and doing a set of statistical analyses can give us insight into what the expected proficiency levels for a construction can be. However, it should be acknowledged that doing this for the test items of a general proficiency exam such as TOEFL or IELTS would be laborious and expensive. In any case, as Biber (2006) shows, examination boards for TOEFL have supported the proficiency exam with corpora for a few decades now.

Second, final achievement tests can also be corpus-driven. Corpora do not only come in the shape native speaker or second language learner corpora, but rather in the shape of a classroom-based corpus (Csomay, 2005). In this case, the textbooks and language materials used in teaching the language can be compiled in a corpus. This corpus would also need to have a spoken component of the language used in the classroom, i.e., how many times did the teacher use *the Xer the Yer* construction and so on. This can be laborious, indeed. Thus, following Barker (2012), CEFR levels can be used to do a norming of the item-frequency and knowledge needed for each level (www.EnglishProfile.org) in accordance with the teaching material. The example they provide of the English Grammar Profile can be argued to include constructional knowledge, i.e., *may well*, *may as well*, and *may X*. Thus, test designers could, in theory, use the language teaching material and CEFR levels to approximate the target constructions to be tested in each level by using their relevant frequency information. But, there is indeed more research needed to specifically pinpoint them.

Reimagining rubrics

Studies suggest that while assessing the productive components, e.g., writing or speaking, assessors struggle with separating grammar from lexis (Ruegg, 2015; Ruegg, Fritz, & Holland, 2011). Römer (2017) demonstrates that the rubrics for the speaking sections of tests can be improved to align the construct of language proficiency with usage-based approaches, especially usage-based construction grammar. By combining categories for grammar and vocabulary under lexicogrammar, assessors can rate the two entities with an

emphasis on the importance of lexis-grammar inseparability by including the use of “nativelike formulaic sequences” (phrases such as ‘be unaware of’ or ‘on the other hand’) as an aspect of the highest level of proficiency” (Römer, 2017).

In the same vein, it is possible to apply this to writing. For instance, Kyle and Crossley (2017) use an automated tool (Kyle, 2016) to assess and compare student essays with a series of statistical analyses against the ratings rated by humans using the data they obtain from the tool. The tool was created based on the COCA. Combining verb argument constructions (VACs) and the type/token frequencies for these combinations, they demonstrate that human raters are sensitive to the type/token frequencies of these combinations and they explain that those essays that “include weakly associated verb–VAC combinations earning lower quality scores and in essays that include strongly associated verb–VAC combinations earning higher quality scores” (Kyle & Crossley, 2017). Weakly associated combinations refer to a verb that typically would not occur in a specific combination with a construction, e.g., I agree the statement (example taken from (Kyle & Crossley, 2017)). There is also evidence that L2 learners’ usage of verb-VAC combinations reflect that of native speakers (Römer et al., 2014).

Conclusion

In this article, we revisited construct validity from usage-based construction grammar and applied its central tenets to the construct of language proficiency in language testing and evaluation. While previous approaches, such as the communicative language testing, are compatible with some of the assumptions of construction grammar, i.e., that language is learned in and through communicative events. However, these approaches do not necessarily merge lexis and grammar. Consequently, this leads to a shift in how we understand construct validity in LTE. Namely, if learning a language is defined as follows:

Language learning takes place with general domain cognitive abilities in social contexts, i.e., pattern recognition, abstraction, generalization, perception, and keeping a record of type and token frequencies to name a few. Linguistic knowledge consists of constructions, i.e., form-meaning pairings, and the mental organization of constructions arises from usage and frequency effects with mechanisms such as entrenchment and statistical preemption. Because speakers strive to speak like everyone else, conventionalized constructions are preferred over unconventionalized constructions.

More specifically, if we define the construct of knowing a language as follows:

In a usage-based approach, “To know a language, one must know its schemata” (Turner, 2018) on the basis of ambient language, i.e., input, and must know form-meaning pairings with their pragmatic and discourse functions. Furthermore, a speaker must have entrenched the transitional probabilities of constructions, i.e., what comes after *give birth* or whether it is

heavy rain or *strong rain*. As Hilpert (2014) summarizes "knowledge of language is to be modeled as knowledge of constructions, and nothing else in addition" since constructions already unify every faculty of language, i.e., form, meaning, discourse, pragmatics, frequency and so on, that was previously separated in previous linguistic traditions (J. R. Taylor, 2012). In conclusion, knowing a language is knowing "constructions all the way down" (Goldberg, 2010).

We advocated for a merging of lexis and grammar in any subsection of a test, e.g., reading, writing and so on, and suggest that social context is important for constructions because constructions carry meaning themselves. We discussed that because frequency effects are at work in language learning, regardless of L1 or L2, construct validity can be ensured by creating test items that are corpus-driven. Finally, rubrics are also affected by this shift in that they should not separate grammar knowledge from vocabulary knowledge in productive sections of a test but rather adopt a unibody rating approach, i.e., lexicogrammatical knowledge. Clearly, there is no need to reinvent the wheel. By attending to new empirical research and a teaching that is informed by this research, it is possible to strive for a more cognitively plausible construct validity. Thus, adopting the view presented in this paper in LTE and specifically rating scales or rubrics can bring about a greater construct validity.

References

- (Ed.), C. of E. (2020). *Common European framework of reference for languages: Learning, teaching, assessment; companion volume*. Strasbourg: Council of Europe Publishing.
- Alderson, J. C., & Kremmel, B. (2013). Re-examining the content validation of a grammar test: The (im)possibility of distinguishing vocabulary and structural knowledge. *Language Testing*, 30(4), 535–556. <https://doi.org/https://doi.org/10.1177/0265532213489568>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Barker, F. (2012). Corpus-Based Testing. In *The Encyclopedia of Applied Linguistics* (p. wbeal0263) (In C. A. C). Oxford: Blackwell. <https://doi.org/https://doi.org/10.1002/9781405198431.wbeal0263>
- Beckner, C., & Bybee, J. (2009). A Usage-Based Account of Constituency and Reanalysis. *Language Learning*, 59, 27–46. <https://doi.org/https://doi.org/10.1111/j.1467-9922.2009.00534.x>
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. (J. Benjamins, Ed.). Amsterdam ; Philadelphia.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14(3), 275–311. <https://doi.org/https://doi.org/10.1075/ijcl.14.3.08bib>
- Boas, H. (2003). Towards a lexical-constructional account of the locative alternation. . Presented at the . Retrieved from. In *Proceedings of the 2001 Western Conference on Linguistics*. Western Conference on Linguistics, University of Texas. Retrieved from https://sites.la.utexas.edu/hcb/files/2011/02/Boas2003a_Locative_Alternation.pdf
- Brown, J. (2000). What is construct validity? *JALT Testing & Evaluation SIG Newsletter*, 4(2). Retrieved from <https://hosted.jalt.org/test/PDF/Brown8.pdf>

- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approach to second language teaching and setting. *Applied Linguistics*, 1, 1–47.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT press.
- Csomay, E. (2005). Interactivity in university classrooms: A corpus-based perspective. In *ICAME 26 & AAACL 6 Conference*. Retrieved from <http://robfelty.com/design/archives/corpusconference/abstracts/abstract18.pdf>
- Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, 2(3), 219–253. <https://doi.org/https://doi.org/10.1075/lab.2.3.01dab>
- Divjak, D. (2019). *Frequency in Language*. Cambridge University Press. <https://doi.org/https://doi.org/10.1017/9781316084410>
- Fulcher, G. (2000). The 'communicative' legacy in language testing. *System*, 28(4), 483–497.
- Goldberg, A. E. (1992). The inherent semantics of argument structure: The case of the English ditransitive construction. *Cognitive Linguistics*, 3(1), 37–74. <https://doi.org/https://doi.org/10.1515/cogl.1992.3.1.37>
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago and London: University of California.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, A. E. (2010). Constructions: A new theoretical approach to language. In L. Hoffmann (Ed.), *Sprachwissenschaft* (pp. 717–729). Berlin, New York: DE GRUYTER. <https://doi.org/https://doi.org/10.1515/9783110226300.6.717>
- Goldberg, A. E. (2019). *Explain me this: Creativity, competition, and the partial productivity of constructions / Adele E Goldberg*. Princeton, New Jersey: Princeton University Press.
- Gries, S. T., & Stefanowitsch, A. (2004). Extending collocation analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics*, 9(1), 97–129. <https://doi.org/https://doi.org/10.1075/ijcl.9.1.06gri>
- Harding, L. (2014). Communicative Language Testing: Current Issues and Future Research. *Language Assessment Quarterly*, 11(2), 186–197. <https://doi.org/https://doi.org/10.1080/15434303.2014.895829>
- Herbst, T. (2020). Constructions, generalizations, and the unpredictability of language: Moving towards collocation. *Constructions and Frames*, 12(1), 56–95. <https://doi.org/https://doi.org/10.1075/cf.00035.her>
- Hilpert, M. (2014). *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge University Press.
- Hymes, D. (1972). *On communicative competence*. *Sociolinguistics* (J. B. Prid). Penguin.
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Georgia State University. Retrieved from http://scholarworks.gsu.edu/alesl_diss/35/
- Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, 34(4).
- Leung, C. (2005). Convivial communication: Recontextualizing communicative competence. *International Journal of Applied Linguistics*, 15(2), 119–144.
- Luoma, S. (2004). *Assessing Speaking (1st ed.)*. Cambridge University Press. <https://doi.org/https://doi.org/10.1017/CBO9780511733017>
- Madlener, K. (2016). Input optimization: Effects of type and token frequency manipulations in instructed second language learning. In H. Behrens & S. Pfänder (Eds.), *Experience Counts:*

- Frequency Effects in Language* (pp. 133–174). De Gruyter.
<https://doi.org/https://doi.org/10.1515/9783110346916-007>
- Mahlberg, M. (2006). Lexical cohesion: Corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics*, 11(3), 363–383.
- McNamara, T. F. (1996). *Measuring second language performance*. Longman Publishing Group.
- Oyinloye, C. A., Adeoye, A., Fatimayin, F., Osikomaiya, M. O., & Fatola, O. L. (2020). The impact of English language proficiency testing on the pronunciation performance of undergraduates in South-West, Nigeria. *Educational Research and Reviews*, 15(9), 530–535.
<https://doi.org/https://doi.org/10.5897/ERR2020.4016>
- Perek, F. (2015). *Argument Structure in Usage-Based Construction Grammar (Vol. 17)*. Amsterdam: John Benjamins Publishing Company. <https://doi.org/https://doi.org/10.1075/cal.17>
- Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review*, 18, 1–2. <https://doi.org/https://doi.org/10.1515/tlir.19.1-2.9>
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481.
- Römer, U. (2017). Language assessment and the inseparability of lexis and grammar: Focus on the construct of speaking. *Language Testing*, 34(4).
<https://doi.org/https://doi.org/10.1177/0265532217711431>
- Römer, U., Roberson, A., O'Donnell, M. B., & Ellis, N. C. (2014). Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal*, 38(1), 115–135. <https://doi.org/https://doi.org/10.2478/icame-2014-0006>
- Ruegg, R. (2015). An Experiment in the Ability of Raters to Evaluate Lexis in Writing. *Language in Focus*, 1(1), 38–50. <https://doi.org/https://doi.org/10.1515/lifijal-2015-0003>
- Ruegg, R., Fritz, E., & Holland, J. (2011). Rater Sensitivity to Qualities of Lexis in Writing. *TESOL Quarterly*, 45(1), 63–80.
- Services, E. T. (2011). *TOEFL program history (No. 6)*. Princeton, New Jersey: Educational Testing Services. Retrieved from https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v6.pdf
- Services, E. T. (2021). *Using TOEFL iBT® Test Scores for Selecting International Teaching Assistants (No. 10)*. Princeton, New Jersey: Educational Testing Services. Retrieved from https://www.ets.org/s/toefl/pdf/toefl_ibt_insight_s1v10.pdf
- Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, 24(1), 99–128. <https://doi.org/https://doi.org/10.1177/0265532207071513>
- Shohamy, E. (1995). Performance Assessment in Language Testing. *Annual Review of Applied Linguistics*, 15, 188–211. <https://doi.org/https://doi.org/10.1017/S026719050002683>
- Shohamy, E. (1996). Competence and performance in language testing. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 136–151). Cambridge ; New York: Cambridge University Press.
- Sinclair, J. (1997). *Corpus, concordance, collocation (4. impr)*. Oxford University Press.
- Sinclair, J. (2014). Trust the Text Language, Corpus and Discourse.
- Skehan, P. (2013). Tasks and language performance assessment. In M. Bygate & M. Swain (Eds.), *Researching Pedagogic Tasks* (pp. 177–195). Routledge.
<https://doi.org/https://doi.org/10.4324/9781315838267-19>
- Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford; New York: Oxford University Press.
- Taylor, L. B. (Ed. . (2011). *Examining speaking: Research and practice in assessing second language speaking*. Cambridge ; New York: Cambridge University Press.
- Timpe, V. (2012). Strategic decoding of sociopragmatic assessment tasks—An exploratory think-aloud validation study. *University of Hawai'i Second Language Studies Paper*, 30(2).

- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Troike, M. S. (2006). *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/s0272263107070301>
- Turner, M. (2018). The Role of Creativity in Multimodal Construction Grammar. *Zeitschrift Für Anglistik Und Amerikanistik*, 66(3), 357–370. <https://doi.org/https://doi.org/10.1515/zaa-2018-0030>
- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.
- Weir, C. J., Vidaković, I., & Galaczi, E. D. (2013). *Measured constructs: A history of Cambridge English examinations, 1913-2012 (Vol. 37)*. Cambridge University Press.