

## **Motion Lexicon: A Corpus-Based Comparison of English Textbooks and University Entrance Exams in Turkey**

**Tan Arda Gedik**

*FAU Erlangen-Nürnberg, MA Student in Linguistics*  
*tangedik@gmail.com*

### **Abstract**

This study analyzed the overlap of motion lexicon, namely manner and path verbs' frequency profiles, in English high school textbooks (9th-12th grade) and English university entrance exams (2010-2019) in Turkey through *AntwordProfiler*, a corpus linguistic tool. The manner verbs were sampled from Levin's study (1993) while the path verbs were gathered from Talmy's book (2001). The frequency of motion verbs in official teaching materials was compared with their frequency in exam materials using SPSS. The results indicate that the mismatch of motion verbs between the textbook and exam corpora is statistically significant in terms of manner verb frequency levels ( $p < .000$ ). While path verbs scored, on average, higher in descriptive statistics in the textbook corpus, there was no statistical significance observed. The findings suggest that whenever the students take English exam, they may be more likely to be under a higher cognitive load and may be forced to develop the negative backwash effect since what is taught is not tested. This, consequently, raises concerns regarding the content validity of exams and other issues related to the reliability and validity of the national English exams. The findings of this study have implications for material developers and test takers.

*Keywords: Corpus Linguistics, Motion Lexicon, Testing, English Language Teaching*

## **Introduction**

As many researchers point out, there seems to be an intertwined connection between L2 success and the quality of the teaching material (Allen, 2008). That is, how well the materials feed the learners in terms of lexical and syntactic input. When typological differences among languages are added to this, the responsibility of teaching materials becomes even more salient (Caluianu, 2016). Clearly, the effectiveness of teaching methods, the design of the activities, and other psychological factors at the moment of learning can also be extended to account for L2 success and its entanglement with teaching material. Nevertheless, this study only focuses on one aspect of the intermediate connection. Lexical sophistication, being an important indicator of early L2 success (Bardel, Gundmunson & Lindqvist, 2012), is one topic that needs attention as its nature can be claimed to be quite similar to motion verb frequency profiling and its influence on proficiency. One such typological difference that becomes prominent in corpus-based studies is Talmy's (2003) satellite/verb-framed language typology. In short, satellite languages (like English or German) prefer encoding motion information into satellites (e.g., particles) while verb-framed languages (like Turkish or French) encode motion information in additional syntactic clauses (e.g., converbials, adverbs etc.). The relationship between how motion is encoded in a satellite or a verb-framed language and lexical diversity has been studied by many scholars in the field (Capelle, 2012; Pavlenko, 2010). Caluianu (2016) indicates that the teaching of English manner verbs to speakers of a verb-framed language dramatically heightens the use of target language constructions and results in "tighter clausal packaging" (Caluianu, 2016, p. 81). A number of studies have studied the overlap ratios of various lexical or syntactic indices among teaching and exam materials and have reported a significant mismatch of the scrutinized indices between the two corpora (Underwood, 2010; Nur & Islam, 2015; Tai & Chen, 2015). However, to the researcher's knowledge, no study has scrutinized the intricate relationship between the motion lexicon of teaching materials and their successive exam materials to this day. Therefore, building on Talmy's (2003) satellite and verb-framed language typology and in the relationship between teaching materials and L2 success, this corpus-based study aims to analyze the overlap of motion lexicon between the English textbooks used in Turkish high schools and the English university entrance exams and its implications.

Building on Gedik and Kolsal's (2020) study, this study deepens the literature to bridge the gap and provide further insight into the following areas: (i) validity and reliability of the English university entrance exams administered in Turkey, (ii) further evidence for the reasons of Gedik and Kolsal's (2020) statistical gap of lexical sophistication/diversity and syntactic complexity between the teaching and exam materials.

## **Literature Review**

### *Typological Differences and Lexicon Size*

Talmy's typological approach to languages has been widely used in many studies. The theory basically states two possible encoding mechanism techniques for languages: (a) satellite languages (such as English, Russian, or German), (b) verb-framed languages (such as Turkish, French, or Spanish). The main difference between these two encoding strategies lies in where languages prefer to encode manner and path information in a clause. Satellite languages tend to encode the manner of motion in the verb itself and give path of motion in satellites (particles). Verb framed languages, however, show a tendency to include the path of motion embedded in the verb and use extra syntactic packaging to convey the manner of motion (e.g., adverbials). The following examples illustrate these typological differences between English and Turkish:

(1) Eve hızlıca koşarak girdi.

Into the house quickly with a running manner entered.

S/he entered the house running quickly.

(2) They sprinted into the house.

As seen in the above examples, Turkish, and other verb framed languages, require further syntactic packaging to indicate the same semantic image. Özçalışkan and Slobin (2003) state that these adverbial or converbial constructions require a heightened cognitive load during production compared to the counterpart speakers of satellite languages. Furthermore, it has been suggested that this typological difference points to an inevitable variation in lexical diversity levels (Verkerk, 2013). In the light of these, it is probable to suggest that materials that are prepared in a satellite language by the speakers of path languages

may lack lexical diversity. Indeed, Capelle (2012) identified this difference in motion lexicon in English-French translations. Similarly, Özçalışkan and Slobin (2003) also pinpointed the same mismatch in motion lexicon among written and oral narratives of English and Turkish. Furthermore, Guitierrez-Clellen and Hofstetter's study (1994) suggest that the syntactic complexity levels increased as the participants encoded more information of time, manner of motion, reason, and place in their narratives. With this in mind, Spanish being a verb-framed language, it can be proposed that if verb-framed languages wish to convey the same amount of information, they would have to pack more clauses in sentences compared to satellite languages.

One thing to note, however, is that these specifications have been valid for productive skills but have not been implied for receptive skills like reading and listening. On the other hand, if verb-framed languages utilize more syntactic clauses per sentence to convey information then it can be argued that a heightened level of syntactic complexity indices will result in a cognitive load on the readers or listeners. Whether this can be applied to teaching and exam materials that are prepared in a satellite language (English) by speakers of a verb-framed language (Turkish) requires further research. Nonetheless, based on what Gedik and Kolsal (2020) report in regard to the lexical sophistication, diversity and syntactic complexity levels, this can be hypothesized to be applicable as the creators of both of these corpora are native speakers of Turkish and the native speakers of English (if there are any) have little say on the corpora. Otherwise, the corpora would have corresponded to one another in terms of indices under scrutiny in the study mentioned above. Therefore, the present study assumes that both the English teaching and English exam materials are prepared by native speakers of Turkish.

### *English Language Teaching and Testing in Turkey*

English language teaching (ELT) is an important area that should be handled with care. Testing, on the other hand, can be thought of as the complementary or binary companion to any field of teaching. As such, some researchers propose that ELT has been a field that has not received enough attention due to various political or practical reasons and has experienced a constant change in teaching and testing and evaluation techniques (Kırkgöz, 2007; Hatipoğlu, 2016). As Choi (2008) puts it, the success and problems of ELT programs and the materials are coupled together and interact with one another. This is applicable in the case of government-imposed books, which allow for little to no

modification (Sheldon, 1988). Under a continual change, the curriculum and the teaching materials would have to adapt themselves, with the same change observed on the examinations and tests.

Turkey has a highly centralized mechanism of testing. That is, Ölçme, Seçme, Yeterlilik Merkezi-ÖSYM- (Measuring, Selection, and Placement Center) is the sole responsible body for the administration, preparation, and scoring of these nationwide university entrance exams of which English is also a part. As Gençoğlu (2017) reports, each year, the Ministry of National Education (MEB) hands out textbooks prepared by MEB free of charge to demonstrate across Turkey. These textbooks are then employed to tutor students for the upcoming university entrance exam. Nonetheless, when one notes the continual change the curriculum and teaching/testing techniques go through, it is possible to suggest that there will inevitably exist discrepancies between the teaching and exam materials. Rightfully, Gedik and Kolsal's study (2020) report on these discrepancies and suggest that unless the two separate material creating teams listen to and collaborate with one another, the results will be devastating for those students who use the government-imposed books due to various factors such as socio-economy, and inaccessibility, leading to a malformed understanding of what learning and communicating in English is. This pattern of mismatch has also been observed in other contexts around the world (Underwood, 2010; Nur & Islam, 2015; Tai & Chen, 2015), thus indicating that this is not just a local issue, but a global issue.

High stakes exams, having become widely popular across the globe (Choi, 2008), impact a number of things. They are known to influence a variety of other domains such as a downsized curriculum (Cheng, 2005), changes in teaching methods (Wall, 2005) and in learning styles (Shih, 2009). They can also generate other outcomes. All of these outcomes have been defined as the washback (or backwash as he refers to it) (Hughes, 1989). Put simply; backwash can be beneficial or devastating, leading to positive or negative backwash effects. These effects have been identified to affect teachers and learners' behaviour (Hawkey, 2006) and teaching materials (Sevimli, 2007). Though the field of backwash effect is far more complex than what is portrayed here, the connection between backwash, teaching materials and learner behavior is sufficient to construct the framework of the study. Moreover, as Hatipoğlu (2016) demonstrates, English university entrance exams in Turkey leave a negative backwash effect on learners' perception of what L2 should be, since they lack productive sections. When what is taught is not tested, as is the case in Turkey for vocabulary items and grammar structures (Gedik & Kolsal, 2020), it leads to a

distorted image of L2 or even a deterioration of certain skills learned during school years. The students may also inquire whether they learn English to communicate and collaborate or to pass the university entrance exam.

### *Lexical Sophistication, L2 Success, and its Connection to the Motion Lexicon Profiling*

Lexical variation can be approached in two ways: (i) lexical sophistication, (ii) lexical diversity. Although many of the studies mentioned above utilized lexical diversity as their starting point, because this study already expands on the given literature of the same two corpora, determining the variation of motion lexicon between the corpora is conducted by examining overlap ratios based on manner and path verbs lists. This implies that lexical sophistication as a concept fits the nature of this paper more suitably than lexical diversity.

Lexical sophistication, in simple terms, is how much a corpus overlaps with the first one thousand words (K1), the first two thousand words (K2), and academic word lists (AWL) in English. Measuring lexical sophistication, however, has been a debated issue. Laufer and Nation (1995) suggest that calculating advanced/sophisticated words in a text is one way to measure it.

Yet, the question of what is and is not sophisticated arises. Bardel, Gundmunson, and Lindqvist (2012) propose that employing words frequency profiles, that is how frequently it is used in a corpus, is an approach to overcome this question. Other researchers such as Hyltenstam (1988), Laufer and Nation (1995), Read (2000), Vermeer (2004) also agree with this approach. Bardel, Gundmunson, and Lindqvist (2012) advise that in order to determine a text's sophistication levels, it is required to calculate the ratio of advanced words (based on K1, K2, and AWL). The researchers also argue that using lexical sophistication is a salient tool when it comes to pinpointing non-native speakers' L2 vocabulary size command and proficiency and that it can establish a base knowledge for L2 testing (Bardel, Gundmunson & Lindqvist, 2012). This argument not only covers the production based corpora but also the teaching and exam materials. Based on this, the following equation can be proposed: as the students work their way through low-frequency words (sophisticated words), their proficiency and command of vocabulary size in L2 grows. This argument, although not covered by literature to the researcher's knowledge, can be extended to the motion lexicon, and specifically motion lexicon frequency profiling, as they are, in nature, quite similar to one another. That is, if

the students are exposed to, depending on their native language's encoding preferences, manner or path verbs, then one can claim that the overall proficiency and command of their L2 will grow as well as their motion lexicon size will improve.

Identifying lexical sophistication levels is an automated process, which is called lexical frequency profiling. First carried out by Laufer and Nation in 1995, since then, numerous researchers have used the same technique to discover lexical sophistication levels of any corpus. *AntwordProfiler* (Anthony, 2014) is one automated software that can detect lexical sophistication levels based on any .txt file created by the user. It has been used in many studies and has been proven to be reliable (Kwary, Artha & Amalia, 2018; Du, 2019; Beauchamp and Constantinou, 2020). In the light of this literature, the researcher compiled the manner and path of motion verbs using Levin's (1993) and Talmy's study (2003) and created two separate .txt files for each group to run through *AntwordProfiler* against the corpora. This would enable the researcher to profile the motion lexicon frequency levels of both the textbook and exam corpora.

The present study deals with the following research question:

(i) Is there a statistically significant mismatch of the manner and path of motion verbs between the textbook corpus and the exam corpus?

## **Methodology**

The study employed a number of procedures to answer the research question as reliably as possible. The corpora in this study was retrieved from Gedik and Kolsal's study (2020), who had already conducted a variety of steps to clean the corpora of any mistakes that would potentially skew with the results. The textbooks were selected from each year in high schools (9th-12th grade) and were released by these publishing houses; (MEB) Relearn, Teenwise, Progress: 9th grade; Count Me In, Gizem: 10th grade; Sunshine, Silverlining: 11th grade; Count Me In: 12th as well as the workbooks released for the textbooks. These textbooks are still in use by the MEB at high schools. While exams were sampled from years 2010-2019, all of which produced and administered by ÖSYM The textbook corpus contained a total token number of 301.255. The ten exams, on the other hand, were all created by ÖSYM and had a total token number of 66.913.

While assembling motion verbs, Levin's (1993) study was selected as the base for manner verbs. Path verbs, on the other hand, were collected from Talmy's (2001) study. The total number of manner verbs was 227, while it was 20 for path verbs. During the study, the following procedures were followed: (a) copy and paste each verb on a separate .txt file, (b) check both files for any typos and correct if there is any, (c) use AntwordProfiler (2014) to check both corpora for overlap ratios, (d) upload the corpora on SketchEngine to check for concordances in the verb category, (e) delete manner and path verbs that are not used in a verb position, (f) employ the cleaned corpora for another round of overlap test, (g) import the .csv files into SPSS, (h) get results for descriptive statistics and the independent t-test, (i) interpret the results.

## Results

### *Manner Verbs*

To scrutinize the difference in overlap ratios in manner verbs among both corpora, SPSS was utilized. The following results met the assumptions of equal variance and normality. The textbook corpus displayed a mismatch in the ratio of manner verbs compared to the exam corpus. As illustrated in Figure 1, the textbook corpus achieved a higher score (*MManVer*: .6593, *SDManVer*; .13367) in its utilization of manner verbs than the exam corpus (*MManVer*: .4086, *SDManVer*; .05336). This difference in means was also further demonstrated by the independent t-samples test. The results for that test reported that the difference between the corpora in their use of manner verbs was statistically significant (*ManVer*:  $p < .000$ ). Descriptive statistics suggest that manner verbs were used more frequently in the textbook corpus. Cohen's *d* (Cohen, 2013) gives insight into the effect size of the differences between the two corpora regarding manner verb use. The effect size for the differences was found to be 2.4%. This percentage indicates the amplitude of the gap among the corpora, and according to McLeod (2019), although there is a statistically significant difference, this difference is trivial.



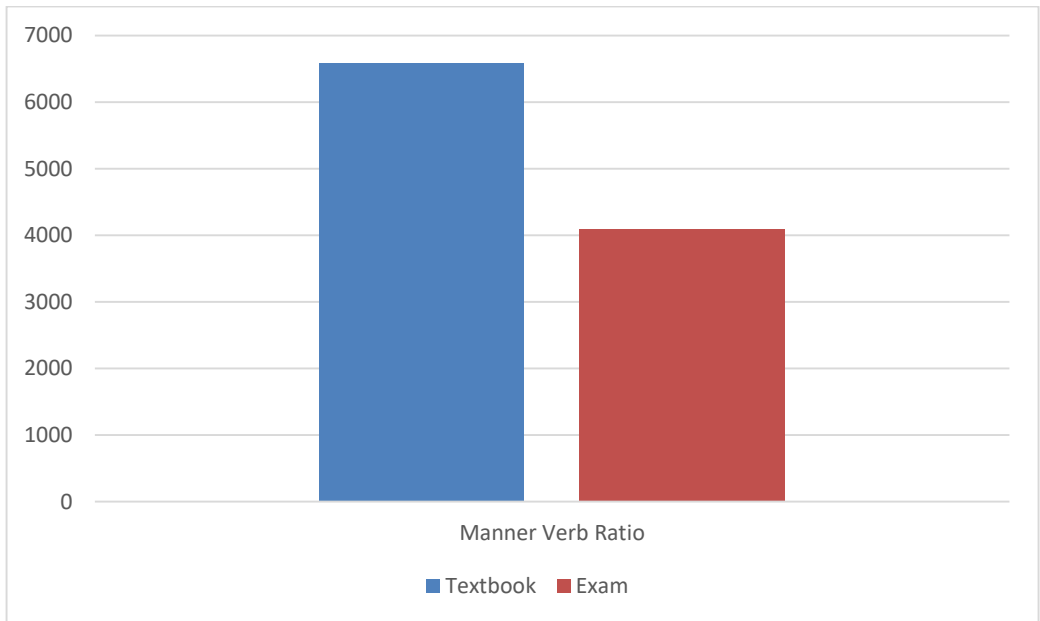
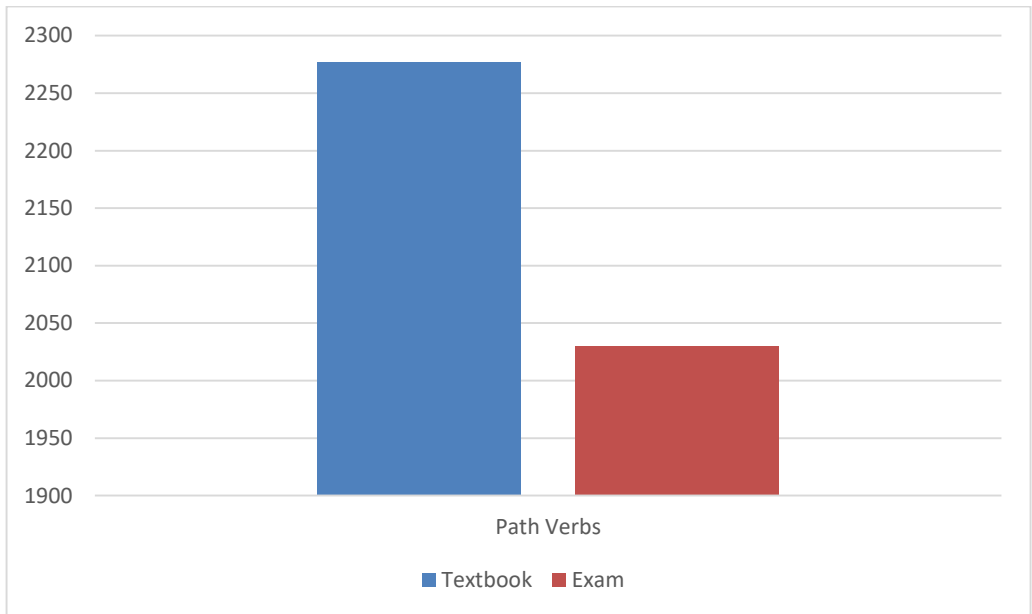


Figure 1. Manner Verb Overlap

### Path Verbs

Unlike manner verbs, the results for path verb ratios displayed no statistical difference ( $PatVer: p > .05$ ). It is evident from the descriptive statistics that the textbook corpus ( $MPatVer: .2277$ ,  $SDPatVer: .04419$ ) is ever so slightly richer (or more sophisticated) in terms of path verbs compared to the exam corpus ( $MPatVer: .2030$ ,  $SDPatVer: .08667$ ). Nevertheless, no statistical significance was observed ( $p = .383$ ) and Cohen's  $d$  (Cohen, 2013) effect size suggests that the amplitude of the gap between the corpora was 3.5%, once again, pointing to a trivial difference in between. Figure 2 illustrates the descriptive results of path verbs in the corpora. The implications of these findings are discussed in the next section.



*Figure2.* Path Overlap

## Discussion and Conclusion

This research paper scrutinized the overlap ratios of motion lexicon, namely manner and path verbs' frequencies, in two corpora in Turkey. These corpora were English high school textbooks between 9th and 12th grade and their accompanying workbooks which are currently in use and national English university entrance exams that were administered between the years 2010-2019.

Descriptive statistics suggest, whatever the lexical or syntactic unit under scrutiny was, the textbook corpus demonstrated richer motion lexicon levels compared to the exam corpus. This difference in levels may have been caused by the gap in token numbers across the corpora, nevertheless, because downscaling the sample size of one of the corpora might not give the best insight into the current situation of English language teaching, and subsequently, English language testing and evaluation in Turkey, it was simply impossible. Therefore, it is important to consider this limitation in further research studies. Nonetheless, as for the frequency of manner verbs in the textbook corpus, it displayed a statistically significant mismatch when compared to the exam corpus. As for path verbs, there was no statistically significant mismatch, even though descriptive statistics show greater results in the textbook corpus.

Practically interpreting these results in the light of previous literature is quite important for applied linguists, teachers, textbook preparation teams, and examination offices in Turkey. For once, the typological difference between the two languages, namely English and Turkish, is an important point to keep in mind. This difference, that is how the two languages encode motion information and how this influences overall lexical diversity levels of a language and cognitive load (Özçalışkan & Slobin, 2003) manifests itself in the violation of content validity of exams. The first evidence lies in the descriptive statistics. Even if textbooks teach students manner verbs, which is what teachers would want if they want their students to achieve native-like command of any L2, unless these manner verbs are recycled and tested within the exam material, they will not stick with the students. This will inevitably lead to a shrinkage of the motion lexicon and consequently a loss of overall L2 proficiency. Research done by Babanoğlu in 2018 proves the exact same idea, where the researcher found that the native speakers of Turkish always scored lower in terms of motion lexicon compared to the native speakers of German in English essays (Babanoğlu, 2018). Turkey, as previously mentioned, teaches English to students to pass the exam, not to communicate. In order for positive backwash effects to be activated, what we teach and what we test across these seemingly disconnected materials need to attune to one another. Only then can we speak of reliability, validity, and content validity being present in the exams. Another point is, if students are given the chance to package more meaning in a sentence in their classroom environment for four years (meaning lower cognitive load/lower syntactic (item) packaging) then during the exam, when they are expected to perform a higher load of cognitive load due to higher syntactic packaging is not feasible. In other words, in class, students may have been taught how to explain someone running at full speed over a short distance using the manner verb “sprint”. But, if the students are expected to decode “they ran away very fast for a couple hundred meters” in the exam, then clearly, this will take students more time to process. Time may not seem important at first glance, but these minor differences accumulate and in a setting where time tick tocks against the exam taker, it becomes a salient part. Of course, Özçalışkan and Slobin (2003) mention that this cognitive load might be prevalent in productive skills. And in ÖSYM exams, there are no production based sections. However, we still know that mean length of sentence and mean length of T-units, which increase as the syntactic item packaging lowers, will put the brain under a cognitive load, even for receptive skills. This could be one of the explanations as to why the exam corpus is syntactically more complex

in Gedik and Kolsal's study (2020).

If we want to claim our exams are valid and reliable, we first need to handle the issues of content validity with care. Students come from a variety of socio-economic backgrounds. With this in mind, their previous exposure to reading long paragraphs and consequently possessing a wide range of lexicon is predetermined by their background. This can be suggested as one obstacle on the way to achieving equal grounds. Previously, it was thought that the exams were fine pieces of work that strived to push students to the native-like command of L2. However, with the present study, it seems as if exams create just as negative backwash and a ground for inequality as textbooks, if they lack content validity. This negative backwash effect is rooted in how students are taught a larger sized motion lexicon, only to be forgotten later on, and not tested in the exams. As an implication of this mismatch across the two corpora, students will inevitably be under a heavier cognitive load when taking the exam, both because the exams can be claimed to have lower syntactic packaging, leading to higher mean length of sentence and T-unit levels, and also because of already proven mismatch in lexical sophistication, diversity, and syntactic complexity (Gedik & Kolsal, 2020). In order to overcome these obstacles which lead to raise concerns about the (content) validity, and reliability of the exams, both ends of the textbook and exam preparation teams must convene and discuss the content they teach and test.

Nonetheless, this study has a number of limitations. Firstly, the corpora are small in number and are not close in terms of token. Secondly, this study does mere quantitative analysis and may not give the entire insight into the issue of motion lexicon overlap between the corpora. Future studies can focus on collecting more samples for the corpora and analyze the issue of motion lexicon qualitatively to examine other syntactic properties of the issue.

### **Acknowledgments**

I would like to thank Xiaoli Yu, Halide İslamoğlu, Fikriye Beyza Dilbaz, and Yağmur Su Kolsal at Middle East Technical University for coming together and compiling these corpora back in 2019. Without their initial help, this research study would not be here.

## References

- Allen, H. W. (2008). Textbook materials and foreign language teaching: Perspectives from the classroom. *The NECTFL Review*, 62, 5-28.
- Anthony, L. (2014). AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- BABANOĞLU, M. P. (2018). Motion Verbs in Learner Corpora. *Gaziantep University Journal of Social Sciences*, 17(1).
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects Of Lexical Sophistication In Advanced Learners' Oral Production: Vocabulary Acquisition and Use in L2 French and Italian. *Studies in Second Language Acquisition*, 34(2), 269-290. doi:10.1017/S0272263112000058
- Beauchamp, D. & Constantinou, F. (2020). Using corpus linguistic tools to identify instances of low linguistic accessibility in tests. *Research Matters: A Cambridge Assessment publication*, 29, 10-16.
- Cappelle, B. (2012). English is less rich in manner-of-motion verbs when translated from French. *Across Languages and Cultures*, 13(2), 173-195.
- Cheng, L. (2005). *Changing Language Teaching Through Language Testing: A washback study*. Cambridge, CUP.
- Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39-62.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Du, W. (2019). Analysis on the Development of Lexical Complexity in Chinese Science Students' English Writing. *Noble International Journal of Social Sciences Research*, 4(7), 116-120.
- Gedik, T. A., & Kolsal, Y. S. (2020). A Corpus-based Analysis of High School English Textbooks and English University Entrance Exams in Turkey. DOI: 10.13140/RG.2.2.24530.99527
- Gençoğlu, C. (2017, October). *REPUBLIC OF TURKEY MINISTRY OF NATIONAL EDUCATION*. COMCEC, Ankara, Turkey.
- Gutierrez-Clellen, V. F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech, Language, and Hearing Research*, 37(3), 645-654.
- Hatipoğlu, Ç. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, 232, 136- 144.
- Hawkey, R. (2006). *Impact theory and practice: Studies of the IELTS test and Progetto Lingue 2000*. Cambridge: Cambridge University Press.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual & Multicultural Development*, 9(1-2), 67-84.

- Kirkgoz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *RELC Journal*, 38(2), 216-228.
- Kwary, Artha & Amalia (2018). Lexical Word-Class Distributions in Research Articles of Four Subject Areas. *Studies about Languages*, 33, 108-118.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16(3), 307-322.
- Levin, B. (1993). *English Verb Classes and Alternations*. University of Chicago Press.
- McLeod, S. A. (2019, July 10). What does effect size tell you? Simply psychology: <https://www.simplypsychology.org/effect-size.html>
- Nur, S., & Islam, M. (2018). The (Dis) Connection between Secondary English Education Assessment Policy and Practice: Insights from Bangladesh. *International Journal of English Language Education*, 6(1), 100-132.
- Özçalışkan, Ş., & Slobin, D. I. (2003). Codability effects on the expression of manner of motion in Turkish and English. In *Studies in Turkish Linguistics*. Istanbul: Bogaziçi.
- Pavlenko, A. (2010). Verbs of motion in L1 Russian of Russian-English bilinguals. *Bilingualism*, 13(1), 49.
- Read, J. (2000). *Assessing vocabulary*. New York: Cambridge University Press.
- Sevimli, S. (2007). The washback effect of foreign language foreign language component of the university entrance examination on the teaching and learning context of English language groups in secondary education (A case study). Unpublished MA Thesis, Gaziantep University.
- Sheldon, L. E. (1998). Evaluating ELT textbooks and materials. *ELT Journal* 42(4), 237-246
- Shih, C. (2009). How tests change teaching: A model for reference. *English Teaching: Practice and Critique*, 8 (2), 188-206.
- Tai, S., & Chen, H.-J.. (2015). Are teachers test-oriented? A comparative corpus-based analysis of the English entrance exam and junior high school English textbooks. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference*, Padova, Italy (pp. 518-522). Dublin: Research-publishing.net. <http://dx.doi.org/10.14705/rpnet.2015.000386>
- Talmy, L. (2003). *Toward a Cognitive Semantics. Volume 2: Typology and Process in Concept Structuring*. Cambridge, MA: MIT Press
- Underwood, P. (2010). A comparative analysis of MEXT English reading textbooks and Japan's National Center Test. *RELC Journal*, 41(2), 165-182.
- Verkerk, A. (2013). Scramble, scurry and dash: The correlation between motion event encoding and manner verb lexicon size in Indo-European. *Language Dynamics and Change*, 3(2), 169-217.
- Vermeer, A. (2004). Vocabulary size in Dutch L1 and L2 children. *Vocabulary in a second language: Selection, acquisition, and testing*, 173-189.

Wall, D. (2005). The impact of high-stakes testing on classroom teaching: A case study using insights from testing and innovation theory. *In Studies in Language Testing, Vol. 22.* Cambridge: Cambridge University Press.

