

DRAGANA BOŽIĆ LENARD¹

JOSIP JURAJ STROSSMAYER UNIVERSITY OF OSIJEK
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND
INFORMATION TECHNOLOGY
DEPARTMENT OF CORE COURSES

GABRIELA CHMELÍKOVÁ²

SLOVAK UNIVERSITY OF TECHNOLOGY
FACULTY OF MATERIALS SCIENCE AND TECHNOLOGY IN TRNAVA
CENTRE OF LANGUAGES, HUMANITIES AND ACADEMIC SPORTS

APPLYING PREDICTIVE ANALYTICS IN ESP COURSES BASED ON STUDENTS' WRITING

ABSTRACT. Fierce competition of student admission to strong reputation higher education institutions as well as getting employment in respected companies has built the need to profile potential candidates and select those that best meet one's requirements. Our research aimed to predict students' success potential, i.e. grades achieved in their English for Specific Purposes courses. Total of 292 students studying at the Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Croatia and 150 students studying at the Faculty of Material Science and Technology in Trnava, Slovakia voluntarily participated in the research by submitting their short biographies. The biographies were analyzed with the software for computational analysis (LIWC) whose output, in the form of raw numbers, was entered in an application specifically designed for this purpose. Based on the input data, the application calculated students' grades in the aforementioned courses which were later compared to the students' actual grades. The research has proven the application's high efficiency since it correctly predicted students' grades in 75% of the cases and in additional

¹ dragana.bozic@ferit.hr

² gabriela.chmelikova@stuba.sk

This paper was submitted on August 14th, 2019 and accepted for publication at the meeting of the Editorial Board held on September 16th, 2019.

12%, the grades were approximately predicted, i.e. the predicted grade was one grade higher/lower than the actual grade.

KEYWORDS: students' biographies; ESP; final grades; LIWC; application; success prediction.

INTRODUCTION

Regardless of one's field of study, every university student has to engage oneself in writing tasks ranging from short in-class essays to bachelor/master theses. In spite of being provided with professional guidance and classes teaching students how to write effectively, writing remains a daunting task for the majority of students because they are aware that they will be evaluated based on their writing. Writing is indeed an important component of one's academic and professional career and this arduous and multifaceted process of mastering the skill requires a coordination of numerous cognitive and knowledge skills such as linguistic abilities, discourse awareness, sociocultural knowledge, management strategies, etc. (Kellogg and Whiteford, 2009). There is a growing tendency of higher education institutions and companies requiring admission essays, in addition to one's CV, irrespective of a candidate's field of study/work (Walfish and Moreira, 2005; Rabow, Wrubel and Rachel, 2009; Turner and Nicholson, 2011). The underlying concept is for the management to get a rough idea on potential candidates; learn about their background, previous experience, emotional states, and current concerns. Given the time-consuming and space-challenging nature of interviews, self-introductory essays are an excellent profiling opportunity for the management to make wise and educated decisions.

THEORETICAL BACKGROUND

The use of data analytics for the purpose of identifying patterns and developing statistical models to predict one's success in higher education is still in its infancy. However, the use of predictive analytics to forecast one's academic outcome may be very informative and useful for management and students' advisors. According to Picciano (2014), the use of predictive analytics as a foundation for

evidence-based decision-making meets institutional management and stakeholders' needs. Predictive analytics used in the context of higher education intersects with Tinto and Pusser's (2006) institutional model comprising the core elements of feedback, support and involvement as illustrated in Figure 1.

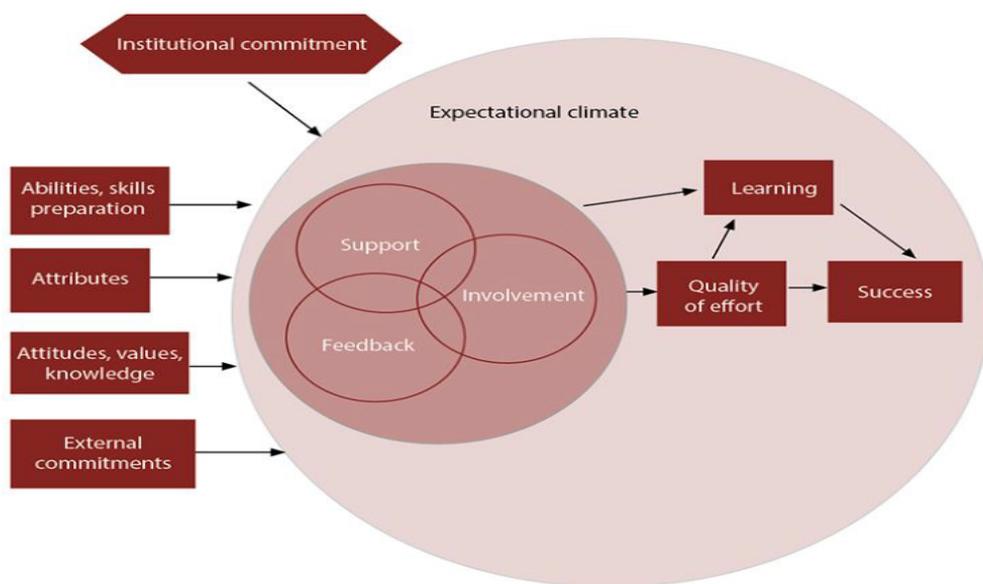


FIGURE 1: TINTO AND PUSSEY'S TRIAD

Predictive analytics creates opportunities to provide timely feedback to management, enhance support due to performance monitoring and increase involvement of all engaged parties. Taking these steps in higher education can possibly lead to increasing students' success.

Computerized text analyses have revolutionized essays scoring and assessment (Elliot, 2003; Burstein, Marcu and Knight 2003; Burstein, Chodorow and Leacock, 2004; Deane, 2013). Automatic natural language processing tools aim to recognize errors, speed up the assessment process, and provide diagnostic feedback. However, those tools focus on students' linguistic knowledge and (morpho)syntax, and the current labor market as well as competitive higher education institutions require more. Consequently, an ever-increasing body of research (Crossley and McNamara, 2012; Deane, 2013; Robinson, Navea and Ickes, 2013; Pennebaker et al., 2014; McNamara et al., 2015; Crossley, Muldner and McNamara,

2016; Kyle and Crossley, 2017; Božić Lenard, Liermann-Zeljok and Ferčec, 2018; Kim and Crossley, 2018; Kim, Crossley and Kyle, 2018) focus on profiling students and predicting their academic success potential based on how and not what they write. Generally, the aforementioned studies show that higher achievers use more complex vocabulary and verbose sentences with fewer verb base forms thus being profiled as thinking in a categorical way which is more sought after by universities and employers. Other indicative categories are positive emotions (Waugh and Fredrickson, 2006) and pronoun *we* (Pennebaker and Chung, 2007) because they are indicators of one's empathy, inclusion, affiliation, and supportive relationship which employers, especially ones organizing their business in teams, value. Other relevant studies will be elaborated on in the respective subchapters. Collectively, these studies highlight the need to take a new direction in students' assessment which will be more time-saving and targeted towards one's needs.

METHODOLOGY

This paper is a follow-up research based on the recommendations put forward by Božić Lenard, Liermann-Zeljok, and Ferčec (2018). The authors conducted the research employing a top-down approach. Their deductive reasoning worked from studying students' essays and categorizing words they believed were related to one's grades into several categories. Upon their computational, statistical, and qualitative analyses, they came up with the hypotheses as to which word category is associated with one's success or the lack of success for that matter in the ESP classes. We replicated a part of the aforementioned research and tested the hypotheses using an application specifically designed for these purposes.

Our research was multi-fold. Firstly, we collected 442 short (200–250 words) biographical essays. Namely, 292 undergraduate students studying at the Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Croatia and 150 undergraduate students studying the Faculty of Material Science and Technology in Trnava, Slovakia voluntarily participated in the research. These two groups of students were chosen because the authors teach at the respective institutions. One of the authors was a one-month research grant holder and spent time at the other author's institution where they jointly conducted the research. We

employed a convenience sampling method and recruited all students enrolled in the English for Specific Purposes classes in the summer semester of the academic year 2018/2019 and 2017/2018 (relates to Croatian students only). Not all students decided to submit their essays and participate in the research, yet the vast majority did. Ideally, we wanted to even the sample of essays at two institutions but since the comparison was not one of our research aims and we did not spend the same amount of time conducting research at two institutions, we could not manage that. However, the students studying at all branches at the aforementioned institutions participated in the research, thus contributing to the population sample diversity.

Using every available tool (online/offline dictionaries), the students were asked to present themselves in whichever way they wished. Some students focused more on their leisure time, some on their academic success, some covered both topics. We did not want to instruct them on what to write about because their choice of the prism to present themselves is a variable studied by this research. Each essay was searched for spelling mistakes, corrected and technically prepared for phase two in which the essays were computationally analyzed by the software for linguistic analysis *Linguistic Inquiry and Word Count* (LIWC). The essays were not graded because this was not our research purpose. LIWC has an internal dictionary comprising 4,500 words and word stems which were (re)categorized by expert judges during several stages. During the analysis, LIWC analyzes a word by a word and categorizes it into one or more preset categories. To exemplify, *happy* would be categorized in the positive emotions and adjective categories. The categories range from linguistic (verbs, adverbs, different pronoun types) and psychological (words related to different emotions) to conversational (words related to achievement, leisure, money) and punctuation (periods, commas) categories. LIWC provides an output in the form of raw numbers which need to be further analyzed.

As previously stated, we replicated the so far methodology by Božić Lenard, Liermann-Zeljask, and Ferčec (2018). In comparison, we decided to employ inductive reasoning by using their specific LIWC category results and compared them to students' grades to examine a LIWC category range for a specific grade. Using these results, we designed an application which calculates one's grade based on input LIWC raw results. To put it differently, in order to use our application, a student's essay needs to be analyzed with

LIWC whose raw results as well as grade ranges for the LIWC categories suggested by Božić Lenard, Liermann-Zeljak, and Ferčec (2018) are implemented in the application. Upon LIWC analysis, the application computes a grade a student is supposed to acquire in English for Specific Purposes courses.

Our research aimed to test the application and examine if final ESP grades proposed by the application are in line with the actual ESP grades the participants acquired in the examined period. We wished to examine the success rate of the application and take a closer look at how each of the LIWC categories behaves in terms of grade prediction with the final aim of potentially revising LIWC categories and grade ranges.

RESULTS AND DISCUSSION

When evaluating our results, we classified them in three categories, namely the correctly predicted app grade, the app grade which is one grade higher/lower than the actual grade and the app grade which greatly differs from the actual grade.

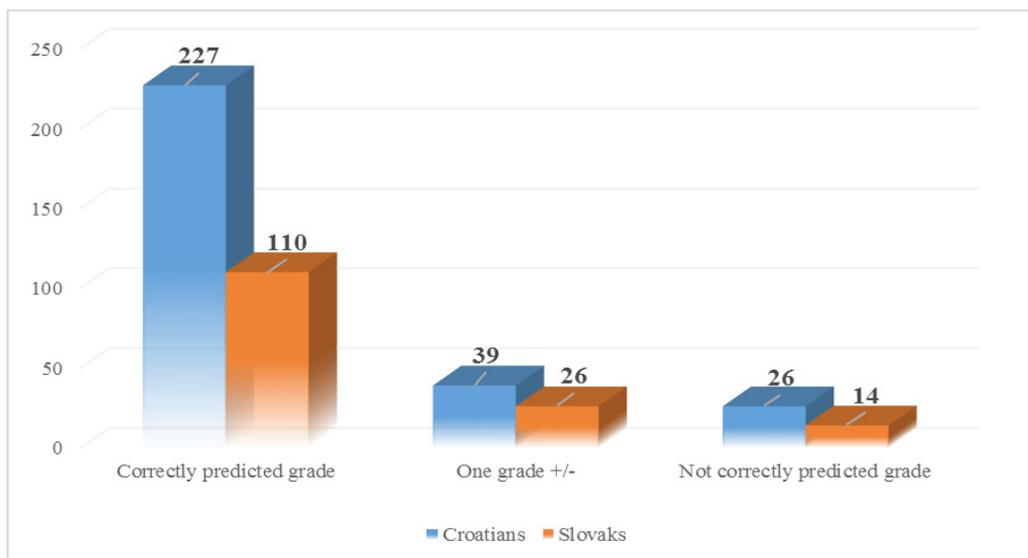


FIGURE 2: APP TESTING RESULTS

As illustrated in Figure 2, 77.8% of the Croatian and 73.4% of the Slovakian students' grades were correctly predicted. Additional

13.3% and 17.3% of the Croatian and Slovakian students' grades respectively were missed by one grade, i.e. 8.9% of the Croatian and 9.3% of the Slovakian students' grades were incorrectly predicted. If we add the first two categories' percentages, we get 91.1% of the Croatian and 90.7% of the Slovakian students' grades predicted within one grade deviance, which is an excellent and a very promising result. A final grade depends on multiple variables (homework and in-class assignments, projects, speaking/presentation exercises, revision exams, oral exam, etc.) so predicting students' grades within one grade deviance is almost natural and expected. Yet, much higher percentages were acquired in the correctly predicted grade category, which proves that the application is successful and points to subtle refining of the LIWC categories' grade ranges in order to increase the correctly predicted grade percentages.

In the following subchapters, we will discuss the LIWC categories used in this research, compare them with previous research using similar categories and elaborate on our observations.

WORDS PER SENTENCE AND PERIOD

A words per sentence category is a descriptor category intertwined with a period category, i.e. LIWC calculates the number of words in a sentence whose ending is marked with a period. In cases when a sentence ended with an exclamation or a question mark, we changed them into a period for the calculation purposes because there are fewer exclamation and question marks and they would be calculated in comparison to other punctuation categories, which would greatly skew the results, hence the punctuation category adjustment.

Previous research has proven the relationship of the words per sentence category and students' academic achievement. In a series of research, Crossley and McNamara (2011, 2012, 2016), Varner, Roscoe and McNamara (2013), Crossley, Roscoe and McNamara (2014), McNamara et al. (2015), Crossley, Kyle and McNamara (2016), Crossley, Muldner and McNamara (2016), Kyle and Crossley (2017), Kim and Crossley (2018) have demonstrated that text elaboration is a strong predictor of essay quality and higher essay grading. Successful writers compose longer texts and provide readers with an increased amount of relevant information thus guiding readers to better understand essay arguments. Generally, longer texts mean more information which equals to better argumenta-

tion and clarification and a writer who can do that is perceived as a knowledgeable and credible expert on a topic. On the other hand, other research (McNamara, Crossley and McCarthy, 2010; McNamara, Crossley and Roscoe, 2013; Rass, 2015; Vinogradova, Lyashevskaya and Panteleeva, 2017) point to inadequacy of counting words and relating them to more successful essays. They found that less proficient university-level writers tend to compose longer sentences but simply because they repeat words more than proficient writers. Moreover, Rass (2015) reported that Palestinian Arab students over repeat phrases and accumulate redundant words before stating the main point.

Our research found merit in both approaches. The majority of our participants, whom the application predicted higher grades, did use lengthy sentences during which they adequately adopted transitional phrases, elaborations, and syntactic complexity. However, there were students predicted to get a poor grade (C or even D) who used extremely long sentences thus potentially skewing their final grade results because the application would suggest grade A in this category. The reason for such result is in carrying out only an automated linguistic analysis. We checked the essays and corrected misspellings but we did not alter syntax and the aforementioned cases are the ones where students did not compose complex sentences or were verbose for the purpose of clarification but merged multiple sentences into one due to not having enough linguistic (syntactical) knowledge. In order to overcome this problem, one might consider proofreading essays and correct all linguistic levels mistakes.

SIX-LETTER WORDS

Similar to the previous, a six-letter word category is a descriptor one, i.e. LIWC searches for all six or more letter words and calculates their usage percentage in comparison to the total number of words per essay. We were aware of the British and American English spelling distinction and trying to avoid one simple word grapheme making a difference (colour vs. color), we proofread the essays using the American English spelling.

Research in the field of vocabulary profiling are fairly recent and were initiated by technological advances two decades ago. Early researchers (Coxhead and Nation, 2001; Morris and Cobb, 2004) started conducting word frequency studies aiming to find consist-

ent patterns of good academic performance based on the expressive use of academic vocabulary. Numerous later research has indicated a positive correlation of complex vocabulary and students' grades. Research using automated essay scoring (Daller and Xue, 2009; Hulstijn, Van Gelderen and Schoonen, 2009; McNamara, Crossley and McCarthy, 2010; Meurers, 2013; McNamara, Crossley and Roscoe, 2013; Crossley, Roscoe and McNamara, 2014; Kim, Crossley and Kyle, 2018) proved that higher scored essays contain less frequent and more complex words. Similarly, research involving expert judges (Roessingh and Elgie, 2009; Leo, 2012) also demonstrated that professors tend to grade essays with the highest grades when they contain longer, less concrete and ambiguous words.

Our research is in line with the aforementioned, i.e. students who used complex, lengthy, and infrequent words were profiled to be more successful in their ESP classes. Productive lexical knowledge was realized through lexical sophistication in writing and concomitantly syntactic complexity. To paraphrase, both lexical and syntactic complexity are strong predictors of successful ESP students. However, there is a skewness potential when studying this category and using natural language processing tools. Even though frequency is generally considered as one of the best indicators of language processing, in this type of research it may not be the best option because study participants can (unintentionally) overly reuse the same long obscure word(s) and consequently skew results. The potential problem can be overcome by taking a range (dispersion, entropy, and contextual diversity) into account during the text screening and preparation phase.

PERSONAL PRONOUNS AND PRONOUN I

The most commonly used words in the English language are pronouns. In spite of being short and almost imperceptible, these function words can reveal a lot about a writer and reflect the way one thinks. Numerous research has been conducted with the purpose of studying the distribution of personal pronouns in an array of registers.

In an extensive computerized text analysis study, Pennebaker et al. (2014) analyzed over 50,000 college admission essays looking for function words. They discovered a negative correlation of students' academic achievement and the use of personal pronouns thus cor-

roborating Robinson, Navea, and Ickes' (2013) results. In recent studies by Demmans, Phirangee, and Hewitt (2017), and Božić Lenard, Liermann-Zeljak, and Ferčec (2018), it was discovered that less successful students shared more personal stories, used more personal pronouns and felt stronger sense of community. Some psycholinguistic research (Nisbett et al. 2001; Heylighen and Dewaele, 2002; Pennebaker, 2013) interpreted their data in terms of profiling people. They claim that high use of personal pronouns is associated with greater focus on other people and a narrative thinking style which is realized through telling personal stories. Also, such people tend to be dynamic thinkers who focus on describing actions, changes, and involved people.

Basically, all the aforementioned research arrived at the same conclusion—if students think, talk, and/or write about other people, they will be academically less successful. Our own research has contributed to the body of previous research and proved that these small words do predict academic success. Poor students chose to share their and their families' and friends' personal stories while describing themselves thus denoting mutual awareness and engagement. Usually, they wrote about their leisure time, hobbies they share and things they jointly do. In a lot of cases, they focused more on other people or their relationship with other people than they exclusively wrote about themselves. On the other hand, more successful students focused on themselves, their own hobbies, their future creating a social distance and not sharing a common space (Baker, 2010).

The category involving the pronoun *I* includes all forms (subject, object, (in)dependent possessive, reflexive) of this pronoun and even though it is undoubtedly intertwined with the personal pronouns category, though not as words per sentence and period categories, we wanted to give it a more proper inspection. Pennebaker was a pioneer in studying pronouns and his research (2013) has inspired other researchers to pay these short words more attention. Some researchers (Chung and Pennebaker, 2007; Kacewitz et al., 2014; Biesen et al., 2015; Božić Lenard, Liermann-Zeljak, and Ferčec, 2018) hypothesized that the use of the first-person pronoun is related to higher grades and social status. If people are aware of their knowledge and power, they tend to focus less on themselves and more on their tasks. Our research corroborated this hypothesis because it showed that more successful students were less self-oriented and simultaneously performed their task of self-introduc-

tion. Less successful students used very short and syntactically simple sentences in the active voice thus overly using the first-person pronoun while more successful students used the passive voice whose acquisition is done in later learning stages. Even though our research did not find a correlation between the students' grades and the use of the pronoun *we*, it would be interesting to examine such correlation.

ARTICLES

Despite being governed by the rules and seeming purposeless to study, articles can be very revealing because they are connected to the use of nouns and one's ability to categorize things. The use of articles is somewhat dichotomous to the use of personal pronouns because articles and nouns are indicators of a categorical thinking style. There is a growing body of research studying the correlation between the use of nouns/articles and students' success in EFL/ESP classes. Researchers in the foreign language writing (Robinson, Navea and Ickes, 2013; Taguchi, Crawford and Wetzel, 2013; Parkinson and Musgrave, 2014; Pennebaker et al., 2014; Córdoba Cubillo and Solís Hernández, 2015; Wang and Slater, 2016; Choi, 2017; Wang and Beckett, 2017) put forward the idea that more successful students and proficient writers use noun phrases and articles which proved to be one important difference between excellent and mediocre students.

Our research results are in accordance with the aforementioned since they showed a positive correlation between the use of nouns/articles and students' academic success pointing to more successful students thinking in a categorical way and focusing on their task not involving other people or describing events. However, there is an issue with this category which needs to be addressed. Our research participants are non-native English speakers who do not use articles properly, i.e. they very frequently use the zero article or omit articles due to the lack of article knowledge. Naturally, those students who are more successful in ESP use articles more so the results are not skewed; yet, one might consider proofreading students' essays for articles to tackle this issue. Another way to address this issue is for LIWC designers to include the category of nouns *per se*. If neither of these two suggestions are adopted, comparative studies involving native and non-native English language speakers are impossible.

ACHIEVEMENT AND LEISURE

The categories of achievement and leisure include lexical words related to the notions of achievement (acquir*, challeng*, master*, organis*, practice*, proficien*, skill*, etc.) and leisure (Nintendo*, football*, party*, ski*, tequila, tv*, mall*, karaoke*, etc.). With an exception of Božić Lenard, Liermann-Zeljak, and Ferčec's (2018) study, these categories did not receive researchers' attention so our results might inspire analyses.

Contrary to what one might intuitively think, students graded with B were the group that focused on emphasizing their achievements the most. It seems possible that excellent students are aware of their professional and private achievements and need not to emphasize it. On the other hand, students obtaining very good grades are always perceived as less successful than students obtaining excellent grades even though differences are subtle and may be (and frequently are) related to factors other than one's intelligence. It is likely that very good students feel the difference and try to compensate on this underappreciation by explicitly accentuating their success. Poorer students (the ones graded with Cs and Ds) probably believe that they are not professionally successful, which is the reason for their lack of reporting it. Dichotomous to the category of achievement is the category of leisure activities and our research has corroborated it. Almost all less successful students (the ones graded with Cs and Ds) decided to present themselves through their leisure activities and hobbies while some more successful students (the ones graded with As and Bs) did that. While doing that, all students wrote about their sport activities and outdoor hobbies; however, while doing that, poorer students involved other people whereas more successful students focused on themselves even in cases of group activities. It seems probable that the higher-achievers perceived the task of self-introductory essays as something formal and focused on their professional rather than private lives. It would be interesting to study the two categories' correlation and its association with students' grades because we noticed that some students report their leisure activities and hobbies' achievements.

CONCLUSIONS

Using previous similar research results cited in the respective *Results and discussion* subchapters, this research aimed to test the application designed to predict students' success potential in ESP classes. A sample of 442 undergraduate students studying at the Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Croatia and the Faculty of Material Science and Technology in Trnava, Slovakia participated in the research by submitting their self-introductory essays computationally analyzed and uploaded in the designed application. The results point to the application's high prediction efficacy (75.6% correctly predicted grades and additional 15.3% grades predicted with one grade deviance). More precisely, the results showed that the higher-performing students were more elaborative in their descriptions, used more transitional phrases, and fewer but more syntactically complex sentences. Verbosity was accompanied by lexical sophistication and more obscure and lengthy words. The higher-achievers were proven to be more task-oriented using more noun phrases while lower-achievers included other people in their frequent description of joint leisure activities, used simple active sentences which abounded in the first personal pronoun *I* and other personal pronouns thus demonstrating the features of a dynamic thinking and verbal style.

Predictive analytics and the application *per se* can serve as a valuable tool to guide management and educators to engage themselves in necessary curriculum and teaching method changes. From another pedagogical perspective, predictive analytics might advance students' success by providing insights on what can be improved on by employing additional or alternative learning methods (Siemens and Long, 2011; Dietz-Uhler and Hurn, 2013), working with an academic advisor or discussing individualized feedback with relevant personnel. Needless to say, management and higher education institution/stakeholder employees are an integral part of this support-feedback-involvement triad (Tinto and Pusser, 2006) so all parties' involvement is of the utmost importance for the overall students' success enhancement.

LIMITATIONS OF THE RESEARCH

There are several weak points of our research. We did not proof-read the students' essays for stylistic and syntactical mistakes so there is a skewness potential. Since a small sample of students participated in the research, we cannot make generalisations on the application's prediction efficacy; yet, it does give us some useful insights on predicting students' ESP class potential.

RECOMMENDATIONS FOR FURTHER RESEARCH

An immediate way to proceed is to conduct a large-scale study including more ESP/EFL participants studying at different higher education institutions and in different countries. Researchers could include other (LIWC) variables into account some of which we had suggested in the *Results and discussion* subchapters. New studies might set a new direction in education assessment beneficial for faculty management and employers looking for skillful and talented individuals.

ACKNOWLEDGMENTS

This research was supported in part by the National Scholarship Programme of the Slovak Republic (NSP). Ideas expressed in this paper are those of the authors and do not reflect the views of NSP. We thank our colleagues Dr. Ľudmila Hurajová, Dr. Emília Mironovová and Jarmila Blahová, MA from the Faculty of Materials Science and Technology in Trnava, Slovakia. We appreciate their great assistance in our research in terms of promoting it among their students who voluntarily participated in it. We would also like to show our gratitude to Tomislav Rešicki, MComp for sharing his application designed as a part of his Master thesis at the Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Croatia.

REFERENCES

- Baker, C. (2010). The impact of instructor immediacy and presence for online student affective learning, cognition, and motivation. *Journal of Educators Online*, 7 (1), 1-30.
- Biesen, J. N., Schooler, D. E., Smith, D. A. (2015). What a Difference a Pronoun Makes I/We Versus You/Me and Worried Couples' Perceptions of Their Interaction Quality. *Journal of Language and Social Psychology*, 35 (2), 180-205.

- Božić Lenard, D., Liermann-Zeljak, Y., Ferčec, I. (2018). A redefined approach to student assessment based on essay writing in ESP. *Collection of Papers of the Faculty of Philosophy in Priština*, 48 (3), 5–36. doi:10.5937/ZRFFP48-18665
- Burstein, J., Chodorow, M., Leacock, C. (2004). Automated essay evaluation: The Criterion online writing system. *AI Magazine*, 25, 27–36.
- Burstein, J., Marcu, D., Knight, K. (2003). Finding the WRITE Stuff: Automatic Identification of Discourse Structure in Student Essays. *IEEE Intelligent Systems: Special Issue on Natural Language Processing*, 18 (1), 32–39.
- Choi, M. (2017). *Use of English Articles by Korean Students. Culminating Projects in English*, 109. https://repository.stcloudstate.edu/engl_etds/109.
- Chung, C., Pennebaker, J. (2007). The Psychological Functions of Function Words. In: Fiedler, K. (ed.) (2007). *Social Communication*. New York: Psychology Press, 343–359.
- Coxhead, A. & Nation, I. S. P. (2001). The specialized vocabulary of English for academic purposes. In: J. Flowerdew, & M. Peacock (Eds.), *Research Perspectives on English for Academic Purposes* (252–267). Cambridge: Cambridge University Press.
- Crossley, S. A., Kyle, K., McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *The Journal of Second Language Writing*, 32, 1–16.
- Crossley, S. A. & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, 7 (3), 351–370.
- Crossley, S. A., McNamara, D. S. (2012). Predicting second language writing proficiency: The role of cohesion, readability, and lexical difficulty. *Journal of Research in Reading*, 35 (2), 115–135.
- Crossley, S. A. & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh- Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life-Long Learning*, 21, 170–191.
- Crossley, S. A., Muldner, K., McNamara, D. S. (2016). Idea generation in student writing: Computational assessments and links to successful writing. *Written Communication*, 33 (3), 328–354.
- Crossley, S. A., Roscoe, R., McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31 (2), 184–215.
- Córdoba Cubillo, P., Solís Hernández, M. (2015). The influence of grammar on English learners' writing accuracy at the school of modern languages. *Káñina, Rev. Artes y Letras, Univ. Costa Rica*, 2, 107–121.
- Daller, M.H., Xue, H. (2009). Vocabulary Knowledge and Academic Success: a Study of Chinese Students in UK Higher Education. In: B. Richards M.

- H.Daller, D. Malvern, P. Meara, J. Milton, J. Treffers-Daller (Eds), *Vocabulary Studies in First and Second Language Acquisition* (179–193). London: Palgrave Macmillan.
- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18, 7–24.
- Demmans, E., C., Phirangee, K. Hewitt, J. (2017). *The Interplay Between Students' Usage of Pronouns and Community Levels in Online Courses*. Presented at the American Educational Research Association (AERA) San Antonio, TX, USA: Annual Meeting.
- Dietz-Uhler, B. & Hurn, J. E. (2013). Using learning analytics to predict (and improve) students success: a faculty perspective. *Journal of Interactive Online Learning*, 12 (1), 17–26.
- Elliott, S. (2003). Intellimetric: From Here to Validity. In: M.Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (71–86). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Heylighen, F. & Dewaele, J. M. (2002). Variation in the contextuality of language: an empirical measure. *Foundations of Science*, 6, 293–340.
- Hulstijn, J. H., Van Gelderen, A., Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, 30 (4), 555–582.
- Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33 (2), 125–143.
- Kellogg, R., Whiteford, A. (2009). Training advanced writing skills: The case for deliberate practice. *Educational Psychologist*, 44, 250–266.
- Kim, M., Crossley, S. A. (2018). Modeling Second Language Writing Quality: A Structural Equation Investigation of Lexical, Syntactic, and Cohesive Features in Source-Based and Independent Writing. *Assessing Writing*, 37, 39–56.
- Kim, M., Crossley, S. A., Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, 102 (1), 120–141.
- Kyle, K., Crossley, S. A. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, 34 (4), 513–535.
- Leo, K. (2012). Investigating Cohesion and Coherence Discourse Strategies of Chinese Students with Varied Lengths of Residence in Canada. *Tesl Canada Journal*, 29 (6), 157–179.
- Meurers, D. (2013). Natural language processing and language learning. In: C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (1–13). New York, NY: Blackwell Publishing Ltd.

- McNamara, D. S., Crossley, S. A., McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication*, 27 (1), 57–86.
- McNamara, D. S., Crossley, S. A., Roscoe, R. (2013). Natural Language Processing in an Intelligent Writing Strategy Tutoring System. *Behavior Research Methods*, 45 (2), 499–515.
- McNamara, D. S., Crossley, S. A., Roscoe, R., Allen, L., Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23 (1), 35–59.
- Morris, L., Cobb, T. (2004). Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System*, 32, 75–87.
- Nisbett, R. E., Peng, K., Choi, I., Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, 108, 291–310.
- Parkinson, J., Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, 14 (2), 48–59.
- Pennebaker, J. W. (2013). *The Secret Life of Pronouns: What Our Words Say About Us*. New York: Bloomsbury Press.
- Pennebaker, J. W. & Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. In: H. Friedman & R. Silver (Eds.), *Handbook of health psychology* (263–284). New York: Oxford University Press.
- Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., Beaver, D. I. (2014). When Small Words Foretell Academic Success: The Case of College Admissions Essays. *PLoS ONE*, 9 (12), <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4281205/>
- Picciano, A. G. (2014). Big data and learning analytics in blended learning environments: benefits and concerns. *International Journal of Artificial Intelligence and Interactive Multimedia*, 2 (7), 35–43.
- Rabow, M. W., Wrubel, J. R., Rachel, N. (2009). Promise of professionalism: Personal mission statements among a national cohort of medical students. *Annals of Family Medicine*, 7, 336–342.
- Rass, R. A. (2015). Challenges Face Arab Students in Writing Well-Developed Paragraphs in English. *English Language Teaching*, 8 (10), 49–59.
- Robinson, R.L., Navea, R., Ickes, W. (2013). Predicting final course performance from students' written self-introductions; A LIWC analysis. *Journal of Language and Social Psychology*, 32, 481–491.
- Roessingh, H., Elgie, S. (2009). Early language and literacy development among young English language learners: Preliminary insights from a longitudinal study. *TESL Canada Journal*, 26, 24–45.

- Siemens, G., Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46 (5), 31–40.
- Taguchi, N., Crawford, W., Wetzel, D. Z. (2013). What linguistic features are indicative of writing quality? A case of argumentative essays in a college composition program. *TESOL Quarterly*, 47, 420–430.
- Tinto, V., Pusser, B. (2006). *Moving from theory to action: Building a model of institutional action for student success*. Paper presented at the National Symposium on Postsecondary Student Success, Washington, DC. Retrieved from https://nces.ed.gov/npec/pdf/Tinto_Pusser_Report.pdf
- Turner, R., Nicholson, S. (2011). Reasons selectors give for accepting and rejecting medical applicants before interview. *Medical Education*, 45, 298–307.
- Varner, L. K., Roscoe, R. D., McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research*, 5 (1), 35–59.
- Vinogradova, O. I., Lyashevskaya, O. N., Panteleeva I. M. (2017). Multi-level student essay feedback in a learner corpus. *Computational Linguistics and Intellectual Technologies - Proceedings of the International Conference "Dialogue 2017"*, 1 (16), 373–386.
- Walfish, S., Moreira, J. (2005). Relative weighting of admission variables in marriage and family therapy graduate programs. *American Journal of Family Therapy*, 33, 395–402.
- Wang, S. & Slater, T. (2016). Syntactic Complexity of EFL Chinese Students' Writing. *English Language and Literature Studies*, 6 (1), 81–86.
- Wang, S., Beckett, G. H. (2017). "My Excellent College Entrance Examination Achievement" — Noun Phrase Use of Chinese EFL Students' Writing. *Journal of Language Teaching and Research*, 8 (2), 271–277.
- Waugh, C.E., Fredrickson, B.L. (2006). Nice to know you: Positive emotions, self-other overlap, and complex understanding in the formation of new relationships. *The Journal of Positive Psychology*, 1, 93–106.

ДРАГАНА БОЖИЋ ЛЕНАРД

СВЕУЧИЛИШТЕ ЈОСИПА ЈУРЈА ШТРОСМАЈЕРА У ОСИЈЕКУ
ФАКУЛТЕТ ЕЛЕКТРОТЕХНИКЕ, РАЧУНАРСТВА И ИНФОРМАЦИЈСКИХ ТЕХНОЛОГИЈА

ГАБРИЕЛА ЧМЕЛИКОВА

СЛОВАЧКИ ТЕХНИЧКИ УНИВЕРЗИТЕТ
МАТЕРИЈАЛНО-ТЕХНОЛОШКИ ФАКУЛТЕТ У ТРНАВИ
ЦЕНТАР ЗА ЈЕЗИКЕ, ХУМАНИСТИКУ И СПОРТ

РЕЗИМЕ

ПРИМЕНА ПРЕДИКТИВНЕ АНАЛИТИКЕ У ЕНГЛЕСКОМ
ЗА ПОСЕБНЕ НАМЕНЕ НА ОСНОВУ СТУДЕНТСКИХ ТЕКСТОВА

Оштра конкуренција приликом уписа на реномиране универзитете и запошљавања у угледним компанијама створила је потребу за профилисањем потенцијалних кандидата са циљем одабира најбољих. Наше истраживање има за циљ да предвиди потенцијал успешности студената, то јест њихове оцене из курсева енглеског за посебне намене. Укупно 292 студента са Факултета електротехнике, рачунарства и информацијских технологија у Осиеку у Хрватској и 150 студената са Материјално-технолошког факултета у Трнави у Словачкој добровољно су учествовали у истраживању и доставили своје кратке биографије. Њихове биографије анализирани су софтвером (LIWC), а добијени резултати, у форми необрађених бројки, унети су у апликацију посебно креирану за ове потребе. На основу унетих података, апликација је рачунала студентске оцене из наведених курсева које су затим поређене са стварним оценама студената. Истраживањем је потврђен висок степен ефикасности апликације будући да је тачно предвидела оцене у 75% случајева, док су код још 12% оцене приближно прогнозиране, то јест предвиђена оцена била је за један виша или нижа од стварне.

Кључне речи: биографије студената; енглески за посебне намене; оцењивање; LIWC; примена, предвиђање успеха.



Овај чланак је објављен и дистрибуира се под лиценцом Creative Commons Ауторство-Некомерцијално Међународна 4.0 (CC BY-NC 4.0) | <https://creativecommons.org/licenses/by-nc/4.0/>.
This paper is published and distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial International 4.0 licence (CC BY-NC 4.0) | <https://creativecommons.org/licenses/by-nc/4.0/>.