



<http://jates.org>

Journal of Applied  
Technical and Educational Sciences  
jATES

ISSN 2560-5429



## Proctoring as a journey to quality education? A critical review of the literature

Michal Cerny<sup>1</sup>, Dominika Solcova<sup>2</sup>

<sup>1</sup> Masaryk University, Faculty of Arts, Arne Novaka 1, Brno, 602 00, Czech Republic, [mcerny@phil.muni.cz](mailto:mcerny@phil.muni.cz)

<sup>2</sup> Masaryk University, Faculty of Arts, Arne Novaka 1, Brno, 602 00, Czech Republic, [solcovadominika@phil.muni.cz](mailto:solcovadominika@phil.muni.cz)

**Abstract:** *The paper aims to analyse the current professional discourse and to map the current views and attitudes of proctoring online exam formats across the globe. The entire theoretical delineation focuses mainly on college or university formats. We introduce the international context in the first part and then discuss the analysis of the texts found in the review study at Scopus and Web of Science databases. The study results in a comprehensive summary of the positive and negative implementation aspects of proctoring. We also reach a new categorisation of the different methods used to control proctoring online exams. The study should provide synthesising findings with easy applicability. Our results may be valuable for educational institutions or others interested in online learning and online proctoring processes, not only when considering implementation.*

**Keywords:** *online proctoring; online exams; exam security; online student authentication; distance education; online learning*

### 1. Introduction

Proctoring is not new, but it is associated with extraordinary dynamics. The first studies in the Web of Science database on this topic date back to 1972. In it, Farmer et al. (1972) analysed the academic performance of psychology students who were proctored and those who were not. This study brings to the forefront of its interest the fundamental question associated with this phenomenon - does proctor increase the quality of education? Studies over time have answered this question differently. Looking at the Web of Science, since the early 1990s, proctoring has been the subject of an order of magnitude of studies per year until 2009, when the topic became topical.

After a drop in interest in 2019 (from 66 to 29 studies), we can see a renewed interest logically accelerated by the COVID-19 pandemic, which brought not only the topic of e-learning as a basic form of education but also the question of monitoring quality and fairness. Universities and other actors in the educational process have begun to think intensively again about how to

make distance education secure and associated with the same level of credibility as regular f2f. This demand has offered, on the one hand, a naive implementation of proctoring into different systems; on the other hand, it has opened up a new deeper field of reflection on what role proctoring has in the educational process and what effects it is associated with.

## 2. Background

Fenu et al. (2018) point out that proctoring cannot just focus on monitoring students in exams but should be concerned with monitoring students' presence in the classroom. The goal should be more than authenticating the examinee but monitoring and controlling their education process. This is to be done by working with biometric data that can be collected and analysed online. Also, Grande de Prado et al. (2021) consider proctoring as a biometric identification of the learner, even if only during exams, and consider this process as one of the central themes of the discussions on e-learning, stressing the importance of planning such a measure.

Raje and Stitzel (2020) see proctoring as part of preventing copying and cheating on tests. They work with the example of chemistry tests, which involve, among other things, rote learning that can be easily circumvented by reinforcing correct answers. Knowledge of these concepts is essential for further study of chemistry, so the authors analyse various options - from watermarking to time-limited windows for completing tests - to prevent cheating.

Lee (2020) returns to a question by Farmer et al. (1972) about educational effectiveness and proctoring. He concludes (using a sample of 1762 students) that online and offline proctoring is unlikely to have any measurable effect on the quality of knowledge. Therefore, the mindset of educational institutions needs to be changed to focus more on significant and practical concepts of education.

Wuthisatian (2020) points out that the issue of proctoring can be complex - the choice of a particular proctoring environment and method has many psychological and technical dimensions that must be considered. In his view, protecting academic integrity alone cannot be a full-fledged reason for introducing proctoring - students' poorer performance on a test does not automatically mean they are copying. Instead, it points to other problems such as time and technical demands, stress, reluctance to take the test, scrutiny, etc. Labayen et al. (2021) discuss the ideal of proctoring as a completely passive solution for the student that will be technologically reliable, simple and scalable. On the one hand, the discourse reducing proctoring to a purely technological problem responds to Wuthisatian's (2020) findings but, at the same time, fails to reflect the psychological, social and ethical dimensions of proctoring.

Saunders and Weible (1999) offer two interesting considerations in their essayistic text about proctoring. Firstly, proctoring may not only be promoted by an educational institution seeking to ensure its academic integrity but may also originate from students' needs. These needs are based on the requirement to confirm the quality of one's education. A crucial factor is how online education is viewed in general - according to the authors, where it has the character of a distinct educational form, the students' need for proctoring is fundamentally less than for those who view online education only as a correspondence course.

The above studies clearly show that proctoring is - as they mention - a controversial and debated topic. However, at the same time, we need more data, a thorough overview, and a sufficiently robust empirical base for its critical analysis. Therefore, this study focuses on analysing available recent studies on the topic of proctoring in university settings to offer a comprehensive and plastic view of the issue, which could then be used, for example, for sociological or ethical analyses and discussions.

### 3. Methodology

We acquired information by creating a short review study conducted through the Web of Science and Scopus databases to understand the current research discourse. The search was conducted in July and August 2022 in two iterative processes to increase the relevance of the results, as proctoring is a highly diversified topic in terms of concept and meaning. We only looked at texts written in English with publication dates between 2020 and 2022. Later results were discarded due to the global pandemic situation, as the diversity of these texts would have been too high within the topic under study. A final limitation of the query was restricting the type of publications to journal articles and conference papers with a free open license.

#### 3.1. Phase 1

The primary phase of the search queries focused on using the keyword phrases online learning, security, and covid-19, which yielded the results of the status or necessary changes in security at a distance education cause of the pandemic situation. The search query's key term was security, at the centre of which the technology-security component of proctoring is stated. Thus, the resulting search query looked as follows:

TITLE-ABS-KEY ( "online learning" AND security AND covid-19 ) AND ( LIMIT-TO ( OA , "all" ) ) AND ( LIMIT-TO ( PUBYEAR , 2022 ) OR LIMIT-TO (

PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR LIMIT-TO ( DOCTYPE , "cp" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )

The search query yielded 20 results, and in the process, three texts were excluded for not fulfilling the thematic requirements after a deeper study. The first one (Lam Lam & Dongol, 2020) described the security of blockchain-based e-learning platforms. The second one dealt with blockchain in the context of online student assessment (Alshahrani, 2022), and the last one (Rahmani et al., 2021) again provided insights into the general scope of e-learning development through artificial intelligence and blockchain. Thus, proctoring topics have been discussed here only marginally or not at all. The initial phase keyed 17 papers.

### *3.2. Phase 2*

The secondary part of the search strings went into more detail about direct proctoring terminology, including the related aspect of test security, which, in contrast to the first search cycle, symbolises the psychological-security facet of guarded test quality. The chain displayed 21 additional results due to this fact. Our form of the chosen methodology thus encompasses different approaches and states of mind that equally interpret the problematic phenomenon. The wording of the second query is as follows:

TITLE-ABS-KEY ( "test security" AND "online proctoring" ) AND ( LIMIT-TO ( OA , "all" ) ) AND ( LIMIT-TO ( PUBYEAR , 2022 ) OR LIMIT-TO ( PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR LIMIT-TO ( DOCTYPE , "cp" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )

### *3.3. Summary*

In a short enumeration table, we provided a complete overview of the number of publications in both phases, including dependence on the relevant Web of Science and Scopus databases.

Table 1. A proportional number of publications in phases

	WoS	Scopus	Overall
Phase 1	9	8	17
Phase 2	12	9	21
Overall	21	17	38

The complete list of all the documents considered comprises 38 publications. It is included in the following parts of the text as a tabular statement. In the table, we provided basic bibliographic information about each paper; the authors, the year of publication, or the type of paper, which informs about the nature of the text. Furthermore, we have chosen to reflect in the table the proctoring statements that are important to us, which we will comment on in more detail in the results of the literature search. These metrics were; relationships to proctoring, geographic scope, and preferred review methods. As an associated part of the table used to assess the quality of the retrieved articles, we also include parameters in the number of citations in the Web of Science database as of 12 November 2022. The texts are typically arranged alphabetically and were all freely available in the full text within the database at the time of the review study.

#### 4. Results

The results presented below show that proctoring is discussed across various continents, except for Africa and South America, which will be influenced mainly by developing countries, which face more elementary problems in the field of education, usually beyond the level of the security sector. Of the 38 documents, the most significant number of publications are associated with Asia, 17; then Europe, 10; North America, 8; and lastly, Australia, where we reflected three publications. The Turkish state became a problematic concept in terms of the breakdown because it is located at the level of the Asian and European continents in terms of area. Still, considering the geographical location of the institution of the written publications, we decided to include Turkey in the Asian continent for most of this analysis.

In the table below, the relatively high diversity of the topics discussed is also visible through the titles, marking both the novelty and the controversy of the concept because we find differentiated views on implementing proctoring even within a single national border. The most

frequent selection studies proposing a new secure proctoring framework 7; studies considering the advantages or disadvantages of proctoring 6; articles comparing the effectiveness of specific proctoring systems; five or comparing proctored and unproctored exams 4. The studies resorted to an analysis of the literature in three cases and obtained student opinions on the concept in another three points. The positively commenting studies mainly discussed reducing or minimising cheating, while opponents of proctoring often referred to secure test settings that effectively reduce cheating techniques.

Thus, our selection focuses on 38 recent papers that, without depending on the number of citations, can adequately demonstrate the current state of the proctoring in published studies and responses. We present the single items crucial to our understanding by attaching a summary table. A detailed analysis of the findings will be discussed in a forthcoming chapter.

Table 2. A complete list of the reflected literature

<b>Authors + year of study</b>	<b>Type of Study</b>	<b>Quotations</b>	<b>State</b>	<b>Relationship to proctoring</b>	<b>Control methods</b>
Ahmad, I., AlQurashi, F., Abozinadah, E., & Mehmood, R. (2021)	application	0	Saudi Arabia	positive	identification
Alshammari, M. T. (2020)	application	0	Saudi Arabia	positive	identification locking
Arnò, S., Galassi, A., Tommasi, M., Saggino, A., & Vittorini, P. (2021)	theoretical	8	Italy	negative	identification locking monitoring
Balash, D. G., Kim, D., Shaibekova, D., Fainchtein, R. A., Sherr, M., & Aviv, A. J. (2021)	empirical	4	USA	neutral	locking monitoring
Becker, B., van Rijn, P., Molenaar, D., & Debeer, D. (2022)	empirical	2	Germany/ Netherlands	neutral	xxx
Bergmans, L.,	empirical	0	Netherlands	negative	locking

Bouali, N., Luttikhuis, M., & Rensink, A. (2021)					monitoring
Butler- Henderson, K., & Crawford, J. (2020)	theoretical	37	Australia	positive	identification locking
ÇELİKBAĞ, M. A., & Delialioğlu, Ö. (2021)	empirical	0	Turkey	neutral	xxx
Conijn, R., Kleingeld, A., Matzat, U., & Snijders, C. (2022)	empirical	2	Netherlands	negative	xxx
Farland, M. Z., & Childs-Kean, L. M. (2021)	theoretical	1	USA	neutral	xxx
Fiano, K. S., Medina, M. S., & Whalen, K. (2021)	theoretical	4	Oklahoma (USA)	positive	xxx
Garg, M., & Goel, A. (2022)	theoretical	1	Indie	positive	identification locking
González, G. CS; Infante Moro, A.; Infante Moro, JC (2020)	empirical	37	Spain	positive	xxx
Hébert, C. (2021)	theoretical	0	Canada	negative	xxx
Howard, D. (2020)	empirical	3	Idaho (USA)	neutral	xxx
Humbert, M., Lambin, X., & Villard, E. (2022)	empirical-application	0	France	positive	xxx
Hussein, M. J., Yusuf, J., Deb, A. S., Fong, L., & Naidu, S. (2020)	empirical	14	Fiji	positive	identification locking monitoring
Indi, C. S., Pritham, K. C. S. V., Acharya, V., & Prakasha,	application	5	Indie	positive	monitoring

K. (2021)					
Infante Moro, A., Infante Moro, J. C., Gallardo Pérez, J., & Martínez López, F. J. (2022)	empirical	3	Spain	positive	xxx
Jadi, A. (2021)	application	3	Saudi Arabia	positive	identification locking monitoring
Khalil, M., Prinsloo, P., & Slade, S. (2022)	theoretical	0	Norway	neutral	xxx
Kilinc, H., Okur, M. R., & İlker, U. S. T. A. (2021)	empirical	0	Turkey	positive	identification locking
Labayen, M., Veá, R., Florez, J., Aginako, N., & Sierra, B. (2021)	application	6	Spain	positive	identification locking monitoring
Langenfeld, T. (2020)	theoretical	16	Iowa (USA)	neutral	locking
Lee, J. W. (2020)	empirical	7	China	neutral	identification
Lee, K., & Fanguy, M. (2022)	empirical	6	England/ South Korea	negative	locking monitoring
Li, M., Luo, L., Sikdar, S., Nizam, N. I., Gao, S., Shan, H., ... & Wang, G. (2021)	application	13	USA	neutral	xxx
Long, D. T. (2021)	application	0	Vietnam	positive	identification
Masud, M. M., Hayawi, K., Mathew, S. S., Michael, T., & El Barachi, M. (2022)	application	1	United Arab Emirates	positive	identification monitoring
Middleton, K. V. (2022)	theoretical	0	Washington (USA)	neutral	xxx
Muzaffar, A.	theoretical	16	Pakistan/Sa	neutral	identification



W., Tahir, M., Anwar, M. W., Chaudry, Q., Mir, S. R., & Rasheed, Y. (2021)			udi Arabia		
Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020)	theoretical	31	Iowa (USA)	neutral	xxx
Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021)	theoretical	10	Indie	neutral	identification locking monitoring
Raman, R., Vachharajani, H., & Nedungadi, P. (2021)	empirical	19	Indie	positive	identification locking monitoring
Selwyn, N., O'Neill, C., Smith, G., Andrejevic, M., & Gu, X. (2021)	empirical	10	Australia	negative	identification monitoring
Shaushenova, A., Zulpykhar, Z., Zhumasseitova, S., Ongarbayeva, M., Akhmetzhanova, S., Mutalova, Z.,...& Zueva, A. (2021)	empirical	1	Kazakhstan	positive	xxx
Topuz, A. C., Saka, E., Fatsa, Ö. F., & Kurşun, E. (2022)	theoretical	1	Turkey	neutral	identification locking monitoring
Valizadeh, M. (2022)	empirical	1	Turkey	neutral	xxx

## 5. Analysis

The literature search results identified principal differences in perceptions of the relationship with proctoring depending on geographic location. A generalisation of the results is provided in the adjacent graph, which visualises an increased positive aspect predominantly in Asian countries; research from American universities tends not to show their opinion in their

investigations and considers both aspects equally, while Europe demonstrates the most fragmented views of both positive and negative aspects of implementation within the selected literature sample. This could also symbolise the attitudes of the countries in question in prioritising educational strategies with an emphasis on the level of security or social/psychological components.

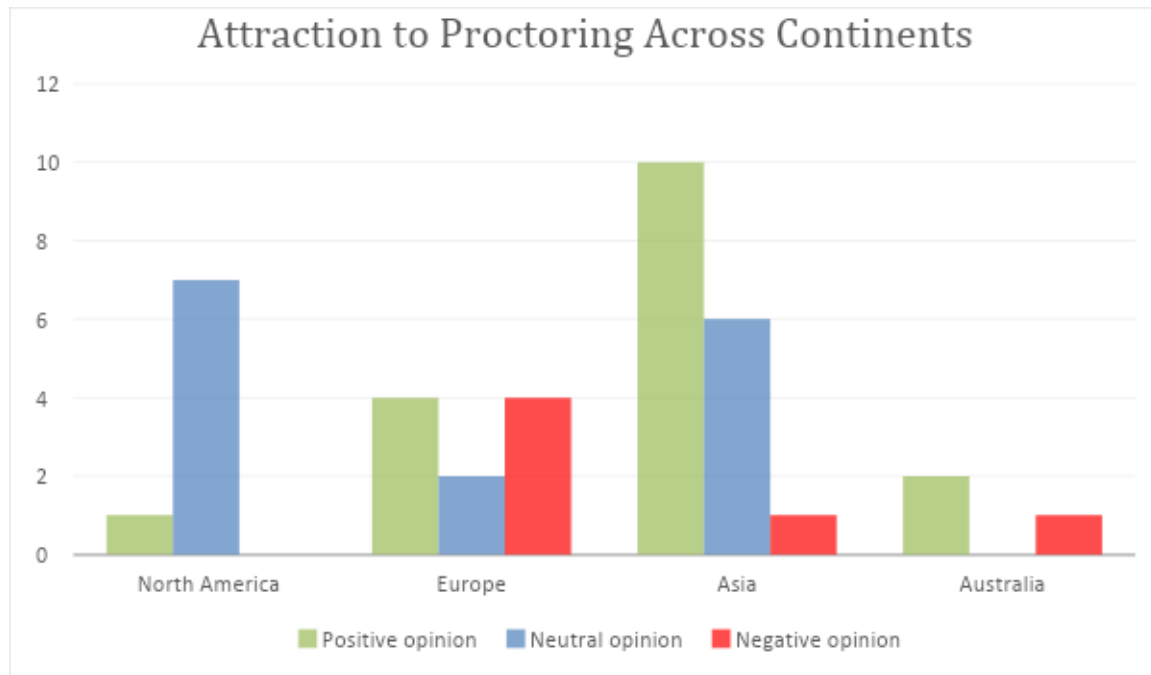


Fig. 1. Attraction to Proctoring Across Continents

### *5.1. Positive aspects of proctoring*

Within the solid sample, it can be said that texts with affirmative colouring, representing proctoring as a necessary prevention against cheating, predominate. Butler-Henderson & Crawford (2020) define proctoring as the most dominant topic area, noting that students enjoy and also quickly adopt online formats that reduce various forms of anxiety compared to traditional instruction. Garg & Goel (2022) support preventing the detection of dishonesty. Their reviews identified diverse types of dishonest student behaviour across online education, from collaborating with others to looking up topics online. They also identified the causes of increased academic dishonesty in the online environment, summarising the main factors as the lust for good grades, parental pressure, lack of time, or simple personality. Valizadeh (2022) investigated students' attitudes toward the perception of cheating in online education, where more than half of the respondents stated that affair is more straightforward in the online

environment. According to his results, the most common cheating techniques were through Google, so avoiding the copy-and-paste function was one of the common suggestions.

Proctoring is also presented in positive texts as a reductionist tool to achieve a greater degree of educational quality and a level of safety against copying and other unethical behaviour across the university. Shaushenova et al. (2021) show significant differences between proctored and unproctored examinations, with the number of offences in proctored solutions found to be significantly lower. Similarly, Humbert et al. (2022) support the idea of protected online exams. Their study investigated the effectiveness of intervention methods in the form of warning messages when cheating was detected to encourage a perception of sophistication among students.

Other texts within the security framework appeal more to the emphasis of the notion of academic integrity, which is representative of the character of the university through a well-set teaching environment. Hébert (2021) argues that online proctoring is a defence against the devaluation of university degrees in the marketplace, as cheating can significantly tarnish a university's brand. However, he does not notably support the concept of proctoring, instead advocating the need for a revolution in online assessment practices that do not monitor the student but rather support value. Some universities, however, take the concept of proctoring and academic integrity as an advantage, where a student pays for a proctored exam to prove the high credibility of the authenticity of the certifications (Scheduling a proctored exam, c2022). However, Khalil et al. (2022) argue that there is little evidence that academic integrity plays a vital role in the adoption of proctoring solutions, but rather that it is a matter of maintaining professional standards and minimising cheating, as mentioned earlier.

Most studies supporting proctoring have sought to present empirical designs or directly tested application models of their trial safety frameworks that they consider highly reliable and potentially successful based on their analyses or pre-tests. Each of these frameworks applied different authentication methods. Labayen et al. (2021) proposed a system based on continuous authentication processes, using passive monitoring capabilities via webcam, microphone, and keystroke. In his model, Jadi (2021) prohibits the operation of other applications besides the allowed auxiliary components. At the same time, students underwent facial checkpoint scanning - eyes, nose, chin, mouth, eyebrows - throughout the trial. Long (2021) bases his model on taking a picture of the face, extracting features, and classifying them. The model considers basic expressive features, neutral expressions of people, and differences in lighting or poses. Masud et al. (2022) mention the effectiveness of video capture based on eye tracking,

head movement, mouth opening, and identity tracking. The average efficiency of these frameworks varies from 90-100 %.

Topuz et al. (2022) write about semi-automated monitoring methods, video and voice analysis, screen recording, and copy/paste blocking as the most effective security proctoring features. In contrast, Indi et al. (2021) say that a highly reliable proctoring system can have even minimal requirements by simply sensing eye gaze and head position. So, in principle, retrospective or simultaneous tracking of the student's image is the most commonly used phenomenon done through proctoring or automated AI formats. In some of the cases mentioned above, we can also notice more widespread constraining formats that interfere with the compositions of the computer settings.

### *5.2. Negative aspects of proctoring*

Our study identified four key points that capture the negative differentiation of proctoring implementations. At the same time, proctoring raises concerns for many authors due to its lack of privacy protection and interference with human psychological factors. Furthermore, there are doubts about the relevance and effectiveness of the systems and doubts about the appropriateness of the technical radicalisation of the examiners' processes.

Nigam et al. (2021) categorised the problems of proctoring into two parts, technological and human, with the latter group respecting humans' psychological and socio-cultural aspects. The psychological safety of the student appeared to be a widely consulted topic in general, as we found many references to increasing stress hormones due to proctoring online exams. In Australia, for example, proctoring has caused controversy and media alarm. Selwyn et al. (2021) state that proctoring was generally only deployed there because of "technological appropriateness," and now it appears that proctoring is an underdeveloped technology causing much student stress or digital resignation. The same view is taken in a Dutch study where the authors Conijn et al. (2022) highlight that proctored exams increase test anxiety and stress levels. Balash et al. (2021) investigated students' attitudes toward online proctoring, finding that a significant intrusion on privacy and comfort mostly exacerbates students' psychological states. Students often cited webcam and microphone recording concerns because the device monitored their room.

From a statistical point of view, it is also crucial that many studies expressed an opposing sight to the study mentioned above (Shaushenova et al., 2021), demonstrating a difference between supervised and unsupervised exams. Çelikbağ & Delialioğlu (2021) show that no statistically

significant differences were found between the exams, with only one case of an unproctoring exam having a higher pass rate. Lee (2020) comments on the equal idea of no difference between exam environments. He also says that proctoring is unlikely to be related to student performance. Questioning of results also occurred because students were usually informed in advance of the proctoring format and, therefore, could prepare better for the exam and did not have to resort to copying (Howard, 2020). Thus, these results provide evidence that proctoring is unlikely to significantly impact students' scores, although the effects of divergent research are mixed. Also, these findings cannot be much related within the framework of intercultural ethical practices, as all the studies mentioned above reflect research from Asian countries.

Proctoring denial has also often been resorted to by studies that have inquired about the effectiveness of individual systems, the likelihood of errors, and the correct detection of cheating techniques. Bergmans et al. (2021) say that proctoring systems work more like some placebo effect because the sensitivity of the Proctoria system is very low. The system could not detect cheating students, so the study evaluated online proctoring as a questionable tool to ensure reliability.

The last and significant fourth identification point detected was the shift away from technology implementation of safety depending on the student's oriented needs. Lee & Fanguy (2022) describe online proctoring as teacher-centred, diminishing the values of social equity and deteriorating student-teacher relationships. The common thread among these concepts is that teachers should respect student privacy and provide well-secured test exams that naturally, without technical solutions, reduce opportunities for cheating and, therefore, false test pass rates. Nguyen et al. (2020) show that cheating can be minimised by thoughtfully composing the course or test operation. In their work, they recommend creating tests through essay questions. Other recommendations arise in a study by Becker et al. (2022), where it was reported that ordering items in a test by difficulty dramatically affects students' final scores. Students with more complex questions at the beginning and more straightforward questions at the end performed significantly worse than when the test was organised the other way around - so they consider it relevant to keep the order of questions fixed to reduce the variance in scores. These findings can be explained primarily by a relatively secure testing design that respects a strict time limit, appropriate question design, and other test settings (Langenfeld, 2020).

In the category of student-centred needs, we can also include efforts to reduce student anxiety, which was reduced in the American study by Farland & Childs-Kean (2021) by transferring classified assessments to non-classified activities. However, these solutions, we believe,

already go beyond proctoring, as it is a general definition of psychological safety in online educational formats.

### *5.3. Identification of different proctoring formats*

In addition to the positive and negative phenomena of proctoring discussed above, mainly due to the proctoring models described in detail, we were able to identify different ways of achieving safety based on differentiated emphases in constraint prioritisation. We refer to these distinct proctoring control methods as identification, monitoring, and locking. All types of proctoring can be used synergistically.

The identification methods of proctoring solutions are limited to human verification features, which can occur at any time during the test, before or after the examination. The most common authentication methods are biometrics through face recognition using a webcam or voice analysis using a microphone (Labayen et al., 2021). However, fingerprint identification can also be classified as this type.

The monitoring functions use identical restriction means, such as webcam or microphone monitoring. However, monitoring methods are already focused only on minimising cheating techniques. Thus, student monitoring occurs not to verify the identity of the correct person but because of the potential threat of exam cheating and resorting to cheating methods. Thus, the goal is usually to monitor various metrics, such as eye, mouth, and head movement or the intensity of sounds in the home or other testing environment (Nigam et al., 2021).

The last type of technical restriction is called locking devices, which temporarily block a user's device, features, or software to prevent collusion. These include prohibiting certain keyboard buttons, such as the print screen or the copy and paste function. Blocking the browser or new browser windows is also another ordinary method used (Lee & Fanguy, 2022)

Our studies show that proctoring is a highly controversial topic which institutions have approached differently—positive views on implementing proctoring consider academic integrity, minimising cheating, and creating new security frameworks. The negative side of the concept is written mainly through an emphasis on respecting the principles of privacy and most students' psychological moods or feelings. A substitute for technical proctoring concepts may be to address the complete change of grading procedures of online assessments by converting them to non-classified actions or by emphasising test-taking designed according to value-based

social principles for students. We conclude our analysis by providing one possible breakdown of proctoring methods into identification, monitoring, and locking.

## 6. Conclusion

The review study shows that proctoring is a question of technological security and a complex socio-pedagogical-technical-ethical problem. On the social level, questioning approaches emphasise the loss of trust. On the other hand, we can identify an unambiguous effort to gain surveillance in a sense already contemplated by Foucault (1975). Depilation here is not only linked to the educational content (Deacon, 2006) but also to the form of education used (Nicoll, 2008; Fawns & Schaepkens, 2022) as a process of surveillance (Venter & Van Niekerk, 2011).

At the pedagogical level, the question of whether the presence of proctoring increases the quality of educational outcomes prevails. There are many studies on this topic, but they do not offer a clear-cut answer. If positive effects exist, they are associated with other factors. However, the studies conducted are currently unable to capture and can be referred to as environmental, probably related to specific educational cultures and practices. At the same time, however, we can observe a debate about the appropriateness of using other forms of testing where cheating has no effect, such as open-book formats (Gharib et al., 2012; Brightwell et al., 2004; Eurboonyanun et al., 2021; Jaap et al., 2021).

A significant number of studies have focused on the technical side of the issue (Selwyn et al., 2021; Kharbat & Abu Daabes, 2021), although they are aware that this is only one aspect. There are views that the ideal technical solution will make proctoring a universal educational tool, as well as sceptical remarks directed towards the dysfunctionality or unreliability of particular solutions or approaches. The development of quality proctoring tools is a crucial issue, but at the same time, we must bear in mind the necessity of anchoring it in a broader context. The goal of development is to provide a system that is non-disruptive to users, secure and stable.

The ethical question (Coghlan et al., 2021; Shaikh et al., 2021) poses the fundamental dilemma of the relationship between the student and the teacher, or more broadly, the person (student and teacher) and the institution. Proctoring can be seen as a tool of control, discipline, and breach of trust. The institution, in its implementation, places the teacher in the role of a clerk in a bureaucratic apparatus, as Arendt describes it, against which it is impossible to rebel, but which at the same time destroys a very fundamental dimension of relationality, trust and freedom.

We believe that the above comments and analyses show that proctoring, in the current social climate and technological maturity of the solution, cannot be considered an ideal and fully functional solution but instead can be seen as an emergency solution where it is not possible to choose another course of action.

### Acknowledgements

The article was created thanks to the financial support of the Ministry of Education of the Czech Republic within the project MUNI 3.2.1 in the specific objective: SC C2 Safety of distance learning, project number NPO\_MUNI\_MSMT-16606/2022.

### References

Ahmad, I., AlQurashi, F., Abozinadah, E., & Mehmood, R. (2021). A Novel Deep Learning-based Online Proctoring System using Face Recognition, Eye Blinking, and Object Detection Techniques. *International Journal of Advanced Computer Science and Applications*, 12(10).

Alshammari, M. T. (2020). An Adaptive Framework for Designing Secure e-Exam Systems. *International Journal Of Computer Science And Network Security*, 20(5), 189-196. <https://www.webofscience.com/wos/woscc/full-record/WOS:000577876400024>

Arnò, S., Galassi, A., Tommasi, M., Saggino, A., & Vittorini, P. (2021). State-of-the-Art of Commercial Proctoring Systems and Their Use in Academic Online Exams. *International Journal Of Distance Education Technologies*, 19(2), 41-62.

Balash, D. G., Kim, D., & Shaibekova, D. (2021). Examining the Examiners: Students' Privacy and Security Perceptions of Online Proctoring Services. In *Seventeenth Symposium on Usable Privacy and Security* (pp. 633-652). USENIX Association. <https://www.usenix.org/conference/soups2021/presentation/balash>

Becker, B., van Rijn, P., Molenaar, D., & Debeer, D. (2021). Item order and speededness: implications for test fairness in higher educational high-stakes testing. *Assessment & Evaluation in Higher Education*, 47(7), 1-13. <https://doi.org/10.1080/02602938.2021.1991273>



Bergmans, L., Bouali, N., Luttkhuis, M., & Rensink, A. (2021). On the Efficacy of Online Proctoring using Proctorio. In *Proceedings of the 13th International Conference on Computer Supported Education* (1) (pp. 279-290). SCITEPRESS – Science and Technology Publications. <https://doi.org/DOI: 10.5220/0010399602790290>

Brightwell, R., Daniel, J. H., & Stewart, A. (2004). Evaluation: Is an open book examination easier?. *Bioscience Education*, 3(1), 1-10.

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education*, 159. <https://doi.org/https://doi.org/10.1016/j.compedu.2020.104024>

ÇELİKBAĞ, M. A., & Delialioğlu, Ö. (2021). Proctored vs Unproctored Online Exams in Language Courses: A Comparative Study.

Coghlan, S., Miller, T., & Paterson, J. (2021). Good proctor or "big brother"? Ethics of online exam supervision technologies. *Philosophy & Technology*, 34(4), 1581-1606.

Conijn, R., Kleingeld, A., Matzat, U., & Snijders, C. (2022). The fear of big brother: The potential negative side-effects of proctored exams. *Journal Of Computer Assisted Learning*, 1-14. <https://doi.org/DOI: 10.1111/jcal.12651>

Deacon, R. (2006). Michel Foucault on education: a preliminary theoretical overview. *South African Journal of Education*, 26(2), 177-187.

Eurboonyanun, C., Wittayapairoch, J., Aphinives, P., Petrusa, E., Gee, D. W., & Phitayakorn, R. (2021). Adaptation to open-book online examination during the COVID-19 pandemic. *Journal of surgical education*, 78(3), 737-739.

Farland, M. Z., & Childs-kean, L. M. (2021). Stop tempting your students to cheat. *Currents In Pharmacy Teaching*, 13(6), 588-590. <https://doi.org/10.1016/j.cptl.2021.01.035>

Farmer, J., Lachter, G. D., Blaustein, J. J., & Cole, B. K. (1972). The role of proctoring in personalised instruction1. *Journal of Applied Behavior Analysis*, 5(4), 401–404. <https://doi.org/10.1901/jaba.1972.5-401>

Fawns, T., & Schaepkens, S. (2022). A matter of trust: Online proctored exams and the integration of technologies of assessment in medical education. *Teaching and Learning in Medicine*, 1-10.

Fenu, G., Marras, M., & Boratto, L. (2018). A multi-biometric system for continuous student authentication in e-learning platforms. *Pattern Recognition Letters*, 113, 83–92. <https://doi.org/10.1016/j.patrec.2017.03.027>

Fiano, K. S., Medina, M. S., & Whalen, K. (2021). The Need for New Guidelines and Training for Remote/Online Testing and Proctoring. *American Journal Of Pharmaceutical Education*, 85(8), 805-808.

Foucault, M. (1975). Discipline and punish. A. Sheridan, Tr., Paris, FR, Gallimard.

Garg, M., & Goel, A. (2022). Systematic literature review on online assessment security: Current challenges and integrity strategies. *Computers & Security*, 113, 2-13. <https://doi.org/10.1016/j.cose.2021.102544>

Gharib, A., Phillips, W., & Mathew, N. (2012). Cheat Sheet or Open-Book? A Comparison of the Effects of Exam Types on Performance, Retention, and Anxiety. *Online Submission*, 2(8), 469-478.

González González, C., Infante Moro, A., & Infante Moro, J. C. (2020). Implementation of E-Proctoring in Online Teaching: A Study about Motivational Factors. *Sustainability*, 12(8) <https://doi.org/10.3390/su12083488>

Grande de Prado, M., García-Peñalvo, F. J., Corell, A., & Abella García, V. (2021). Evaluación en Educación Superior durante la pandemia de la COVID-19. *Campus Virtuales*, 1(10), 49-58.

Hébert, C. (2021). Online Remote Proctoring Software in the Neoliberal Institution: Measurement, Accountability, and Testing Culture. *Education*, 27(1), 23-40.

Howard, D. (2020). Comparison of Exam Scores and Time Taken on Exams Between Proctored On-Campus and Unproctored Online Students. *Online Learning*, 24(4), 204-228. <https://doi.org/10.24059/olj.v24i4.2148>

Humbert, M., Lambin, X., & Villard, E. (2021). The role of prior warnings when cheating is easy and punishment is credible. *Information Economics And Policy*, 2022(58). <https://doi.org/https://doi.org/10.1016/j.infoecopol.2021.100959>

Hussein, M. J., Yusuf, J., Deb, A. S., Fong, L., & Naidu, S. (2020). An Evaluation of Online Proctoring Tool. *Open Praxis*, 12(4), 509-525. <https://search.informit.org/doi/abs/10.3316/informit.620366163696963>

Indi, C. S., Pritham, K. C. S. V., Acharya, V., & Prakasha, K. (2021). Detection of Malpractice in E-exams by Head Pose and Gaze Estimation. *International Journal Of Emerging Technologies In Learning*, 16(8), 47-60. <https://doi.org/10.3991/ijet.v16i08.15995>

Infante Moro, A., Infante Moro, J. C., Gallardo Pérez, J., & Martínez López, F. J. (2022). Key Factors in the Implementation of E-Proctoring in the Spanish University System. *Sustainability*, 14(13), 8112. <https://doi.org/10.3390/su14138112>

Jaap, A., Dewar, A., Duncan, C., Fairhurst, K., Hope, D., & Kluth, D. (2021). Effect of remote online exam delivery on student experience and performance in applied knowledge tests. *BMC Medical Education*, 21(1), 1-7.

Jadi, A. (2021). New Detection Cheating Method of Online-Exams during COVID-19 Pandemic. *International Journal Of Computer Science And Network Security*, 2(4), 123-130

Khalil, M., Prinsloo, P., & Slade, S. (2022). In the nexus of integrity and surveillance: Proctoring (re) considered. *Journal Of Computer Assisted Learning*, 1-14. <https://doi.org/DOI:10.1111/jcal.12713>

Kharbat, F. F., & Abu Daabes, A. S. (2021). E-proctored exams during the COVID-19 pandemic: A close understanding. *Education and Information Technologies*, 26(6), 6589-6605.

Kilinc, H., Okur, M. R., & Usta, I. (2021). The Opinions of Field Experts on Online Test Applications and Test Security During the COVID-19 Pandemic. *International Journal Of Assessment Tools In Education*, 8(4), 975–990. <https://doi.org/10.21449/ijate.875293>

Labayen, M., Veal, R., Florez, J., Aginako, N., & Sierra, B. (2021). Online Student Authentication and Proctoring System Based on Multimodal Biometrics Technology. *IEEE Access*, 9, 72398–72411. <https://doi.org/10.1109/ACCESS.2021.3079375>

Langenfeld, T. (2020). Internet-Based Proctored Assessment: Security and Fairness Issues. *Journal Of Educational Measurement*, 39(3), 24-27. <https://doi.org/10.1111/emip.12359>

Lee, J. W. (2020). Impact of proctoring environments on student performance: Online vs offline proctored exams. *Journal Of Asian Finance, Economics And Business*, 7(8), 653-660. <https://doi.org/10.13106/jafeb.2020.vol7.no8.653>

Lee, K., & Fanguy, M. (2022). Online exam proctoring technologies: Educational innovation or deterioration?. *British Journal Of Educational Technology*, 53, 475-490. <https://doi.org/DOI:10.1111/bjet.13182>

Li, M., Luo, L., Sikdar, S., Gao, S., Shan, H., Kruger, M., Kruger, U., Mohamed, H., Xia, L., & Wang, G. (2021). Optimised collusion prevention for online exams during social distancing. *Npj Science Of Learning*, 6(1), 1-9. <https://doi.org/10.1038/s41539-020-00083-3>

Long, D. T. (2021). A Facial Expressions Recognition Method Using Residual Network Architecture for Online Learning Evaluation. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 25(6), 953-962. <https://doi.org/10.20965/jaciii.2021.p0953>

Masud, M. M., Hayawi, K., Mathew, S. S., Michael, T., & El Barachi, M. (2022, February). Smart Online Exam Proctoring Assist for Cheating Detection. In *International Conference on Advanced Data Mining and Applications* (pp. 118-132). Springer, Cham.

Middleton, K. V. (2022). Considerations for future online testing and assessment in colleges and universities. *Educational Measurement: Issues And Practice*, 41(1), 51-53. <https://doi.org/10.1111/emip.12497>

Muzaffar AW et al. (2021). Systematic Review of Online Exams Solutions in E-Learning: Techniques, Tools, and Global Adoption. *IEEE Access*, 9, pp. 32689-32712, doi: 10.1109/ACCESS.2021.3060192.

Nguyen, J. G., Keuseman, K. J., & Humston, J. J. (2020). Minimise Online Cheating for Online Assessments During COVID-19 Pandemic. *Journal of Chemical Education*, 97, 3429-3435. <https://doi.org/10.1021/acs.jchemed.0c00790>

Nicoll, K. (2008). Discipline and e-learning. In *Foucault and lifelong learning* (pp. 184-197). Routledge.

Nigam, A., Pasricha, R., Singh, T., & Churi, P. (2021). A Systematic Review on AI-Based Proctoring Systems: Past, Present and Future. *Education And Information Technologies*, 26(5), 6421-6445.

Raje, S., & Stitzel, S. (2020). Strategies for Effective Assessments while Ensuring Academic Integrity in General Chemistry Courses during COVID-19. *Journal of Chemical Education*, 97(9), 3436–3440. <https://doi.org/10.1021/acs.jchemed.0c00797>

Raman, R., B, S., G, V., Vachharajani, H., & Nedungadi, P. (2021). Adoption of online proctored examinations by university students during COVID-19: Innovation diffusion study. *Education And Information Technologies*, 26(6), 7339-7358. <https://doi.org/10.1007/s10639-021-10581-5>

Saunders, G., & Weible, R. (1999). Electronic courses: Old wine in new bottles? *Internet Research*, 9(5), 339–347. <https://doi.org/10.1108/10662249910297741>

*Scheduling a proctored exam.* (c2022). Acadia University: Ranked One Of The Top Undergraduate Universities In Nova Scotia. Dostupné 30 November 2022, in: <https://www2.acadiau.ca/online/current-students/exams.html>

Selwyn, N., O'Neill, C., Smith, G., Andrejevic, M., & Gu, X. (2021). A necessary evil? The rise of online exam proctoring in Australian universities. *Media International Australia*, 1-16. <https://doi.org/DOI: 10.1177/1329878X211005862>

Shaikh, S., Shahid, M., Ali, M., & Ali, A. (2021). Ethical consideration of Online Exam Supervision. *Pakistan Journal of Ethics*, 1(1), 9-13.

Shaushenova, A., Zulpykhar, Z., Zhumasseitova, S., Ongarbayeva, M., Akhmetzhanova, S., Mutalova, Z., Niyazbekova, S., & Zueva, A. (2021). The influence of the proctoring system on

the results of online tests in the conditions of distance learning. *Ad Alta: Journal Of Interdisciplinary Research*, 11(2), 250-256.

Topuz, A. C., Saka, E., Fatsa, Ö. F., & Kurşun, E. (2022). Emerging trends of online assessment systems in the emergency remote teaching period. *Smart Learning Environments*, 9(1), 1-21.

Valizadeh, M. (2022). Cheating in online learning programs: learners' perceptions and solutions. *Turkish Online Journal Of Distance Education (Tojde)*, 23(1), 195-209. <https://doi.org/10.17718/tojde.1050394>

Venter, E., & Van Niekerk, L. J. (2011). Reconsidering the role of power, punishment and discipline in South African schools. *Koers*, 76(2), 243-260.

Wuthisatian, R. (2020). Student exam performance in different proctored environments: Evidence from an online economics course. *International Review of Economics Education*, 35, 100196. <https://doi.org/10.1016/j.iree.2020.100196>

### About Authors

**Michal CERNY** is an assistant professor in the Department of Information Studies and Library Science, where his research focuses on technology in education and its philosophical and sociological reflection. As a senior researcher, he led a project on information literacy and now works as a researcher in projects on personal learning environments and immersive virtual reality learning. He received his PhD in Social Pedagogy when his work focused on the changing philosophy of education in the context of technological change.

**Dominika SOLCOVA** works as a researcher at the Department of Information Studies and Librarianship, Faculty of Arts, Masaryk University in Brno. She has been involved in the project National Recovery Plan for restoring the safety of distance education in the Czech Republic, where she informs about security aspects abroad. She specialises in the topics of distance education, online courses, online tutoring, cyber security, psychological safety, and proctoring.



<http://jates.org>

Journal of Applied  
Technical and Educational Sciences  
jATES

ISSN 2560-5429



**Towards Developing an Immersive Virtual Reality  
Applications for Supporting Vision Screening –  
A User Study**

Are Dæhlen<sup>1</sup>, Ilona Heldal<sup>1</sup>, Jozsef Katona<sup>1,2,3</sup>

<sup>1</sup>Western Norway University of Applied Sciences, Electrical Engineering and Mathematical Sciences,  
Department of Computer Science, Bergen, Norway, [are.daehlen@gmail.com](mailto:are.daehlen@gmail.com)

<sup>1</sup>Western Norway University of Applied Sciences, Electrical Engineering and Mathematical Sciences,  
Department of Computer Science, Bergen, Norway, [ilona.heldal@hvl.no](mailto:ilona.heldal@hvl.no)

<sup>1</sup>Western Norway University of Applied Sciences, Electrical Engineering and Mathematical Sciences,  
Department of Computer Science, Bergen, Norway, [Jozsef.Katona@hvl.no](mailto:Jozsef.Katona@hvl.no)

<sup>2</sup>University of Dunaujvaros, Informatics Institute, Department of Software Development and Application,  
Dunaujvaros, Hungary, [katonaj@uniduma.hu](mailto:katonaj@uniduma.hu)

<sup>3</sup>Obuda University, Trefort Ágoston Center for Engineering Education, Institute of Electrophysics, Kandó  
Kálmán Faculty of Electrical Engineering, Budapest, Hungary, [katona.jozsef@kvk.uni-obuda.hu](mailto:katona.jozsef@kvk.uni-obuda.hu)

**Abstract:** Functional Vision Problems (FVPs) are problems related to eye musculature and/or eye coordination rather than visual acuity. Such problems are rarely diagnosed through standard vision control, and can get worse with age if not properly treated. However, a thorough vision screening of many requires enormous resources from society, which is not possible today. This paper illustrates current challenges to screen functional vision problems (FVPs) using immersive VR applications. For this, a laptop-based program for screening oculomotor difficulties (OMDs), which is a particular type of FVP, is transferred to a head-mounted display (HMD) with integrated eye-tracking technology (ET). The program records the participants' small eye movements via ETs, e.g., fixation stability, saccades, and smooth pursuits during task performance, and allows examining irregularities pointing to OMDs. The paper illustrates the iterative development of the immersive VR program based on continuous feedback from a vision expert. Data was collected from a vision expert and 7 end-users. The results include user experience and usability evaluation of the immersive VR program compared to the laptop version. It discusses the necessity for having vision expertise during specific faces of the development process and presents critical issues for further development of applications utilizing eye movements for more accurate support for vision screening, e.g., issues regarding measuring depth perception and hand-eye coordination.

**Keywords:** Head mounted display (HMD); Virtual Reality (VR); Eye-tracking (ET); User Experience Questionnaire (UEQ); Functional Vision Problem (FVP); Oculomotor difficulty (OMD);

## 1. Introduction

Vision problems can severely impact a person's daily life, and often degrade eyesight over time. This makes it hard for the affected party to notice something is wrong themselves, and we should attempt to provide mechanisms for obtaining screening services to those with limited access to specialized aid (Preslan, 1996). Standardized screening of vision is performed on children at age 4-5 in most European countries (Wallace, 118), however, FVPs can often be missed during these screenings (Willhelmsen, 2015) (Ali, 2020) (Heldal, 2021). As FVPs can lead to diseases such as amblyopia and strabismus, it is imperative that we provide a thorough screening of FVPs during standardized screening procedures (see e.g., (Lazarus, 2021)). Contemporary research shows that technologies such as VR and ET can have great effects for the detection and rehabilitation of vision problems (Nowak, 2018) (Backus, 2020) (Mishra, 2020).

A functional vision screening is performed to determine if the users' eyes are coordinated correctly. This is time-consuming and resource-intensive (Beauchamp, 2010). There are promises from technologies to support screening from accessible technologies with possibilities to measure eye coordination (Wallace, 2018) (Ali, 2021).

A previous study developed a computer-supported program (C&Look) for screening functional vision to determine such vision problems by registering fixation stability, saccades, and smooth pursuits (Wilhelmsen, 2015) (Wallace, 2018). This application was designed to complement manual vision screening by experts (Eide, 2019). The C&Look program needs a laptop computer and an eye-tracker (ET), and indicates possible oculomotor problems, a coordination problem between the two eyes. Its records movements from both eyes separately when the user is performing structural tasks on a computer screen. The objects (stimuli) may move with saccadic or smooth movements in a horizontal, vertical, or diagonal direction. By recording the gaze points following the stimuli on the screen, the eye gaze and coordination between the left and the right eye can be measured. Since many functional vision problems can be improved by training, a follow-up study developed programs including games supporting training to correct the problems (Heldal, 2021).

This study investigates how an ET-based immersive VR software can complement the current vision screening battery in relation to a computerized method and its current challenges. The first step towards this ambition is transferring the C&Look application (Eide, 2019) to immersive VR. Such an application would better support detection of functional vision



problems which is a highly special problem, with many promises from added values from virtual reality and considering cognitive aspects (Wijkmark, 2021).

## 2. Related literature

The laptop screening and training program (C&Look) faces challenges regarding the limited computer screen sizes where the stimuli can be represented, or regarding head movements during examination, and understanding how the gaze works for perceiving depth. This prohibits measuring important eye movements which can be approximated with manual screening, e.g., convergence or peripheral vision (Eide, 2019). If one could measure these aspects and address these challenges, vision screening could be achieved easier and could produce more evidence, i.e. measurements to illustrate problems and improvements.

Utilizing VR to better understand vision is challenging for several research teams (Bennett, 2019) (Lambooi, 2009) (Grassini, 2021). Such cases indicate a disconnect between researchers and vision experts, where cooperation with specialists from the domain of vision science could help direct research and development. By utilizing expert opinions from both fields, related research can achieve higher confidence in their results and measurements.

The primary goal of Human-Computer Interfaces (HCI) is to provide interfaces that enable efficient communication between computers and their users. Brain-computer interfaces (BCIs) are able to provide a kind of alternative communication channel between the human brain and some device that is intended to be used, controlled or controlled. Among other things, BCI interfaces can contribute to the investigation of human attention (Katona, 2014), memorization and indirectly the process of human learning, thereby serving as a kind of support and predictive system to increase the efficiency of human learning. (Katona & Kovari, 2018a, 2018b) In addition to BCI, the information processing process can also be examined using ETs, and even complex cognitive processes such as programming can be analyzed. More and more complex program systems result in more and more complex source codes, where possible error correction, maintenance and further development present challenges to developers, and may even mean the end of the application. For this reason, programmers use more and more new technologies, but the efficiency of development could be further increased if they could objectively examine which techniques can be used with less cognitive load for individuals, and even more so for development groups. From the eye movement parameters, it can be concluded that, for example, in which cases the use of the Language-Integrated-Query (LINQ) query syntax (Katona, et al., 2020, Katona, 2021b) means a lower mental load, as well as what

algorithm description tools the developers can use easier to interpret (Katona, 2022), and the quality of the source code (Katona, 2021a) and its readability can also be objectively measured.

In relation to the importance of considering cognitive issues in VR for training and examination, cognitive aspects of recognition, learning, decision making and problem solving are adequately supported (Heldal, 2004). Literature within the field of visual perception can help possible VR screening programs take advantage of these unique benefits, and an interdisciplinary approach during development has the possibility to improve this aspect further (Wallace, 2018).

While VR has seen growing interest in multiple fields of research, its usage within the domain of vision therapy remains limited. Bibliometric analysis by Ali et al. (Ali, 2022a) shows that VR has seen increased attention within research, for example by utilizing games, see Ali et al (Ali, 2022b), although publications related to vision therapy and screening remain limited compared to other fields.

### 3. Study design

#### 3.1. Methodology

The design science research paradigm was used for the project, combined with a testing battery for data collection proposed by Heldal (Heldal, 2004). This includes a cyclical development strategy, where the application is revised based on results and feedback from evaluation with a domain expert regarding issues important for problem solving (task performance), the utilized social context (influences from the place and people involved), and technical issues (challenges from the applied technologies). By constructing an early prototype VR vision screening application and testing it on a vision expert, future development was influenced through direct feedback collected via interviews. This led to changes to the user interface, an implementation of a task replay system, and less focus on attempting to measure attention in VR. New development instead focused on utilizing peripheral vision, depth, and the added sense of presence in VR, as suggested by the vision expert.

A usability and user experience test was then performed on the revised application to determine how it compares to other available products on the market. This final evaluation test also focused on comparing the new VR solution to the original laptop version of C&Look. Data collection methods used during this test were open-ended questions, interviews, observations, User Experience Questionnaires (UEQs), and we asked some questions focusing on the users' presence during the VR experience. A total of 7 participants took part in the test, with each

participant performing the same predefined screening tasks in each application. These participants were of varying ages, between 14 to 35, with different technological backgrounds and genders.

### 3.2. Technology Description

To transfer C&Look into an immersive VR application using eye behavioural data, and to investigate how to complement manual vision screening and C&Look, the prospects and limitations of using a head-mounted display (HMD) are analysed. The HMD used is a Varjo VR2-Pro, as it is one of the leading VR headsets with embedded ET together with hand controls (see Fig. 1).



Fig 1. The used Varjo HMD.

The sampling rate of integrated ET in the Varjo HMD is 100Hz. Varjo also allows investigating hand and eye coordination, and problems with how the eye is functioning is highly correlated with other functions, e.g. balance or hearing (Ali, 2021).

The 2-dimensional version of C&Look provides high-quality data when testing for oculomotor problems (OMD) but lacks the possibility of capturing important tasks performed during special performance that is included in manual vision screenings. The laptop used for testing the original version of C&Look had a screen size of 14 inches and utilized a Tobii 4C mobile eye-tracker with a sampling rate of 90Hz.

## 4. Application Overview

The foundation for implementing C&Look in VR is designed to be easy to use, similar to the computer application, and include possibilities to add new tasks to it. These applications point to several problems of understanding measurements in space for using VR in general, depending on where the user is looking, where they are in space and how they comprehend objects (Lambooi, 2009) (Grassini, 2021) as well as problems related to understanding depth (Wijkmark, 2021). A current shortcoming of the computer C&Look application is the

functionality to approximate eye movements for depth perception. While this seems like a suitable task for VR since one can measure gaze, target position and depth, the problem lies in the accuracy of the ET integrated in the HMD. Therefore, the first step was to plan a screen into the VR environment. The program requires a scalable solution. Since the recording of eye-movement data is standardized, this makes the process of storing data easy. However, even for this step, the collection of eye-tracking data for the current technology can be inconsistent. While with closer targets, the errors are smaller, but the possible measurements may not measure accurately enough for increased distances needed to experience various depth. The measures of depth at the current stage are not reliable. Small changes to the actual gaze vector get significantly larger when objects are further away. The small errors from the eye tracker turn into data not close enough to the desired target. Inconsistency in the eye-tracker also leads to loss of data for some frames, reducing the reliability of data visualization even further.

A similar issue was discovered when attempting to measure hand-eye coordination using the hand-tracking technology of the Varjo VR2-Pro. While satisfactory on big hand movements and suitable for navigating menus, precise hand positioning and collision detection for this hand-tracking technology is inconsistent. This leads to unreliable data when attempting to measure small movements.

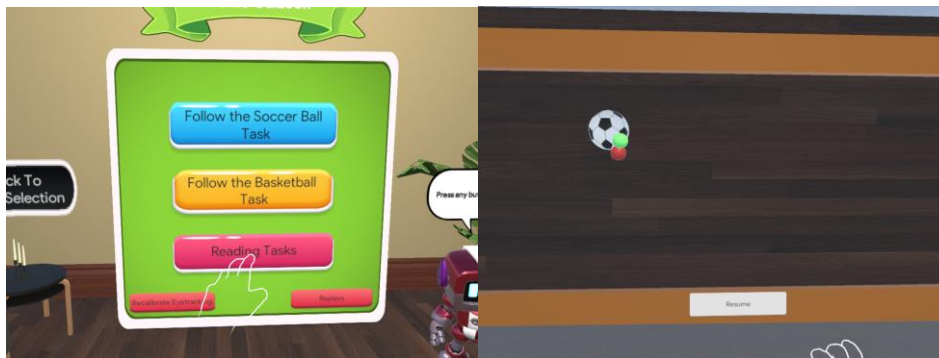


Fig. 2. A screenshot from the virtual reality C&Look with gaze points enabled.

Fig 2 contains two screenshots taken while the application is running and recording eye tracking data for the person using the application (see Fig. 1). The first screenshot is from the main menu where the user can select which task to perform. The second screenshot is from a replay of a task following a soccer ball where gaze points are being displayed. The red ball is a visual representation of the gaze point of the left eye, and the green ball is the gaze point of the right eye.

## 5. Results

After collecting data from the participants during the final test, UEQ answers for both the laptop and VR version were analysed using the UEQ benchmark proposed by Schrepp et al. (Schrepp, 2017). This resulted in scores in 6 categories ranging from -3 to 3. These scores led to classifications of either bad, below average, above average, good or excellent for each category. Fig 3 shows the scores for each category, while Table 1 depicts their final rankings.

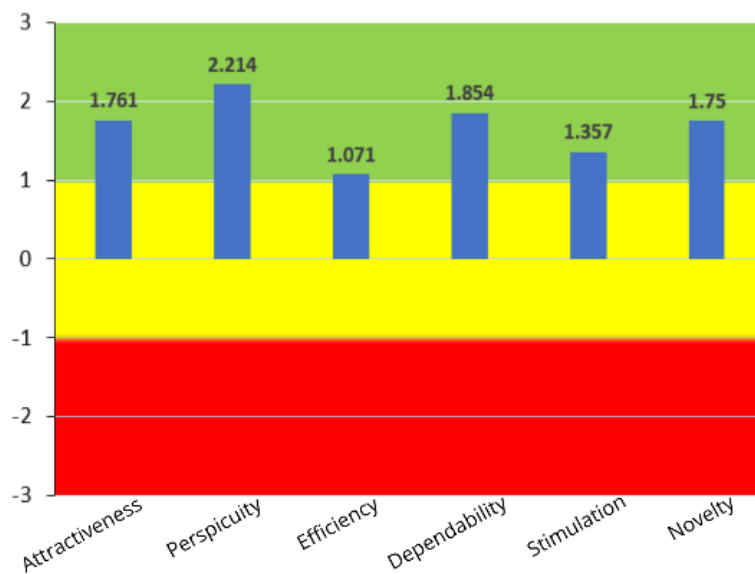


Fig 3. UEQ Scores from C&Look in VR

Table 1. UEQ classification from C&Look in VR

Category	Attractiveness	Perspicuity	Efficiency	Dependability	Stimulation	Novelty
Value	1.761	2.214	1.071	1.854	1.375	1.75
Classification	Excellent	Excellent	Below Average	Excellent	Above Average	Excellent

These results show that efficiency and stimulation are current weakness in the VR application which needs to be addressed, as their classification results in below and above average. Compared to the scores produced by the laptop application, these areas are the only ones to receive a lower classification. This indicates a need for better optimization of the application, as well as the addition of more interactive features to increase stimulation. These points were echoed in interviews, observations, and other data collection methods used during the evaluation.

To evaluate the users' experiences in the VR application, we collected data about their experiences in the two environments using the application on the laptop and the VR technologies. These questionnaires ask each participant to rate elements of the application on a scale from 1 to 7, where 7 feels very similar to being examined professionally, and 1 the opposite. Figure 4 shows average scores of experienced presence related to the users' overall experiences, engagement, involvement, the focus of attention, their experiencing time, and how they related the experiences in the application compared to real memories when they had eye screening. The VR version scores higher in almost every single category, indicating the added feeling of presence. "Compared to real memories" is the only exception, as those who had already been examined before founding the low confidence in results and lower quality replay function distracting.

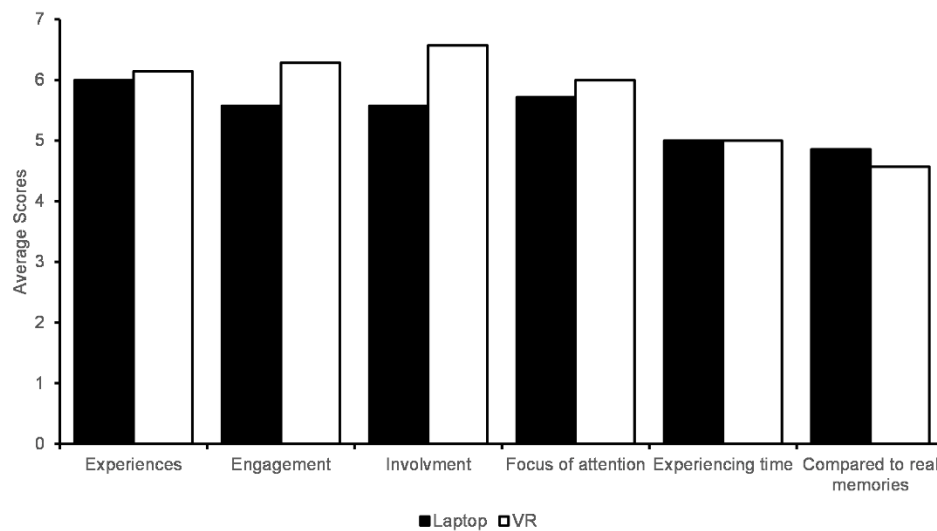


Fig 4. Average scores of presence from C&Look on a laptop and in VR.

## 6. Discussion and Limitations

The main goal of this project was to examine the possibilities of using immersive VR to assist a vision screening in detecting functional vision problems. This is done by developing a vision screening suite for VR and comparing it to a previously developed and utilized laptop solution developed for the same purpose.

Today, measuring small eye movements combining HMD and ET via applications meets several hinders for measuring exact distances in space, especially eye-movements for identifying depth, but also understanding peripheral vision. It is not a direct correlation between the capability of HMD and eye-tracking and understanding the space through the eyes. Therefore, as a first step, we implemented C&Look in immersive VR, allowing it to work on a

visualized plan-screen in the HMD and perform the eye measurements only based on that screen. Experimental development of depth perception and hand-eye coordination testing was also performed during this project, however, these solutions proved unreliable and difficult to implement due to low quality ET and hand-tracking data. Depth perception testing has been proven to be compatible with VR by VividVision (Backus, 2020) but would require a larger development team and more time.

Interdisciplinary cooperation with a vision expert proved effective when choosing a direction for future development after the initial prototype. This communication was lost in later stages of the project. For the future this should be prioritized to a greater extent, having the possibility to lead to exact demands for quality and confidence of measurements. When data quality is poor, other possible avenues for screening in VR could be proposed by someone with a background in vision science.

Several limitations were discovered during this study. Firstly, the eye-tracking hardware of the Varjo VR-2 Pro lacks the ability to provide accurate positions of gaze points while returning no missing data. This gives less confidence to the results produced by the application. Secondly, the current database transaction methods in the application rely on a separate server and custom API. This was the result of a lack of Unity SDKs compatible with PostgreSQL, resulting in significant technical debt if the current solution is to be developed further. Thirdly, access to only one HMD for testing led to a smaller participant pool than suggested by Schrepp et al. (Schrepp, 2017). Testing of each participant lasted between 1-2 hours and a larger pool of participants would necessitate being able to test multiple subjects at the same time. The final limitation comes from the lack of available resources and prior research related to the Varjo VR2-Pro HMD. Very few publications and online resources related to this HMD exist, making development for this environment time-consuming and difficult when utilizing features such as eye-tracking and gaze point visualization.

## 7. Conclusion

This study presents a prototype version of a VR application that, if developed further, can help assess functional vision problems. Results from user experience and usability testing show that the current version suffers from a low score in efficiency, while performing well in most other categories. These efficiency problems, as well as other limitations, are closely related to the used HMD, so utilizing a more researched and reliable HMD should be prioritized for the future. Test participants report an added sense of presence in VR, highlighting the unique

immersive elements provided by HMDs. Experimental development for accurate depth perception and hand-eye coordination measurements highlight the need for high precision data concerning both gaze and hand positioning. More communication with researchers from the vision science domain have the ability to identify such problems early, while also being able to propose alternative directions when data quality suffers. The current prototype lays the foundation for developing a VR application that can help vision experts diagnose functional vision problems in a real-world screening.

### Acknowledgements

The authors would like to thank Qasim Ali for much help with technical implementation and tests, Eva Bjånes for helping with domain expertise from vision science and the EU project EMPOWER, no. 101060918 for financial support. We also thanks thank the students participating in the evaluations.

### References

- Ali, Q., Heldal, I., Eide, M. G., Helgesen, C. G., & Wilhelmsen, G. B. (2020). Using Eye-tracking Technologies in Vision Teachers' Work—a Norwegian Perspective. In 2020 International Conference on e-Health and Bioengineering (EHB) (pp. 1-5). IEEE.
- Ali, Q., Heldal, I., & Helgesen, C. G. (2022a). A bibliometric analysis of virtual reality-aided vision therapy. *Studies in Health Technology and Informatics*.  
<https://doi.org/10.3233/shti220781>
- Ali, Q., Heldal, I., Helgesen, C. G., Krumina, G., Costescu, C., Kovari, A., Katona, J., & Thill, S. (2021). Current challenges supporting school-aged children with Vision Problems: A Rapid Review. *Applied Sciences*, 11(20), 9673. <https://doi.org/10.3390/app11209673>
- Ali, Q., Heldal, I., Helgesen, C. G., & Dæhlen, A. (2022b). Serious Games for Vision Training Exercises with Eye-Tracking Technologies: Lessons from Developing a Prototype. *Information*, 13(12), 569.
- Backus, B. T. (2022, May). Standardized Stereoacuity Testing in Vr: The Vivid Vision Stereoacuity Test. I heart VT 2020.
- Beauchamp, G. R., Ellepola, C., & Beauchamp, C. L. (2010). Evidence-based medicine: The value of vision screening. *American Orthoptic Journal*, 60(1), 23–27.  
<https://doi.org/10.3368/aoj.60.1.23>
- Bennett, C. R., Bex, P. J., Bauer, C. M., & Merabet, L. B. (2019). The assessment of visual function and Functional Vision. *Seminars in Pediatric Neurology*, 31, 30–40.  
<https://doi.org/10.1016/j.spen.2019.05.006>



Eide, M. G., Heldal, I., Helgesen, C. G., Birkeland Wilhelmsen, G., Watanabe, R., Geitung, A., Soleim, H., & Costescu, C. (2019). Eye-tracking complementing manual vision screening for detecting oculomotor dysfunction. 2019 E-Health and Bioengineering Conference (EHB). <https://doi.org/10.1109/ehb47216.2019.8969956>

Grassini, S., & Laumann, K. (2021). Immersive Visual Technologies and human health. European Conference on Cognitive Ergonomics 2021. <https://doi.org/10.1145/3452853.3452856>

Katona, J. (2014, November). Examination and comparison of the EEG based Attention Test with CPT and TOVA. In 2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI) (pp. 117-120). IEEE.

Katona, J. (2021a). Clean and dirty code comprehension by eye-tracking based evaluation using GP3 eye tracker. *Acta Polytechnica Hungarica*, 18(1), 79-99.

Katona, J. (2021b). Analyse the readability of LINQ code using an eye-tracking-based evaluation. *Acta Polytech. Hung*, 18, 193-215.

Katona, J. (2022). Measuring cognition load using eye-tracking parameters based on algorithm description tools. *Sensors*, 22(3), 912.

Katona, J., & Kovari, A. (2018a). Examining the learning efficiency by a brain-computer interface system. *Acta Polytechnica Hungarica*, 15(3), 251-280.

Katona, J., & Kovari, A. (2018b). The evaluation of bci and pebl-based attention tests. *Acta Polytechnica Hungarica*, 15(3), 225-249.

Katona, J., Kovari, A., Heldal, I., Costescu, C., Rosan, A., Demeter, R., ... & Stefanut, T. (2020, September). Using eye-tracking to examine query syntax and method syntax comprehension in LINQ. In 2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000437-000444). IEEE.

Heldal, I. (2004). The usability of collaborative virtual environments: Towards an evaluation framework. *Doktorsavhandlingar vid Chalmers Tekniska Hogskola*.

Heldal, I., Helgesen, C., Ali, Q., Patel, D., Geitung, A. B., & Pettersen, H. (2021). Supporting school aged children to train their vision by using serious games. *Computers*, 10(4), 53. <https://doi.org/10.3390/computers10040053>

Lambooi, M., Fortuin, M., Heynderickx, I., & IJsselsteijn, W. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A Review. *Journal of Imaging Science and Technology*, 53(3). <https://doi.org/10.2352/j.imagingsci.technol.2009.53.3.030201>

Lazarus, R. (2021, October 19). Convergence insufficiency. *Optometrists.org*. Retrieved February 6, 2023, from <https://www.optometrists.org/vision-therapy/vision-therapy-for-children/convergence-insufficiency-2/>

Mishra, S., Kim, Y.-S., Intarasirisawat, J., Kwon, Y.-T., Lee, Y., Mahmood, M., Lim, H.-R., Herbert, R., Yu, K. J., Ang, C. S., & Yeo, W.-H. (2020). Soft, wireless periocular wearable electronics for real-time detection of eye vergence in a virtual reality toward Mobile Eye Therapies. *Science Advances*, 6(11). <https://doi.org/10.1126/sciadv.aay1729>

Nowak, A., Woźniak, M., Pieprzowski, M., & Romanowski, A. (2018). Towards amblyopia therapy using mixed reality technology. Proceedings of the 2018 Federated Conference on Computer Science and Information Systems. <https://doi.org/10.15439/2018f335>

Oates, B. J. (2006). *Researching Information Systems and computing*. Sage.

Pai, Y. S., Tag, B., Outram, B., Vontin, N., Sugiura, K., & Kunze, K. (2016). GazeSim. ACM SIGGRAPH 2016 Posters, 1–2. <https://doi.org/10.1145/2945078.2945153>

Preslan, M. W., & Novak, A. (1996). Baltimore Vision Screening Project. *Ophthalmology*, 103(1), 105–109. [https://doi.org/10.1016/s0161-6420\(96\)30753-7](https://doi.org/10.1016/s0161-6420(96)30753-7)

Schrepp, M., Hinderks, A., & Thomaschewski Jörg. (2017). Construction of a benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(4), 40. <https://doi.org/10.9781/ijimai.2017.445>

Wallace, D. K., Morse, C. L., Melia, M., Sprunger, D. T., Repka, M. X., Lee, K. A., & Christiansen, S. P. (2018). Pediatric eye evaluations preferred practice pattern ®: I. Vision screening in the primary care and community setting; II. Comprehensive ophthalmic examination. *Ophthalmology*, 125(1), P184-P227.

Wijkmark, C. H., Metallinou, M. M., & Heldal, I. (2021). Remote virtual simulation for incident commanders—cognitive aspects. *Applied Sciences*, 11(14), 6434–6434. <https://doi.org/10.3390/app11146434>

Wilhelmsen, G. B., Aanstad, M. L., & Leirvik, E. I. B. (2015). Implementing vision research in special needs education. *Support for learning*, 30(2), 134-149.

### About Authors

**Are Dæhlen** is a PhD Candidate at the Department of Computer Science, Electrical Engineering and Mathematical Sciences at Western Norway University of Applied Sciences. He received an MSc degree from Western Norway University of Applied Sciences in Software Engineering, about using virtual reality, eye-tracking technologies, and serious games to support functional vision problem screening of children. His current research focuses on utilizing eye-tracking technology to define more supportive serious games for children diagnosed with an autism spectrum disorder.

**Ilona Heldal** is a PhD is Professor (Chair) of Informatics within the area of Interactive Systems at the Faculty for Engineering and Science, Western Norway University of Applied Sciences. Her research experiences were gained at Chalmers University of Technology, Sweden (PhD in usability evaluation of virtual environments) and post-doctoral work in technology management at Chalmers, Uppsala University, University of Skövde, and University West, Sweden. Her research focuses on designing, developing, and implementing new technologies

---

such as virtual environments/virtual reality technology, serious games, eye-tracking technologies for training, education, and work for specific user requirements and environments.

**Jozsef Katona** Jozsef Katona is an Associate Professor of Informatics within the area of Human-Computer Interfaces and a Head of Department at University of Dunaujvaros, as well as an Associate Professor at Obuda University and a Postdoctoral Fellow at the Faculty for Engineering and Science, Western Norway University of Applied Sciences furthermore. His current research focuses on software engineering, eye-tracking systems, cognitive psychology and education.



<http://jates.org>

Journal of Applied  
Technical and Educational Sciences  
jATES

ISSN 2560-5429



## The relationship between the supervisors and the supervisee: Experiences of PhD completers

Tehmina Khan <sup>1</sup>, Eva Mikuska <sup>2</sup>

<sup>1</sup> University of Surrey, School of Law, Stag Hill Campus, Guilford, GU2 7XH, [t.khan@surrey.ac.uk](mailto:t.khan@surrey.ac.uk)

<sup>2</sup> University of Chichester, College Line, Chichester, PO19 6PE, UK, [e.mikuska@chi.ac.uk](mailto:e.mikuska@chi.ac.uk)

**Abstract:** *There is a growing pool of literature addressing doctoral students' experiences, especially how the supervisory relationship prompted their poor wellbeing during their study. The purpose of this article is to offer additional critical discussion on PhD completers' experiences, and to examine the influence that their supervisors had on completing their thesis. A qualitative approach to research is applied to capture the experiences of eight PhD students, who were also senior lecturers in various higher education institutions, and who recently completed their studies in the United Kingdom. Data collected from semi-structured interview were analysed using a constructionist narrative approach, where stories were co-constructed between the participants and the researchers. Our research findings show that for some of these completers, success was not merely based on their individual determination, nor was the role of supervisors, but support from their family and friends. This research offers an insight into the positive and negative experiences of PhD completers which may help supervisors and other academic colleagues in universities in England and globally, on how to support students and how to be part of a successful PhD completion team.*

**Keywords:** *PhD completers; Supervisory role; Emotionality;*

### 1. Introduction

A PhD in the UK context is a research degree and is the highest award available at university level. The PhD study is based around a substantial research project on an area of academic interest, typically up to 100,000 words in length, written as a thesis which then must be defended in an oral examination in front of a panel of experts. Students are assigned a supervisor (or supervisors) and the duration of a PhD is typically three years full-time and up to six years part-time (Philips and Johnson, 2022). According to HESA (2019/20) the uptake of PhD studies in the UK increased by 4,500 from 98,525 in 2015/16 to 102,030 in 2018/19. This was a gradual, yet steady increase. However, this figure dipped in 2019/20 by 9% to 101,350. This interruption in the growth of PhD candidates can be contributed to the Covid-19 pandemic, when individuals were concerned about protecting their finances, as opposed to investing in PhD studies.

Sverdlik's et al. (2018) research in the USA reported a similar trend, that there was an increase in enrolments to the PhD programme between 1998 and 2010. Their study focused on determining the factors that motivated students to succeed and the factors that featured in students dropping out of the programme. They noted that the dropout rate over the last 50 years on the programme was high, at over 50%. Sverdlik's et al. (2018) research concluded that the single most influencing external factor, which determined the success of the PhD student, was the relationship that they had with their supervisor.

Generally, the context of research on students in higher education has mainly focused on undergraduates, especially relating to motivation, achievement and their well-being (Kwan, 2009; Mikuska, 2016). Although, there is a growing pool of literature addressing PhD and doctoral students' experiences, this topic is largely overlooked; therefore, this article offers an insight into eight PhD completers, who were also senior lecturers in various higher education institutions, in England. As part of the participants roles at the university, they were offered an opportunity to enrol on the PhD programme. With a competitive labour market (Morrison, 2012), they all took up this opportunity in order to gain higher and specialist qualification as well as to preserve their current job.

## 2. What the literature tells us

Kwan (2009) stated that one of the major tasks in a PhD undertaking is reading in preparation for the thesis. Students need to be prepared to read extensively and thoroughly to collect ideas that can inform their studies and that can be used to construct the literature review and other parts in their thesis. However, undertaking PhD studies while being employed full-time can be a source of difficulties. For example, Ali and Kohun (2006) examined the postgraduate experience affecting the progress and completion of research degrees, and stated that there are two main factors influencing students' experiences: university factors and student factors. They suggested that university factors typically include fitting in with the supervisor and institutional (particularly departmental) expectations and regulations. On the other hand, student factors often include demographic characteristics, disciplinary background, and ability, as well as students' personal life constructions (for example financial support, living arrangements, number of dependents, etc.). Phillips and Johnson (2022) also identified factors that can be readily categorised as external and internal to the student. External factors are all relationships and structures, which involve individuals, resources, and institutions outside the student that may either directly or indirectly impact PhD progress. These include supervision, personal and

social lives, the department and socialisation, and financial support opportunities. Conversely, internal factors concentrate on inner processes (for example psychological or/and mental processes), which are directly associated with academic work, specifically motivation, writing skills, self-regulatory strategies, and academic identity. An additional factor was highlighted by Heath's (2010) quantitative study on PhD students' views of supervision. This study illustrated the decreasing frequency of formal meetings between students and supervisors; however, a shift towards an increase of meetings by the end of the supervision was observed (Heath, 2010).

Ali and Kohen's (2006) study identified that the supervisory relationship between the student and supervisor(s) is one of the most important elements of success to a PhD completer. They reported that the relationship is multi-layered and covered further aspects such as commitment and respect for timeframes, through to reputation, job opportunities and academic relationships of the supervisor. Similarly, Ridgway (2022) drew upon autoethnography to document her PhD experience, and reported the importance of a good relationship between her and her supervisor. Ridgway was consumed with the emotions that she felt through bereavement of losing her parents during her PhD journey. Ridgway (2022) stated that the PhD journey is a 'slow and steady race' (p.3) where time does not stop, and the world continues revolving outside of the student's PhD journey. This includes the external factors that may include the loss of loved ones which may affect the supervisory relationship, as grief affects individuals in different ways.

Ridgway (2022) further argued that some students challenged with bereavement may become withdrawn and display a lack of interest in completing the programme. Whereas, other students may throw themselves into the process in memory of their loved one and dedicate the PhD qualification to them. Either-way, there would be a shift in the students' engagement with the programme. Mikuska's (2014), Sisson and Jackman's (2019) as well as Hunter and Devine's (2016) research highlighted the importance of emotional experiences of research students in Higher Education (HE) as it has an impact on student's achievements, progression and completion. Sisson and Jackman (2016) for example, argued how good and poor psychological wellbeing fluctuates across the timespan of the PhD. While Hunter and Devine (2016) stated that those students who report lower psychological wellbeing are more likely to consider withdrawing from their studies. McCray and Joseph-Richard (2020) argued that participants (completers) in their study had been able to access the universities emotional wellbeing support services. These students had self-referred themselves for this support. However, universities neglected to identify this need and continued to focus on the wellbeing of undergraduate

students. Similarly, Beasy's et al. (2019) research suggested that many PhD students experienced 'shallow support'. Thus, highlighting the importance of developing understanding of the factors that contribute to completing the study, as well as to promoting wellbeing in this population.

### **3. Methodology**

In this study, the aims were to investigate PhD completers' experiences, and to examine the influence that their relationship with their supervisors had on completing their thesis. A qualitative methodological approach was deployed, and after ethical permission was granted by the university ethics committee, careful reference was made to the ethical guidelines of the British Education Research Association [BERA] (2018); for example, anonymity was exercised, no traceable information was used and data was securely stored.

#### *3.1. Methods of data collection*

To address our aims, we collected narratives, using semi-structured interviews, with PhD completers who were, at the time of the interview, working in eight different Higher Education Institution (HEI). The use of semi-structured interviews proved to be an appropriate choice, as it provided us with rich and detailed data. As a support we designed a draft topic guide for us, which formed part of the deliverables of the project. The interview questions were open to allow the participants to tell their stories (Bruner, 2004). The interviews were conducted through face-to-face meetings and through online platforms such as Microsoft Teams and Zoom; they were recorded, transcribed and analysed under two main themes: i) positive experiences of PhD completers ii) negative experiences of PhD completers. Each interview lasted, on average, one hour.

#### *3.2. Researchers*

It is also important to address our doctoral experiences. As researchers, we found it difficult to balance between analysing the narratives on the experiences of PhD completers and, at the same time, avoiding making judgements. This was due to our own doctoral experiences and some of the small stories resembling our experiences. Denzin (2014) warned that this kind of subjective reflexivity can be a trap, as it may produce dramatic conceptions of the meaning making and interpretation of the data. Denzin (2014) further suggested that there is no research free from the knowledge of the researcher and therefore, it requires the researcher to be aware of their

own subjectivity and how it is created. To overcome this challenge, we followed the suggestion made by Drake (2010: 88) to apply ‘reflexivity in action’. Therefore, we as the researchers, needed to reflect on ‘what frames [our] seeing’ and critically look at what we chose to make visible in our analysis of the data.

#### **4. Data analysis – small stories**

Narrative analysis was applied by grouping the themes under the two main aims of the project. We employed a constructionist narrative approach where stories are understood as co-constructed or dialogically constructed between the participant and the researcher with elements of constant change. Esin et al. (2014: 205) stated that “rather than reading them [the stories] as finished products of particular circumstances they may change over time.” Esin et al. (2014: 204) further argued that data analysis is conceptualised as an ongoing complex process through which multiple layers of “told small stories” focusing on “the participants’ self-generating meaning” are considered. An important part of the approach was the power relation between the researched and researcher, the data and its interpretation. In order to address how narratives were co-constructed, Bamberg and Georgakopoulou’s (2008) concept of small stories were employed. Their argument was to understand ‘small stories’ by moving away from a more traditional metanarrative approaches as they act as a structural and linear ordering of events. In contrast, small stories are conceptualised as ‘interactive engagements’ where researchers construct meaning making enabling a focus on how interviewees are positioned by the narrator (Bamberg and Georgakopoulou, 2008). Interactive engagements facilitate co-construction on multiple levels; including how the researched co-construct their narrative together with the listener/researcher. However, only the interaction between the two is important, but emphasis is played on how the story is co-constructed by the researcher in connection with their own life experiences. On a wider level, co-construction includes how narratives connect to the metanarratives in society (Harris et al., 2001). Nutbrown (2011: 243) further pointed out that the use of small stories enables the researcher to take a ‘reflexive stance’ ‘providing an opportunity, self-consciously, to ask questions, give account, wonder, push and prod’. Therefore, employing a small story approach was a way of addressing how our qualitative data was understood by us, the researchers.



## 5. Positive experiences of PhD completers

This part of the paper addresses the positive experiences of PhD student completers, which gives an insight to why participants complete their thesis despite facing difficulties.

*After the very first meeting with my supervisor, she went out for a run and said she thought my PhD had legs...it could run somewhere! That comment I still reflect on and that has a positive effect on my mental health. If I reflect further, the thesis has kept running, it has never stopped, I have never stopped believing in it and completing it. It is my race. I have ownership over every run I have ever done, and over every second, every word of the PhD... I have learned that the PhD is not anybody else's race, it is mine and having that realisation dawned on me again probably about half way through.*

This small story showed that this particular student was able to latch onto a vital initial comment made by the supervisor, as a guide, focus and inspiration, throughout the PhD journey. This endorsement of the research became an unshakable view, forming the positive foundations. Regardless of any future comments, this student had carved out a coping mechanism for this long research journey, based on the research having legs to carry them through the process and over the finish-line. The feeling of slow and 'steady race' has been addressed by Ridgway (2022) as well as how students develop resilience (McCray and Joseph-Richards, 2020). The story indicates that while the journey is 'lonely', the supervisor's comment made a significant impact on this student, helping them to build resilience. The following small story shows a similar experience:

*The characteristics of the relationship between me and my primary supervisor are energy, positive feedback and a sense of humour...all elements of a good run. Hearing positive comments as I leave the office after PhD supervision are essential and I model my own supervisory role in the same way.*

In this small story, a very effective way of instilling motivation into the process by the supervisor was discussed. Both humour and a positive last note, upon leaving a setting are tools used to provide motivation and encouragement to students. However, the following small story alluded to a less positive relationship with the supervisor(s). Yet, positive experiences can be harnessed at the final stage of the PhD journey. This case demonstrates Sisson and Jackman (2016) suggestion of good and poor psychological wellbeing that fluctuates across the timespan of the PhD:

*My only positive experience was my viva. The feedback they [examiners] gave me, their, professional approach, debates and constructive conversations, thoughts exchange was really great. It was everything that my supervisors couldn't and didn't offer.*

Here, this particular student was able to salvage some meaning to the external process of guidance and direction. They side-stepped their supervisory relationship, by reflecting on the key stage of the research, the viva, which would provide the most beneficial feedback that they needed to complete.

Phillips and Johnson (2022) identified relationships as important external factors that helps students to complete their study, such as forming a lifelong friendship:

*When I started the programme, I met some lovely colleagues, who became live long friends.*

Similarly, another participant has talked about the importance of social interaction with other PhD students seeking encouragement and professional support as well as from family and friends. It was reported that for example, informal chats about the research and which kind of support they have received from their supervisors, played a significant factor in completing the thesis. This is when co-construction of the stories between the researchers and researched becomes unavoidable:

*I loved meeting new people, people with similar passion, with whom you then form a strong bond, and with some of them a good friendship, an understanding ... someone you can tell your secrets about yourself, your doubts in your ability, but also to trust to offload all those feelings, positive but also negative, was so important to me. Someone who listens, who understand what I'm talking about. And then the encouragement that 'you can do it, you are so clever ...'*

This small story resonates with Nutbrown's (2011) self-reflective qualitative research approach and the importance of reflexivity and the need for more reflective enquiry within the field education. Building on Nutbrown's (2011) suggestion, we reflected on our own doctoral journey as we re-lived our experiences by listening to this small story. Both of us had shared similar experiences, whereby we utilised other benefits available in the PhD programme, such as collaborations outside of the process, which allowed us to keep a part of this important PhD journey alive. Due to shared experiences, our friendship was born. The encouragement and support we provided each other gave us the strength to continue with our study. Therefore,

through co-constructions and co-reflections our life experiences intersect with those of our participants and their narratives.

## 6. Negative experiences of PhD completers

The data revealed more negative experiences than positive ones. Some students struggled to understand and manage their relationship with their supervisor. They desperately tried to construct the relationship as value.

*I don't want to make a complaint, you know... I have always seen a supervisor as like a critical friend, they were testing you, when you were asking the questions; 'why did you say that?', or 'why did you do that?' you know, but I didn't really gain from that, they cast unnecessary questions you know... I just wanted a general direction and I will work out what to do by myself. So, I did most of the things on my own, as often I waited months and months for feedback. For me personally this is how it worked.... It took me probably longer like this by about 2-3 years.*

This small story appeared to be a common thread in the data, whereby unnecessary delay was caused in receiving feedback from the supervisor, which hindered progression or earlier completion. This small story resonates with the study carried out by Beasy et al. (2019) which suggested that many PhD students experienced shallow support. However, disclosure from one participant, in this study, about the unequal relationship they had with their supervisor was alarming.

*She put me in a very awkward and compromising position early in the supervisory relationship. Knowing my background, she asked me to help her with her legal applications. I felt threatened. From her character and personality. I knew that if I refused to help her, this would have an adverse impact on the supervisory relationship and my work.*

This particular example showed the issue of power in the supervisory relationship, when the personal life takes over the professional. This situation demonstrated the supervisor's insecurities, and the need to reconstruct the superior role in the relationship. It could be argued that here the supervisors doubted their own ability to conduct the supervisory role adequately by questioning the student's ability to be good at their role outside of the PhD research.

*My supervisor was nothing short of a nightmare, I was traumatised by the experience. I was angered by the process and what I had to endure. There were no*

*checks and balances, and little opportunities to feedback to programme leaders. The programme itself was mismanaged with staff departing and course leaders changing frequently.*

This small story showed very powerful emotions and highlighted the inefficiencies of the programme. This highlighted the struggles that the university suffered in its ability to deliver a robust programme. Professional programmes at PhD levels are often neglected, as they are seen as self-serving and self-sufficient. Safeguarding is not a requirement at HE level, so PhD students are not afforded that protection.

*In three months I changed supervisors twice. First, there were supervisors 1 and 2 and then supervisor number 3 came in in place of number 1. And now there are 2 and number 4. As you can imagine I often felt despised and I couldn't understand why. The atmosphere became more and more sinister, as in one of those science fiction movies where everyone all of a sudden start acting weird and you can't think of any explanation. I felt scared and angry. There was time I couldn't stop crying. I thought about quitting my PhD, since nobody seemed to care except me and a few friends who encouraged me to keep going.*

In this small story the thought of withdrawing from the programme arose. This correlated with Hunter and Devine's (2016) study which suggested that those students who report lower psychological wellbeing are more likely to consider withdrawing from their studies. This participant had experienced a number of emotions, initially, stemming from abandonment from their supervisors, which they found difficult to process. This resulted in them feeling ignored and disregarded, which Mikuska (2014) addressed in her study reporting on the importance of emotions in HE and the impact it can make on students' wellbeing.

The most emotionally driven small story was about how this participant managed anger.

*It took me 9 minutes to burn 9 years of my hard work. Literarily. I lit on fire all the paperwork I found that was to do with my (f...g] PhD. It was perfect emotional healing process for me, this is how I processed painful PhD experiences.*

By the end of the PhD journey, so much tension and unpleasant memories had built up, that the only way to deal with these emotions was likely to be destructive. It was explained that to deal with anger, the paperwork had to be burnt. This symbolic destruction released emotions that had built up and erased unwanted memories, which made way for the much-needed healing process.

## 7. Conclusions

Like most relationships, the supervisory relationship is complex and not without its impediments. However, the dependency of students upon their supervisors makes the relationship, unbalanced, unequal and fragile. The key to a successful supervisory relationship is for the supervisor to prioritise the future direction of the research for the student and steer away from redundant relationships and factors that prolong insufficiency and distrust. Our recommendation is that universities need to invest into the supervisory role, by introducing a rigorous selection process, compulsory continuing training for future and present supervisors, and an opportunity for students to provide feedback on their supervisors throughout their PhD journey. Therefore, more research is needed to highlight the issue of supervision in order to detect training needs.

Students should be encouraged to share, in a safe space, their PhD experiences, and in particular, their supervisory experiences with their peers. This would offer support and a platform to exchange ideas and solutions to resolve issues and enhance the PhD experience. This would also create a self-devised safeguarding community for PhD students outside of the HE institution. This added layer of protection may lead to an increase in the uptake of PhD research studies. Hopefully, future supervisors reading this paper will be mindful of their supervisory position and the impact that they have on PhD students during their PhD journey.

## Acknowledgements

This research is not part of any funded research. We would like to thank all the participants for their time and for sharing their stories with us.

## References

- Ali, A., & Kohun, F. (2006). Dealing with isolation feelings at IS doctoral programs. *International Journal of Doctoral Studies*, 1, 21-33. <https://doi.org/10.28945/58>
- Bamberg, M. & Georgakopoulou, A. (2008). Small stories as a new perspective in narrative and identity analysis, *Text and Talk*, 28(3), 377-396. <https://doi.org/10.1515/TEXT.2008.018>
- Beasy, K., Emery, S. & Crawford, A. (2019). Drowning in the shallows: an Australian study of the PhD experience of wellbeing. *Teaching in Higher Education*. 26(6), 1-17. <https://doi.org/10.1080/13562517.2019.1669014>
- Bruner, J. (2004). Life as narrative. *Social Research*, 71 (3), 691-710. <https://doi.org/10.1353/sor.2004.0045>

- Drake, P. (2010) Grasping at methodological understanding: a cautionary tale from insider research. *International Journal of Research and Method in Education*, 33(1), 85-99. Available at:
- Denzin, N. (2014) Writing and/as analysis or performing the world. In Flick, U. [Ed] *The SAGE Handbook of Qualitative Data Analysis* (569-584). London: Sage.
- Harris, A., Carney, S. & Fine, M. (2001). Counter work: introduction to under the covers: theorising the politics of counter stories. *International Journal of Critical Psychology*, 4, .6-18. <https://doi.org/10.1177/00380385093572>
- Hunter, K. H. & Devine, K. (2016). Doctoral Students' Emotional Exhaustion and Intentions to Leave Academia. *International Journal of Doctoral Studies* 11, 35-61. <https://doi.org/10.28945/3396>
- Kwan, B. S. C. (2009). Reading in Preparation for Writing a PhD Thesis: Case Study s of Experiences. *Journal of English for Academic Purposes*. 8(1), 180-191.
- McCray, J. & Joseph-Richard, P. (2020) Towards a model of resilience protection: factors influencing doctoral completion. *High Education*. 80, 679–699. <https://doi.org/10.1007/s10734-020-00507-4>
- Mikuska, E. (2014). The relationship between higher education, emotion and gender: a qualitative study using text and interviews. In *SAGE Research Methods Cases Part 1*. SAGE Publications. <https://dx.doi.org/10.4135/978144627305013510241>
- Mikuska, E. (2016). Mature students' emotional well-being, personal lives and achievements. In Trodd, L. [ED] *The Early Years Handbook for Students and Practitioners: An essential guide for the foundation degree and levels 4 and 5*. (pp. 3-15) London: Routledge. <https://doi.org/10.4324/9781315773209>
- Morrison, A. R. (2012). 'You have to be spoken': Student's vies on employability within the graduate labour market. *Journal of Education and Work*. 27 (9), 179-198. <https://doi.org/10.1080/13639080.2012.742178>
- Nutbrown, C. (2011) 'A box of childhood: small stories at the roots of a career', *International Journal of Early Years Education*, 19(3-4), 233-248. <https://doi.org/10.1080/09669760.2011.629491>
- Phillips, E. & Johnson, C. (2022). eBook: *How to Get a PHD. A Handbook for Students and Their Supervisors*. Maidenhead: Open University Press
- Ridgway, A. (2022) *Love, loss and a doctorate: an autoethnography of grieving while writing a PhD*. *Higher Education Research & Development*, 1-14. <https://doi.org/10.1080/07294360.2021.2019202>
- Sisson, K. & Jackman, P. (2019) From enrolment to completion: An exploration of psychological wellbeing in doctoral students during the PhD journey. In: *International Conference on the Mental Health & Wellbeing of Postgraduate Researchers*, May 16-17 2019, Brighton, UK.

Streeton, R., Cooke, M. & Campbell, J. (2004). Researching the researchers: using a snowballing technique. *Nurse Researcher*, 12(1), 35.

<https://link.gale.com/apps/doc/A122660692/AONE?u=anon~c9c3a3c5&sid=googleScholar&xid=558e0150>

Sverdlik, A., Hall, N. C., McAlpine, L., & Hubbard, K. (2018.) Journeys of a PhD student and unaccompanied minors. *International Journal of Doctoral Studies*, 13, 361-388.

<https://doi.org/10.28945/4113>

HESA: Higher Education Student Statistics: UK, 2019/20. <https://www.hesa.ac.uk/news/27-01-2021/sb258-higher-education-student-statistics>

### About Authors

**Tehmina KHAN**, qualified as a solicitor (lawyer) in 2000 and sits as a Magistrate. She entered the world of academia in 2004. Tehmina worked as a Senior Law Lecturer for over 19 years, she was awarded her PgCert in 2005, followed by a LL.M in 2011 and the EdD in 2019. She became a SFHEA in 2022. She now works at the University of Surrey, School of Law, as a Senior Lecturer. Her research interests include gender and race inequalities, forced marriage, teachers' identities, generally addressing teaching and learning practices in education and responses to government policy in these areas.

**Eva MIKUSKA**, works at the University of Chichester as a Senior Lecturer and Programme Leader; she gained MA LTHE in 2012 and the EdD in 2021, and became a SFHEA in 2021. Eva joined the University of Chichester in 2010, Institute of Education, and Social Sciences Department. Her research interest focuses on exploring the role of emotion in professional practices in addition to examining the gender discourses in the Early Childhood Education Care field. She is a trustee for TACTYC, and a member for the ECSDN and SEFDEY networks.



<http://jates.org>

# Journal of Applied Technical and Educational Sciences jATES

ISSN 2560-5429



## Business applications of natural language processing

Zsolt Krutilla<sup>1,2</sup>, Attila Kovari<sup>3,4,5,6</sup>

<sup>1</sup> Department of Software Development and Applications, Institute of Computer Engineering, University of Dunaujváros, Táncsics Mihály street 1/A, Dunaujváros 2401, Hungary, [krutillazs@uniduna.hu](mailto:krutillazs@uniduna.hu)

<sup>2</sup> Doctoral School of Applied Informatics and Applied Mathematics, Óbuda University, Bécsi street 96/B, Budapest 1034, Hungary

<sup>3</sup> Institute of Digital Technology, Faculty of Informatics, Eszterházy Károly Catholic University, Eszterházy tér 1, 3300 Eger, Hungary [kovari.attila@uni-eszterhazy.hu](mailto:kovari.attila@uni-eszterhazy.hu)

<sup>4</sup> Department of Natural Sciences, Institute of Engineering, University of Dunaujváros, Táncsics M. u. 1/A, 2400 Dunaujváros, Hungary, [kovari@uniduna.hu](mailto:kovari@uniduna.hu)

<sup>5</sup> Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Bécsi street 96/B, 1034 Budapest, Hungary, [kovari.attila@uni-obuda.hu](mailto:kovari.attila@uni-obuda.hu)

<sup>6</sup> GAMF Faculty of Engineering and Computer Science, John von Neumann University, Izsáki u. 10, 6000 Kecskemét, Hungary [kovari.attila@gamf.uni-neumann.hu](mailto:kovari.attila@gamf.uni-neumann.hu)

**Abstract:** Natural Language Processing (NLP) is becoming increasingly important for researchers doing research in the field of information technology. However, NLP itself is not only used in the academic world, but is also a topic of interest in the competitive world, as it offers many opportunities for automation. The aim of this paper is to analyze and summarize the articles published on NLP from several perspectives, with a particular focus on the development and practical application of the technology, as well as to explore possible future research opportunities, directions and its potential role in industry. Among the practical applications of NLP, research and development in the field of education is increasingly being observed. In order to identify future uses, the underlying textual stocks of the technology should be considered. From this point of view, we can focus on the insurance and banking sectors, where there is a series of administrative processes where centralized paper processing is the incoming data set for the process and where a database dating back several years may be available. The research has confirmed that there are still many research opportunities in the field of NLP, as even the most modern methods cannot be applied without error and full context coverage. Moreover, the field of application can be extended considerably, both from the technical and from the economic and educational point of view.

**Keywords:** *natural language processing; deep learning; artificial intelligence;*

## Géppel történő természetes nyelvfeldolgozás üzleti alkalmazása

Krutilla Zsolt<sup>1,2</sup>, Kővári Attila<sup>3,4,5,6</sup>

<sup>1</sup> Szoftverfejlesztési és Alkalmazási Tanszék, Informatikai Intézet, Dunaujvárosi Egyetem, Táncsics Mihály út 1/A, Dunaujváros 2401, Magyarország, [krutillazs@uniduna.hu](mailto:krutillazs@uniduna.hu)



<sup>2</sup> Alkalmazott Informatikai és Alkalmazott Matematikai Doktori Iskola, Óbudai Egyetem, Bécsi út 96/B, Budapest 1034, Magyarország

<sup>3</sup> Digitális Technológia Intézet, Informatikai Kar, Eszterházy Károly Katolikus Egyetem, Eszterházy tér 1, 3300 Eger, Magyarország kovari.attila@uni-eszterhazy.hu

<sup>4</sup> Műszaki Intézet, Dunaujvárosi Egyetem, Táncsics M. u. 1/A, 2400 Dunaujváros, Magyarország kovari@uniduna.hu

<sup>5</sup> Kandó Kálmán Villamosmérnöki Kar, Óbudai Egyetem, Bécsi út 96/B, 1034 Budapest, Magyarország kovari.attila@uni-obuda.hu

<sup>6</sup> GAMF Műszaki és Informatikai Kar, Neumann János Egyetem, Izsáki u. 10, 6000 Kecskemét, Magyarország kovari.attila@gamf.uni-neumann.hu

---

**Absztrakt:** A számítógéppel történő természetes nyelvi feldolgozás, azaz az NLP (Natural Language Processing) egyre jelentősebb szerephez jut. Jelen tanulmány célja az NLP témában megjelent cikkek elemzése és több szempontú összefoglalása, különös tekintettel a technológia fejlesztésére és gyakorlati alkalmazására, valamint esetleges jövőbeli kutatási lehetőségek, irányvonalak feltárására és az iparban betöltött lehetséges szerepére. A jövőbeli felhasználási terület meghatározásához figyelembe kell vennünk a technológia alapjául szolgáló szöveges állományokat. Ezen aspektusból előtérbe helyezhetjük a biztosítási és banki szektorokat is, melyek esetében olyan ügyintézési folyamatok sorozata zajlik le, ahol központi papírfeldolgozás képezi a folyamat bejövő adatállományát és több évre visszamenőleges adatbázis állhat a rendelkezésünkre. A kutatás során beigazolódott az, hogy az NLP területén még számos kutatási lehetőségek előtt állunk, ugyanis a jelenleg legmodernebb módszerek sem alkalmazhatók hiba nélkül és teljes kontextus-lefedettséggel. Ezen felül az alkalmazási terület is jelentősen kiterjeszhető, mint a műszaki, mint a gazdasági és oktatási terület aspektusából.

**Kulcsszavak:** természetes nyelvi feldolgozás; mélygépi tanulás; mesterséges intelligencia;

---

## 1. Bevezető

Jelen cikk célja a természetes nyelv számítógépes feldolgozásának jelenlegi és jövőbeni alkalmazási területek lehetőségeinek feltárása, valamint az alkalmazott módszerekhez kapcsolódóan megjelent tudományos cikkek elemzése. Napjainkban egyre nagyobb teret hódít az informatikában a számítógépes természetes nyelvi feldolgozás, azaz az NLP (Natural Language Processing). Az NLP mint technológia ötlete azonban nem a 21. század. században fogalmazták meg, de már az 1950-es években is folytak releváns kutatások és születtek elméletek arra vonatkozóan, hogy egy gép hogyan tudja értelmezni az emberi beszédet. Az 1960-as években King, Masterman, Ceccato és Yngve olyan fordítógépen dolgoztak, amely magát az NLP technológiát alapozta meg. (Wilks, 2005) Az IT megjelenésével a RENDEZVOUS rendszer volt az első természetes nyelvi interfész (NLI) adatbázis. Az 1970-es években két transzformációs NLP megoldási módszer volt elterjedtebb, a generatív és a statisztikai algoritmus alapú módszer. A generatív eljárás esetében a két nyelv szerkezetét, szabályrendszerét veszi alapul, és meghatározza a két szabályrendszer közötti kapcsolatot. A statisztikai módszer esetében a tanítási minta valószínűségi szabályai alkotják az összefüggést.

(Kovács, 2009) Az NLP eredete tehát az ötvenes évekre vezethető vissza, jóval a számítógépek megjelenése előtt. Manapság elsősorban a Deep Learning eljárásokat alkalmazzák az NLP során, tekintettel arra, hogy ma már igen nagy pontosságú neurális hálózati modellek, valamint hatalmas adatbázisok és adattárházak állnak rendelkezésünkre a modellek betanításához, de korábban inkább a statisztikai és különféle matematikai modellek használata volt jellemző. A jelenlegi nyelvi modelleknek azonban megvannak a maga hátrányai is, különös tekintettel arra, hogy bár nagy adatbázis áll rendelkezésünkre a modellek tanítására, ezek korán sem nevezhetők tiszta adatoknak, sok esetben gépi generált tanulói adatokat használnak, amelyek torzítják az eredményt a „valódi beszéd” aspektusából. Végső célunk, hogy hosszú távon megtaláljuk azt az NLP technológiát, amely az emberhez hasonlóan képes elemezni és értelmezni a természetes beszédet.

Az NLP alatt olyan számítógépes megoldásokat értünk, amelyek képesek a mindennapi természetes szövegek nyelvi elemzésére, annak érdekében, hogy számos feladathoz vagy alkalmazáshoz emberszerű nyelvi feldolgozást érjenek el. A meghatározás azonban korántsem tökéletes, több megjegyzést is fűzhet hozzá. Mindenekelőtt nézzük meg közelebbről a „számítógépes megoldásokat”, mert számos módszer vagy technika közül választhatunk egy-egy adott típusú nyelvi elemzés elvégzéséhez. Másodszor a „természetes szövegek” jelentését is érdemes kiemelni, mivel ez azt jelenti, hogy a szövegek bármilyen nyelven, műfajban, írott vagy beszélt szövegek lehetnek. Az egyetlen követelmény az, hogy olyan nyelven legyenek, amellyel az emberek kommunikálnak egymással.

Fogalmilag mi emberek hajlamosak vagyunk nyelvi elemzési szinteket használni, mert mindegyik szint más-más jelentéstípust közvetít. Az NLP-rendszerek a nyelvi elemzés különböző szintjeit vagy szintek kombinációit használják, és ez a különböző NLP-alkalmazások közötti különbségekben is megmutatkozik. Ez sok zavart okoz a nem szakértők körében, hogy mi is az NLP valójában, mivel egy olyan rendszer, amely bármilyen Ezen elemzési szintek részhalmaza NLP-alapú rendszernek mondható, tehát a különbség valójában az lehet, hogy a rendszer „gyenge” vagy „erős” NLP-t használ. (Liddy, 2001)

## 2. Módszertan

Az elemzés célja, hogy az NLP témájához kapcsolódóan megjelent tudományos cikkeket tanulmányozzuk. Ehhez meg kell határozni az adott témát és ki kell tűzni a célokat. A kutatás lényegi része az NLP tudományág történetének és fejlődési szakaszainak, valamint a jelenlegi

alkalmazási területeknek az áttekintése. Ennek megfelelően került kiválogatásra a hazai és elsősorban a nemzetközi szakirodalom.

### 2.1. *Kutatás módszertana*

A cikkek kiválasztásához és az irány meghatározásához szisztematikusan a klasszikus kutatási módszertan alkalmazásával, az empirikus tapasztalatok és a kvalitatív elemzések eredményeinek ötvözése került alkalmazásra. A tudományági sajátosságok szempontjából a tárgyalt téma külön témakörökre bontható, így az alkalmazott módszerek témája, a tudományág eredete és fejlődése (valamint szakaszai), a nyelvészet és az informatika kapcsolata, mély tanulási alkalmazások fejlesztése, a matematikai nyelvi modellek fejlesztése, valamint a tanítási adatok struktúrái és tárolási technológiai kategóriákra bontható. Ezeknek a témáknak a közös elemzése azonban évekig is eltarthat, így ennek megfelelően a hatókör a tudományág fejlesztésére és az alkalmazott technológiák körére került redukálásra.

A szakirodalom áttanulmányozása során a keresők nemcsak releváns tudományos cikkeket, könyveket hoztak fel, hanem egyéb a témakörhöz köthető tanulmányokat is, melyek azonban nem relevánsak a jelen témakör aspektusából. A kutatási terület nagysága és aktualitása miatt az eredményeket szükséges volt leszűkíteni a releváns információkat tartalmazó eredményekre, amelyek elsősorban az NLP fejlesztésére, kezdetleges algoritmusaira és azokra a területekre koncentráltak, melyek egészen a deep learning (gépi mélytanulás) megoldásokig vezettek.

## 3. NLP módszerek és fejlődési szakaszai

Az NLP technológia terén nem csak általánosságban vehető szavakról és mondatokról beszélhetünk, hanem az úgynevezett stopszavakat is érdemes figyelembe vennünk. A stopszavak olyan csekély jelentéssel bíró szavak, amelyeket eltávolítanak az indexekből és a lekérdezésekből. Az NLP-ben a stopszavak negatív mintáknak is tekinthetők, amelyeket az LSTM neurális hálózatok tanításakor használnak. A stopszavaknak kevés jelentése lehet a gyakoriság vagy a fogalom szempontjából, továbbá a nagyon gyakori szavak eltávolítása nem befolyásolja a dokumentumok rangsorolását és jelentését sem. Ha a szavaknak kevés fogalmi jelentésük van, akkor eltávolíthatók, függetlenül attól, hogy gyakoriságuk magas vagy alacsony a gyűjteményben. Valójában különösen akkor fontos ezeket a szavakat eltávolítani, amikor alacsony a gyakoriságuk, mert ezek a szavak befolyásolják leginkább a dokumentumok rangsorolását. Szövegelemzés tekintetében elmondható, hogy szakirodalom általában kétféle szövegelemzési technikát tárgyal, a kulcsszószámlálást és a térképelemzést. Ezeken felül (vagy

ezzel egyidőben) dokumentumindexelésről is beszélhetünk, melynek során a lekérdezés feldolgozás tipikus megközelítése a következő. Először egy tokenizálási folyamat megy végbe, majd a stopszavak eltávolításra kerülnek. Ezen túlmenően a természetes nyelvi feldolgozási technikák azonosíthatják a kifejezéseket vagy az osztott összetételeket. (Hiemstra és Jong, 2001)

Az első módszer, a kulcsszószámlálás, jellemzően számítógép segítségével méri automatikusan azon kulcsszavak gyakoriságát, amelyekről a kutatók korábban megállapították, hogy egy érdekes konstrukcióhoz kapcsolódnak (a mai napig ez a módszer volt a legszélesebb körben használt technika). (Duriau, Reger, Pfarrer, 2007) A technika népszerűsége abból adódik, hogy a kulcsszavak gyakorisága objektív és intuitív mutatója az gyűjtemény méretének vagy relatív fontosságának. (Knoke, 1982)

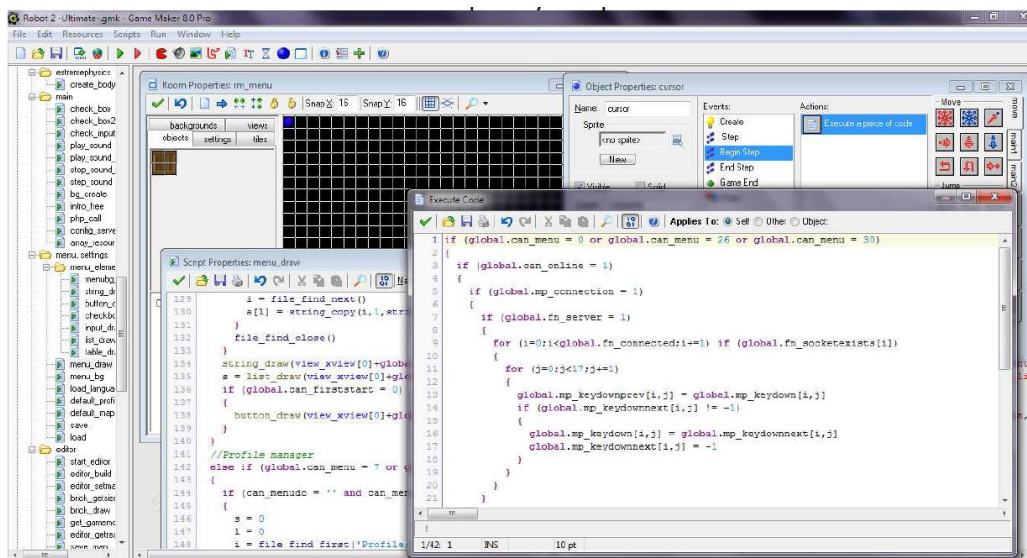
A második módszer, az úgynevezett „térképezési elemzés”, melynek során a fogalmak közötti relációs kapcsolatok jelennek meg a szövegekben (többnyire ez inkább az fenti módszert kiegészítő technika). (Axelrod, 1976) A szövegelemzőket általában arra képezik ki, hogy azonosítsák a releváns fogalmakat és összefüggéseket a szövegekben, és addig olvassák át a dokumentumokat, amíg el nem jutnak a szövegek mögött meghúzódó kognitív kapcsolatok megbízható reprezentációjához. (Huff, Narapareddy, Fletcher, 1990)

Bár mindkét technika elfogadott szövegelemzési módszerré vált a szakirodalomban, a két hagyományos technika nyilvánvaló kompromisszumot jelent. A kulcsszavak megszámlálásával a kutatók szűk mértékegységeket vonhatnak ki a szöveg tartalmából, de nem tudják kihasználni azt a részletes és értékes információt, amely a szövegek szavainak tágabb szerkezetében rejlik. (Carley, Kathleen, Palmquist, 1992) A gépi segítséggel támogatott NLP módszerek legújabb fejlesztései új lehetőségeket kínálnak a szöveges adatok elemzésére és értelmezésére. (Grimmer, 2013) A modern NLP lehetővé teszi a kutatóknak, hogy számítási algoritmusok segítségével mélyebb jelentésstruktúrákat vonjanak ki nagy mennyiségű szövegből, és ennek segítségével megnyílik az elemzés lehetősége.

Összefoglalva elmondható, hogy a szövegbányászat és a természetes nyelvi feldolgozás a „természetes nyelv” megértését és elemzését jelenti számítógépes algoritmusok és programok segítségével, és fontos kutatási irány a mesterséges intelligencia alkalmazási területén. Folyamatos és kiterjedt gépi kutatásokkal. tanulási és adatbányászati algoritmusok, a meglévő szövegbányászati technológiák jó eredményeket értek el az automatikus absztrakcióban, az automatikus kérdésválaszolásban, a webes relációs hálózatelemzésben és az anafora felbontásban (Zeng et al, 2015).

#### 4. Az NLP szerepe a gyakorlatban

Az NLP általában olyan oktatási szoftverrendszerek és oktatási stratégiák fejlesztésére összpontosít, amelyek támogathatják a természetes nyelvek használatát az oktatásban (pl. e-Assessment és Text Adapter). Az NLP-vel rendelkező szoftverrendszerek természetes körülmények között képesek azonosítani a nyelvtanulás folyamatát. (Gy. Molnár; P. Nyíró, 2016)



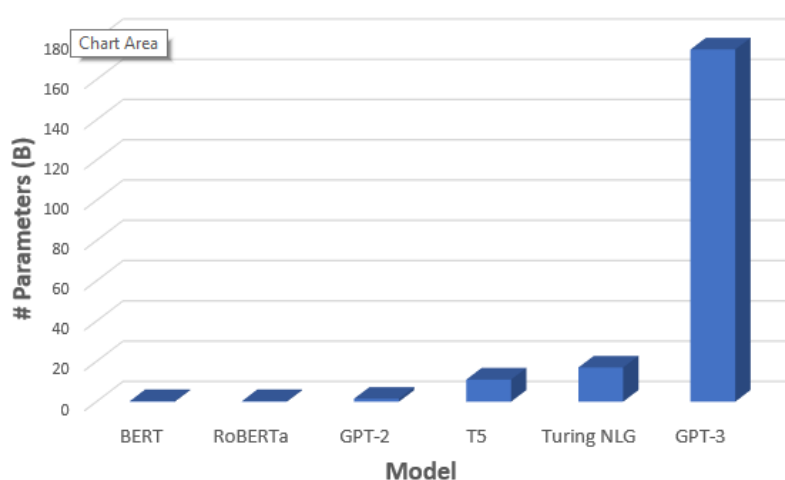
1. ábra - Játékfejlesztés a kurzusokon alkalmazott Game Maker 8.0 használatával. (Gy. Molnár; P. Nyíró, 2016)

Az NLP hatékony módszer egy olyan rendszer kifejlesztésére is, amely különféle szavakon, mondatokon és szövegeken keresztül dolgozza fel a nyelvi bemenetet a természetes környezetben. A természetes nyelvi feldolgozás különféle nyelvtani szabályokat és nyelvi megközelítéseket is használ, mint például a származékokat, a nyelvtani igeidőket, a szemantikai rendszert, a lexikont, a korpuszt, a morfémákat, az igeidőket stb.

Ezek a hatékony megközelítések felhasználhatók az oktatási környezetben is, hogy a tanulók jobban megértsék az oktatási anyagokat. Az NLP a nyelvtanulás területén is széles körben ismert megoldás az egész világon, és sikeresen alkalmazzák a nyelvoktatási rendszerek fejlesztésének hatékony eszközeként. Hasonlóan hatékony és eredményes tanulás érhető el vizuális alapú, problémamegoldó, algoritmikus gondolkodással (Francisti et al, 2021). A legtöbb tanulmányban az angol a leggyakoribb nyelv, ami azt mutatja, hogy az NLP-t hatékonyan használják a nyelvtanulásban. Az NLP hatékony megközelítés az arab országok oktatási rendszerének javítására is. (Habash, 2010)

Az elmúlt évtizedben a mesterséges intelligencia területén a technikai fejlődés jelentős áttörést hozott az AI-technológia kutatásával és alkalmazásával foglalkozó valamennyi területen, így az NLP területén is. Az első, mondhatni kézzelfogható áttörésre a számítógépek megjelenéséig kellett várni, amelyek képesek voltak elméleteket a gyakorlatba átültetni és sikerüket bizonyítani.

A Moore-törvénynek köszönhetően a számítási teljesítmény évtizedek óta tartó exponenciális növekedése lehetővé tette az olyan statisztikai megközelítéseket, mint a GPT-3. Ezek a modellek az írott nyelv tömeges példáival való találkozás révén tanulják meg a nyelvet. (ELKINS, Katherine; CHUN, Jon, 2020)



2. ábra - GPT-3: The New Mighty Language Model from OpenAI (forrás: <https://towardsdatascience.com/gpt-3-the-new-mighty-language-model-from-openai-a74ff35346fc>)

A GPT-3 (Generative Pre-trained Transformer) egy harmadik generációs, autoregresszív nyelvi modell, amely mély tanulást használ az emberhez hasonló szöveg előállítására. Vagy egyszerűbben fogalmazva, ez egy olyan számítási rendszer, amelyet arra terveztek, hogy szavak, kódok vagy más adatok szekvenciáit generálja egy forrásbemenetből, az úgynevezett promptból kiindulva. A GPT első, 2018-as iterációja 110 millió tanulási paramétert használt. Egy évvel később a GPT-2 már 1,5 milliárdot használt belőlük. Azonban mindezen számok eltörpülnek a jelenlegi GPT-3 175 milliárd paraméteréhez képest. A modell tanítása a Microsoft Azure mesterséges intelligencia szuperszámítógépén (Scott 2020) történik. A modell felhasználási területe igen széles skálán mozog, beleértve az összegzést, a fordítást, a nyelvtani javítást, a kérdések megválaszolását, a chatbotokat, az e-mailek összeállítását és még sok más is. (Floridi, Chiriatti, 2020)

## 5. Következtetés, további kutatási lehetőségek

A kutatás során beigazolódott az az elképzelés, hogy a tudományág korai szakaszában és a számítógépek megjelenése előtt a matematikai modellezés és a statisztikai valószínűségekre alapuló modellek teret hódítottak a területen. Az információs technológia (számítógép) megjelenésével a módszerek és az alkalmazott eljárások is megváltoztak, úgymond követték a technológiai fejlődést, és egyre inkább hasznosították a technológiában rejlő erőforrásokat. A számítástechnika folyamatos fejlődésével megjelentek az adatbázisok, amelyek új alapokra helyezték az NLP tudományát. Az adatvezérelt szövegelemzés korszaka kezdődött el, amikor lehetővé vált nagy adathalmazok tárolása, kezelése és elemzése. Új eszköz állt a kutatók rendelkezésére, amely még nem nevezhető mesterséges intelligenciának, de már a logikai megközelítés sokkal inkább hasonlított a mai LSTM (Long-Short-Term-Memory) neurális modellhez, mint a statisztikai elemzéshez. Lehetőség nyílt a korábbi adatok alapján történő szövegelemzésre, megfigyelésre és azonosításra. Az NLP tudományágában ez a korszak tekinthető az egyik legjelentősebb szakasznak, mondhatni paradigmaváltásnak.

Bár a korábbi eredmények alapján az NLP mint tudományág még az adatvezérelt módszer korszakában is jelentős eredményeket ért el, de az igazi áttörést a mesterséges intelligencia megjelenése és sikeres megvalósítása hozta meg. Az újonnan megjelenő mély tanulási módszer új perspektívába helyezte az NLP-kutatással foglalkozó tudósok munkáját, új megközelítések és módszerek jelentek meg. A tanulmány során betekintést nyertünk a természetes nyelv gépi feldolgozásának kiindulópontjába, feltárta, hogy a diszciplína elmélete jóval az első számítógépek megjelenése előtt született, és annak ellenére, hogy egyre nagyobb áttöréseket érnek el a deep learning módszer segítségével, még mindig nem áll a rendelkezésünkre egy olyan eljárás vagy módszertan, amelyet valóban az NLP probléma megoldására terveztek. Az NLP a mai napig tudományosan megoldandó probléma, amelyre még mindig nincs pontos és egységes megoldás.

A GPT-3 modellek megjelenésével egy sokkal hatékonyabb nyelvi elemző eszköz áll a rendelkezésünkre és a fordítógépek mellett egyre nagyobb az igény a természetes nyelvi elemzésre is, amely nemcsak az emberi nyelven írt szövegek gépi értelmezésére ad lehetőséget, hanem egy olyan kommunikációs eszközt is biztosít számunkra, amely új szintre emelheti a számítógépes kommunikációt (J. Katona, 2021a), (J. Katona, 2021b), (J. Katona, 2022). Képzelnünk el egy olyan programozási feladatot, melyet már nem emberek programoznak, hanem „természetes beszéd útján felmondjuk” a szoftverrel szemben támasztott elvárásainkat a

számítógépnek, melyet a gép képes értelmezni, majd a fejlesztést önállóan elvégezni a szóbeli specifikációnak megfelelően. Természetesen a természetes nyelv ilyen mélységű felhasználásáig több problémát is meg kell oldanunk, azonban az ilyen mélységű természetes beszéd értelmezése már egy másik szintet képvisel, az értelem szintjét.

Egy másik (és természetesen sokkal közelebbi) felhasználási területe lehet az NLP-nek a banki és biztosítási dokumentumok értelmezése és a további munkavégzés (feladatlépés) elvégzése. Jelenleg ezen a két területen igen nagy jelentőséggel bírnak a szöveges dokumentumok értelmezése, így az NLP egy új szintre emelheti azirodai munkavégzés folyamatát.

Az NLP aspektusából ugyancsak izgalmas terület lehet a jogrendszerek labirintusában történő kiigazodás és jogesetek azonosítása akár a magánszférában, akár a bírósági döntéshozás területén. A jogi szabályozás tekintetében az NLP technológia által a döntéshozás akár automatizálható folyamattá válhat, azonban semmiképp sem hanyagolhatjuk el a területen esetlegesen betölthető jelentőségüket, szerepüket (Riczu, Krutilla, 2021).

### **Irodalomjegyzék**

Axelrod, R. (1976). *The cognitive maps of political elites. Structure of decision*. Princeton University Press: Princeton, NJ.

Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social forces*, 70(3), 601-636.

D. a. JHK Knoke (1982). *Network analysis*, Beverly Hills: SAGE Publishing.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational research methods*, 10(1), 5-34.

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a Writer's turing test?. *Journal of Cultural Analytics*, 5(2), 17212.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681-694.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.

Habash, N. Y. (2010). Introduction to Arabic natural language processing. *Synthesis lectures on human language technologies*, 3(1), 1-187.

Hiemstra, D., & Jong, F. D. (2001). *Statistical Language Models and Information Retrieval: natural language processing really meets retrieval*.

Huff, A. S., Narapareddy, V., & Fletcher, K. E. (1990). Coding the causal association of concepts. *Mapping strategic thought*, 311-325.



- Katona, J. (2021a). Clean and dirty code comprehension by eye-tracking based evaluation using GP3 eye tracker. *Acta Polytechnica Hungarica*, 18(1), 79-99.
- Katona, J. (2021b). Analyse the Readability of LINQ Code using an Eye-Tracking-based Evaluation. *Acta Polytech. Hung*, 18, 193-215.
- Katona, J. (2022). Measuring Cognition Load Using Eye-Tracking Parameters Based on Algorithm Description Tools. *Sensors*, 22(3), 912.
- Kovács, L. (2009). Szemantikai gráf alapú mondatelemző modul kidolgozása IS-NLI értelmezőhöz, VI. Magyar Számítógépes Nyelvészeti Konferencia, pp. 356-359.
- Liddy, E. D. (2001). *Natural language processing*.
- Molnár, Gy et al (2016). A gyakorlati programozás tanításának játékfejlesztésen alapuló, élménypedagógiai alapú módszerének bemutatása. Karlovitz János Tibor (szerk.), *Pedagógiai és szakmódszertani tanulmányok*, 89-98.
- Riczu, Z., & Krutilla, Z. (2021). The impact of optical character recognition artificial intelligence on the labour market. *International Journal of Engineering and Management Sciences*, 6(4).
- Wilks, Y. (2005). The history of natural language processing and machine translation. *Encyclopedia of Language and Linguistics*, 9.
- Zeng, Z., Shi, H., Wu, Y., & Hong, Z. (2015). Survey of natural language processing techniques in bioinformatics. *Computational and mathematical methods in medicine*, 2015.

### Rövid szakmai életrajz

**Krutilla Zsolt** a Dunaújvárosi Egyetemen egyetemi tanársegédként dolgozik az Informatika Intézetnél. 2007 óta dolgozik az OTP Bank Nyrt.-nél, 2021 óta pedig mint főosztályvezető. Dunaújvárosi Egyetemen szerzett mérnökinformatikus (szoftvertechnológia) és okleveles mérnöktanár (informatika) diplomát. Kutatási területe a mesterséges neurális hálózat alapú, Magyar nyelvű természetes-nyelvi (NLP/NLU) és a szótáralapú szöveganalitikai modellek vizsgálata banki back-office környezetben.

**Kővári Attila** az Eszterházy Károly Katolikus Egyetem egyetemi tanára. Informatika és neveléstudományok területén habilitált. Kutatási területe az alkalmazott informatika, adatfeldolgozás, adatelemzés valamint a digitális pedagógia, IKT alkalmazása az oktatásban és kapcsolódó témaköröket öleli fel.