Early detection of dropout factors in vocational education: A large-scale case study from Finland

Sonsoles López-Pernas [1,2], Riina Kleimola [2], Sanna Väisänen [2] and Laura Hirsto [2]

[1] *School of Computing, University of Eastern Finland, Joensuu, Finland*
[2] *School of Applied Educational Science and Teacher Education, University of Eastern Finland, Joensuu, Finland*

Abstract
The aim of this study is to analyze which factors from students' admission data can predict dropout in initial vocational education and training (VET) in Finland. The sample included 15,523 students in different fields of VET that started an initial VET between 2014 and 2021 in a large-size vocational school in Finland. The results of fitting a logistic regression model to the admission data showed that students who started a VET program after basic education were more likely to drop out, as well as students who combined their studies with a job or job-seeking. Our findings pave the pathway for further research to implement support measures for decreasing dropout that are tailored to each specific "risk group".

Keywords
vocational education and training (VET), dropout, learning analytics, prediction

1. Introduction and background

It has been suggested that students should acquire skills for life-long learning through their studies, as well as the ability to self-regulate [1]. Tynjälä [2] argued that a fast change in working life has made lifelong learning and learning in the workplace necessary. Furthermore, self-regulated learning is seen as significant for workplace learning [3] and it is related to academic achievement [4]. Vocational education and training (VET) has a significant role in promoting opportunities of life-long learning for both young and adults. However, despite a strong emphasis placed on VET in education and economic policies worldwide, it is not without challenges [5], [6]. In many countries, dropping out of VET, especially among young people, has been a target of concern as it may have negative consequences not only for individuals but also for the whole society [7].

These challenges are evident also in Finland, the context of this study. In the academic year 2019–2020, 12.3% of upper secondary initial VET students in Finland interrupted their studies without continuing them in education aiming at a qualification or degree [8]. In the last decade, there have been several attempts to improve study completion and to prevent dropouts in the Finnish VET, for example, through a large-scale national retention programme implemented in 2011–2015 (see lapaisy.fi). The programme aimed to develop more proactive and individualized operating models for guidance and student care, to make use of appropriate pedagogical solutions that would support the study completion, and to facilitate the provision of labor-intensive learning environments [9, p. 39]. Despite some promising early results reported [10], a recent study by Vehkasalo [11] investigating the programme's effects revealed that the programme has not been successful in terms of increasing graduation or decreasing student attrition in Finnish VET. Based on highly detailed register data, the study highlighted that although the completion and dropout rates have shown somewhat favorable development in recent years, this is likely due to nationwide macroeconomic fluctuations and a new, tightened criteria for youth unemployment benefits rather than programme initiatives [11]. Thus, preventing attrition and promoting the study completion in the Finnish VET continue to require

Proceedings of the Finnish Learning Analytics and Artificial Intelligence in Education Conference (FLAIEC22), Sep 29-30, 2022, Joensuu, Finland
EMAIL: sonsoles.lopez@uef.fi (A.1); riina.kleimola@uef.fi (A.2); sanna.m.vaisanen@uef.fi (A.3); laura.hirsto@uef.fi (A.4)
ORCID: 0000-0002-9621-1392 (A.1); 0000-0003-2091-2798 (A.2); 0000-0002-2981-912X (A.3); 0000-0002-8963-3036 (A.4)

further exploration and development of comprehensive support measures. In particular, there seems to be a call for better monitoring of students who are at risk of leaving education early (e.g., [7]) and for prediction of possible dropouts. Learning analytics (LA) has offered a promising approach to address these issues.

The field of LA emerged over a decade ago with the goal of understanding and optimizing learning and the environments in which it occurs. The massive amount of data generated by learners when using online learning systems has allowed to, for example, get insights into students' learning strategies [12], map students' collaboration patterns [13], and predict performance [14], and dropout [15]. Existing research in this field has mainly focused on higher education [16]. A possible reason is that blended and online learning have been long established at universities and therefore the availability of data is greater than at other stages of education. Moreover, researchers in learning analytics often employ data from their own university courses which makes the process of data collection and retrieval highly convenient. In turn, there is a paucity of LA studies in VET [16] since most of the learning is hands-on and students leave no or only little digital trace of their progress. Therefore, researchers in VET need to rely on coarser data such as the information provided by students in the application and admission process. Although the potential of such data is not as high, it does not require additional data collection processes and it allows to profile students from the beginning.

There has been considerable interest in LA research in the possibilities of predicting student dropout. Such research has experimented with various methods for prediction. For example, Rovira et al. [17] used course grades and course ranking to develop predictions of student dropout with machine learning techniques among university students of Computer Science, Law and Mathematics. According to them, they were able to develop a system with which they could quite reliably predict students' dropping out of their studies and their final grade based on their mean grades after their first year of studies. Furthermore, Dardiri, Dwiyanto, and Utama [18] showed through their systematic review, that some computational methods are useful for predicting various kinds of problems in vocational education, such as Naïve Bayes, Artificial Neural Network (ANN), and C4.5. They also suggest that deep learning may have a significant role in solving problems in vocational education in a well-organized way. Pradeep and Thomas [19] used educational data mining (EDM) techniques to predict college students' dropout, and they suggest the use of various classification techniques to identify the weak students who are likely to have challenges in their academic achievement. They used various classification techniques like induction rules and decision-tree for predicting. With respect to the input variables for predictive models of EDM applied to school dropout, according to Shahiri, Husain and Rashid [20], the most used seemed to have been Cumulative Grade Points Average (CGPA), quizzes, lab work, class test, and attendance.

The aim of this study was to analyze which factors from the admission data can predict dropout in initial VET in Finland. Instead of following the EDM approach, which often deals with building predictive machine learning models that are used as black boxes and, as such, are hard to interpret by practitioners, we follow the LA path in which "understanding" is key. Therefore, we chose logistic regression to identify factors that predict dropout and analyze them in a way that is understandable for both practitioners and educational researchers. In the next section, we describe in detail the context and methods of this study, followed by the results and discussion.

2. Methods

2.1 Study context

In Finland, initial VET implemented at an upper secondary level is targeted for both young and adult learners who wish to develop basic vocational skills and competences required for entry level jobs or further studies. Learners may apply for initial vocational qualifications after completing basic education and getting a graduation certificate. However, more particular student selection criteria are decided by each education provider [21]. Approximately half of the students who have completed the basic education apply to VET and half of them continue to general upper secondary education [22].

Initial VET qualifications typically last three years although the duration may vary depending on the individual students' previously acquired competences [21]. There is a strong emphasis placed on work-based learning (WBL), the forms of which are individually determined for the student in the personal competence development plan [21]. Students are expected to practice and demonstrate their competence in practical assignments in authentic settings, both in schools and working-life [21]. When all the studies included in the personal competence development plan have been successfully completed, the student will be given a certificate for the entire qualification or for one or more qualification units [21].

2.2 Data and Methods

We extracted the admission data from all of the students that started an initial VET between 2014 and 2021 in a large-size vocational school in Finland, offering study pathways in the different fields of VET. The data were collected through the system that the school uses for managing its students, processes and operations from the admission to graduation phase.

The sample included a total of 15,523 students, of whom 10,350 (66.68%) completed their VET studies and 5,173 (33.32%) dropped out. In this study, a student is regarded as a dropout if he/she resigns from the initial VET qualification. The data available for the students were the following: gender, age, native language (Finnish or other), employment status, and educational background. Employment status could be one of the following: employed, unemployed, or other. The students who have defined their employment status as 'other' are generally people outside the labour force, such as full-time students, pensioners or conscripts. Regarding educational background, students could have none or several of the following: (general upper) secondary school, matriculation exam, vocational education and training (VET), university (either research-driven university or university of applied sciences), and other, usually referring to a foreign qualification or a degree that is not part of the Finnish education system. Figure 1 shows the distribution of variables among students.

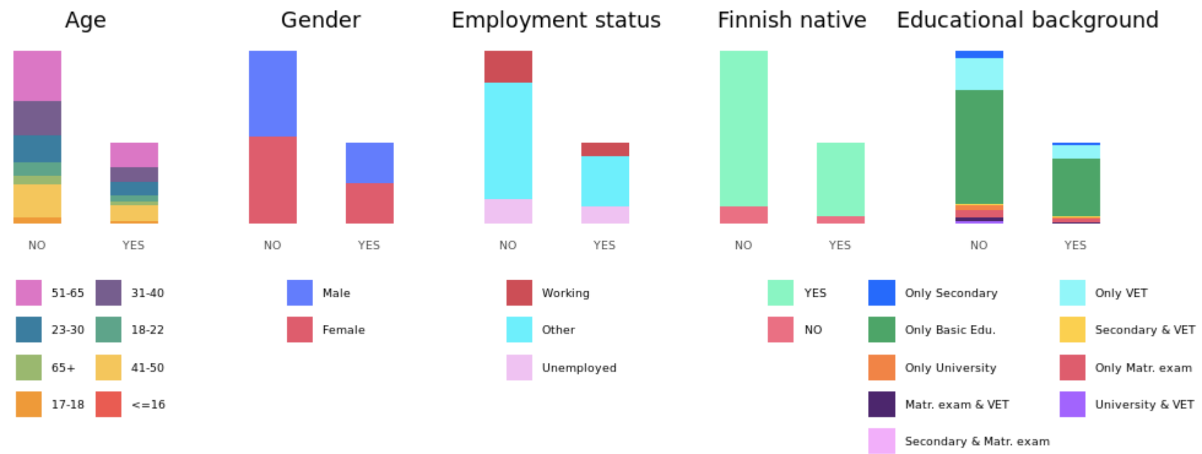

Figure 1: Distribution of variables for students who dropped out (YES) and students who did not (NO)

We fitted a logistic model (estimated using Maximum Likelihood) to find out which variables from the admission data were predictors of dropout. Categorical variables were converted into binary. We performed step AIC feature selection using the MASS R package [23] both forward and backwards to select only those variables that resulted in the smallest AIC value for the model and avoid overfitting. The selected features for the final model were age, employment status, and educational background. We used the ggstatsplot R package to graphically represent the model [24].

3. Results

Using admission data from VET students, we fitted a logistic model to predict dropout with age, employment status, and educational background. The model's intercept is at -0.75 (95% CI [-0.86, -0.64], $p < .001$). Within this model, the effect of age was statistically significant and positive, and the smallest in magnitude (β = 3.00e-03, 95% CI [7.12e-04, 5.29e-03], $p = 0.010$; Std. $\beta = 0.04$, 95% CI [0.01, 0.08]). This indicates that the effect of age is almost negligible when predicting dropout.

The effect of being unemployed was statistically significant and positive ($\beta = 0.86$, 95% CI [0.76, 0.95], $p < .001$; Std. $\beta = 0.86$, 95% CI [0.76, 0.95]), and so was the effect of being employed ($\beta = 0.26$, 95% CI [0.14, 0.37], $p < .001$; Std. $\beta = 0.26$, 95% CI [0.14, 0.37]), being the baseline the status of 'Other' (mainly students who were outside the labour force, such as full-time students). These results indicate that students who are fully devoted to studying and not to working or to job-seeking are more likely to complete their studies than the latter.

Regarding educational background, the effect for all levels of education was statistically significant and negative: university ($\beta = -0.58$, 95% CI [-0.75, -0.41], $p < .001$; Std. $\beta = -0.58$, 95% CI [-0.75, -0.41]), VET ($\beta = -0.44$, 95% CI [-0.54, -0.34], $p < .001$; Std. $\beta = -0.44$, 95% CI [-0.54, -0.34]), Matriculation Exam ($\beta = -0.33$, 95% CI [-0.47, -0.18], $p < .001$; Std. $\beta = -0.33$, 95% CI [-0.47, -0.18]), General Upper Secondary Education ($\beta = -0.57$, 95% CI [-0.74, -0.40], $p < .001$; Std. $\beta = -0.57$, 95% CI [-0.74, -0.40]), and Other ($\beta = -0.53$, 95% CI [-0.62, -0.44], $p < .001$; Std. $\beta = -0.53$, 95% CI [-0.62, -0.44]). This indicates that students with any additional educational experience besides basic education were less likely to drop out. The model is represented in Figure 2 and the statistics are described in Table 1.

Table 1
GLM logistic model parameters

Variable	Estimate	Std. Error	Statistic (z)	p-value
(Intercept)	−0.750	0.055	−13.566	< 0.001***
Age	0.003	0.001	2.570	0.01*
Employment status: Working	0.857	0.049	17.557	< 0.001***
Employment status: Unemployed	0.256	0.057	4.506	< 0.001***
Edu. background: University	−0.581	0.087	−6.708	< 0.001***
Edu. background: VET	−0.439	0.051	−8.653	< 0.001***
Edu. background: Matriculation Exam	−0.325	0.075	−4.361	< 0.001***
Edu. background: Secondary school	−0.566	0.086	−6.610	< 0.001***
Edu. background: Other	−0.527	0.045	−11.737	< 0.001***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

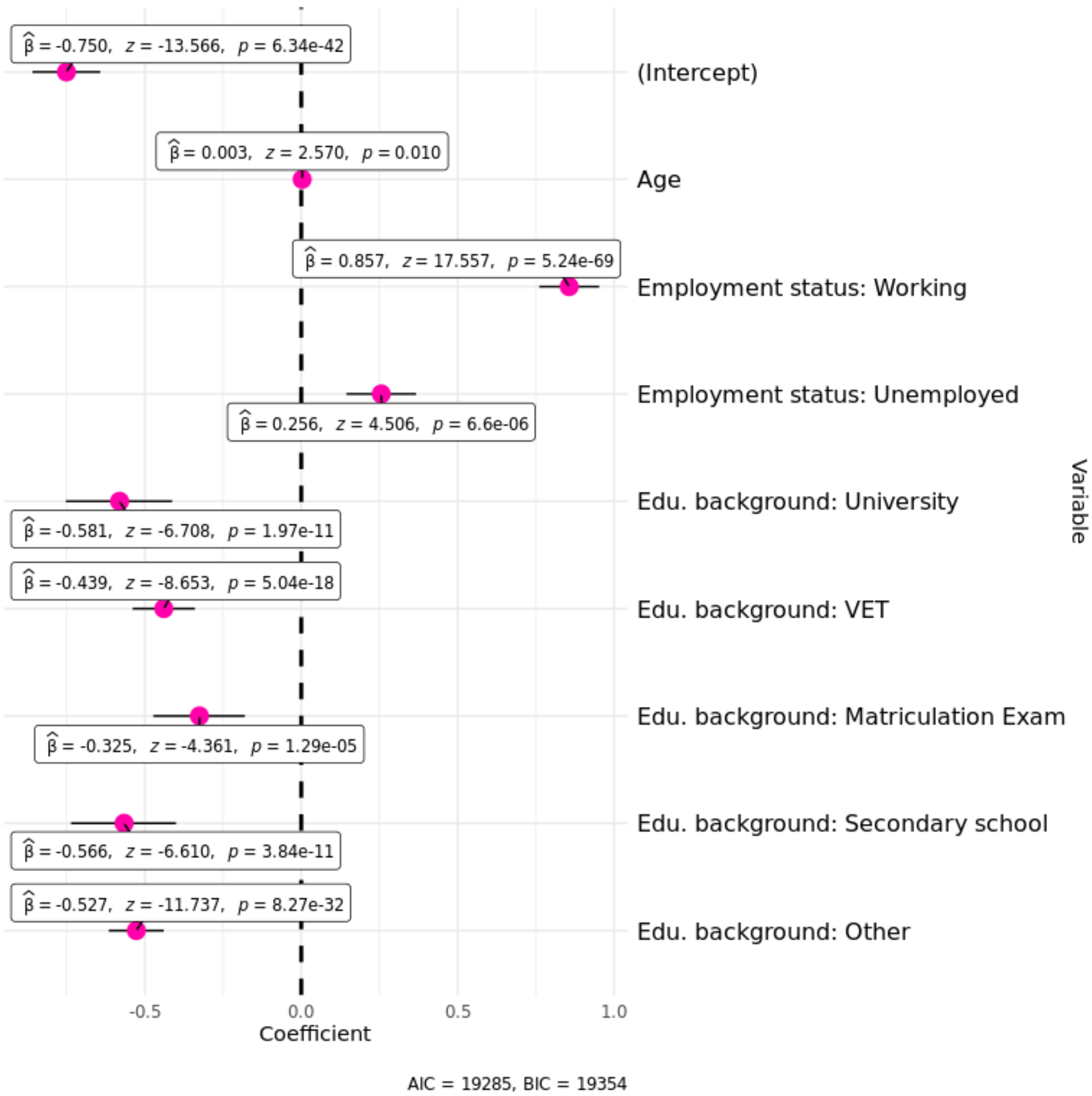

Figure 2: Results of the GLM logistic regression model

4. Discussion

This study aimed to analyze which factors from the admission data can predict dropout in initial VET in Finland. The results suggest that the VET students who might benefit from extra support to avoid dropout are the ones that are working or seeking a job, who might have difficulties combining their studies with other duties. A possible measure to prevent such students from dropping out might be giving them more flexibility and time to complete their tasks. Moreover, since any previous educational experience negatively predicts dropout, students who enroll in VET right after basic education might be the ones that are in greater need of support. In general, VET is considered to require a certain level of self-regulation skills from the students due to, e.g., the growing amount of independent work, high flexibility of how to proceed with studies, more emphasis on workplace learning and individualized competence study paths (e.g., [25]), and not everyone has the capacity to do so after basic education. In addition, students who come from basic education might not have the right perception of the studies that they apply to, which would risk the completion thereof. Since self-regulated learning has shown to be meaningful for academic achievement (see [4]) it may be necessary to support students that lack this ability in order to decrease students' dropout.

Finding ways to support VET students in completing their studies has become increasingly important, as the current COVID-19 pandemic has significantly disrupted learning opportunities, especially in VET [26], [27]. LA and EDM methods have proven useful in predicting factors that are related to dropout. The findings of this article represent a starting point for further research on dropout prevention in VET targeted at the specific factors identified herein. In the future, it would also be interesting to explore additional sources of data, especially those including more fine-grained information of students' progress throughout the program. This represents a challenge for VET since, as mentioned earlier, it is often heavily based on physical activities, on-site teaching, and also workplace learning. Thus, it would require some additional effort from students and instructors to log their activities online. The availability of such data would be useful not only for dropout prediction but also for a better monitoring of the students by the teachers and by themselves. Some vocational schools have already carried out experiments in this regard, but more detailed research results are not yet available.

Lastly, this study is not without limitations. First, the admission data available included less information about the students compared to other earlier studies. However, the results indicate that the variables chosen are statistically significant with relatively narrow confidence intervals. Moreover, using step AIC feature selection allowed us to filter out the variables that only added noise and preserve the ones with significant predictive power. As mentioned earlier, having fine-grained data of students' daily activity is challenging in VET due to its often physical nature. The availability of such data would probably add significantly to our results. Regarding the methods used, compared to prior research, in which algorithms involving neural networks and decision trees were employed, the logistic regression used in this study provides a simpler and less sophisticated method. This would be a drawback if our intention was to create a predictive model that could be used in the admission process to automatically flag students from the beginning. However, the aim of our study was to detect the factors that might predict dropout rather than a blind prediction thereof. More accurate and sophisticated predictive methods (e.g., neural networks) are often hard to interpret by education researchers and practitioners, which would mean that the reason why a student might be flagged as in risk of dropout would be not easily detected, making offering adequate support more challenging.

Acknowledgments

This article was supported by funding from Business Finland through the European Regional Development Fund (ERDF) project "Utilization of learning analytics in the various educational levels for supporting self-regulated learning (OAHOT)" (Grant No. 5145/31/2019).

References

[1] D. Murdoch-Eaton and S. Whittle, "Generic skills in medical education: developing the tools for successful lifelong learning," *Medical Education*, vol. 46, no. 1, pp. 120–128, 2012.

[2] P. Tynjälä, "Perspectives into learning at the workplace," *Educational Research Review*, vol. 3, no. 2, pp. 130–154, 2008.

[3] A. Littlejohn, C. Milligan, R. P. Fontana, and A. Margaryan, "Professional Learning Through Everyday Work: How Finance Professionals Self-Regulate Their Learning," *Vocations and Learning*, vol. 9, no. 2, pp. 207–226, 2016.

[4] C. Mega, L. Ronconi, and R. De Beni, "What makes a good student? How emotions, self-regulated learning, and motivation contribute to academic achievement," *Journal of Educational Psychology*, vol. 106, no. 1, pp. 121–131, 2014.

[5] S. Böhn and Viola Deutscher, "Dropout from initial vocational training – A meta-synthesis of reasons from the apprentice's point of view," *Educational Research Review*, vol. 35, no. 100414, p. 100414, 2022.

[6] H. Yi et al., "Exploring the dropout rates and causes of dropout in upper-secondary technical and vocational education and training (TVET) schools in China," *International Journal of Educational Development*, vol. 42, pp. 115–123, 2015.

[7] Cedefop, *Leaving education early: putting vocational education and training centre stage.* Volume I, Investigating causes and extent. Publications Office, 2016.

[8] Statistics Finland, “Official Statistics of Finland (OSF): Discontinuation of education [e-publication],” 2022 [Online]. Available: http://www.stat.fi/til/kkesk/2020/kkesk_2020_2022-03-17_tie_001_en.html
[9] Ministry of Education and Culture, “Koulutus ja tutkimus vuosina 2011–2016: Kehittämissuunnitelma,” 2012.
[10] S. Ahola, L. Saikkonen, and L. Valkoja-Lähteenmäki, “Ammatillisen koulutuksen läpäisyn tehostamisohjelma, arviointiraportti,” *Helsinki: Opetushallitus*, 2015.
[11] V. Vehkasalo, “Dropout prevention in vocational education: Evidence from Finnish register data,” *Nordic Journal of Vocational Education and Training*., vol. 10, no. 2, pp. 81–105, 2020.
[12] S. López-Pernas and M. Saqr, “Bringing Synchrony and Clarity to Complex Multi-Channel Data: A Learning Analytics Study in Programming Education,” *IEEE Access*, vol. 9, pp. 166531-166541, 2021
[13] M. Saqr and S. López-Pernas, “The Curious Case of Centrality Measures: A Large-Scale Empirical Investigation,” *Journal of Learning Analytics*, vol. 9, no. 1, pp. 13-31, 2022.
[14] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, “Predicting Student Performance using Advanced Learning Analytics,” in *Proceedings of the 26th International Conference on World Wide Web Companion*, Perth, Australia, 2017, pp. 415–421.
[15] H. Aldowah, H. Al-Samarraie, A. I. Alzahrani, and N. Alalwan, “Factors affecting student dropout in MOOCs: a cause and effect decision-making model,” *Journal of Computing in Higher Education*, vol. 32, no. 2, pp. 429–454, 2020.
[16] E. Gedrimiene, A. Silvola, J. Pursiainen, J. Rusanen, and H. Muukkonen, “Learning Analytics in Education: Literature Review and Case Examples From Vocational Education,” *Scandinavian Journal of Educational Research*, vol. 64, no. 7, pp. 1105–1119, 2020.
[17] S. Rovira, E. Puertas, and L. Igual, “Data-driven system to predict academic grades and dropout,” *PLOS ONE*, vol. 12, no. 2. p. e0171207, 2017
[18] A. Dardiri, F. A. Dwiyanto, and A. B. P. Utama, “An integrative review of computational methods for vocational curriculum, apprenticeship, labor market, and enrollment problems,” *International Journal of Advances in Intelligent Informatics*, vol. 6, no. 3. p. 246, 2020
[19] A. Pradeep and J. Thomas, “Predicting College Students Dropout using EDM Techniques,” *International Journal of Computer Applications in Technology.*, vol. 123, pp. 26–34, 2015.
[20] A. M. Shahiri, W. Husain, and N. A. Rashid, “A Review on Predicting Student’s Performance Using Data Mining Techniques,” *Procedia Computer Science*, vol. 72, pp. 414–422, 2015.
[21] Cedefop, *Vocational education and training in Finland: short description.* Publications Office, 2019.
[22] Ministry of Education and Culture, Finnish National Agency of Education, “Finnish VET in a nutshell,” 2019.
[23] W. N. Venables, B. D. Ripley, and S. Isbn, “Statistics Complements to Modern Applied Statistics with S Fourth edition by,” 2002
[24] I. Patil, “Visualizations with statistical details: The ‘ggstatsplot’ approach.” 2018 [Online]. Available: http://dx.doi.org/10.31234/osf.io/p7mku
[25] Nylund, M. & Virolainen, M, “Balancing ‘flexibility’ and ‘employability’: The changing role of general studies in the Finnish and Swedish VET curricula of the 1990s and 2010s.” *European Educational Research Journal*, vol. 18, no. 3, pp. 314–334, 2019.
[26] Cedefop, “Digital gap during COVID-19 for VET learners at risk in Europe. Synthesis report on seven countries based on preliminary information provided by Cedefop’s Network of Ambassadors tackling early leaving from VET,” Cedefop, 2020 [Online]. Available: https://www.cedefop.europa.eu/files/digital_gap_during_covid-19.pdf
[27] International Labour Organization and World Bank, “Skills development in the time of COVID-19: Taking stock of the initial responses in technical and vocational education and training,” 2021 [Online]. Available: https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/WCMS_766557