The **European Centre for the Development of Vocational Training** (Cedefop) is the European Union's reference centre for vocational education and training, skills and qualifications. We provide information, research, analyses and evidence on vocational education and training, skills and qualifications for policy-making in the EU Member States.
Cedefop was originally established in 1975 by Council Regulation (EEC) No 337/75. This decision was repealed in 2019 by Regulation (EU) 2019/128 establishing Cedefop as a Union Agency with a renewed mandate.

Europe 123, Thessaloniki (Pylea), GREECE
Postal: Cedefop service post, 570 01 Thermi, GREECE
Tel. +30 2310490111, Fax +30 2310490020
Email: info@cedefop.europa.eu
www.cedefop.europa.eu

Please cite this publication as:
Hippe, R.; Pokropek, A. and Costa, P. (2022). *Digital capacity building in VET using the SELFIE tool: cross-country validation*. Luxembourg: Publications Office. Cedefop working paper, No 15. http://data.europa.eu/doi/10.2801/000855

Luxembourg: Publications Office of the European Union, 2022

PDF ISBN 978-92-896-3450-2
ISSN 1831-2403
doi:10.2801/000855
TI-BA-22-008-EN-N

Acknowledgements

We gratefully acknowledge support from the European Commission's Joint Research Centre (JRC). The substantial part of the work was conducted while Ralph Hippe was a scientific officer at the JRC Human Capital and Employment unit.

This publication was produced by Cedefop, Department for VET and skills, under the supervision of Antonio Ranieri.

Contents

1. **INTRODUCTION** **4**
2. **THE SELFIE TOOL** **7**
   - 2.1. The overall tool 7
   - 2.2. SELFIE for VET 9
   - 2.3. Data 12
3. **METHODS 14**
   - 3.1. CFA modelling 14
   - 3.2. Model structure 14
   - 3.3. Reliability analysis 16
   - 3.4. Measurement invariance 17
   - 3.5. Validity 19
4. **RESULTS 20**
   - 4.1. Students 20
     - 4.1.1. Model structure 20
     - 4.1.2. Reliability analysis 22
     - 4.1.3. Measurement invariance 23
   - 4.2. Teachers 24
     - 4.2.1. Model structure 24
     - 4.2.2. Reliability analysis 27
     - 4.2.3. Measurement invariance 28
     - 4.2.4. Validity 29
5. **GENERAL DISCUSSION** **31**

ACRONYMS **33**

REFERENCES **34**

Tables and figures

Tables

1. SELFIE structure, items and number of respondents 10
2. Number of respondents by country used in this study 12
3. measures for different model specifications: students' sample................. 22
4. Reliability analysis based on bifactor model: s students' sample 22
5. Reliability analysis based on observed scores: students' sample 23
6. Measurement invariance analysis: students' sample 24
7. Fit measures for different model specifications: teachers' sample 27
8. analysis based on the bifactor model: teachers' sample 27
9. Reliability analysis based on observed scores: teachers' sample 28
10. Measurement invariance analysis: teachers' sample 29
11. OLS regression models: standardised beta coefficients........................... 29

Figures

1. Example of an excerpt from the SELFIE report 8
2. Number of VET students and share of VET in selected countries, 2018... 13
3. Four-dimensional model for students 20
4. Five-dimensional (four areas plus general factor) bifactor model for students 21
5. Eight-dimensional model for teachers 25

CHAPTER 1.
Introduction

The vocational education and training (VET) sector has a key role to play in equalising inequalities in society and offering all learners optimal training conditions for a job that suits them best. Digital learning has a range of potential benefits for students, such as furthering inclusion by supporting disabled students, flexibility in the learning pace, and more independence (JISC, 2021). As the economy is becoming increasingly digitalised, the value of digital and job-specific digital skills is increasing, and companies are trying out new e-learning solutions (Jansseon et al., 2018). Nevertheless, as the recent ET2020 Working Group on VET emphasises, VET has been under-researched, despite being part of a larger innovation ecosystem (Cedefop, 2017; European Commission, 2020a). In addition, a particular gap exists around research on how to integrate digital capacity into vocational schools, though more efforts have been undertaken recently (Broek and Buiskool, 2020). This is very relevant as VET can be a crucial element in boosting digitalisation both in the education sector and in the economy and in increasing the digital capacity of learners (European Commission, 2020a). The importance and impact of VET in the education system highlights this further: in the European Union (EU-27), around half of all upper secondary students are in VET (48%), with more than 10 million students learning in upper secondary vocational schools in 2018 (Eurostat, 2020).

The COVID-19 outbreak has brought into the spotlight the key importance of using digital technologies for teaching and learning in the classroom. However, this is nothing new, as the European Skills Agenda in 2016 already emphasised that 'digital skills are needed for all jobs, from the simplest to the most complex' (European Commission, 2016, p. 2). As a consequence, the new (2020) European Skills Agenda for sustainable competitiveness, social fairness and resilience emphasises that '[…] many people do not have the required level of digital skills or are in workplaces or schools lagging behind in digitalisation' (European Commission, 2016, p. 3). This is a great challenge, as 90% of jobs in Europe require at least basic digital skills, but one citizen in five (22%) does not have any digital skills (European Commission, 2018). For this reason, the European Commission released the renewed Digital Education Action Plan (DEAP) 2021-27 in September 2020 (European Commission, 2020b), which includes continued use of SELFIE in schools; SELFIE was one of the 11 priority actions of the initial DEAP released in January 2018 ([1]). In addition, the Council Recommendation on vocational education and training (VET) for sustainable competitiveness, social fairness and resilience also supports 'qualitative and effective digitalisation of VET provision in both school-based and work-based learning through promoting the use of European competence frameworks and self-assessment tools' (European Commission, 2020c, p. C 417/9), explicitly referring to SELFIE.

Self-evaluation tools can contribute to school improvement research and practice. Chapman and Sammons (2013) present a review of the key debates related to school self-evaluation and which principles and processes are associated with it. School self-evaluation can influence quality in several ways: preparation for inspection; raising standards; professional development; building school capacity to respond to and manage change. For instance, Schildkamp, Visscher and Luyten (2009) showed that self-evaluation results can affect the professional development of teachers. There is also evidence supporting the positive effects of school self-evaluation on school quality and student achievement (Hofman; Dijkstra and Hofman, 2009). These authors suggest that there is a positive association between school self-evaluation policies and both accountability and the desire for improvement. Ilomäki and Lakkala (2018) created a model aiming at improving schools with digital technology which uses the self-evaluation of teachers and students. Their findings indicated that the model worked particularly well for those elements mainly related to the responsibility for leadership inside a school.

In recent years, several tools, online and paper-based, aimed to support schools to self-evaluate or reflect on their digital use or capacity (e.g. Balaban; Redjep and Calopa, 2018; Tanhua-Piiroinen and Viteli, 2017). In Europe, Kampylis et al. (2016) analysed nine online tools used for improving schools' digital capacity, primarily focusing on digital infrastructure and the frequency of technology use. These tools provide information about the digital capacity of the schools based only on school leaders or on school leaders and teachers.

The European Commission's SELFIE (Self-reflection on effective learning by fostering the use of innovative educational technologies) tool supports schools' digital capacity-building and provides a 360° view of how digital technologies are used at the school and at home for training; it involves school leaders (SL), teachers (T) and also students (S) in the process. SELFIE is an online

([1]) The European Commission Communication School development and excellent teaching for a great start in life from 2017 already specifically called for the development of 'a self-assessment tool on digital capacity so that schools in the EU can, on a voluntary basis, self-evaluate where they stand in relation to common criteria and are supported in developing and improving their effective use of technologies for digital age learning' (European Commission, 2017, p. 4).

customisable tool that can help schools to improve their digital capacity (European Commission, 2020a). This tool anonymously gathers the views of students, teachers and school leaders on how technology is used in their school (Castaño-Muñoz et al., 2018). Participation in SELFIE is free of charge and any school in the world can use it. It was launched in October 2018 and is now available in over 30 languages. It has been developed by the European Commission in collaboration with a team of education experts, ministries of education and research institutes from across Europe (European Commission, 2021a).

The aim of this paper is to validate the SELFIE tool in the VET context and check the cross-comparability of the tool among different countries. The paper is structured as follows: first, we provide more details on the characteristics of the SELFIE tool, before moving on to describe in more detail the data and the psychometric methods used. Then, we show the results of our analyses and discuss them. Finally, the last section of the paper concludes and indicates directions for future research.

CHAPTER 2.
The SELFIE tool

2.1. The overall tool

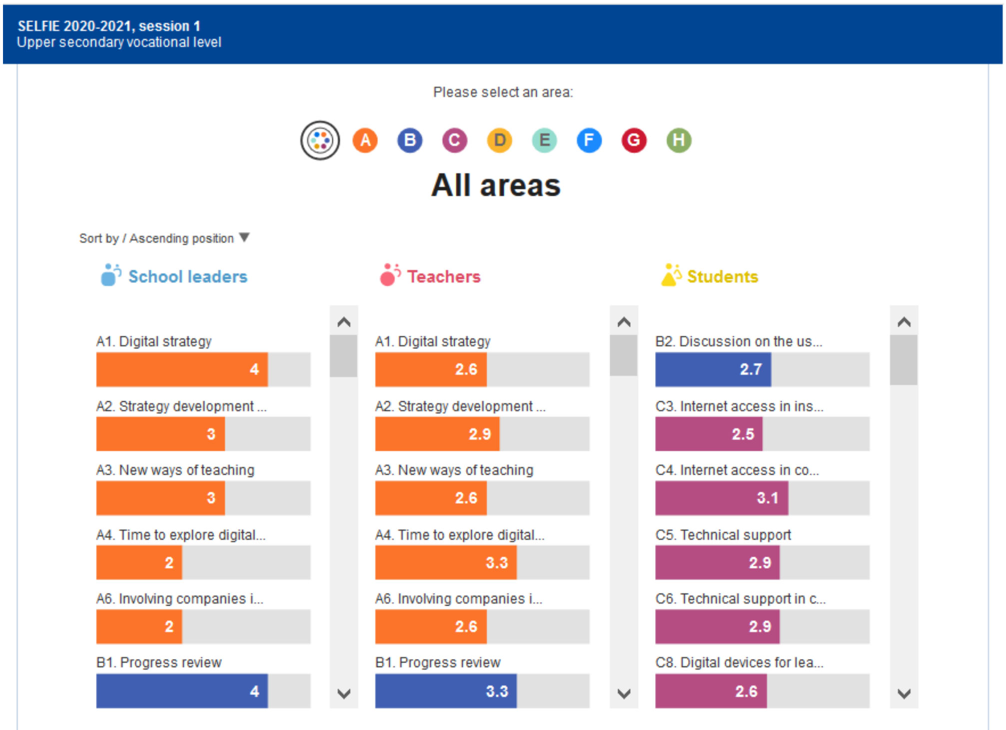
SELFIE has been designed to be used by schools offering general education (primary: ISCED 1, lower and upper secondary: ISCED 2 and 3), upper secondary vocational schools (ISCED 3: VET) and post-secondary non-tertiary education institutions (ISCED 4: PSNTE).

The process of developing the tool involved a range of quality checks, including both expert consultations and data analysis. It followed a participatory design and involved also mixed-method research. The theoretical underpinning of SELFIE is the European Framework for Digitally Competent Educational Organisations, DigCompOrg, which was published in 2015 (Kampylis; Punie and Devine, 2015). Following a meta-analysis of 15 existing tools to inform the practical design of the new tool, expert consultations involved the ET2020 Working group on digital skills and competences, the organisation of a dedicated SELFIE community workshop, and a user consultation with more than 5 000 school leaders, teachers and students. Based on the results of these consultations, successful pilot testing took place in 2017 in 14 European countries with more than 650 schools (Castaño-Muñoz et al., 2018). This pilot phase included 28 case studies (with qualitative research performed within the framework of focus groups and interviews), 14 country reports and analysis of the collected pilot data (e.g. thematic and descriptive analyses, item analysis and confirmatory factor analysis). Finally, a validation and scoping workshop once again consulted the opinion of many internationally recognised experts and paved the way for the final version of the tool, translated into all 24 official EU languages, which was eventually released in 2018.

The version for the VET sector has been jointly developed with all other levels, so that it has followed the same quality checks and tests. However, to adapt the general questionnaire to the specific needs of VET schools, the European Centre for the Development of Vocational Education and Training (Cedefop) and the European Training Foundation (ETF) have contributed their expertise on VET to ensure the relevance of the questionnaire. The overall assessment was that the questionnaire for general upper secondary schools needs only small changes to be a good fit for the VET sector (Hippe, 2020).

Once the participants have completed the SELFIE questions, schools using the tool receive a customised report with aggregated results in a visual and interactive way, providing insights into what is working well and what could be improved in the school (Figure 1). The report belongs to the school – and only the school can access it – and it can help the school community to discuss its approach to embedding technology and developing digital skills for staff and students.

Figure 1. **Example of an excerpt from the SELFIE report**



Source: European Commission (2021b).

Every school can adapt the questionnaire and even add new questions so that SELFIE suits its needs. There are several types of items in SELFIE: a set of core items that are the same and obligatory for all respondents, every school and school level; a set of optional items that school coordinators can choose from; a set of up to 10 items that the school coordinator can add to adapt the questionnaire to the needs and context of their school; additional items about the use of digital technologies for teaching and learning inside and outside school; a few demographic questions; and a set of items that are only for upper secondary vocational schools and directed towards their specific context.

2.2. SELFIE for VET

In this paper we focus on data from vocational schools based on the 36 core items answered by SELFIE participants and the two VET-specific items. These items are composed of short statements and questions with five answer options: 1: strongly disagree – In my experience, this is not true at all; 2: disagree; 3: slightly agree; 4: agree; 5: strongly agree – In my experience, this is very true. In addition, in all items there are also the answer options 'Not applicable' or 'Prefer not to say'.

The core and VET items for teachers ([2]) are organised into eight areas (Costa; Castaño-Muñoz and Kampylis, 2021): leadership (LE), infrastructure and equipment (IE), continuing professional development (CPD), pedagogy: supports and resources (PS), pedagogy: implementation in the classroom (PI), assessment practices (AP), student digital competence (DC) and collaboration and networking (CN).

The leadership area relates to the role of leadership in the school-wide integration of digital technologies and their effective use for the school's core work: teaching and learning. The area collaboration and networking relates to measures that schools may consider to support a culture of collaboration and communication for sharing experiences and learning effectively within and beyond the organisational boundaries. The area infrastructure and equipment is about having adequate, reliable and secure infrastructure (such as equipment, software, information resources, an internet connection, technical support or physical space). The continuing professional development (CPD) area looks at whether the school facilitates and invests in the CPD of its staff at all levels. The area pedagogy: supports and resources, refers to the preparation of using digital technologies for learning by updating and innovating teaching and learning practices. The area pedagogy: implementation in the classroom, relates to the implementation of digital technologies for learning in the classroom, by updating and innovating teaching and learning practices. The assessment practices area relates to measures that schools may consider in order to shift the balance gradually from traditional assessment towards a more comprehensive repertoire of practices (student-centred, personalised and authentic). The area on Student digital competence includes questions on a set of skills, knowledge and attitudes that enable the confident, creative and critical use of digital technologies by the students. Table 1 presents an overview of the tool together with the samples used in this study.

([2]) The students' questionnaires have a much lower number of questions than the teachers' questionnaire and prior empirical evidence and theory supports the four-dimensional structure for students (areas: student digital competence; infrastructure and equipment; assessment practices; and pedagogy: implementation in the classroom).

Table 1. **SELFIE structure, items and number of respondents**

Item Code	Item title	Item (for T)	N of S	N of T
LE		**Leadership**		
LE1	Digital strategy	In our school, we have a digital strategy	0	11 981
LE2	Strategy development with teachers	Our school leaders involve us teachers in the development of the school's digital strategy	0	12 122
LE3	New ways of teaching	Our school leaders support me in trying out new ways of teaching with digital technologies	0	12 253
CN		**Collaboration and networking**		
CN1	Progress review	In our school, we review our progress in teaching and learning with digital technologies	0	12 110
CN2	Discussion on the use of technology	In our school, we discuss the advantages and disadvantages of teaching and learning with digital technologies	0	12 259
CN3	Involving companies in strategy	In our school, companies we collaborate with are involved in the development of the school's digital strategy	0	10 908
IN		**Infrastructure and equipment**		
IN1	Infrastructure	In our school, the digital infrastructure supports teaching and learning with digital technologies	73 026	12 393
IN2	Digital devices for teaching	In our school, there are digital devices for me to use for teaching	72 578	12 397
IN3	Internet access	In our school, there is access to the Internet for teaching and learning	73 198	12 401
IN4	Technical support	In our school, technical support is available in case of problems with digital technologies	0	12 380
IN5	Data protection	In our school, there are data protection systems in place	0	11 652
IN6	Digital devices for learning in school	In our school, there are school-owned/managed digital devices for students to use when they need them	0	12 248
INv	Database of training opportunities	In our school, students have access to a database of in-company training opportunities	69 797	11 361
CO		**Continuing professional development**		
CO1	CPD needs	Our school leaders discuss with us our CPD needs for teaching with digital technologies	0	12 243
CO2	Participation in CPD	I have opportunities to participate in CPD for teaching and learning with digital technologies	0	12 243
CO3	Sharing experiences	Our school leaders support us to share experiences within school about teaching with digital technologies	0	12 255
PS		**Pedagogy: supports and resources**		12 365
PS1	Online educational resources	I search online for digital educational resources	0	12 186
PS2	Creating digital resources	I create digital resources to support my teaching	0	11 755
PS3	Using virtual learning environments	I use virtual learning environments with students	0	12 246

Item Code	Item title	Item (for T)	N of S	N of T
PS4	Communicating with the school community	I use digital technologies for school-related communication	0	0
PI		**Pedagogy: implementation in the classroom**		
PI1	Tailoring to students' needs	I use digital technologies to tailor my teaching to students' individual needs	73 130	12 058
PI2	Fostering creativity	I use digital technologies to foster student's creativity	72 943	11 965
PI3	Engaging students	I set digital learning activities that engage students	73 151	12 004
PI4	Student collaboration	I use digital technologies to facilitate student collaboration	73 162	11 803
PI5	Cross-curricular projects	I engage students in using digital technologies in cross-curricular projects	72 688	11 403
AP		**Assessment practices**		
AP1	Assessing skills	I use digital technologies to assess students' skills	72 301	11 594
AP2	Timely feedback	I use digital technologies to provide timely feedback to students	72 727	11 550
AP3	Self-reflection on learning	I use digital technologies to enable students to reflect on their own learning	72 184	11 371
AP4	Feedback to other students	I use digital technologies to enable students to provide feedback on other students' work	0	11 160
DC		**Student digital competence**		
DC1	Safe behaviour	In our school, students learn how to behave safely online	72 968	11 972
DC2	Responsible behaviour	In our school, students learn how to behave responsibly when they are online	72 983	12 017
DC3	Checking quality of information	In our school, students learn how to check that the information they find online is reliable and accurate	72 988	11 969
DC4	Giving credit to others' work	In our school, students learn how to give credit to others' work they have found online	72 732	11 892
DC5	Creating digital content	In our school, students learn to create digital content	72 975	11 915
DC6	Learning to communicate	In our school, students learn to communicate using digital technologies	73 055	12 066
DCv	Skills for vocational qualification	In our school, students develop digital skills related to their vocational qualification	72 715	11 815

NB: N = number of observations, T = teachers, S = students. Items refer to questions asked to T. The same items for S are usually very similar, with some adaptations where needed to match the items to this respondent group. Details can be found in the SELFIE overview questionnaire available in the SELFIE online tool.

Source: Extract from European Commission (2020d).

2.3. Data

For this study, we use data extracted from the SELFIE tool database (European Commission, 2020d) ([3]). More specifically, these data stem from the first six sessions ([4]) that SELFIE has been in operation, i.e. between October 2018 and August 2020. Data were collected automatically by the SELFIE tool when respondents finished filling out the online questionnaires. SELFIE is a tool that is free to use by any school and participation is voluntary. This means that the sample is not random: respondents have self-selected to participate. This has important implications, and general conclusions on the digital capacities of schools in a given country cannot or can only be taken very cautiously as the results are not representative. Nevertheless, they offer a detailed (but imperfect and non-representative) view, contributing to gathering knowledge on the needs of schools in digital education.

For this study we applied the criterion of a minimum of 10 VET schools in a given country participated in a study. This gives us the following 11 countries: Belgium, Bulgaria, Germany, Hungary, Italy, Moldova, Montenegro, Portugal, Serbia, Spain, and Turkey. Although SELFIE surveys students, teachers and school leaders, in this analysis we focused only on the first two groups. The number of school leaders in particular countries was too low (in most of the countries the number of participating school leaders is less than 100) to perform reliable estimates of psychometric models for each country and assess the comparability of the tool.

Table 2. **Number of respondents by country used in this study**

Country	T	S	N of schools
Belgium	772	3 599	37
Bulgaria	1 909	14 581	147
Germany	456	2 915	21
Spain	1 330	7 073	123
Italy	443	2 792	25
Hungary	1 378	9 662	45
Italy	443	2 792	25

([3]) We report how we determined our sample size, all data exclusions, all manipulations, and all measures in this study.

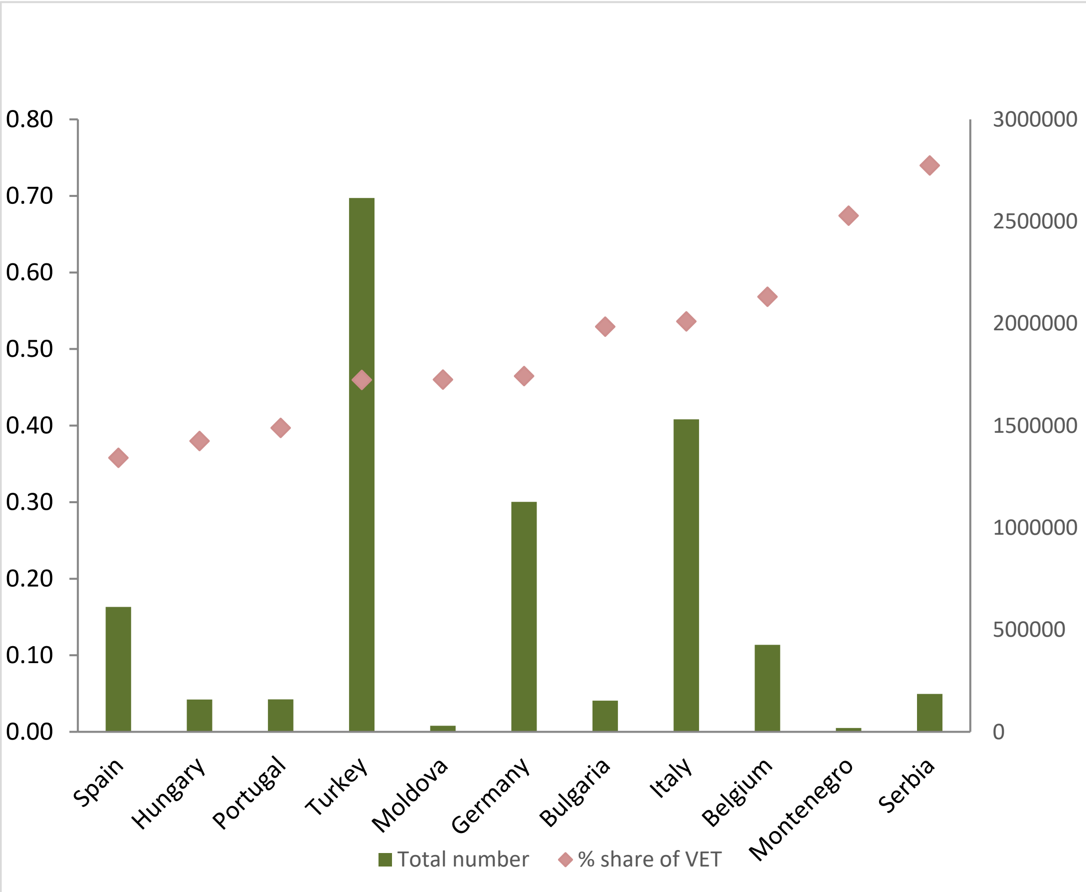
([4]) SELFIE divides the school year into three sessions: August to December, January to April, and April to August. The reason that the first data stem from October 2018 is due to the fact that the tool was officially launched at this time.

Country	T	S	N of schools
Portugal	143	841	13
Moldova	503	4 697	13
Montenegro	590	3 037	21
Serbia	4 471	21 996	146
Turkey	459	3 080	18
Total	12 454	74 273	609
Total (T + S)	86 727		609

Source: Authors.

We have 86 727 VET teachers and students in our database (Table 2). VET plays an important role in all of the countries in our sample, as the share of VET students in all upper secondary students is between 35% and 75% (Figure 2; Eurostat, 2020).

Figure 2. **Number of VET students and share of VET in selected countries, 2018**



NB: Share of VET refers to the share of upper secondary VET in upper secondary education.
Source: European Training Foundation (2020), Eurostat (2020).

CHAPTER 3.
Methods

3.1. CFA modelling

To investigate the psychometric properties of SELFIE, we used confirmatory factor analysis (CFA) modelling and the multiple group confirmatory factor analysis (MG-CFA) framework for students and teachers. We kept the standard assumptions of CFA modelling, that observed indicators are continuous and the relationship between indicators and the unobserved latent traits are linear. This is a simplification because responses in SELFIE were measured on a 5-point Likert. However, simulation studies have shown that when the number of categories is at least four and the distribution of responses are normal, the use of CFA modelling for continuous indicators is justified (Beauducel and Herzberg, 2006; DiStefano, 2002; Johnson and Creech, 1983; Muthén and Kaplan, 1985; Pokropek et al., 2019). We decided to use linear models for two reasons. First, models with continuous indicators reduce the complexity of the estimation and avoid convergence problems. Second, testing measurement invariance for continuous data is less problematic compared to the categorical indicators where the competing needs of identification and invariance constraints might cause problems in robust assessment (Wu and Estabrook, 2016). For analysis, we used the *lavaan* R package (Ros, 2012). We used full information maximum likelihood estimation that accounts for missing data.

3.2. Model structure

To investigate the structure of the SELFIE tool in VET settings we used a strategy based on comparisons of CFA models. As the SELFIE tool was designed, constructed and pre-tested using an imposed theoretical structure, there is no need for exploratory analysis, and we focused on testing whether the model fits the predefined specification focusing on the question of whether the theoretical structure fits the data. To do this we tested a series of CFA models.

M1) One-dimensional CFA model. The first model assumes that responses are directly related to the underlying unidimensional factor, reflecting general digital capacity, and do not differentiate between different aspects of capacity. This model is unlikely to fit the data well because SELFIE was constructed and pre-tested using an imposed theoretical structure, but it should be treated as the reference for strict unidimensionality.

M2) Multi-dimensional CFA model. This model assumes that multiple different but correlated latent digital capacities best describe patterns of responses to the SELFIE. This model would suggest the need for using scores from all the areas of digital capacity to describe the response pattern fully.

M3) Bifactor CFA model. The bifactor model considers the interconnected but independent role of area-specific abilities and a common general factor of digital capacity (g factor). It implies that an underlying g factor shapes general digital capacity, but also that specific sets of capacity contribute to within-individual differences in general digital capacity (Gustafsson and Balke, 1993). The model assumes the latent structure where each item is directly related to the general factor and to one of a set of mutually orthogonal specific factors, representing the portion of the variance in items, which are not explained by the general factor and are specific to a given set of items (Reise; Moore and Haviland, 2010). Those specific factors may have a dual nature. On the one hand, they can relate to the specific dimensions substantially significant to the entire structure. On the other hand, they may reflect differences in measurement tools, and therefore be considered as disturbing factors. In this application, we assume that the g factor reflects general digital capacity, and the specific factors reflect areas of capacity (Reise; Bonifay and Haviland, 2013, 2018; Rodriguez; Reise and Haviland 2016a; 2016b).

We also tested models where we allow for extra dependencies between specific items. Local item dependencies, that is the correlations between items that are not explained by the model, could introduce noise dimensions and, if ignored, could reduce the model fit, resulting in problems with the estimate (Brown, 2015), and the assessment of measurement invariance (Braeken and Blömeke, 2016). We used correlated error terms conservatively based on substantive justification and empirical testing using Lagrange multiplier statistics (modification indices) along with the expected parameter change statistics. This approach pointed to four pairs of correlated residuals in refined models (dc1 and dc2; dc5 and dc6; in1 and in2; cn3 and inv) for both the student and teacher samples. In this paper, we refer to the models with correlated errors (both multidimensional and bifactor) as refined models.

We use four standard fit measures to assess the model fit: comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA) and standardised root mean square residual (SRMR), applying RMSEA/SRMR <0.08 and CFI/TLI ≥0.90 threshold criteria as a definition of acceptable fit for multifactor rating tools when analyses are performed at the item level and there are multiple factors (for more details, Kline, 2005; Marsh et al., 2004).

3.3. Reliability analysis

As full reliance on comparisons of model fit has been criticised for being overly simplistic (e.g. Berge and Sočan, 2004; Bentler, 2009), we supplemented comparisons with statistical indices primarily derived from the bifactor model, to identify the sources of the common variance in responses to questionnaire items (Reise; Bonifay and Haviland, 2018; Gignac and Kretzschmar, 2017). We used the explained common variance (*ECV*) index (Reise; Bonifay and Haviland, 2018), the Omega bifactor model-based reliability indices (Raykov, 1997; Reise; Bonifay and Haviland, 2013; Reise at al., 2013), construct replicability H-index (Hancock and Mueller, 2001) and factor determinacies (Gorsuch, 1983). We also calculated Haberman's (2008) proportional reduction in mean squared error (*PRMSE*) for subscale scores.

ECV could be computed for all factors and is the proportion of common variance of the items in each factor which is due to that factor. ECV for the general factor is the common variance explained by the general factor divided by the total common variance. It indicates the relative 'strength' of the general factor. Reise et al. (2013, p. 22) suggest that 'the presence of some multidimensionality is not severe enough to disqualify the interpretation of the tool as primarily unidimensional' when the percent of uncontaminated correlations is low, that is the percentage of correlations not involving the specific factors, and ECV values on the general factor are above 0.6 (Reise et al., 2013) or above 0.7 according to stricter criteria (Rodriguez; Reise and Haviland, 2016).

Omega indicates how much of the variance in standardised (unit-weighted) scores can be attributed to all common factors, that is all factors related to a set of items. For the general factor, Omega is calculated from all items. For specific factors, Omega is calculated only with the items belonging to the respective domain. The OmegaH indicates how much reliable variance of the standardised total scores can be attributed to each factor (Reise et al., 2013). For the general factor, the higher the OmegaH, the more the general factor is the dominant source of systematic variation. OmegaH for a general factor higher than 0.80 indicates unidimensionality because most of the reliable variance is due to a single common factor. OmegaH for subscales is the proportion of subscale score variance attributable to a group factor after removing the reliable variance due to the general factor (Rodriguez; Reise and Haviland, 2016a, 2016b). The ratio of Omega to OmegaH describes how much of the reliable variance in total scores is accounted for by the general factor *g* and the other specific factors. High Omega and OmegaH indicates multidimensionality. Omega and OmegaH and their ratios are computed for each of the orthogonal factors. When the ratio for a subscale is low, much of the reliable variance of the subscale scores can be attributed to the general factor. If the ratio is high there is substantial unique subscale reliable variance.

Construct replicability H-index informs on whether the CFA model is suitable and replicable across potential implementations. The values of the index represent the correlation between a factor and an optimally weighted item composite score. High H values suggest a well-defined latent variable, which is likely to be stable across potential future studies. Low H values suggest a poorly defined latent variable, which is likely to change across potential future studies (Rodriguez et al., 2016a; 2016b).

Factor determinacies (FD) could be interpreted as expected correlations between factor scores and latent factors. Gorsuch (1983) suggested that FD should be higher than 0.90 to justify the appropriate use of factor scores in subsequent analysis.

PRMSE indicates the relative importance of specific dimensions over the general dimension in explaining variability in response patterns to test items (Haberman, 2008). PRMSE for a specific subscale is also an estimate of the 'degree to which the measurement error on a subscale is reduced based on the subscale reliability' (Reise; Bonifay and Haviland, 2013, p. 131). The PRMSE ratio indicates the extent to which separate scaling increases or reduces the amount of information conveyed in the scale. If the PRMSE ratio for a specific subscale is greater than 1.0, the corresponding dimension is considered to add information beyond that provided by the general factor. If the PRMSE ratio is less than 1.0, the addition of specific dimensions does not provide additional information. At the conceptual level, PRMSE is like the Omega coefficients. However, PRMSE is computed on observed scores and does not assume that the factors are orthogonal or an underlying bifactor model. PRMSE is an additional check for robustness since it does not depend on the validity of the model specification and is not affected by estimation issues (Haberman; Sinharay and Puhan, 2008). Together with PRMSE, we also present a classical indicator of reliability, Cronbach's alpha coefficient (Cronbach, 1995), which could be used for facilitating comparisons with similar tools.

3.4. Measurement invariance

Measurement tools are used to assess (or allow a self-reflection of) different groups of individuals. The responsibility of measurement tool constructors is to assure that the tool provided treats the groups equally and fairly. This means that the measurement tool functions in the same way for different groups, and respondents with the same value of measured latent trait receive the same score regardless of a group membership. Technically, this desired feature is referred to as measurement invariance (MI).

Multi-group confirmatory factor analysis (MG-CFA) was used (Byrne et al., 1989) to investigate the measurement invariance of the SELFIE tool. In this approach sequential tests are employed to impose increasingly restrictive equality constraints on parameters of interest across comparison groups (MG-CFA; Byrne et al., 1989; Horn and McArdle, 1992; Jöreskog, 1971; Meredith, 1993). Following the previous studies (Cheung and Rensvold, 2002; Horn and McArdle, 1992; Vandenberg and Lance, 2000), we tested five models:

(a) MI1 configural invariance model: the same factor structure is imposed on all groups;
(b) MI2 weak invariance model: the factor loadings are constrained to be equal across groups;
(c) MI3 scalar (strong) invariance model: the factor loadings and intercepts are constrained to be equal across groups;
(d) MI4 partial scalar invariance model: most of the factor loadings and intercepts are constrained to be equal across groups, but for items that exhibit great differences between item parameters in different groups the equality constraints are relaxed (Pokropek et al., 2019 for discussion);
(e) MI5 residual invariance model: the factor loadings, intercepts and residual variances are constrained to be equal across groups.

The proper fit of the first model tells us whether a measurement tool could be treated as comparable across countries. A good fit of the second model legitimises comparisons between correlations and regression coefficients in different countries in in-depth analysis. The third and fourth models suggest the possibility of valid comparisons of the factor means. All previous models legitimise comparisons based on SEM modelling or using unbiased factor scores estimates, while a good fit of model five ensures direct comparisons of scores based on the simple sum of the respondents' answers (Davidov et al., 2014 for a detailed discussion).

For assessing measurement invariance, we used CFI and RMSEA fit indices. The difference between the CFI values or RMSEA values, respectively, was computed for the restricted model against the less restricted model. A change of CFI by 0.010 accompanied by a change of RMSEA by 0.015 is usually considered as a significant reduction in model fit and hence non-invariance (Chen, 2007). This criterion was used also in this study.

3.5. Validity

The SELFIE tool analysed in this paper collects detailed information on the use of digital technology for teaching and learning in VET schools. It was designed for best use of the respondent time to generate valid responses with reasonable workload; therefore, only limited additional information was collected in the questionnaires. This means that a full validity study could not be performed within the scope of the current study. We conducted only introductory basic external validity, relying on the teachers' responses and four pieces of information which are collected within the SELFIE self-reflection exercise: teachers' age; years of experience; percentage of teaching time spent using digital technologies in the classroom; and general satisfaction with the SELFIE tool, based on the responses to the question 'How likely is it that you would recommend SELFIE to a colleague?'. This will allow us to check to what extent the results of our study can be generalised to and across other contexts.

We assume that age, controlled for experience, should be negatively related to the SELFIE areas, while experience, controlled for age, is positively related. The distinction between experience and age is especially important while studying digital capacities in education. On the one hand, previous studies showed that teachers' age is usually negatively correlated with different aspects of digital competences (Hatlevik and Arnseth, 2012; Tomczyk, 2009; Carlo et al., 2019; Krumsvik et al., 2016). On the other hand, teaching experience is positively related to various aspects of digital competences and teaching (Hatlevik and Arnseth, 2012; Tomczyk, 2009; Carlo et al., 2019) or not significantly related (Krumsvik et al., 2016). We assume that positive attitudes towards the self-reflection tool are an indirect indicator of positive attitudes towards digital technologies in education. Therefore, they should be positively associated with the digital capacity measured by SELFIE. We also assume that three areas (Pedagogy: supports and resources; Pedagogy: implementation in the classroom; Assessment practices) should be more related to age, experience and positive attitudes towards SELFIE. Those three areas are directly related to individual work, while others relate more to school institutional settings.

To investigate these relationships, we used ordinary least squares regression models with country fixed effects. For each SELFIE area, factor scores were obtained and used as the dependent variables in modelling. Standardised beta coefficients with statistical significance were presented to investigate conditional relations between SELFIE areas and criterion variables.

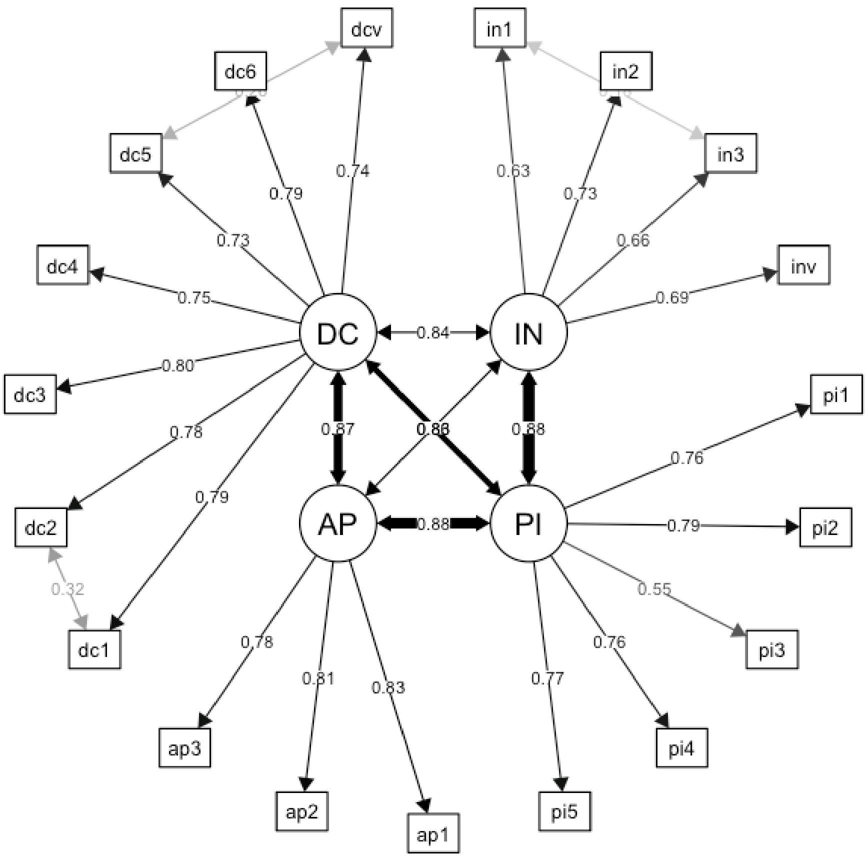
CHAPTER 4. Results

4.1. Students

4.1.1. Model structure

In order to investigate the structure of the SELFIE tool for VET students, Figure 3 and Figure 4 present a graphical representation of the 4D four-dimensional (M2) and bifactor (five-dimensional: four SELFIE areas plus general factor ([5]) (M3)) models for students, respectively. On the graph, standardised factor loadings (single-ended arrow), residual correlations (grey double-ended arrows) and correlations between factors (black double-ended arrows) are presented. Factor loadings are depicted in circles and observed indicators by rectangles.

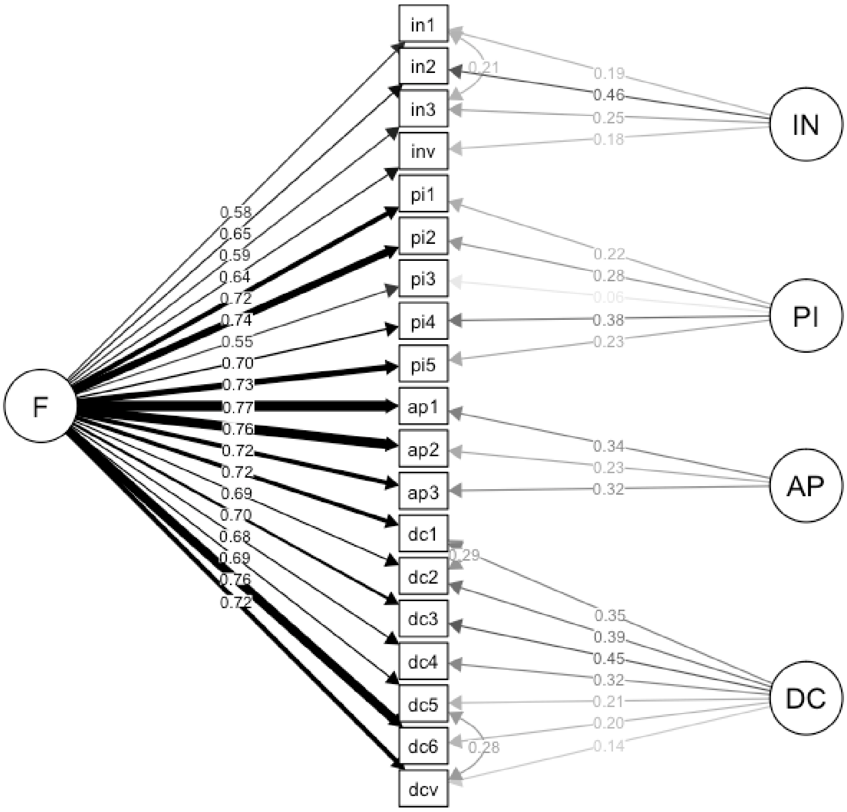
Figure 3. **Four-dimensional model for students**



Source: Authors.

([5]) Schools' digital capacity self-reflected by students.

Figure 4. **Five-dimensional (four areas plus general factor) bifactor model for students**



Source: Authors.

High factor loadings (in all cases higher than 0.5 and in most cases higher than 0.7) of the multidimensional solution (M2) show that the overall structure fits well to the data. However, all four factors are very highly correlated: in all cases higher than 0.8 and in most cases the correlations are close to 0.9. This means that around 70% to 80% of the variation in one factor could be explained by another factor. This is reflected in the bifactor model representation (M3) where all items load strongly on the general factor representing one single dimension (in all cases it is higher than 0.5, in most cases higher than 0.7), while factors for specific loadings are much lower (in all cases not higher than 0.5 and in most cases not higher than 0.3). This structure of factor loadings suggests that the digital capacity of students is explained well by one single dimension.

Table 3 presents the fit measures of the unidimensional model (M1, baseline model), multidimensional model (M2) and bifactor (M3) as well as the refined multidimensional (M2´) and bifactor models (M3´), which are models with residual correlations for four pairs of items (dc1 and dc2; dc5 and dc6; in1 and in2; cn3 and inv).

Table 3. **Fit measures for different model specifications: students' sample**

Model/Fit	CFI	TLI	SRMR	RMSEA	BIC	AIC
M1 – Unidimensional	0.953	0.945	0.032	0.061	3 415 689	3 415 298
M2 – 4-D	0.969	0.964	0.024	0.049	3 403 621	3 403 222
M3 – Bifactor	0.982	0.976	0.017	0.040	3 394 338	3 393 820
M2´ – 4-D refined	0.984	0.981	0.019	0.036	3 392 335	3 391 908
M3´ – Bifactor refined	0.991	0.988	0.013	0.028	3 387 542	3 386 997

Source: Authors.

The results clearly show that the bifactor model fits best according to all fit measures. This suggests essential unidimensionality of the measurement tool. For a more in-depth investigation we now explore the bifactor indices and reliability measures presented in the next section.

4.1.2. Reliability analysis

Table 4 shows the reliability analysis based on the bifactor model. ECV for the general factor is 0.84; this is a high value and, together with PUC (6) = 0.77, could be recognised as a strong indicator of essential unidimensionality. This is also confirmed by low values of ECV for specific factors. High Omega coefficients confirm high reliability of the general factor (0.957) together with high reliabilities of specific factors (from 0.786 to 0.917). However, low OmegaH coefficients (from 0.101 to 0.158) and Omega ratio (from 0.118 to 0.200) for subscales indicate that only a small portion of variation in student responses could be attributable to the specific factors.

Table 4. **Reliability analysis based on bifactor model: students' sample**

Factor/measure	ECV	Omega	Omega H	Omega Ratio	H	FD
General factor – digital capacity	0.836	0.957	0.916	0.957	0.949	0.961
Infrastructure and equipment	0.222	0.786	0.158	0.200	0.338	0.649
Pedagogy: implementation	0.133	0.853	0.101	0.118	0.288	0.618
Assessment practices	0.150	0.849	0.126	0.148	0.248	0.613
Student digital competence	0.165	0.917	0.122	0.133	0.470	0.760

Source: Authors.

(6) PUC refers to Percent of Uncontaminated Correlations.

The Omega ratios indicate that between 12% and 20% of the variation in responses is domain/factor-specific. The indices H and FD are high for the general factor, at 0.949 and 0.961 respectively, but low for specific factors. H indices for specific factors are not higher than 0.470 and FD indices not higher than 0.760.

Table 5 indicates a reliability analysis based on observed score analysis. The results correspond to the bifactor indices analysis. We observe high reliability for general factor with Alpha being 0.946 while the reliability of sub-scores is substantially lower (from 0.781 to 0.907). Overall reliability of sub-scores is acceptable although the PRMSE analysis shows that sub-scores bring no or very little additional information above the general score. Only sub-scores for assessment practice and student digital competence areas provide some additional information above the general dimension, although the gain is only 1% and 6%, respectively.

Table 5. **Reliability analysis based on observed scores: students' sample**

Factor/ measure	Alpha	PRMSEs	PRMSEx	Added Value
General factor – digital capacity	0.946	NA	NA	NA
Infrastructure and equipment	0.781	0.781	0.782	No
Pedagogy: implementation	0.832	0.832	0.847	No
Assessment practices	0.827	0.827	0.815	Yes (1%)
Student digital competence	0.907	0.907	0.857	Yes (6%)

NB: NA stands for non-applicable.
Source: Authors.

4.1.3. Measurement invariance

Table 6 presents our measurement invariance analysis for the students. The results clearly show that metric invariance holds with only a marginal change of fit compared to the configural model. The scalar model, on the other hand, fits substantially worse than the metric model with ΔCFI greater than 0.015. This result suggests some non-invariance in the intercept. We explored modification indices and found that non-invariance is primarily the problem of four items: in1, in3, pi3, pi4. After releasing those items to be freely estimated in each group, the partial scalar model fits reasonably well with ΔCFI being 0.010 and ΔRMSEA being 0.003 compared to the metric model. The residual invariance model (even in partial settings with no constraints on four items) fits substantially less than the scalar (and metric) model.

Table 6. **Measurement invariance analysis: students' sample**

Model/Fit	CFI	RMSEA	ΔCFI	ΔRMSEA
MI1 – Configural	0.934	0.058	NA	NA
MI2 -Weak – Metric (loadings)	0.932	0.058	0.002	0.000
MI3 – Scalar (loadings and intercepts)	0.915	0.064	0.017	0.005
MI4 – Partial scalar	0.922	0.061	0.010	0.003
MI5 – Residuals (partial)	0.904	0.066	00.018	0.005

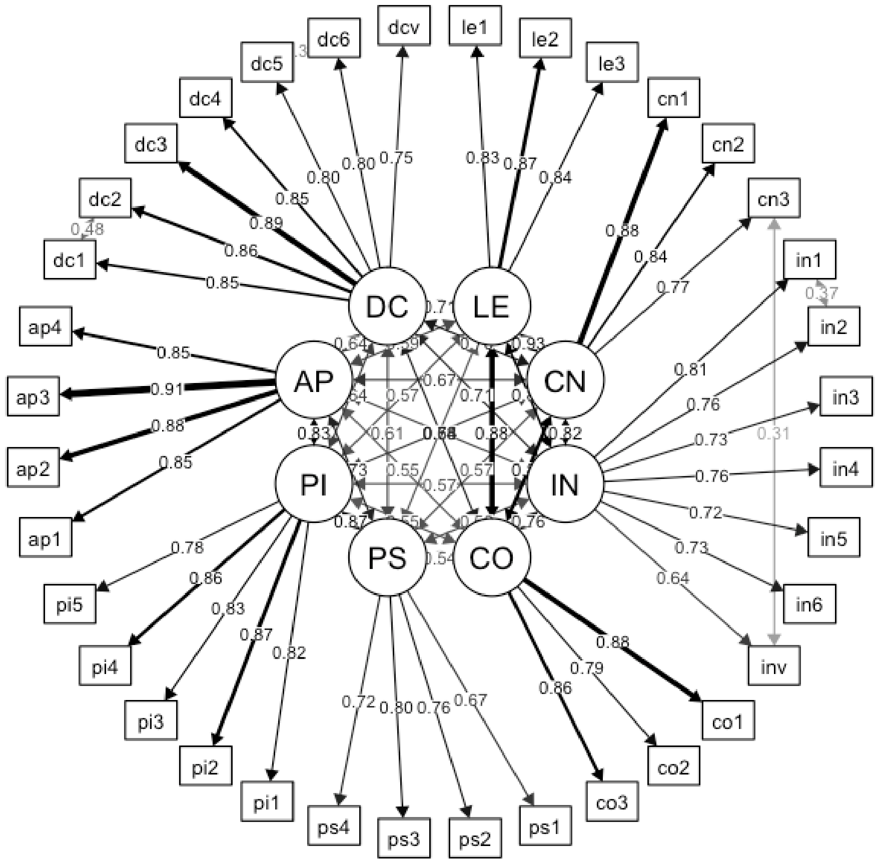
Source: Authors.

4.2. Teachers

4.2.1. Model structure

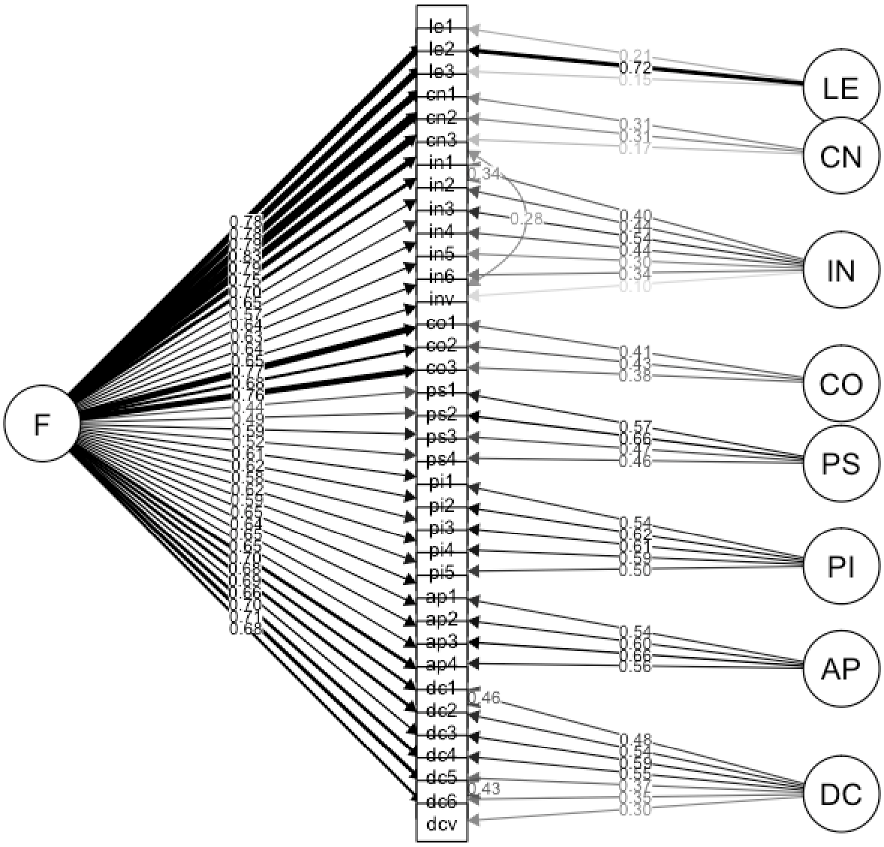
In this part we analyse the results for teachers. Figure 5 and Figure 6 present a graphical representation of the 8D eight-dimensional (M2) and bifactor models (M3) for teachers. The graphs are analogues to those presented in Figure 3 and 4, but instead of four substantive dimensions we now have eight dimensions.

Figure 5. **Eight-dimensional model for teachers**



Source: Authors.

Figure 6. **Nine-dimensional bifactor model for teachers**



Source: Authors.

In Figure 5, the factor loadings for all dimensions are high (in all cases higher than 0.7 and in most cases higher than 0.8). This confirms that the multidimensional solution fits to data well. Factor loadings are correlated but not as high as in the students' version of the tool. In most cases those correlations are between 0.5 and 0.6. This means that, in most cases, only around 25 to 36% (7) of variation in one factor could be explained by another factor (one exception is high correlation between CO and LE). This is reflected in bifactor model representation where all items load strongly on the general factor but also relatively strongly on other dimensions (Figure 6). This structure of factor loadings indicates that the digital capacity of schools measured by teachers is multidimensional in nature and could not be explained well by one single dimension.

Table 7 shows the fit measures of the unidimensional model (M1, baseline model), multidimensional model (M2) and the bifactor model (M3) as well as the

(7) $0.5^2 = 0.25$ and $0.6^2 = 0.36$.

refined multidimensional and bifactor models (models with residual correlations: dc1 and dc2; dc5 and dc6; in1 and in2; cn3 and inv).

Table 7. **Fit measures for different model specifications: teachers' sample**

Model/Fit	CFI	TLI	SRMR	RMSEA	BIC	AIC
M1 – Unidimensional	0.790	0.756	0.101	0.149	420 978	420 671
M2 – 8-D	0.932	0.925	0.040	0.060	685 777	685 071
M3 – Bifactor	0.911	0.899	0.070	0.070	691 529	690 766
M2´- 8-D refined	0.950	0.944	0.037	0.052	681 112	680 378
M3´- Bifactor refined	0.926	0.916	0.070	0.064	687 507	686 716

Source: Authors.

The results are clear, indicating that the best fit of all the fit measures has the eight-dimensional model which suggests multidimensionality. For a more in-depth investigation, we now explore the bifactor indices and reliability measures presented in the next section.

4.2.2. Reliability analysis

Table 8 presents a reliability analysis based on the bifactor indices. The ECV for the general factor is 0.66. This is a relatively low value and, together with PUC = 0.88, indicates substantial dimensionality. ECV values for specific factors indicate that each specific factor accounts for 11% to 53% of the common variance in the responses related to each factor. The Omega coefficient shows a high reliability of the general factor together with high reliabilities of specific factors. OmegaH and Omega ratios indicate that a substantial portion of the variation in responses could be attributed to specific dimensions. The last four specific factors (Pedaogy: supports; Pedagogy: implementation; Assessment practices; Student digital competence) explain a great amount of variation in responses above and beyond the general factor. This picture is confirmed by a relatively high value of H and FD indices for those specific factors.

Table 8. **Reliability analysis based on the bifactor model: teachers' sample**

Model/Fit	ECV	Omega	OmegaH	Omega Ratio	H	FD
General factor – digital capacity	0.664	0.981	0.920	0.937	0.970	0.971
Leadership	0.232	0.915	0.160	0.175	0.497	1.000
Collaboration and networking	0.110	0.877	0.092	0.105	0.205	0.621
Infrastructure and equipment	0.273	0.898	0.213	0.238	0.586	0.834

Model/Fit	ECV	Omega	OmegaH	Omega Ratio	H	FD
Continuing development	0.236	0.882	0.209	0.236	0.379	0.753
Pedagogy: supports	0.529	0.835	0.439	0.526	0.642	0.848
Pedagogy: implementation	0.470	0.919	0.431	0.469	0.711	0.897
Assessment practices	0.451	0.928	0.417	0.450	0.682	0.904
Student digital competence	0.323	0.942	0.292	0.310	0.690	0.892

Source: Authors.

The analysis based on observed scores indicates that the subscales of the SELFIE VET tool based on the teachers' questionnaire are highly reliable (Alpha between 0.81 and 0.93) and bring additional information to justify using sub-scores of the scales (Table 9). According to PRMSE, subscales add between 14% and 33% of additional information to what a single total score could provide.

Table 9. **Reliability analysis based on observed scores: teachers' sample**

Factor/measure	Alpha	PRMSEs	PRMSEx	Added Value
General factor – digital capacity	0.969	NA	NA	NA
Leadership	0.870	0.870	0.684	Yes (21%)
Collaboration and networking	0.867	0.867	0.744	Yes (14%)
Infrastructure and equipment	0.890	0.890	0.690	Yes (22%)
Continuing development	0.873	0.873	0.634	Yes (27%)
Pedagogy: supports	0.811	0.811	0.559	Yes (31%)
Pedagogy: implementation	0.912	0.912	0.669	Yes (27%)
Assessment practices	0.928	0.928	0.625	Yes (33%)
Student digital competence	0.932	0.932	0.703	Yes (25%)

Source: Authors.

4.2.3. Measurement invariance

Table 10 indicates the measurement invariance analysis based on the MI1 to MI5 models. The results are very similar to the ones based on the student questionnaire, with only a marginal change of fit compared to the configural model. The results clearly show that metric invariance holds while scalar invariance does not.

ΔCFI is 0.013 while ΔRMSEA is 0.009. We explored modification indices and found that non-invariance is mainly a problem of six items: co2, in6, pi3, ps3, ap2 and ap4. After releasing those items to be free, the partial scalar model fits reasonably well compared to the metric model, with ΔCFI being 0.010 and ΔRMSEA being 0.003. The residual invariance model (even in partial settings with no constraints on six items) fits substantially less well than the scalar (and metric) model.

Table 10. **Measurement invariance analysis: teachers' sample**

Model/Fit	CFI	RMSEA	ΔCFI	ΔRMSEA
MI1 – Configural	0.935	0.058	NA	NA
MI2 – Weak – Metric (loadings)	0.932	0.058	0.002	0.000
MI3 – Scalar (loadings and intercepts)	0.919	0.067	0.013	0.009
MI4 – Partial Scalar	0.922	0.061	0.010	0.003
MI5 – Residuals (partial)	0.904	0.066	0.018	0.005

Source: Authors.

4.2.4. Validity

In order to check the external validity of the SELFIE tool for VET schools, Table 11 presents the OLS models considering the independent variables: teachers' age (age), years of experience (experience), percentage of teaching time spent using digital technologies in the classroom (digital) and teachers' general satisfaction with the SELFIE tool (recommend) and their relationship with each one of the SELFIE area scores.

Table 11. **OLS regression models: standardised beta coefficients**

Areas/ subscales	(1) LE	(2) CN	(3) IE	(4) CO	(5) PS	(6) PI	(7) AP	(8) DC
Age	-0.011	-0.040**	-0.004	-0.007	-0.078**	-0.064**	-0.044**	-0.015
	(-1.01)	(-3.79)	(-0.37)	(-0.63)	(-8.15)	(-6.42)	(-4.34)	(-1.45)
Experience	-0.002	0.000	0.003	-0.006	-0.008	-0.002	-0.003	0.005
	(-0.16)	(0.05)	(0.28)	(-0.63)	(-0.88)	(-0.16)	(-0.35)	(0.47)
Digital	0.190**	0.184**	0.215**	0.179**	0.428**	0.395**	0.341**	0.238**
	(20.01)	(19.06)	(21.52)	(18.93)	(49.27)	(43.87)	(37.47)	(24.66)
Recommend	0.198**	0.195**	0.160**	0.201**	0.181**	0.177**	0.167**	0.189**
	(21.39)	(20.74)	(16.46)	(21.70)	(21.37)	(20.04)	(18.77)	(20.05)
N	10 528	9 712	9 685	10 724	10 373	9 921	9 806	9 986
adj. R-sq	0.1576	0.2011	0.1431	0.1478	0.3059	0.2834	0.2789	0.1679

NB: *t* statistics in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Authors.

The results indicate that there is a positive association between the teachers' time spent using digital technologies in the classroom and the scores of all areas of the SELFIE tool. The same is true for the teachers' recommendation of the SELFIE tool. The strongest association is found for the variable measuring the percentage of time spent using digital technologies in the classroom, in the areas Pedagogy: supports and resources, Pedagogy: implementation in the classroom and Assessment practices. However, the results also show that teachers' age is negatively associated with scores in the areas Collaboration and networking, Pedagogy: supports and resources, Pedagogy: implementation in the classroom and Assessment practices. The strongest association is found in the area Pedagogy: supports and resources.

CHAPTER 5.
General discussion

In this paper, we analysed a large, newly collected and unexplored cross-country data set from the SELFIE self-reflection tool to support the digital capacity development of schools. It has focused on the VET schools included, which play a crucial role in upskilling and the provision of digital competences to students. Overall, we found that the tool's questionnaires for VET are highly reliable, discriminatory and the validity of the construction of the questionnaires has been confirmed using psychometric methods. Our results have verified the scientific soundness and psychometric validity of the SELFIE tool for the VET sector, which underlines the usefulness of the tool for measuring the digital readiness of VET schools. The tool contributes to evidence-based decision-making in education, both at the school and, in aggregated form, also at regional or national level. The scientifically validated properties of the tool distinguish it from some other tools in the area, and the European approach provides additional value added by providing a common language between schools when discussing the development of digital education in classes.

Our analysis suggests that the shortened version of the tool for students could be successfully used as an indicator of overall digital capacity. Subscales of the shortened version of the SELFIE tool are too imprecise to indicate information on specific areas but all together very reliably describe the overall digital capacity of a school being evaluated by its students. The SELFIE tool for the group of teachers, that is in the regular form of the tool, is more complex and allows for the reliable (self-) assessment of many aspects of digital capacity, but also provides a reliable composite score of digital capacity. Measurement invariance analysis reveals that direct comparability between countries holds for both versions of the tool.

The small validity study carried out, while not presenting definitive arguments for validity, shows that the general pattern is in line with expectations. The small scale of the validity study is its main limitation; future studies should investigate the validity in a more direct and detailed way. The direct relationship between digital capacities measured by SELFIE and digital abilities (measured by knowledge tests) would be of most interest. In addition, studies that could relate SELFIE usage to school effectiveness in teaching digital technologies could bring more validity arguments that the SELFIE tool lacks.

While a complete set of information and studies on the relevance of the tool is not yet available, its good alignment with the theoretical background, the excellent reliability and psychometric properties, and the favourable reception of the tool by schools and education authorities encourage the further development of the tool and extending its reach to more schools and user groups. For this reason, the European Commission has also been developing a further extension of the VET tool by adapting it to the work-based learning (WBL) sector (Hippe, 2020). This adaptation includes the addition of in-company trainers as a fourth respondent group, enabling schools also to involve training companies in the SELFIE exercise and improve their collaboration with them. The SELFIE WBL pilot phase was completed in December 2020, involving around 35 000 participants in 150 VET schools and 300 companies from nine European countries from within and outside the EU (ET 2020 Working Group on VET). The same scientific criteria should be applied to ensure that this new addition to the SELFIE tool adheres to the same standards that we have been able to verify in this paper.

Following the literature on the subject, we expect that a tool like SELFIE could generate an important push for education institutions to understand their digital capacities and educational practices and should therefore be further developed, improved and tested. We believe that the proposed tool is well suited for self-evaluation purposes and its good psychometric properties let us show that the tool could also be used for research purposes: assessing the digital capacities of schools, teachers, school leaders and students in a non-invasive, efficient way.

Acronyms

AIC	Akaike information criterion
AP	assessment practices
BIC	Bayesian information criterion
CFA	confirmatory factor analysis
CFI	comparative fit index
CN	collaboration and networking
CPD	continuing professional development
DC	student digital competence
DigCompOrg	European Framework for Digitally Competent Educational Organisations
ECV	explained common variance
FD	factor determinacies
ISCED	international standard classification of education
LE	leadership
MG-CFA	multi-group confirmatory factor analysis
MI	measurement invariance
PI	pedagogy: implementation in the classroom
PRMSE	proportional reduction in mean squared error
PS	pedagogy: support and resources
PSNTE	post-secondary non-tertiary education
PUC	percent of uncontaminated correlations
RMSEA	root mean square error of approximation
SELFIE	Self-reflection on effective learning by fostering the use of innovative educational technologies
SL	school leavers
SRMR	standardised root mean square residual
TLI	Tucker-Lewis index
VET	vocational education and training

References

[URLs accessed 21.9.2022]

Balaban, I.; Redjep, N.B. and Calopa, M.K. (2018). The analysis of digital maturity of schools in Croatia. *International Journal of Emerging Technologies in Learning*, Vol. 13, No 6, pp. 4-15.

Beauducel, A. and Herzberg, P.Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, Vol 13, No 2, pp. 186-203.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, Vol. 74, No 1, p. 137.

Braeken, J. and Blömeke, S. (2016). Comparing future teachers' beliefs across countries: Approximate measurement invariance with Bayesian elastic constraints for local item dependence and differential item functioning. *Assessment and Evaluation in Higher Education*, Vol. 41, No 5, pp. 733-749.

Broek, S. and Buiskool, B.J. (2020). *Adapting the SELFIE tool for work-based learning systems in vocational education and training: a feasibility study*. Luxembourg: Publications Office. JRC technical report. https://data.europa.eu/doi/10.2760/934724

Brown, T.A. (2015). *Confirmatory factor analysis for applied research: 2nd edition.* Guilford publications.

Byrne, B.M.; Shavelson, R.J. and Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological bulletin*, Vol. 105, No 3, p. 456.

Castaño-Muñoz, J. et al. (2018). Within-school differences in the views on the use of digital technologies. In: Europe evidence from the SELFIE tool. *EDULEARN18 proceedings, IATED Academy*, pp. 10417-10426.

Cedefop. (2017). *Key capacity in vocational education and training 2018/S 010-017323: contract award notice.* https://www.cedefop.europa.eu/files/award_notice_key_competences.pdf

Chapman, C. and Sammons, P. (2013). *School self-evaluation for school improvement: what works and why?* Reading: CfBT Education Trust. https://www.educationdevelopmenttrust.com/our-research-and-insights/research/school-self-evaluation-for-school-improvement

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling*, Vol. 14, No 3, pp. 464-504.

Claro, M. et al. (2018). Teaching in a digital environment (TIDE): defining and measuring teachers' capacity to develop students' digital information and communication skills. *Computers and Education*, Vol. 121, pp. 162-174.

Costa, P.; Castaño-Muñoz, J. and Kampylis, P. (2021). Capturing schools' digital capacity: psychometric analyses of the SELFIE self-reflection tool. *Computers and Education*, Vol. 162, 104080. https://www.sciencedirect.com/science/article/pii/S0360131520302785

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, Vol. 16, No 3, pp. 297-334.

Davidov, E. et al. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, Vol. 40, pp. 55-75. https://doi.org/10.1146/annurev-soc-071913-043137

DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural equation modeling*, Vol. 9, No 3, pp. 327-346.

ET 2020 Working Group on VET (2020). SELFIE greetings by Ralph Hippe, 31 December 2020, Yammer group [non-public].

European Commission (2016). *Communication from the Commission […]: a new skills agenda for Europe: working together to strengthen human capital, employability and competitiveness.* COM(2016) 381 final. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52016DC0381andrid=3

European Commission (2017). *Communication from the Commission […]: school development and excellent teaching for a great start in life*. COM(2017) 248 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2017%3A248%3AFIN

European Commission (2018). *Digital Education Action Plan.* https://ec.europa.eu/education/education-in-the-eu/digital-education-action-plan_en

European Commission (2020a). *Innovation and digitalisation: a report of the ET 2020 Working Group on Vocational Education and Training*. https://europa.eu/!dF68qC

European Commission (2020b). *Digital Education Action Plan 2021-2027*. https://ec.europa.eu/education/sites/default/files/document-library-docs/deap-communication-sept2020_en.pdf

European Commission (2020c). Council Recommendation of 24 November 2020 on vocational education and training (VET) for sustainable competitiveness, social fairness and resilience. *Official Journal of the European Union, C 417, 2.12.2020, p. 1-16.* https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32020H1202(01)&from=EN

European Commission (2020d). *SELFIE database. https://education.ec.europa.eu/selfie*

European Commission (2021a). *SELFIE.* https://ec.europa.eu/education/schools-go-digital_en

European Commission (2021b). *SELFIE: discover your school's digital potential.* https://schools-go-digital.jrc.ec.europa.eu/

European Training Foundation (2020). *Key indicators on education, skills and employment 2020.* https://www.etf.europa.eu/sites/default/files/2020-11/kiese_2020_0.pdf

Eurostat (2020). *Pupils enrolled in upper secondary education by programme orientation, sex, type of institution and intensity of participation.* [Data set identifier: educ_uoe_enrs4]. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=educ_uoe_enrs04

Gignac, G.E. and Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence*, Vol. 62, pp. 138-147.

Gliem, J.A. and Gliem, R.R. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales*. Midwest research-to-practice conference in adult, continuing, and community education.

Gorsuch, R. L. and Venable, G. D. (1983). Development of an 'age universal' IE scale. *Journal for the scientific study of religion*, Vol. 22, No 2, pp. 181-187.

Gustafsson, J. E. and Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate behavioral research*, Vol. 28, No 4, pp. 407-434.

Haberman, S. J. (2008). When can sub-scores have value*? Journal of Educational and Behavioral Statistics*, Vol. 33, No 2, pp. 204-229.

Haberman, S.; Sinharay, S. and Puhan, G. (2008). Reporting sub-scores for institutions. *British Journal of Mathematical and Statistical Psychology*, Vol. 62, No 1, pp. 79-95.

Hatlevik, O. E. and Arnseth, H. C. (2012). ICT, teaching and leadership: How do teachers experience the importance of ICT-supportive school leaders? *Nordic Journal of Digital Literacy*, Vol. 7, No 1, pp. 55–69.

Hippe, R. (2020). *Make the most of digital technologies with the SELFIE tool for VET schools and work-based learning.* https://ec.europa.eu/education/schools-go-digital/selfie_news/make-the-most-of-digital-technologies-with-the-selfie-tool-for-vet-schools-and-work-based-learning_en

Hofman, R.H.; Dijkstra, N.J. and Hofman, W.H.A. (2009). School self-evaluation and student achievement. *School Effectiveness and School Improvement*, Vol. 20, No 1, pp. 47-68, https://doi.org/10.1080/09243450802664115

Horn, J.L. and McArdle, J.J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental aging research*, Vol. 18, No 3, pp. 117-144.

Hu, L.T. and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling*, Vol. 6, No 1, pp. 1-55.

Ilomäki, L. and Lakkala, M. (2018). Digital technology and practices for school improvement: innovative digital school model. *Research and Practice in Technology Enhanced Learning*, Vol. 13, No 1, 25. https://doi.org/10.1186/s41039-018-0094-8

Jansseon, S. et al. (2018). *Betriebe und Arbeitswelt 4.0: mit Investitionen in die Digitalisierung steigt auch die Weiterbildung.* IAB-Kurzbericht, No 26. https://doku.iab.de/kurzber/2018/kb2618.pdf

JISC (2021). *JISC digital insights: investigating student and staff expectations of the digital environment.* https://digitalstudent.jiscinvolve.org/wp/digital-learner-stories/

Johnson, D.R. and Creech, J.C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, Vol. 48, No 3, pp. 398-407.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika,* Vol. 36, No 4, pp. 409-426.

Kampylis, P. et al. (2016). Supporting schools to go digital: from a conceptual model towards the design of a self-assessment tool for digital-age learning*. 9th International Conference of Education, Research and Innovation*, pp. 816–825, Sevilla: IATED Academy.

Kampylis, P.; Punie, Y. and Devine, J. (2015). *Promoting effective digital-age learning: a European framework for digitally competent educational organisations.* Luxembourg: Publications Office. JRC technical report. https://data.europa.eu/doi/10.2791/54070

Kline, T.J. (2005). *Psychological testing: a practical approach to design and evaluation.* Sage Publications.

Krumsvik, R. J. et al. (2016). Upper secondary school teachers' digital competence: Analysed by demographic, personal and professional characteristics. *Nordic Journal of Digital Literacy*, Vol. 11, No 3, pp. 143-164.

Marsh, H.W.; Hau, K.T. and Wen, Z. (2004). In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's findings. *Structural Equation Modeling*, Vol. 11, No 3, pp. 320-41.

McDonald, R. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Erlbaum.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, Vol. 58, No 4, pp. 525-543.

Muthén, B. and Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, Vol. 38, No 2, pp. 171-189.

Pokropek, A.; Davidov, E. and Schmidt, P. (2019). A Monte Carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance*. Structural Equation Modeling*, Vol. 26, No 5, pp. 724-744.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, Vol. 21, No 2, pp. 173–184.

Reise, S. P.; Bonifay, W. E., and Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, Vol. 95, No 2, pp. 129–140.

Reise, S.P.; Bonifay, W. and Haviland, M.G. (2018). Bifactor Modeling and the Evaluation of Scale Scores. In: Irwing, P.; Booth, T. and Hughes, D.J. *The Wiley handbook of psychometric testing: a multidisciplinary reference on survey, scale and test development*, pp. 675-707, Chichester: Wiley Blackwell.

Reise, S.P.; Moore, T.M. and Haviland, M.G. (2010). Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. *Journal of personality assessment*, Vol. 92, No 6, pp. 544-559.

Reise, S. P. et al. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: a bifactor perspective. *Educational and Psychological Measurement*, Vol. 73, No 1, pp. 5-26.

Rodriguez, A.; Reise, S.P. and Haviland, M.G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, Vol. 98, No 3, pp. 223-237. http://dx.doi.org/10.1080/00223891.2015.1089249

Rodriguez, A.; Reise, S.P. and Haviland, M.G. (2016b). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychological Methods*, Vol. 21, No 2, pp. 137-150.

Rosseel, Y. (2012). lavaan: an R package for structural equation modelling. *Journal of Statistical Software*, Vol. 48, No 2, pp. 1–36.

Schildkamp, K.; Visscher, A. and Luyten, H. (2008). The effects of the use of a school self-evaluation tool. *School Effectiveness and School Improvement*, Vol. 20, No 1, pp. 69-88. https://dx.doi.org/10.1080/09243450802605506

Tanhua-Piiroinen, E. and Viteli, J. (2017). Opeka and Ropeka, the self-assessing services for teachers and principals. In: Lavoue, E. et al. (eds.). *Data driven approaches in digital education: 12th European conference on technology enhanced learning, EC-TEL 2017,* pp. 602–605.

Ten Berge, J.M. and Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, Vol. 69, No 4, pp. 613-625.

Tomczyk, Ł. (2019). What do teachers know about digital safety? *Computers in the Schools*, Vol. 36, No 3, pp. 167-187.

Vandenberg, R. J. and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, Vol. 3, No 1, pp. 4-70.

Wu, H. and Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, Vol. 81, No 4, pp. 1014-1045.