

EUROPEAN TEACHERS IN THE PISA 2015 SAMPLE: A MULTILEVEL MODELLING APPROACH TO ICT INFRASTRUCTURE AND TEACHER TRAINING

PROFESORADO EUROPEO EN PISA 2015: UN ENFOQUE DE MODELACIÓN MULTI NIVEL PARA LA INFRAESTRUCTURA DE LAS TIC Y LA FORMACIÓN DOCENTE

Leo Van Waveren
Christine Sälzer

ABSTRACT

Drawing on the background questionnaires in PISA 2015 for teachers and schools, this paper aims to investigate the interdependencies and influences of teacher training and ICT infrastructure on their ability and willingness to employ modern media in a classroom setting. A multi-level approach is used to account for the data structure.

Our analyses show that students' background and ICT interests provide limited predictive power concerning their test performance in the PISA 2015 sample from five European countries. Moreover, teachers and principals view shortages in staff and educational material quite differently in the background questionnaires. By these findings, new research approaches may be better suited to understand the influences teachers and school characteristics have on the learning outcomes of students in large-scale assessments. This also highlights potential problems when an analysis draws on joining different sources for a combined description.

Key words: PISA 2015, ICT, teacher education, multilevel modelling

RESUMEN

Basándose en las encuestas de PISA 2015 para los maestros y colegios, este artículo pretende investigar las interdependencias e influencias de la formación de los maestros y la infraestructura de TIC en su habilidad y voluntad para emplear los medios modernos en el entorno escolar. Se ha utilizado un enfoque de niveles múltiples para justificar la estructura de datos.

Nuestro análisis demuestra que el entorno de los estudiantes y acciones de TIC ofrecen un poder de predicción limitado en relación con el rendimiento de las pruebas de PISA 2015 de cinco países europeos. Además, los maestros y los directores ven escasez de personal y material educativo bastante diferente en las encuestas del entorno. Sobre la base de esas conclusiones, se puede adaptar unos enfoques nuevos de investigación para entender mejor las influencias que tienen los maestros y las características escolares en resultados del aprendizaje de los estudiantes en las evaluaciones a gran escala. Esto también resalta los problemas potenciales cuando un análisis recurre reuniendo fuentes diferentes para una descripción combinada.

Palabras clave: PISA 2015, TICs, Formación docente, modelo multinivel

Fecha de recepción: 17 de septiembre de 2019

Fecha de aceptación: 11 de noviembre de 2019

INTRODUCTION

The Programme for International Student Assessment (PISA) has been conducted since 2000 as a three-year cycle by the Organisation for Economic Co-operation and Development, OECD. The main focus of the PISA study is to compare the learning outcomes of students across countries at the age of 15 in the three domains of mathematics, science and reading, with science being the major domain of the PISA 2015 cycle. At the age of 15, students approach the end of compulsory schooling in many countries. Their competence level at this point therefore yields a meaningful prognosis for their further development as active and responsible citizens (Sälzer & Prenzel, 2014). While student performance remains at the heart of the assessment, background questionnaires for students, school principals, parents and teachers are used to shed light on the interdependencies of context factors with the learning outcomes. Questionnaires for students and schools are mandatory in all participating countries, while those for parents and teachers are optional. More recently, the effort put into the background questionnaire reflects the increased attribution to these factors in explaining learning outcomes based on influencing factors like the social background (Watermann, Maaz, Bayer, & Roczen, 2017) and the engagement with ICT (Goldhammer, Gniewosz, & Zylka, 2017). Helmke (2010) remarked that the teacher is a focal point for the instructional quality of lessons in schools. The initial training and professional development of teachers will therefore play a vital role on equipping students with the necessary tools to meet these challenges. And recent international approaches like the DigiCompEdu framework¹ and the TPACK-Model (Koehler, Mishra, Kereluik, Shin, & Graham, 2014) are a visible acknowledgement of the role teachers play in preparing today's and tomorrow's society for changes by the global megatrend of digitalisation (OECD, 2018).

The international data sample of the PISA 2015 cycle includes answers to the teacher questionnaire from a total of 19 countries worldwide. Of these, five European countries chose to administer the teacher questionnaire. This paper takes a closer look at these five European countries (Czech Republic, Germany, Italy, Portugal and Spain) in order to investigate approaches to adequately model teacher variables and evaluate, whether the view of teachers and principals coincide on shared topics from the teacher and school questionnaire. As the global megatrend of digitalization will continue to pose a challenge for society and will undoubtedly influence teaching throughout all educational stages, we take a look at the occurrence of ICT content in the teacher's professional development.

1. THEORETICAL BACKGROUND

From previous PISA cycles, the constructs of reading, mathematical and scientific literacy are well documented (OECD, 2016a). While the construct of "Information and communications technology (ICT)" or "Computer and Information Literacy - CIL" competence was the central topic in the ICILS study of 2013 (the results from the recent cycle of 2018 are to be published at the end of this year), it has also been a recurring theme in the background questionnaires of the PISA study since 2012. The focus here is set on investigating its relationship with the three cognitive domains of mathematics, science and reading. With regard to the role of ICT familiarity and use for student achievement, recent findings differ considerably. What becomes quite obvious is that knowing the characteristics of students participating in large-scale assessments alone does not predict their ICT literacy very well. The interplay of both student and family characteristics as well as school level characteristics seems to be key to investigate how ICT literacy can be predicted, understood and hence scaffolded by teachers and schools.

¹ <https://ec.europa.eu/jrc/en/digcompedu>

Youssef and Dahmani (2008) observe that despite the major role student characteristics play in their achievement, the role of their educational environment and those of their teachers' are not to be neglected as well. More than 10 years ago, they argued to acknowledge the influence ICT may have on the learning outcome. Steffens (2014) for example points out that the results in the PISA 2012 sample do not increase with an increased ICT use.

Furthermore, the ICILS study implies that ICT literacy differs strongly depending on the students' socioeconomic background. Strikingly, female students outperformed their male counterparts, who in turn usually overestimated their own ICT competence. Nearly a third of all German students in the sample did not overcome the lowest two competence levels (Bos et al., 2014), calling the phrase "digital natives" somewhat into question.

Gerick (2018, 161f) points out three major factors in literature to identify clusters of school focusing the approaches to teaching and learning ICT: "the importance of visions and strategies for schools regarding teaching and learning with ICT", "activities to develop the ICT skills of teachers" and "quantity (e.g. availability) and [...] the (perceived) quality of the available ICT infrastructure".

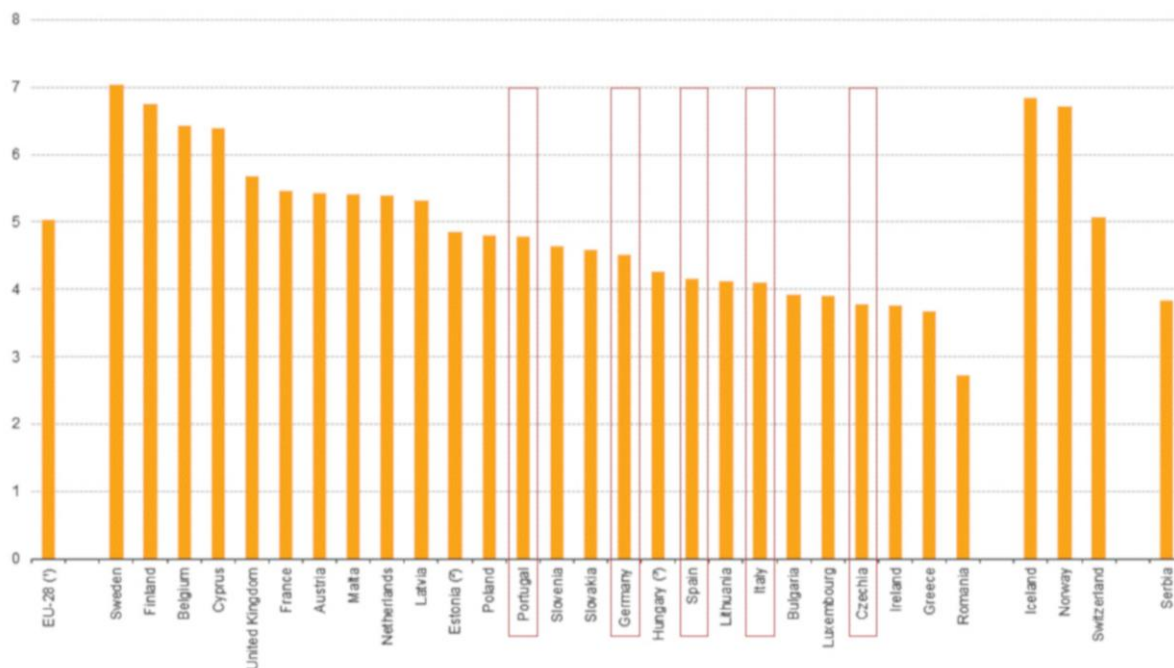
Taking both the student and the school level into account, Wößmann and Fuchs (2004) propose in a multivariate analysis that the school characteristics need to be taken into consideration in addition to student background variables when interpreting the test performance.

When data on different levels of measurement are available, analyses are at this point usually approached via multilevel modelling (H. Goldstein, 2003; Rabe-Hesketh, Skrondal, & Pickles, 2004; Raudenbush & Bryk, 2010). Based on the TIMSS study, Falck, Mang, and Wößmann (2018) argue that mostly students with a high socioeconomic status (SES) profit from ICT learning scenarios and that effects on the learning outcome depend on the kind of intended learning outcome and applied teaching method. Similarly, Rosén and Gustafsson (2016) use data from the 10-Year Trend Study and the PIRLS 2001 and 2006 studies-to gauge the effect of home computer use on reading achievement in grade 4. While controlling for SES and gender, they stipulate a negative causal effect from ICT use at home on reading competence. However, contrary findings are reported by Araya et al. (2015). Their findings from primary schools in Chile show effect sizes around .3 when ICT assisted learning scenarios were employed in low socioeconomic status schools.

Taking a perspective at the system level, Hanushek, Ruhose, and Wößmann (2016) point out a close connection between a country's gross domestic product (GDP) and its performance in large-scale assessments. Taking the public expenditure on education (excluding early childhood educational development) as a proxy, depicts the spending of European countries in 2015 by Eurostat².

² https://ec.europa.eu/eurostat/web/products-datasets/-/sdg_08_10

Figure 1: Educational expenditure in the European Union in 2015 with an added highlighting of the countries under consideration in this article.



Source: PISA, 2015

Zhang and Luman (2016) find in their three-level hierarchical linear modelling approach for the PISA studies from 2000 to 2012 a negative correlation between student achievement in mathematics and science with different types of ICT use. This effect remains stable when controlling for national GDP, school type, school ICT investment and SES. A more differentiated interpretation provides Güzeller and Akin's study (2014) who used the 2006 PISA data to investigate to which extent ICT variables could be used to infer mathematical achievement at an international level. In their findings, the use of "internet/entertainment" and "program/software" were negatively associated, while students' confidence in "internet tasks" and in "high level ICT tasks" were positively correlated. The authors interpreted these findings as evidence for an incomplete integration of ICT into classroom and school environment. Broadening the perspective with regard to different school types, Dinis da Costa and Araújo (2016, p. 61) compared general and VET schools in terms of the digital reading performance of their students in PISA 2012 and infer that a considerable part of variance in the performance of students stems from school-to-school differences. This scaffolds Güzeller and Akin's interpretation stating that ICT needs to be actively integrated in classrooms and learning processes and not just being made available to students. However, when comparing the importance of students' family and school environment in order to explain their ICT literacy, it becomes also obvious that a student's family background has a stronger relationship with ICT literacy than other factors. For example, Yalçın (2018) conducted a two-level latent class analysis on the complete PISA 2015 sample for reading performance. She argues that, from the variables included in the analysis, the student's ICT resources at home are the most influential variable concerning reading skills at the country level. In more detail, Gómez-Fernández and Mediavilla (2018) find in their analysis for Spain's PISA 2015 data a negative correlation between the age when students first interacted with ICT and their outcome in the three main competence domains in PISA. Furthermore, they report that interest in ICT is fostered by engagement in the student's leisure time, but the use in schoolwork and general use has negative effects on the learning process.

Also drawing on the PISA 2015 sample, Juhaňák, Zounek, Záleská, Bárta, and Vlčková (2019, 56f) investigated the relationship between a student's school performance and their ICT use in the Czech Republic. They conclude that ICT availability at home had a significant interaction with school achievement, while the broad availability of an at least basic ICT infrastructure in Czech schools prevented a direct effect on that level (similar findings are presented in Gerick (2018)). To conclude, findings aiming at explaining how student ICT literacy develops depending on their family and school-related environment differ in their level of detail and their conclusions. It is therefore a complex challenge for teachers to determine a valuable strategy for implementing ICT literacy in their teaching.

As already pointed out in the introduction, teachers stand at the hearth of educational processes in the classroom (Helmke, 2010). The close link of job satisfaction and job performance with a congruence of one's interests and the occupational demands is well established (Holland, 1997; Jaensch, Hirschi, & Spurk, 2016; Spokane, Meir, & Catalano, 2000), factors contributing to the preparation of teachers for new(er) challenges in the classroom (e.g., use of ICT during initial education and/or professional development) should be also accounted for.

2. EXPLORATORY ANALYSIS

In order to disentangle the complex interplay of factors contributing to student ICT development, we begin our data analyses with two exploratory steps. We intend to initially a) find supporting evidence to represent the data sample by a sufficiently fitting but parsimonious model, and b) investigate the relationship between teacher and principal ratings as an indication to which extent the different background questionnaires form a comprehensive and conclusive view for school-characteristics.

As the questions identified as relevant to the topic of ICT in the PISA teacher questionnaire have been asked only to a subsample of the teachers, we decided to investigate first, whether a multi-level approach is needed when modelling selected teacher variables for the PISA 2015 sample.

To this end, we limited our first analysis to variables from the teacher questionnaire, which were concerned only with the teachers' point of view. We decided to account for gender in our model, since the ICILS study (Bos et al., 2014) revealed gender differences in the ICT-literacy at the student level.

Our second analysis was aimed to include items from the school questionnaire as well, focussing on the shortages of staff and educational resources for a corroboration on a shared metric between both levels of the principal (school level) and the teachers (teacher level). In addition, the emphasis of professional development from the principal's point of view was used in this step as a stepping stone for the topic of ICT content in teacher's advanced training in the second part of this paper.

2.1. METHODOLOGY

The multi-level modelling was done in Mplus 8.3 (Muthén, Muthén, Asparouhov, & Nguyen, 2018). Modelling factors on different levels accounts for the fact that the observed individuals (level 1) may share more traits with those belonging to the same school and country (levels 2 and 3) respectively (Asparouhov & Muthén, 2008) than with other individuals.

As outlined above, gender was represented in the model (TC001Q01NA). Years of working as a teacher (TC007Q02NA)³ was included as an independent variables in the initial model, to account for the possibility of differences in the "teacher generations".⁴

Furthermore, our first model used two WL-estimates for satisfaction with "the current job environment (SATJOB)" and "teaching profession (SATTEACH)" in addition to the two for the

³ As age (TC002Q01NA) shares a significant correlation above .93, we opted in favour of the teaching experience.

⁴ We decided against a separation into different age groups in accordance with the reasoning of Streiner (2002).

teacher’s view on “[e]ducational material shortage (TCEDUSHORT)” and “[s]taff shortage (TCSTAFFSHORT)” to reflect the school characteristics from their perspective as these indicators to our research interest were available for the full sample.⁵

In the second model discussed here, we additionally included therefore three WLEs from the school-level questionnaire to reflect “Professional development (LEADPD)”, “Shortage of educational material (EDUSHORT)” and educational staff (STAFFSHORT), which could not be accounted for in the lower-level modelling approaches to ascertain comparability of nested models. These newly added variables are used as an approximation of the school characteristics, similar to those discussed by Wößmann and Fuchs (2004).

Kenny (2014) argues that χ^2 -statistics are usually insufficiently sensitive in their differentiation to judge a model’s adequacy, when the number of cases in the sample increases. Therefore, the information criteria AIC, BIC and SABIC were used to evaluate the model fit.

Burnham (2004) pointed out as a rule of thumb regarding the model fit judged by AIC (and similarly Raftery (1995) concerning the BIC). Following these guidelines, a maximum difference of 2 between two models would indicate that both may prove to be suitable descriptions. A difference below 7 would indicate less evidence for the appropriateness of the worse fitting model. If the difference passes a threshold of 10, the model can usually be excluded from further consideration. Furthermore, Burnham (2004) argues that the AIC should only be favoured above the sample size adjusted AIC_c when the person-to-parameter ratio exceeds 40.

When drawing on fit indices to evaluate model fit, Hu and Bentler (1999, p. 27) recommend to use a combination for the cut-off for TLI⁶ (> .95) and SRMR (.09) for model evaluation, while a RMSEA > .06 and a SRMR > .09 produced the least number of Type I and II errors.

2.2. SAMPLE

While this paper draws on the teachers and schools participating in PISA 2015 (overall 108,292 teachers in the sample⁷), the limitation to countries that employed the teacher-questionnaire (19), we decided to limit our analysis to Europe under the assumption of more or less comparable conditions. Therefore, a sum of 33,058 teachers from the remaining five countries (Czech Republic, Germany, Italy, Portugal and Spain) form the basis for the analysis presented here.

From these, 25,152 teachers answered the questions for Sex, Age and Gender. Their distribution on the five countries is given in Table 1. Although the sample includes about twice the number of female teachers, the mean values for age and teaching experience can be assumed to be nearly identical (47 years of age and 20 years of teaching experience each around a standard deviation of roughly 10 years).

Table 1: Distribution of teacher age and job experience between the countries in the sample

Country	Gender	Age Average	Average Years of Teaching Experience	# in Sample
Czech Republic	Female	46 ± 10	20 ± 11	3,818
	Male	46 ± 12	19 ± 11	1,864
Germany	Female	45 ± 11	17 ± 12	3,378
	Male	46 ± 11	17 ± 12	1,956

⁵ Due to the fact that nearly 98 % of the non-missing responses in the sample indicated that the teacher “[p]articipated in professional development in the last 12 months” (PROPDT20), no meaningful interdependencies with other variables could be investigated.

⁶ Kenny (2014) points out that the CFI and TLI are closely related and argues in favour of reporting the TLI.

⁷ <https://www.oecd.org/pisa/data/2015database/>

Spain	Female	45 ± 8	18 ± 9	2,231
	Male	46 ± 9	18 ± 10	1,565
Italy	Female	49 ± 9	21 ± 10	4,422
	Male	49 ± 10	20 ± 11	2,300
Portugal	Female	47 ± 8	22 ± 8	2,603
	Male	47 ± 8	21 ± 9	1,014
Total	Female	47 ± 10	20 ± 10	16,452
	Male	47 ± 10	19 ± 11	8,699

Source: Original material

2.3. RESULTS

The first goal was to establish, whether a multi-level model is statistically superior to a more parsimonious lower-level approach of representing the teacher answers in the questionnaire from PISA 2015. Perhaps unsurprisingly, the values of the information criteria favour a multi-level modelling approach to the teacher's characteristics (cf.

Table 2).

Table 2: Fit of modelling approaches by information criteria to teacher's perception of satisfaction and available resources in the PISA 2015 study.

Modelled on	AIC	BIC	SABIC
Teacher level only	314,959	315,105	315,048
Clustered by country	303,436	303,623	303,550
Clustered by country and school	298,262	298,449	298,376

Source: Original material

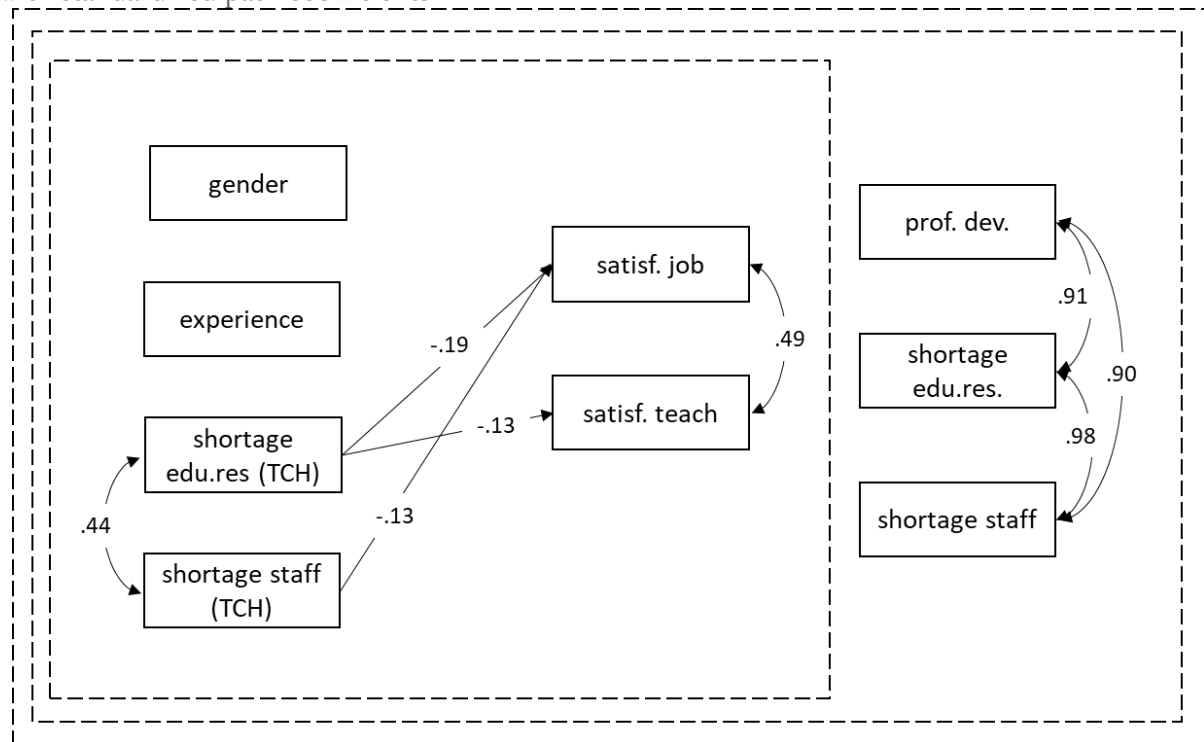
Based on these findings, the second model additionally included the answers from the school questionnaire for the topics of shortages of staff and educational materials together with the principals' rating on professional development.

The resulting model failed to fully satisfy the recommendations by Hu and Bentler (1999). The TLI (> .97), the RMSEA (< .01) and the SRMR on the within-level (< .02) were firmly inside the desirable range, but the between-level SRMR (< .15) indicated a model not fully adequate to represent the full complexity of the data sample.

Only significant path coefficients above a threshold of .10 are depicted in

Figure 2. Not included in the picture is a slightly negative influence by years of experience ($< .06$) on job and teaching satisfaction. Omitted from the figure is also a marginally higher teaching satisfaction for female teachers ($< .04$). Apart from these, the model showed no interdependencies towards gender or teaching experience. From a teacher's perspective, ratings for job and teaching shared a correlation around .4 and shortages for educational material and staff around .5. The latter two were negatively influenced by perceived shortness in staff and educational material ($< -.20$).

Figure 2: (simplified) Model for teacher's view on shortages and satisfaction in the PISA 2015 sample from five European countries, accounting for gender and years of teaching experience with standardized path coefficients.



Source: Original material

Strikingly, the ratings given in the school questionnaire regarding shortages on staff and educational resources showed no significant correlation with the teacher ratings concerning the same circumstances. Furthermore, the ratings for professional development also shared no correlation towards the other teacher ratings in this model. This implies that teachers and principals seem to perceive differently not only shortages but also the topic of professional development.

A first assumption to interpret this discrepancy here would have to be that teachers and principals in general would still share a common conception what constitutes as shortage of educational material and staff. One potential explanation could be that principals may experience less opportunities to teach in general and potentially have a more administrative point of view. Along these lines of thought it would be reasonable to assume, that they experience shortages of educational material in different terms as well.

These findings highlight the challenge of combining different sources into a comprehensive view, especially when at least parts of the constructs may not share an objective metric.

In sum, our exploratory analysis showed that:

- a multi-level approach was appropriate for modelling variables on teacher level
- perceived circumstances from school and teacher questionnaire were not significantly linked in the given ratings.
- Neither gender nor teaching experience had a meaningful influence on reported job or teacher satisfaction.

On a side note the similarities in

Figure 2 regarding the high correlations between the manifest WL-estimates on school-level to those of the three main PISA domains are to be mentioned. Without further evidence, we could not rule out the possibility, that these are a consequence of their original modelling and not necessarily reflect an equivalent close relationship at an individual level (Wu, 2007). For that reason, we abstained from attributing any interpretation to the close link between the three scores.

3. ICT AS PART OF TEACHER TRAINING IN THE PISA 2015 SAMPLE

Already pointed out was the finding from the ICILS study (Bos et al., 2014) that students do not necessarily develop ICT competence during their school career by themselves. It follows that teachers are required to possess ICT competences themselves, be it for the sake of sheer participation in today's society, or from a professional understanding of their vocation (e.g., in supporting the development process of their students in terms of becoming digitally literate citizens). Students need to be prepared to continuously develop new ICT skills, since ICT and digital environments constantly evolve. Consequently, teachers need to be equipped with a set of competences regarding ICT. This opens the question at which point in their professional career these competences are to be acquired.

To this end we extended the previous, low fitting model in Figure 2, which was intended to explore the needed hierarchy structure, by additionally accounting for the participation in ICT related programs as part of the teacher's professional development.

3.1. RESEARCH QUESTION

The PISA 2015 teacher questionnaire also addressed content of the professional development program the teachers took. We used the corresponding variables to investigate how teachers' age is related to their participation in ICT programs after the completion of the initial training and whether these coincided with an increased level of satisfaction regarding job and teaching.

H1₀: Teachers participate in ICT professional development independent of age.

H1₁: Younger teachers participate more often in ICT-oriented professional development programs.

Younger teachers also may have an increased chance that ICT was already part of their initial training, which is less likely for the initial training of older teachers. Although teachers of all ages are expected to participate in professional development, the topics chosen are likely to be based on their current teaching approaches, as teachers mainly select those programs from which they assume to have an immediate relevance to their teaching (Knowles, Holton, & Swanson, 2012). Therefore, it stands to reason that younger teachers also are more likely to opt for advanced training programs that are focussed on these topics, as potentially enhancing already present prior knowledge. It can also be surmised that younger teachers, as "Digital natives" (Palfrey & Gasser, 2008), feel already more familiar with ICT. We therefore investigated whether a younger age coincided with ICT-related trainings.

H2₀: Teachers who experienced ICT related content in their initial training and/or professional development report an equal level of satisfaction regarding job and teaching.

H2₁: Teachers with higher ratings on job and teaching satisfaction are more likely to engage in ICT trainings.

Taking the already mentioned findings of job affordances and satisfaction into consideration (Holland, 1997; Spokane et al., 2000), one should expect that teachers, who are more satisfied with their job (and teaching respectively) are more willing to meet the evolving challenges of ICT, in addition to those already present. We therefore concluded that teachers already comfortable with their current professional career should feel more compelled to engage in ICT training.

3.2. METHODOLOGY

The modelling approach pursued here was also done in Mplus 8.3 (Muthén et al., 2018) and followed the reasoning of the previous section, incorporating (with few adjustments) the same variables. The analysis here included the already mentioned variables for gender, job and teaching satisfaction on the individual level. In addition, age was included via the variable TC002Q01NA⁸. The ICT content in the different stages of teacher training was an extension to the previous model used in the exploratory analysis and was accounted for by the variable “(TC045Q05NB)”.

At the between level the “Number of available computers per student at modal grade (RATCMP1)”⁹ from the school questionnaire was used as an additional reference, since this objective measurement should be independent of divergent perceptions.

In sum, the modelling here can be viewed as an enhancement of the approach shown in Figure 2.

3.3. SAMPLE

From the data sample used in the exploratory analysis close to 60 % of the teachers (concerning country and gender, cf. Table 3) answered the corresponding item and a total number of 16,458 teachers answered the question regarding the occurrence of ICT content in their training and professional development.

Table 3: Fraction of teachers that answered the items concerning the occurrence of ICT in their initial training/professional development from the five European countries.

Country	Gender	# answers to ICT in prof. dev.	% of prev. sample
Czech Republic	Female	2,685	70 %
	Male	1,053	56 %
Germany	Female	2,395	71 %
	Male	1,114	57 %
Spain	Female	1,534	69 %
	Male	964	62 %
Italy	Female	3,026	68 %
	Male	1,437	62 %
Portugal	Female	1,551	60 %
	Male	699	69 %

Source: Original material

Although the gender distribution of teachers who answered the item regarding ICT content in their professional development is not fully equal between the countries, the overall ratio for the teachers varied between 62 and 66 %, making a systematic bias in the sampling unlikely.

3.4. RESULTS

The model was not able to provide evidence against the first Null-hypothesis, because the participation in ICT-related professional development programs did not show a significant correlation with age (or gender for that matter), similar to teaching experience in section 2.3. Interestingly, neither the rating of principals regarding shortages of educational resources nor their view on professional development had an influence on the teachers’ participation in ICT related programs. A negative but low correlation (< .20) between the “Number of available

⁸ In the second exploratory analysis the variable “total years of working as a teacher (TC007Q02NA) was used, but as both share a significant correlation above .93, we opted here in this approach for age as the arguably the more fitting indicator to the question without recalculating the previous findings.

⁹ The variable “Proportion of available computers that are connected to the Internet (RATCMP2)” had to be excluded from the analysis due variance limitations, as nearly all computers were connected to the internet in these five countries.

computers per student at modal grade” and the school rating on shortages of educational material strongly suggest that looking at available ICT equipment alone does not work as a reliable indicator for an overall lack of educational resources. However, it should be noted that on a between level a low correlation of .12 was observed concerning the number of available computers at school and the teacher’s participation in the ICT programs.

We take this as indication that a lack of (computer based) resources provides only a very limited incentive to engage in professional development programs focussed on ICT, and furthermore that the decision process seems to be strongly residing with the teacher and less so with the principal. The timely application of acquired knowledge is needed to let professional development programs have a meaningful impact and effectively change the approaches in the workplace environment in general (I. L. Goldstein & Ford, 2009; Grossman & Salas, 2011). Garet, Porter, Desimone, Birman, and Yoon (2001) showed that professional development for teachers heavily relies on phases of applying the training content to be effective. This is compounded by the findings of the meta-analysis by Blume, Ford, Baldwin, and Huang (2010), as they demonstrated that self-efficacy reduces with time, potentially making later attempts to put the training into effect less likely.

The model was also not able to provide conclusive evidence to reject the second Null-hypothesis, as the participation in ICT-related professional development programs was not significantly associated with the teaching satisfaction. Furthermore, the standardized coefficient for job satisfaction was below .06 and can also be only attributed to a very limited degree with being responsible for the decision to participate. Therefore, a graphical representation of the findings is omitted here, as the inferred conclusion is similar to the one drawn from Figure 2.

It is, however, to be noted that the distribution of participation in ICT related professional development programs in the five countries does not coincide in the assumed order that could be inferred from **¡Error! No se encuentra el origen de la referencia..** A summary in Table 4 shows the distribution between the countries and phases of teacher’s professional development.

Table 4: Distribution of ICT content in teacher training/professional development the five European countries by teacher rating.

	not existing	only part of IE	only part of PD	overlap IE and PD
Portugal	24%	30%	19%	27%
Czech Republic	29%	26%	25%	21%
Italy	31%	28%	24%	16%
Spain	26%	26%	25%	23%
Germany	39%	33%	14%	14%

Source: Original material

Perhaps most strikingly is the considerably lower amount of teachers in Germany not having experienced ICT related training in their professional career, although the German teachers are on average one to two years younger compared to the other four countries (cf. Table 1). It has to be mentioned however that nearly a third of the German teachers also stated that ICT was part of their initial training.

While the sum of the last two columns is just shy of 30 % in Germany, 40 % and more of the teachers of the other four countries stated that ICT was part of their professional development (potentially as overlap to their initial training). Looking at the data from PISA 2015 it has to be mentioned that despite the second highest expenditure on education from the countries taken into consideration here, Germany seems to put less emphasis on ICT in the teacher education than other European countries.

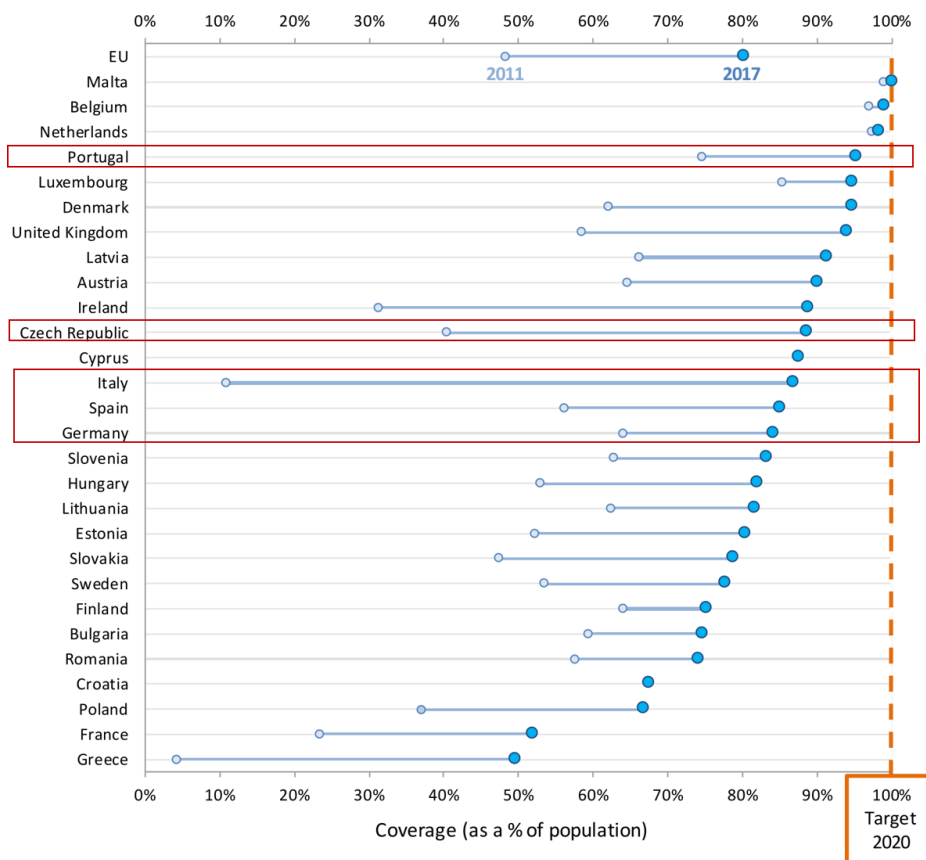
4. DISCUSSION AND FUTURE RESEARCH

As the perception of learning environment from a student’s view and school administration may very well differ, it was no surprise that the model for a combination of student and school ratings showed a misfit beyond the individual level (Figure 2).

While teachers tend to attribute emphasis within the curriculum at a higher rate compared to their students, the ratings overall show the same trend and can be taken as measurements of the same construct (Maier et al., 2015, 236p). From our analysis we tend to draw an opposite conclusion regarding the perceived shortness of staff and educational resources on both the teacher and the school level. Already mentioned in section 2.3, we observed no relevant similarities between the ratings given by the two sides.

If one would take an in-depth look at teacher trainings within the countries, there is an even chance that differences in teacher training and school equipment constitute part of the reasons why the Null-hypothesis could not be rejected, while the multi-level approach accounted for country-specific differences in the variance here.

Figure 3: Comparison of broadband access from the year 2011 and 2017 in European countries compared to the target of complete coverage in 2020 by the European Court of Auditors (2018).



Source: Original material

While Güzeller and Akin (2014) argued for the PISA 2006 sample that ICT has not been fully integrated into the school environment, we take our findings as a strong indication that about a decade later the process is still underway. The resemblance in our findings (e.g., Table 4) to the observance of available infrastructure in the countries supports the assumption that schools need a reasonable infrastructure in the first place to engage with challenges from the process of digitalisation in a meaningful way. Building on that premise would then facilitate an advancement of teachers' professional development to include ICT content, as the use in everyday classroom situations becomes readily available (cf. section 3.4).

At the same time, this stipulates institutions tasked with the initial training of teachers to develop and employ concepts suited to equip teachers with the necessary skills to meet the challenges of digitalisation and ICT bearing environments in their later work life. It is obviously not enough to equip schools with ICT devices; what is necessary is an educated implementation of teaching skills incorporating the use of ICT.

This paper focussed on teacher education and school characteristics in five countries, but due to the nature of the PISA data (i.e., that no direct connection between teacher and students (e.g., by class-ID) was possible. This goes back to the fact that the overall goal of PISA is to gather information about available competences at the population level and not at an individual level. Accordingly, the sampling design was not suited to draw inferences on students' individual learning outcomes within specific learning environments. For empirical evidence on research questions referring to individual learning outcomes, alternative study designs are required. One example for such a study is the German COACTIV study (the acronym stands for Professional Competence of Teachers, Cognitively Activating Instruction, and Development of Students' Mathematical Literacy). COACTIV used a multimethod design including cognitive tests, questionnaires, document analyses and combined multiple data sources that could be uniquely linked to each other. Around 11 % of the German¹⁰ participants in the 2012 Programme for the International Assessment of Adult Competencies (PIAAC) cycle had either no computer experience at all or were unable to complete the IT trial prior to the testing (OECD, 2016b). This is compounded by the roughly 14 % below and 30 % on the lowest of the four competence levels in technology based problem solving (Rammstedt, 2013, p. 68). With regard to the population of 15-year old students tested in PISA, the results from the first cycle of the ICILS study placed around 30 % of students on the lowest two competence levels in the German sample for grade 8 (Bos et al., 2014, p. 131). The "digital divide" (OECD, 2000) may therefore very well also start to affect the assessment part as a whole in a significant way, if low performing individuals are starting to be not able to participate in the evaluation process itself.

Causal inferences can hardly be established from the PISA data. This goes back to a number of limitations that the data has due to the sampling and study design. First, as an international large-scale student assessment, PISA is designed to collect a database for analyses aiming at the system and population level, not a micro-level like schools and individual students (Prenzel & Sälzer, 2019). It is a periodically repeated assessment and hence a cross-sectional design allowing for trend analysis, but not causal conclusions. While trends can be used as indicators for developments and potential risks in educational systems, causal inference require longitudinal studies with multiple points of measurement. Second, PISA's sampling design is not meant to prepare causal analyses. The sampling design follows the general aim of any study and hence, the aims and the procedure of sampling can make a difference with regard to the outcomes measured. When assessing educational systems, the basic idea of randomized sampling may require complex stratified samples (e. g. in order to capture sub-systems within a country, different school types, grouping of students in classrooms etc.) and thus produce differences in the selected approaches. With the focus on

¹⁰ A similar fraction was observed in the Czech Republic, the other three countries considered in this paper did not participate in the PIAAC 2012 study.

students' average performance levels in different countries, the question of how relevant target groups (e.g. teachers) should be included in the sample arises. In PISA, such a sample of teachers has only recently been drawn for the first time and no unique linking between teachers and students in the sense of "who teaches whom" is possible. Due to this, concluding from one level to the other is therefore per se limited. From our point of view, future research could lead to further insights, if learning outcomes could be linked closer to the environment they were developed in (at least at the class level). Data from large-scale assessments like PISA follow and serve certain requirements, but of course they also face limitations. When we want to know more about how teaching ICT-related skills in school works and how ICT literacy can be taught, designs beyond large-scale assessments are needed. Experimental studies, quasi experimental studies or interventions could be worthwhile here. Third, PISA's design using rotated questionnaires which consist of two or more parts and are presented only to a subsample each means in our case that only roughly 50 % of the teachers in the sample were given questions concerning the content of initial training and professional development (TC045Q01NA-TC045Q14NB). From a transnational perspective, both would have been very interesting starting points, which might be worthwhile to pursue when investigating the interdependencies between the various factors, but were beyond the immediate scope of this paper. Moreover, it needs to be taken into consideration that the sample would have been further diminished, potentially preventing the international perspective at the heart of this work. Even though, for the mentioned reasons, our study cannot provide causal suggestions for explaining what works in ICT education, we are able to take a look into the future of large-scale assessments in education and develop a few thoughts to conclude this work and suggest how data from large-scale assessments can be further enriched.

In the process of methodologically developing and improving large-scale student assessments, tests have to be continuously optimized. Taking into account that competence development always takes place in a context, the focus of assessment needs to be broadened in order to capture societal changes and requirements. At the same time, new test formats such as computer-based testing and alternative assessment methods have begun to be developed and will be further elaborated continuously. From a meta-perspective aiming on educational policy-making, researchers should bring together evidence from different large-scale assessments. One key aspect here is to gain evidence on the prognostic validity of such studies, in order to determine whether the competences assumed to be necessary for participating in society actually meet this expectation. Digitalization is an example for such an assumption, which we tried to examine in our paper.

The future of using large-scale assessments as an empirical reference for educational policy making should overcome focussing on single studies as separate databases. Instead, huge and growing datasets need to be re-analysed, the instruments adapted, and additional research components developed continuously. In this regard, international large-scale student assessments can be used as a quasi-panel of multiple datasets (Chmielewski & Savage, 2016). Since student achievement scores and background data are collected across all participating educational systems, data from large-scale assessments are uniquely suited for making cross-national comparisons. With regard to the constant improvement of datasets from large-scale studies, possibilities of causal inferences will be addressed more accurately in the future. National characteristics of educational systems and policies are not randomly assigned to countries and students, but are instead a product of multiple historical and cultural influences. This is especially true for the structure of teacher education, being composed of initial training and professional development. It can therefore not be determined whether a difference in results in international large-scale assessments goes back to a difference in a particular national policy or to other sources. Our interpretation of our analyses therefore remain vague in the sense that we suggest plausible explanations without being able to causally model the postulated relationships of variables. Since at this point none of the international large-scale assessments include longitudinal data for students, a lot of research in the field uses quasi-

experimental designs in order to attempt to make causal inferences with observational data. Recent improvements in both the methodology and the combination of different cycles of large-scale assessments have increasingly allowed researchers to design studies approaching causal inferences by linking different levels of data sources.

BIBLIOGRAPHY

- Araya, R., Gormaz, R., Bahamondez, M., Aguirre, C., Calfucura, P., Jaure, P., & Laborda, C. (2015). Ict Supported Learning Rises Math Achievement in Low Socio Economic Status Schools. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & É. Lavoué (Eds.), *Lecture Notes in Computer Science: Vol. 9307. Design for teaching and learning in a networked world: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015, Toledo, Spain, September 15-18, 2015 : proceedings* (Vol. 9307, pp. 383–388). Cham, Heidelberg, New York, Dordrecht, London: Springer. https://doi.org/10.1007/978-3-319-24258-3_28
- Asparouhov, T., & Muthén, B. (2008). Multilevel Mixture Models. In G. R. Hancock & K. M. Samuelsen (Eds.), *CILVR series on latent variable methodology. Advances in latent variable mixture models: The theme for the inaugural conference, held at the University of Maryland on May 18 and 19, 2006, was Mixture Models in Latent Variable Research* (pp. 27–52). Charlotte, NC: Information Age Pub.
- Blume, B. D., Ford, J. K. [J. Kevin], Baldwin, T. T., & Huang, J. L. (2010). Transfer of Training: A Meta-Analytic Review. *Journal of Management*, 36(4), 1065–1105. <https://doi.org/10.1177/0149206309352880>
- Bos, W., Eickelmann, B., Gerick, J., Goldhammer, F., Schaumburg, H., Schippert, K., . . . Wendt, H. (Eds.). (2014). *ICILS 2013: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern in der 8. Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Burnham, K. P. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261–304. <https://doi.org/10.1177/0049124104268644>
- Chmielewski, A. K., & Savage, C. (2016). Socioeconomic segregation between schools in the US and Latin America, 1970–2012. In G. W. McCarthy, G. K. Ingram, & S. A. Moody (Eds.), *Land and the city* (pp. 394–423). Cambridge, MA: Lincoln Institute of Land Policy.
- Dinis da Costa, P., & Araújo, L. (2016). *Digital reading in PISA 2012 and ICT uses: How do VET and general education students perform? EUR, Scientific and technical research series: Vol. 28291*. Luxembourg: Publications Office.
- European Court of Auditors. (2018). *Broadband in the EU Member States: Despite progress, not all the Europe 2020 targets will be met. Special report: No 12, 2018*. Luxemburg: Publications Office of the European Union.
- Falck, O., Mang, C., & Wößmann, L. (2018). Virtually No Effect? Different Uses of Classroom Computers and their Effect on Student Achievement. *Oxford Bulletin of Economics and Statistics*, 80(1), 1–38. <https://doi.org/10.1111/obes.12192>
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What Makes Professional Development Effective? Results From a National Sample of Teachers. *American Educational Research Journal*, 38(4), 915–945. <https://doi.org/10.3102/00028312038004915>
- Gerick, J. (2018). School level characteristics and students' CIL in Europe – A latent class analysis approach. *Computers & Education*, 120, 160–171. <https://doi.org/10.1016/j.compedu.2018.01.013>
- Goldhammer, F., Gniewosz, G., & Zylka, J. (2017). Ict Engagement in Learning Environments. In S. Kuger, N. Jude, & D. Kaplan (Eds.), *Methodology of Educational Measurement and Assessment. Assessing Contexts of Learning: An International Perspective* (Vol. 34, pp. 331–351). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-45357-6_13
- Goldstein, H. (2003). *Multilevel statistical models* (3. ed.). *Kendall's library of statistics: Vol. 3*. London: Arnold. Retrieved from <http://www.loc.gov/catdir/enhancements/fy0615/2003276198-d.html>

- Goldstein, I. L., & Ford, J. K. [John Kevin]. (2009). *Training in organizations: Needs assessment, development, and evaluation* (4. ed., [Nachdr.]). Belmont, Calif: Wadsworth/Thomson Learning.
- Gómez-Fernández, N., & Mediavilla, M. (2018). *Do Information And Communication Technologies (ICT) Improve Educational Outcomes?: Evidence For Spain in PISA 2015* (IEB Working Paper No. 20). <https://doi.org/10.13140/RG.2.2.21085.87528>
- Grossman, R., & Salas, E. (2011). The transfer of training: what really matters. *International Journal of Training and Development*, 15(2), 103–120. <https://doi.org/10.1111/j.1468-2419.2011.00373.x>
- Güzeller, C. O., & Akın, A. (2014). Relationship between ICT Variables and Mathematics Achievement Based on PISA 2006 Database: International Evidence. *Turkish Online Journal of Educational Technology - TOJET*, 13(1), 184–192. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1018171.pdf>
- Hanushek, E. A., Ruhose, J., & Wößmann, L. (2016). Knowledge Capital and Aggregate Income Differences: Development Accounting for U.S. States. *American Economic Journal: Macroeconomics*.
- Helmke, A. (2010). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts ; Franz Emanuel Weinert gewidmet ; [Orientierungsband]* (3. Aufl.). [Unterricht verbessern - Schule entwickeln]. Stuttgart: Klett [u.a.].
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments. 3rd ed.* Odessa Fla: Psychological Assessment Resources.
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jaensch, V. K., Hirschi, A., & Spurk, D. (2016). Relationships of Vocational Interest Congruence, Differentiation, and Elevation to Career Preparedness Among University Students. *Zeitschrift für Arbeits- und Organisationspsychologie A&O*, 60(2), 79–89. <https://doi.org/10.1026/0932-4089/a000210>
- Juhaňák, L., Zounek, J., Záleská, K., Bárta, O., & Vlčková, K. (2019). The Relationship Between Students' ICT Use and Their School Performance: Evidence from PISA 2015 in the Czech Republic. *ORBIS SCHOLAE*, 12(2), 37–64. <https://doi.org/10.14712/23363177.2018.292>
- Kenny, D. A. (2014). Measuring Model Fit. Retrieved from <http://davidakenny.net/cm/fit.htm>
- Knowles, M. S., Holton, E. F., & Swanson, R. A. (2012). *The adult learner: The definitive classic in adult education and human resource development* (seventh edition). London, New York: Routledge.
- Koehler, M. J., Mishra, P., Kereluik, K., Shin, T. S., & Graham, C. R. (2014). The Technological Pedagogical Content Knowledge Framework. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 101–111). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_9
- Maier, A., Nitzschke, A., Nickolaus, R., Schnitzler, A., Velten, S., & Dietzen, A. (2015). Der Einfluss schulischer und betrieblicher Ausbildungsqualität auf die Entwicklung des Fachwissens. In M. Stock, P. Schlögl, K. Schmid, & D. Moser (Eds.), *Innovationen in der Berufsbildung: Vol. 9. Kompetent - wofür? Life Skills - Beruflichkeit - Persönlichkeitsbildung: Beiträge zur Berufsbildungsforschung* (1st ed., pp. 225–243). Innsbruck: Studien Verlag.
- Muthén, B., Muthén, L., Asparouhov, T., & Nguyen, T. (2018). Mplus: Muthén & Muthén.
- OECD. (2000). *Learning to Bridge the Digital Divide. Education and Skills*: OECD. <https://doi.org/10.1787/9789264187764-en>
- OECD. (2016a). *PISA 2015 Ergebnisse (Band I)*: W. Bertelsmann Verlag. <https://doi.org/10.1787/19963793>

- OECD. (2016b). *Skills Matter: Further Results From The Survey Of Adult Skills. PLAAC*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264258051-en>
- OECD. (2018). *Science, Technology and Innovation Outlook 2018: Adapting to Technological and Societal Disruption*. OECD. https://doi.org/10.1787/sti_in_outlook-2018-en
- Palfrey, J., & Gasser, U. (2008). *Born digital: Understanding the first generation of digital natives*. New York: Basic Books. Retrieved from <http://site.ebrary.com/lib/academiccompletetitles/home.action>
- Prenzel, M., & Sälzer, C. (2019). Large-scale assessments of educational systems. In R. Becker (Ed.), *Research handbooks in sociology. Research handbook on the sociology of education* (pp. 536–552). Cheltenham: Edward Elgar.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167–190. <https://doi.org/10.1007/BF02295939>
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology*, 25, 111–163. <https://doi.org/10.2307/271063>
- Rammstedt, B. (Ed.). (2013). *Erwachsenenbildung 2013/14. Grundlegende Kompetenzen Erwachsener im internationalen Vergleich*. Münster: Waxmann Verlag. Retrieved from http://www.content-select.com/index.php?id=bib_view&ean=9783830979999
- Raudenbush, S. W., & Bryk, A. S. (2010). *Hierarchical linear models: Applications and data analysis methods* (2. ed., [Nachdr.]). *Advanced quantitative techniques in the social sciences: Vol. 1*. Thousand Oaks, Calif.: Sage Publ.
- Rosén, M., & Gustafsson, J.-E. (2016). Is computer availability at home causally related to reading achievement in grade 4? A longitudinal difference in differences approach to IEA data from 1991 to 2006. *Large-scale Assessments in Education*, 4(1), 130. <https://doi.org/10.1186/s40536-016-0020-8>
- Sälzer, C., & Prenzel, M. (2014). Looking back at five rounds of PISA: Impacts on teaching and learning in Germany. *Solsko Polje (The School Field). Evidence from the PISA Study on Educational Quality in Slovenia and Other Countries, XXV*, 53–72.
- Spokane, A. R., Meir, E. I., & Catalano, M. (2000). Person–Environment Congruence and Holland's Theory: A Review and Reconsideration. *Journal of Vocational Behavior*, 57(2), 137–187. <https://doi.org/10.1006/jvbe.2000.1771>
- Steffens, K. (2014). ICT Use and Achievement in Three European Countries: What Does PISA Tell Us? *European Educational Research Journal*, 13(5), 553–562. <https://doi.org/10.2304/eej.2014.13.5.553>
- Streiner, D. L. (2002). Breaking up is hard to do: The heartbreak of dichotomizing continuous data. *Canadian Journal of Psychiatry. Revue Canadienne De Psychiatrie*, 47(3), 262–266.
- Watermann, R., Maaz, K., Bayer, S., & Roczen, N. (2017). Social Background. In S. Kuger, N. Jude, & D. Kaplan (Eds.), *Methodology of Educational Measurement and Assessment. Assessing Contexts of Learning: An International Perspective* (Vol. 40, pp. 117–145). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-45357-6_5
- Wößmann, L., & Fuchs, T. (2004). *Computers and Student Learning: Bivariate and Multivariate Evidence on the Availability and Use of Computers at Home and at School*. Retrieved from CESifo Working Paper Series No. 1321 website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=619101
- Wu, M. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, Vic.: ACER Press.
- Yalçın, S. (2018). Multilevel Classification of PISA 2015 Research Participant Countries' Literacy and These Classes' Relationship with Information and Communication Technologies. *International Journal of Progressive Education*, 14(1), 165–176. <https://doi.org/10.29329/ijpe.2018.129.12>

- Youssef, A. B., & Dahmani, M. (2008). The Impact of ICT on Student Performance in Higher Education: Direct Effects, Indirect Effects and Organisational Change. *RUSC. Universities and Knowledge Society Journal*, 5(1). <https://doi.org/10.7238/rusc.v5i1.321>
- Zhang, D., & Luman, L. (2016). How Does ICT Use Influence Students' Achievements in Math and Science Over Time?: Evidence from PISA 2000 to 2012. *EURASIA Journal of Mathematics, Science & Technology Education*, 12(9), 2431–2449. <https://doi.org/10.12973/eurasia.2016.1297a>

ABOUT THE AUTHORS

Leo Van Waveren

Leo van Waveren's research is focussed learning outcomes, especially in large scale assessments and digitalisation in teacher education. In his PhD thesis, he compares results from trainees in two different occupations in Germany and Switzerland. Since 2018 he is working as a Post-Doc in Christine Sälzer's Research Group on an EU-funded project on criterion referenced testing in an international project.

Contact information: University of Stuttgart, Institute of Educational Science, Department of Pedagogics, Azenbergstr. 16, 70174 Stuttgart, Germany, +49 711-685 81085, leo.vanwaveren@ife.uni-stuttgart.de

Christine Sälzer

Christine Sälzer's main topics of research are educational large-scale assessments, school leaving and support for children with special needs as well as competence modelling. She holds the professorship for educational science at the University of Stuttgart after she completed her habilitation on the use of large-scale assessment results educational policy decision making at the Technical University of Munich. From 2010 to 2017 she was the national project manager for PISA.

Contact information: University of Stuttgart, Institute of Educational Science, Department of Pedagogics, Azenbergstr. 16, 70174 Stuttgart, Germany, +49 711-685 87440, Christine.saelzer@ife.uni-stuttgart.de