Assessing subjective and objective information literacy at upper secondary schools – an empirical study in four German-speaking countries

Sabine Seufert*

Institute of Business Education and Educational Management,
University of St. Gallen,
St. Jakob-Strasse 21, 9000 St. Gallen, Switzerland
Email: sabine.seufert@unisg.ch
*Corresponding author

Katarina Stanoevska-Slabeva

Institute for Media and Communications Management,
University of St. Gallen,
Blumenbergplatz 9, 9000 St. Gallen, Switzerland
Email: katarina.stanoevska@unisg.ch

Josef Guggemos

Institute of Business Education and Educational Management,
University of St. Gallen,
St. Jakob-Strasse 21, 9000 St. Gallen, Switzerland
Email: josef.guggemos@unisg.ch

Abstract: Information literacy (IL) is a key 21st-century skill. This paper contributes to the body of knowledge by specifying the construct of IL. It defines IL and presents a seven-dimensional IL model. On this basis, two types of test instruments are evaluated: a performance test for objective IL (OIL) and a self-assessment instrument for subjective IL (SIL). Both instruments are validated by means of item response theory. We utilise a sample with N = 432 upper secondary students from German-speaking countries. Rasch scalability of the instruments is demonstrated. Reliability equals in the case of OIL for EAP/PV .84 and for WLE .88; in the case of SIL, .75 and .77, respectively. For OIL, we present a proficiency level model to ensure a criterion-based interpretation of test scores. In terms of instructional sensitivity, we provide evidence for a positive influence of students' grade and IL-related content in class on OIL. The correlation between OIL and SIL equals .52.

Keywords: digital natives; information literacy; bifactor model; instructional sensitivity; curriculum-instruction-assessment triad; proficiency level model; secondary education.

Reference to this paper should be made as follows: Seufert, S., Stanoevska-Slabeva, K. and Guggemos, J. (2020) 'Assessing subjective and objective information literacy at upper secondary schools – an empirical study in four German-speaking countries', *Int. J. Learning Technology*, Vol. 15, No. 1, pp.82–103.

Biographical notes: Sabine Seufert is a Full Professor and the Director of the Institute of Business Education and Educational Management at the University of St. Gallen, Switzerland. She is a member of the editorial board of the *International Journal of Learning Technology* and the *International Journal of Advanced Corporate Learning*. Her research interests lie on educational management and digital transformation of education.

Katarina Stanoevska-Slabeva is an Associate Professor at the Institute for Media and Communications Management at the University of St. Gallen, Switzerland. She has received her PhD in Information Management at the University of St. Gallen. Her main research interest concerns digital competences in particular information and data competence, impact, and application of social media, mobile communication as well as digital and platform-based business models.

Josef Guggemos is an Assistant Professor at the Institute of Business Education and Educational Management, University of St. Gallen, Switzerland. He obtained his Master's in Business Research (MBR) and PhD at the Munich School of Management, LMU Munich, Germany. His research focus includes computational thinking, social robots in education, and competencies of teachers in the context of digital transformation.

1 Introduction

The change towards a digital society is deemed to be a megatrend. In light of this, information literacy (IL) is one of the key competences of the 21st century (Binkley et al., 2012). IL may be conducive to social engagement, a self-determined life and lifelong learning (Gapski and Tekster, 2009).

Persons born after 1980 are often referred to as 'digital natives' (Tapscott, 1997; Prensky, 2001; Palfrey and Gasser, 2008). This indicates that they have been using digital devices their entire life. However, there is evidence that using such technologies is not necessarily accompanied by critical reflection (Kirschner and van Merriënboer, 2013). For instance, 'digital natives' effortlessly use search engines and various sources. However, they often deal naively with the information found (Fraillon et al., 2014; OECD, 2015). In light of this, the call for an early integration of IL into school curricula is not surprising (Argelagós and Pifarré, 2012; Fraillon et al., 2014). However, despite the growing awareness about digital natives' IL deficits, efforts to anchor IL systematically as a curricular goal may be regarded as insufficient (Jones-Kavalier and Flannigan, 2006; SERI, 2017).

To tackle this issue, we rely on the curriculum-instruction-assessment triad as a framework. The key idea is that curriculum, instruction, and assessment have to be in alignment for successful educational processes. The curriculum "consists of the knowledge and skills in subject matter areas that teachers teach and students are supposed to learn" [Pellegrino, (2010), p.4]. Instruction "refers to methods of teaching and the learning activities used to help students master the content and objectives specified by a curriculum" [Pellegrino, (2010), pp.4–5]. Assessment addresses "the means used to measure the outcomes of education and the achievement of students with regard to important competencies" [Pellegrino, (2010), p.5]. The paper at hand addresses the

curriculum and assessment. To this end, we first present a model of IL that may act as a curriculum for upper secondary students. In terms of assessment, we show how IL can be measured using a performance test and a self-assessment instrument. In considering self-assessment, we follow the call for more reflection by students in the learning processes (Cope and Kalantzis, 2016). We validate both instruments by means of item response theory (IRT). Moreover, we provide a proficiency level model of objective information literacy (OIL) to allow for a criterion-referenced interpretation of test scores. Concerning instruction, we give an outlook on how IL may be taught to ensure an alignment of curriculum, instruction and assessment.

Both instruments, OIL and subjective information literacy (SIL), should be Rasch scalable. In this case, specific objectivity is ensured, which may be a criterion for test quality (Bühner, 2011). To check for Rasch scalability, we examined person homogeneity, item discrimination and item homogeneity (Bühner, 2011). The following research question structures our analyses:

RQ1 Are the developed OIL and SIL instruments Rasch scalable?

- a Are the test instruments fair [absence of differential item functioning (DIF)]?
- b Is the discrimination of the items identical?
- c Does a seven-dimensional model show a better fit than a one-dimensional one?

To make the results of the OIL test interpretable, it is necessary to develop a proficiency level model (Fraillon et al., 2015, 2019). This allows us to separate students into meaningful levels. Each level is characterised by tasks that students of the respective level are able to systematically solve, and that students at a lower level are not able to systematically solve. Against this backdrop, we raise a second research question:

RQ2 What proficiency levels can be formed in terms of OIL?

Assessment is not an end in itself. Drawing on the curriculum-instruction-assessment triad, the assessment has to be in alignment with the instruction. Instructional sensitivity might be demonstrated because otherwise the instruments would not be able to detect instructional treatments (Naumann et al., 2019). Thus, we formulate a third research question:

RQ3 How does instruction influence SIL/OIL?

Assessing the relationship between SIL and OIL may be important for demonstrating construct validity. OIL and SIL should be positively related (Rosman et al., 2015): the same construct is measured with different methods. However, the strength of the relationship could be gender dependent. Lundeberg et al. (1994) conclude that women and men both overestimate their performance, but men do so more strongly than women. The IEA's International Computer and Information Literacy Study (ICILS) provide empirical evidence on this gender issue (Fraillon et al., 2019). In light of this, we formulate the following research question:

RQ4 What is the relationship between OIL and SIL?

- a How strong is the overall relationship between OIL and SIL (quality of self-assessment)?
- b Do men overestimate their IL more so than women?

2 Review of the literature

Overall, there may be a lack of suitable standardised measurement frameworks and methods for capturing IL, in particular for the secondary level. ICILS focuses on the lower secondary level [for the first time in 2013, for the second time in 2018 (Fraillon et al., 2019)]. Furthermore, prior research only provides insights into certain aspects of IL, such as the evaluation of sources [see, for example, the study by McGrew (2020) as an intervention in civic online reasoning]. In a systematic literature review, Stanoevska-Slabeva et al. (2015) identified seven central IL facets. In this respect, we define IL as the ability and willingness to:

1 identify information needs

2 select information sources

3 access information

4 evaluate

5 use it

6 present it

7 reflect on both the process of information retrieval and processing, as well as the results.

The following section describes these seven facets:

1 Information is required to make decisions and solve problems (Bruce, 1999). 'Information needs', as the first step, addresses the identification of required information in comparison with already available knowledge (ALA, 1989; Bruce, 1992; Rader, 2003; UNESCO, 2003; Livingstone et al., 2005; ACRL, 2000). Evidence for this facet may be the delineation of an ill-defined problem, for example, a 15-minute presentation on the UNESCO.

2 The facet 'information sources' differs substantially between the reviewed studies. The term IL originally comes from the library sciences. Here, the sources of information are classical libraries. In the age of the internet, however, the information-base has changed dramatically. In particular, online encyclopaedias, data from public authorities, and social media are now important sources of information. Social media plays a special role: in contrast to conventional sources, information is not only sought (Jones-Kavalier and Flannigan, 2006), but also jointly generated, presented and shared (Livingstone et al., 2005). Since we aim at a broad understanding of IL, we explicitly consider social media as a source of information. An evidence of this would be the selection of sources that provide up-to-date information, such as the Twitter account of the UNESCO.

3 The 'information access and seeking strategy' is part of all reviewed IL definitions (ALA, 1989; Bruce, 1992; Braun, 1997; Rader, 2003; UNESCO, 2003; Livingstone et al., 2005; Chaka, 2009; ACRL, 2000). Accessing information depends on the information source. Formulating adequate search strings using Boolean operators might be evidence for the competent use of search engines. In the case of social

media, an evidence would be the formulation of an appropriate question in an online forum.

4 The retrieved information is of varying quality, especially if social media is utilised as a source. Hence, 'information evaluation' is necessary before it can be used (ALA, 1989; Marshall, 2006; Miller and Bartlett, 2012; Bruce, 1992; Braun, 1997; Rader, 2003; UNESCO, 2003; Livingstone et al., 2005; Chaka, 2009; ACRL, 2000). The key evaluation criteria in this process are reliability and relevance. An example would be an accurate assessment of the quality of different answers to a question in an online forum.

5 In the phase 'information use', Bruce (1992) and Rader (2003) also address moral aspects. Those may be strongly culture dependent. Against this backdrop, we restrict ourselves to the adherence to rules that apply across national borders. The main aspects of 'information use' are extracting and paraphrasing information from the evaluated sources. An evidence for this may be appropriately citing others' work.

6 Depending on the objective of the search, the obtained information may be presented. 'Information presentation' is particularly relevant when using social media. An imprudent publication of content might cause serious disadvantages for the individual. Our literature review indicated that presentation is rarely an independent facet of IL. An exception is Marshall (2006). The appropriate publication of information on social media channels might be evidence for this facet.

7 Phases 1 to 6 have to be reflected on with regard to the achievement of objectives in order to learn for the future (Rader, 2003). Evidence for 'information processing and reflection' might be the appropriate selection of criteria for assessing the success of phases 1 to 6.

The seven facets do not follow a strict linear sequence. Rather, they are performed iteratively. The process is accompanied by permanent reflection on the procedure and the obtained results. The seventh facet thus forms the centre of the information process and is interconnected with all six other facets. Figure 1 summarises our 7i model of IL.

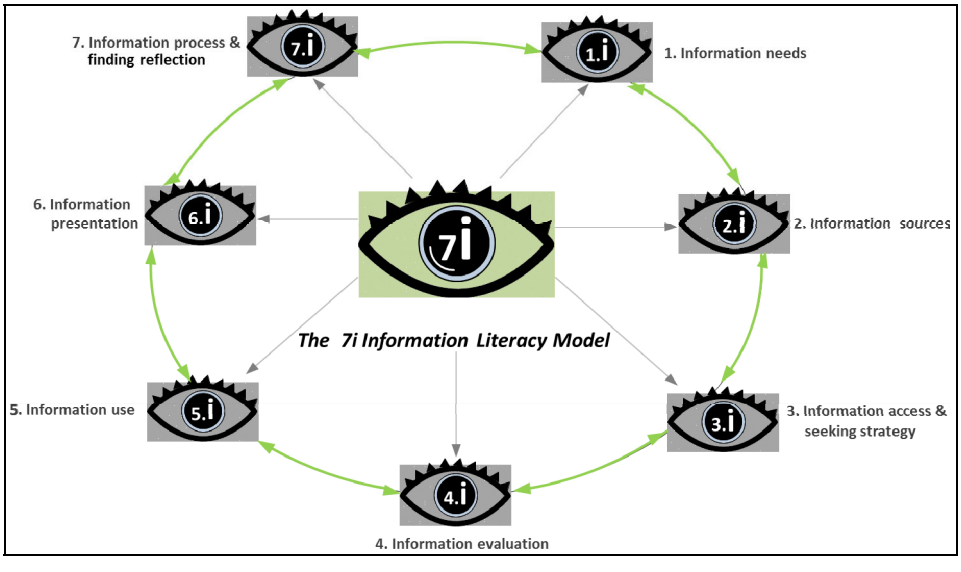
After the delineation of the curriculum, we reviewed available instruments that are suitable for assessing IL. Due to its advantageous psychometric properties (Hartig and Frey, 2013), the instrument should meet the requirements of the IRT. Moreover, a good fit of the instruments with the characteristics of the target group (upper secondary level) is crucial. Instruments from the higher education context are therefore not directly helpful. Overall, two instruments meet our requirements: the work of Balceris (2011) and Marshall (2006).

Balceris (2011) developed 168 test items for measuring IL. They mainly comprise situational judgement tasks (McDaniel and Nguyen, 2001), i.e., they describe an incident where the test taker has to assess the quality of potential actions. This kind of task may be suitable for evaluating IL (Rosman et al., 2016). Balceris (2011) successfully validated his instrument by means of confirmatory factor analyses in a sample of ninth graders at German high schools. However, he neither explicitly takes into account social media as a source, nor does he consider our IL facet 'information presentation'.

Marshall (2006) uses self-assessments to measure IL. His questionnaire comprises 40 items. The items primarily cover whether students think they are able or not to perform certain actions. Like Balceris (2011), Marshall (2006) does not explicitly

consider social media. However, the facet 'information presentation' is included. Marshall (2006) successfully performed reliability analyses (Cronbach's alpha > .85). A critical point is the reliance on self-assessment. The instrument therefore may be unsuitable for summative assessments; however, it may be suitable for formative assessment and as an instrument for critical reflection.

Figure 1 7i IL model (see online version for colours)



Source: Stanoevska-Slabeva et al. (2015)

3 Method

3.1 Developing instruments for assessing IL

The work of Balceris (2011) and Marshall (2006) is the starting point for the development of our test instruments. This is necessary to ensure alignment between curriculum and assessment, i.e., the instruments should exhaustively cover the facets of the 7i model. An international research consortium with members from Austria, Germany, Lichtenstein, and Switzerland was involved in the development and selection process. The aim was to capture IL objectively (OIL) by means of a performance test, and subjectively (SIL) by means of a questionnaire.

3.1.1 Objective information literacy

Since 'information presentation' is not part of the task pool, we have constructed items for this facet. Furthermore, we have developed items in the realm of social media. We paid strict attention to local stochastic independence when designing and selecting items and ensured that all items could be solved independently of each other.

Due to the test validation process, we expected an exclusion of approximately 40% of the initial items (Baethge et al., 2006). Thus, we included 89 items in order to achieve a decent reliability (> .80) even after the removal of 40% of the initial items.

The test comprises two types of situational judgement items:

1 A behaviour in a given situation has to be assessed using a four-point scale ranging from 'entirely appropriate' to 'entirely inappropriate'. The quality of the response is measured using partial credits. Concordance with the expert assessment is coded with '2', the deviation by one step with '1', and all stronger deviations with '0'.

2 In single-choice tasks, behaviours in a concrete situation have to be assessed. Since only one alternative is appropriate, these items are binary coded.

An example for type 1 is depicted in Figure 4, and for type 2 in Figure 5. Due to the selected-response format, all answers can be coded automatically. The objectivity of the coding is therefore optimal.

3.1.2 Subjective information literacy

We have adapted Marshall's (2006) questionnaire analogously to OIL and considered all seven IL facets. This resulted in 29 items. Persons have to assess themselves on a five-point rating scale from 'applies entirely' to 'does not apply at all'. Sample items are 'I can pick up a complex topic and break it down into meaningful, simpler parts', 'I know how to access different social media platforms (e.g., Wikipedia or Facebook) to search for specific information', or 'I can correctly classify comments and information from social media platforms'.

3.2 Sample

To answer our research questions, we used a sample of N = 432 students nested in 23 classes from Germany (24.9%), Austria (21.7 %), Lichtenstein (7.8%) and Switzerland (45.6%). Of the students, 53.3% were female. 17.3% attended the 9th grade, 11.5% the 10th, 14.1% the 11th, 37.3% the 12th, and 19.8% the 13th. 37.8% are from high schools, 55.7% from vocational schools leading to the Abitur, and 6.5% from vocational schools not leading to the Abitur. On average, the test subjects were about 18 years old (SD = 2.02 years).[1] They performed the OIL and SIL instruments as an online questionnaire in Unipark under the supervision of teachers. Although participation was voluntary, only 1.85% values were missing.

3.3 Dealing with missing values and outliers

Missing values occur only in the OIL and SIL items, not in the context and control variables. Due to the (marginal) maximum likelihood method used in the parameter estimation, we considered those as 'not present' (Finch, 2008).

The exclusion of persons as outliers is problematic in the context of test validation (Rost, 2004). However, outliers could lead to biased results when investigating RQ3 with an ordinary least squares regression. Hence, we identified potential outliers using studentised residuals. We then tried to identify the reasons and take appropriate means.

3.4 Ensuring Rasch scalability (RQ1)

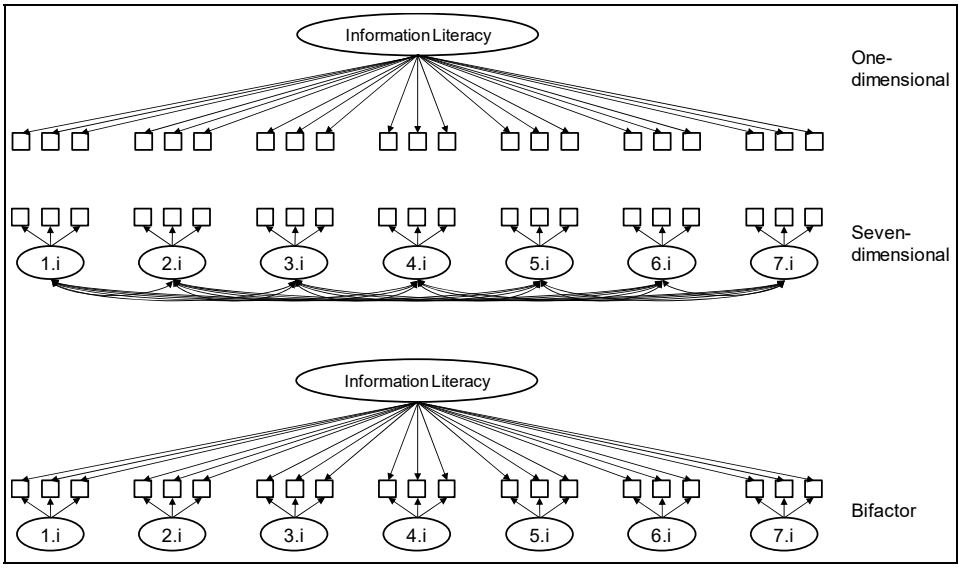
Due to partly correct solutions (OIL) and several possible levels in self-assessment (SIL), we relied on a partial credit model (PCM) (Masters, 1982). All statistical analysis was carried out using R version 3.6.0 (R Core Team, 2019).

3.4.1 Person homogeneity and item discrimination

Using Andersen's (1973) likelihood ratio test (LRT), we checked for person homogeneity (test fairness) by means of DIF analysis. We formed groups on the basis of the median of the raw score of the test (level of IL), type of school, grade, gender, country of origin, and age (above and below average). We performed these analyses for all seven IL facets. In the case of a significant LRT, we conducted post hoc tests to exclude items with significant DIF effects. Following Koller et al. (2012), we chose a significance level of 10% and applied a Bonferroni correction. Since we carried out the LRT six times, we divided the significance level by six. To perform the LRT, we used the eRM package in R (Mair et al., 2018).

To examine item discrimination, we checked the extent to which the empirical item characteristic curves (ICC) differ from that implied by the Rasch model (Wilson, 2005). In line with the PISA study, we set 0.8 to 1.2 as an acceptable Infit (OECD, 2014). To carry out the analyses, we utilised the TAM package in R (Robitzsch et al., 2018).

Figure 2 Illustration of one-dimensional, seven-dimensional and bifactor model



3.4.2 Item homogeneity

The 7i model implies a seven-dimensional structure. First, we examined whether a seven-dimensional model is superior to a one-dimensional solution. In addition, we assessed whether, in addition to the seven dimensions, an overarching factor in the sense of a general IL can be extracted. To this end, we examined whether a bifactor model

(Reise, 2012) shows a better fit to the data than a seven-dimensional model. In a bifactor model, the seven-dimensional structure is taken into account using seven factors. Moreover, the items load on a general factor. By definition, the eight extracted factors are orthogonal. We checked the relative fit of the three models with a LRT (Glas and Verhelst, 1995) and using the information criteria AIC and BIC (Vrieze, 2012). For the described analyses, we relied on the TAM package in R and used a marginal maximum likelihood estimator. Figure 2 illustrates the three models.

3.4.3 Reliability

We gauged the reliability of OIL and SIL by means of two measures (Adams, 2005):

1 The expected a posteriori based on plausible values (EAP/PV) estimation, which shows if the order of the item difficulties could be replicated in another group of students of similar IL (item separation reliability).

2 The weighted maximum likelihood estimation (WLE) reliability which indicates if the estimated IL of the students could be replicated using other items of similar difficulty (person separation reliability).

3.5 Proficiency level model (RQ2)

To form proficiency levels, we applied the approach used by the PISA study [OECD, (2017), pp.276–287]. The cognitive taxonomy of Marzano and Kendall (2007, 2008) acts as a theoretical background. Marzano and Kendall (2007, pp.37–53) separate four levels in terms of cognitive complexity: retrieval, comprehension, analysis and knowledge utilisation. Each level is characterised by specific processes, e.g., comprehension by 'integrating' and 'symbolising'. We use this taxonomy to classify the items of our OIL test. Inspecting the Wright map, we opted for three proficiency levels with a width of 0.5 logit each. We do not distinguish between analysis and knowledge utilisation because only one student in our sample would be on the fourth level. Like in the PISA study, we selected a response probability of 62%. This means that a student at the bottom (top) of a proficiency level solves an item at the bottom (top) of the proficiency level with a probability of 62%. A student at the bottom of a level solves an item at the top of the level with a probability of 50%. A student at the top of a level solves an item at the bottom of the level with a probability of 73%.

3.6 Instructional sensitivity of OIL/SIL (RQ3)

We investigated the influence of school instruction on OIL and SIL by means of a linear regression. The WLE of IL acted as the dependent variable. The WLE are, similar to the PISA studies, standardised to M = 500 and SD = 100 in order to facilitate the interpretation of the regression results.

The test persons' grade acts as a measure of received school instruction. For the extent of IL-related content received in school, we used a binary item from the context questionnaire: 'I have used Wikipedia in class because my teacher wanted me to perform a task with this tool' (M = .71, SD = .45). We also asked about the use of other platforms, such as Twitter. However, such platforms are rarely used in lessons and therefore show a very low variance; hence, we will not consider them any further.

As control variables, we used type of school, country of origin, gender, year of birth, test motivation, and home use of Wikipedia for school purposes. The school type serves as a proxy for cognitive ability and interest structure (general education vs. vocational education). Controlling for the country of origin is particularly important in order to take into account varying instruction intensities due to differing curricula. According to Taylor and Dalal (2017), gender is a predictor for IL and is therefore also inserted as a control variable. Controlling the age of test takers is necessary because it correlates highly with the grade ($r = .58$, $p < .01$). By doing so, it is possible to separate the influence of grade (proxy for received instruction) and age (proxy for life experience) on OIL and SIL. If Wikipedia is used in class, it is likely that the test persons also use this platform beyond the classroom setting. To control for this, we used the binary item 'I have used Wikipedia for school for the last half year (e.g., for research)' ($M = .88$, $SD = .32$). The correlation (Pearson) with class use is .20 ($p < .01$).

Our data was collected in a low-stakes context. A lack of test motivation can impair the validity of the results (Wise and DeMars, 2005). To control for test motivation, we used the idea of the measure 'response time effort' (Wise and Kong, 2005). Since no data are available for the processing times of the individual items, we relied on the total processing time. After inspection of the entire test, we considered a minimum processing time of 15 minutes to be appropriate. In less than 15 minutes, it is not even possible to read all the questions and tick the correct answers. If the processing time is less than this, we classified the person as unmotivated (binary variable). Twenty-three subjects were classified as unmotivated in this way.

The data show a cluster structure (students nested in classes). Since we are only interested in individual effects, we do not use multilevel modelling, but cluster-robust standard errors (McNeish et al., 2017; Pustejovsky and Tipton, 2018).

3.7 Relationship between OIL and SIL (RQ4)

Pearson's correlation served as a global measure of the relationship between OIL and SIL. To check whether men overestimate themselves more than women, and vice versa, we used the difference between OIL and SIL as the dependent variable in a regression design. Both variables are standardised to $M = 500$ and $SD = 100$. The independent variables correspond to those described in the previous section (RQ3).

4 Results

4.1 Testing Rasch scalability (RQ1)

4.1.1 OIL: person homogeneity and item discrimination

Of the 89 items, 48 showed significant DIF effects; we therefore excluded them from the further analysis. The Infit of the remaining items lies between 0.98 and 1.05. Significant deviations from the ICC implied by the Rasch model do not occur ($p > .73$).

4.1.2 OIL: item homogeneity

The seven-dimensional model shows a better fit than the one-dimensional model ($\Delta\chi^2$ = 188, df = 27, p < .01); this is also the case when taking into account the model complexity using AIC and BIC. The bifactor model in turn shows a better fit than the seven-dimensional model ($\Delta\chi^2$ = 632, df = 39, p < . 01); considering model complexity (AIC and BIC) does not change the order. In light of this, we use the bifactor model for further evaluations. The fit values of the models are summarised in Table 1.

Table 1 Fit values for OIL (N = 432)

	One-dimensional PCM	*Seven-dimensional PCM*	*Bifactor PCM*
Deviation	27,599	27,408	26,963
Parameter	69	96	108
AIC	27,734	27,600	27,179
BIC	28,015	27,990	27,619

The EAP/PV reliability of the general factor of the bifactor model is .84, the WLE reliability equals .88. The reliability of the other seven factors ranges between .22 and .66. Therefore, the individual facets are not interpreted any further. The OIL estimated with the one-dimensional model correlates (Pearson) with the general factor of the bifactor model with .93.

4.1.3 SIL: person homogeneity

Of the initial 29 items, 17 were removed due to a lack of person homogeneity. The Infit of all remaining items lies between 0.98 and 1.01. There are no significant deviations from the ICC implied by the Rasch model (p > .68).

4.1.4 SIL: item homogeneity

The seven-dimensional PCM fits better than the one-dimensional model ($\Delta\chi^2$ = 138, df = 20, p < .01), also in terms of AIC and BIC. The seven-dimensional PCM has a better fit than the bifactor model ($\Delta\chi^2$ = 21, df = 9, p < .01); the seven-dimensional model shows a higher BIC, but a lower AIC. The fit values of the three models are summarised in Table 2.

Table 2 Fit values for SIL (N = 432)

	One-dimensional PCM	*Seven-dimensional PCM*	*Bifactor PCM*
Deviation	12,440	12,303	12,324
Parameter	49	69	60[a]
AIC	12,539	12,441	12,444
BIC	12,736	12,720	12,686

Note: [a]The lower number of estimated parameters in the case of the bifactor model can be attributed to the fact that there are no more items of the second IL facet.

Since in the case of the seven-dimensional model, the reliability of the single facets is too low (min = .59, max = .72) and because comparability with the results of the OIL should be ensured, we used the general factor of the bifactor model for the following analyses.

The EAP/PV reliability equals .75 and the WLE reliability .77. The correlation (Pearson) of SIL estimated with the one-dimensional model and the general factor of the bifactor model equals .92.

4.2 Proficiency level model (RQ2)

The Wright map shows a good fit between the ability of the test taker and the difficulty of the items (see Figure 3).

Figure 3 Wright map of OIL

STP	Logit	Students	Items	Level
804	2			
785				
766		x		
747			4.202.2	
728	1.5			
709			1.66.2 3.369.2	
689				
670			4.109.2 7.164.2	
651	1	xx		
632		xxxx	1.62.2 6.374.2	
613		xxx	3.369.1 4.108.2 6.451	III
594		xxxxxxxxxxxxxxxx	6.373.2 7.174.2	
574	0.5	xxxxxxxxx	2.85.2 3.99.2 4.352 5.147	
555		xxxxxxxxxxxxxxxxxxx	1.65.2 3.366.2 5.141.2 5.149	II
536		xxxxxxxxxxxxxxxxxxxxxxxxx	**1.59** 5.443 7.175.2	
517		xxxxxxxxxxxx		
498	0	xxxxxxxxxxxxxxxxxxxxxxxxxxxxx	7.168.2	
479		xxxxxxxxxxxxxxxxxxxxxxx	6.374.1	I
459		xxxxxxxxxxxxxxxxxx	2.83.2 4.201.2 5.139.2 6.450 7.171.2	
440		xxxxxxxxxx	4.108.1 **5.138.2** 5.142.2 5.440 7.163.2 7.170.2	
421	–0.5	xxxxxxxxx	2.85.1 4.202.1	
402		xx	4.111.1 7.172.2	
383		xxxxxxxx	1.62.1 2.83.1 7.173.2	
364		xxxx	1.66.1 **3.99.1**	
344	–1	xxx	5.444	
325		xxx	5.141.1	
306			5.139.1 7.173.1	
287		x	3.366.1 5.142.1 7.172.1	
268	–1.5	xx		
249		x	4.111.2 5.138.1 6.446 7.170.1	
229		x	5.442 6.373.1	
210			6.359 7.163.1 7.168.1 7.171.1 7.174.1 7.175.1	
191	–2		4.201.1 5.438	
172		x	1.65.1	
153				
134				
114	–2.5		4.109.1	
95				
76		x		
57			7.164.1	
38	–3	x		

Notes: Each ‘X’ represents two cases. STP = standardised points: M = 500, SD = 100. The item code is structured as follows: F.N.P, where F indicates the facet of the 7i model, N the internal number of the item, and P the partial credit category (1 and 2). In the case of a dichotomous item, P is omitted. Items in bold are anchor items (see Figures 4 to 6).

Hence, we are able to select suitable anchor items to describe the proficiency levels. Since the response probability is set at 62% and anchor items may be selected from the bottom of a level, these items are located 0.5 logits below the start of the level in the Wright map. For instance, item '3.99.1' is an anchor item for Level I (see Figure 3). In the following section, we will describe the proficiency level model on the content level. The metric is standardised points (STP) with M = 500 and SD = 100.

Below Level I, students are not able to solve basic IL tasks, such as covered by item '3.99.1' (see Figure 4). Of the students in our sample, 17% are below Level I.

- *Level I (440 to 517 STP):* Students show a basic knowledge of the information process. They are able to avoid serious mistakes. Students on this level may search for information on their own, but have to rely on support to achieve meaningful outcomes. An anchor item for Level I can be found in Figure 3. Of the students in our sample, 39% are on Level I.
- *Level II (517 to 594 STP):* Students demonstrate a comprehension of the information process. They are able to select appropriate information sources, extract information, assess reliability as well as relevance, and adequately present their results. For instance, they rely on newsletters if they are interested in the latest news and developments. They also show a basic understanding of copyright issues, for example, concerning the use of social media content. However, they may not be able to carry out complex information processes. An anchor item for Level II is depicted in Figure 5. Of the students in our sample, 32% are on Level II.
- *Level III (above 594 STP):* Students are able to analyse problems in the information process. This means they may independently perform complex information processes. They use very suitable information sources, extract information from complex sources, and choose very suitable ways for presenting their results. An anchor item can be found in Figure 6. Of the students in our sample, 12% are on Level III. We do not distinguish a fourth level. Level IV would comprise, in line with the taxonomy of Marzano and Kendall (2007), finding novel approaches to tackling information processes. However, we do not regard this as necessary for upper secondary students. Moreover, only one student in our sample would be on this level.

Figure 4 Anchor item for Level I

Item 3.99.1
A friend of yours would like to inform herself about a certain topic. She decides to post her questions in a student forum.
How do you rate the following action of your friend within the forum?

- After half an hour she still has not received an answer; she writes another post to draw attention to her question.

Correct answer: rather inappropriate

Figure 5 Anchor item for Level II

Item 5.138.2
Imagine you have to give a presentation in collaboration with three of your classmates on a current topic (e.g., presidential elections in the USA). You decide that everyone will collect information on your topic for the next lesson.
In the next lesson, you will present your results to each other. Your friends choose very different approaches.
How would you assess the individual procedures for organizing information, which should serve as a basis for further presentation?

- A friend has created a summary in her own words from all the information she has found.

Correct answer: entirely appropriate

Figure 6 Anchor item for Level III

Item 1.59
Imagine, your best friend calls you and asks you: "Could you please help me with my research on the Swiss Federal President, since I have to give a presentation about him tomorrow?"
What would you most likely say to help your friend with her problem? Please tick only one answer.

- "What have you already found out?"
- "I'll find out something about the Federal President on the internet immediately and then call you back."
- "Does the Federal President have a profile on a social media platform, for instance Instagram or Twitter?"
- "What exactly is your presentation about?" *(correct answer)*
- "Let's just look together for information on the internet."

Table 3 summarises the proficiency level cut-offs and percentages of students at each level.

Table 3 Proficiency level cut-offs and percentages at each level

Level	*Logit*		*Standard points (STP)*		*Percentages*
	Equal or higher than	*Below*	*Equal or higher than*	*Below*	
Below I		–0.375		440	17
I	–0.375	0.125	440	517	39
II	+0.125	+0.625	517	594	32
III	+0.625		594		12

4.3 Instructional sensitivity of OIL and SIL (RQ3)

Based on the residual analyses, we identified two potential outliers that we inspected. The first person's responses did not show any variance. For the second person, there was no variance from about halfway through the test onwards. We decided to exclude these two observations for the regression analysis. Results for the regression analysis are depicted in Table 4.

Table 4 Influence of control variables and instruction on OIL and SIL (N = 430)

Variable	*(1)*	*(2)*	*(3)*
Intercept	327.06*** (43.16)/ 356.37*** (56.99)	298.18*** (42.25)/ 350.22*** (60.54)	275.73*** (41.97)/ 329.21*** (56.56)
BS_w_Abi	11.32 (16.06)/ 11.82 (20.72)	13.24 (11.63)/ 12.23 (21.21)	15.62 (12.14)/ 13.13 (20.73)
BS_o_Abi	–28.25 (24.84)/ 9.98 (18.26)	–34.73 (20.41)/ 8.60 (17.05)	–29.53 (19.15)/ 9.85 (17.70)
AT	–33.51 (23.55) –1.67 (19.21)	–26.46 (14.68)/ –0.27 (17.18)	–30.47 (15.60)/ 0.15 (18.34)
LI	–26.29 (20.85)/ 8.26 (25.98)	–31.45* (14.68)/ 7.16 (27.08)	–28.47 (14.51)/ 6.80 (25.88)
CH	–14.89 (15.80)/ –36.18 (19.04)	–13.70 (9.60)/ –35.93 (18.85)	–14.60 (10.58)/ –37.23* (18.26)
Female	19.57* (8.53)/ –2.04 (11.81)	19.97* (8.60)/ –1.94 (11.87)	17.11* (8.41)/ –2.68 (11.45)
Test motivation	190.06*** (42.18)/ 167.27** (55.47)	194.38*** (42.47)/ 168.19** (55.90)	192.55*** (40.56)/ 166.66** (54.53)
Age	0.32 (2.67)/ 2.39 (3.57)	–2.76 (3.14)/ –3.04 (3.50)	–2.81 (3.65)/ 3.11 (3.68)
Grade		9.71** (3.49)/ 2.07 (4.83)	9.51** (3.58)/ 2.14 (4.67)
Wiki_school			12.76 (13.05)/ 22.21 (21.92)
Wiki_lessons			20.09* (8.09)/ 4.19 (10.68)
R^2, R^2_{adj}	.29, .27/ .26, .24	.30, .28/ .26, .24	.31, .29/ .27, .25

Notes: Two students from the sample (N = 432) were excluded as outliers. Cluster-robust standard errors with bias-reduced linearisation (Pustejovsky and Tipton, 2018) in parentheses; clustering by school class. Collinearity: VIF < 3.3 for all variables. Abbreviations: BS_w_Abi = vocational school with the possibility of obtaining the Abitur, BS_o_Abi = vocational school without the possibility of obtaining the Abitur, reference category = high school. AT = Austria, CH = Switzerland, LI = Liechtenstein, Germany as reference category. Wiki_school = use of Wikipedia for school, Wiki_lessons = use of Wikipedia in class. *p < .05, **p < .01 and ***p < .001 (two-tailed).

Model 1 considers all control variables. Women show a significantly higher OIL than men (p = .02). In terms of SIL, the difference is not significant (p = .86). A lack of test motivation affects both OIL and SIL negatively (p < .01). Model 2 takes into account the grade as a variable of interest. The influence of the grade on OIL is positive (p < .01): each additional school year leads to an expected increase in OIL of 9.71 STP, corresponding to 0.1 standard deviations. The influence of the grade on SIL is not significant (p = .67). Model 3 considers the use of Wikipedia during the class. It has a positive influence on OIL (p = .01): using Wikipedia in class is associated with a 0.2 standard deviation higher OIL. The influence of the grade on OIL decreases slightly in

the transition from models 2 to 3 (9.71 vs. 9.51 STP). Concerning SIL, the influence of the use of Wikipedia in class is not significant ($p = .69$).

4.4 Relationship between OIL and SIL (RQ4)

The correlation between OIL and SIL is .52 ($t = 12.58$, $df = 428$, $p < .01$). The average difference between OIL and SIL is –11.56 for men (SD = 107.11) and 10.25 for women (SD = 88.39). There is a significant positive influence of the gender 'female' on the difference between OIL and SIL ($\beta = 18.93$, $p = .04$) even when considering all (non-significant) context and control variables.

5 Summary and discussion

5.1 Rasch scalability (RQ1)

After the exclusion of misfitting items, we achieved Rasch scalability. In particular, establishing person homogeneity or fairness of the test instrument is crucial. Only in this case, observed IL differences in subgroups can validly be attributed to latent IL differences. This is important if general proficiency levels should be set (RQ2) and the influence of context variables on IL is investigated (RQ3). Ensuring person homogeneity necessitated numerous (anticipated) item exclusions (OIL: 54%, SIL: 59%). In most cases, we observed a DIF with respect to the country of origin. We inspected all excluded items, but could not always find a plausible reason on the content level. In the case of OIL, content validity is still ensured: all seven facets are represented by at least two PCM items. In the case of SIL, there is no item left for the second facet. Here, new items should be constructed and validated. The reliability of the test instruments is good (OIL: EAP/PV = .84, WLE = .88) or satisfactory (SIL: EAP/PV = .75, WLE = .77). The OIL test shows a decent fit between item difficulty and ability of the test takers. The instrument is therefore complex enough for our primary target group: upper secondary learners.

5.2 Proficiency level model (RQ2)

By means of our proficiency level model, we are able to provide evidence for severe deficits among the sampled upper secondary students; 17% were not able to perform basic tasks of the information process even partially correctly and were prone to making serious mistakes. Another 39% might need support to master typical information processes. With this in mind, it has to be considered that the test takers are not children but, on average, 18 years old. We conclude that instructional measures are very necessary to tackle this issue.

5.3 Influence of instruction on OIL and SIL (RQ3)

Evidence for instructional sensitivity can only be provided for OIL. An additional school year leads to an expected increase of OIL by 9.51 STP. This increase is rather small: the width of the proficiency level equals 77 STP. Hence, on average, it would take about eight school years to increase a student's IL from the bottom to the top of a proficiency

level. However, the results may be consistent with the criticised inadequate curricular integration of IL (Jones-Kavalier and Flannigan, 2006). In this context, the lack of an influence of age on OIL is interesting, even if the grade is not taken into account (model 1 vs. 2). OIL does not seem to improve automatically with increasing life experience.

The relationship between the uses of Wikipedia at home for school purposes and OIL is not significant, but the relationship between the use of Wikipedia in class and OIL is important. This result might be regarded as evidence for instructional sensitivity: the appropriate use of Wikipedia requires an instruction that is likely to be given in class. A general use of Wikipedia without instruction has no significant influence on OIL. However, the observed relationship between the use of Wikipedia in class and OIL should not be interpreted as a causal relationship. An alternative explanation would be that IL-affine teachers particularly promote IL, and at the same time, use Wikipedia more often than non-IL-affine teachers.

Concerning SIL, a possible explanation for the lack of influence of grade and the use of Wikipedia in class on the SIL score would be that 'digital natives' had already developed a relatively persistent view about their IL before the ninth grade. They may rarely adjust this view with increasing age or instruction. Against this background, the SIL instrument might be unsuitable for evaluating IL instruction.

Drawing on all three models, women show a higher OIL than do men. This is contrary to the results of Large et al. (2002), Roy and Chi (2003), and Tsai (2009), all of whom found higher OIL for male test takers. There are no gender differences with regard to SIL. As expected, test motivation substantially influences both OIL and SIL. The school type as proxy for cognitive ability and interest structure has no significant influence on OIL and SIL. This result might suggest that instruction rather than cognitive ability or interest can explain IL. As a robustness check, we carried out multigroup analyses in terms of the country of origin. However, for the variables of interest, there are no significantly different path coefficients.

5.4 Relationship between OIL and SIL (RQ4)

The correlation between OIL and SIL of .52 indicates a satisfactory construct validity because it has to be considered that different methods were used to capture IL (Bühner, 2011): performance test vs. self-assessment. Overall, digital natives have difficulties in accurately assessing their own IL. Against this background, the (sole) use of self-assessments to diagnose IL seems problematic. Since, by means of our test, IL can be measured objectively in a reasonable amount of time, this kind of measurement might be the suitable approach in most contexts.

As the rather low correlation between OIL and SIL shows, it might be difficult for both sexes to assess their own IL accurately. Women underestimate their IL while men overestimate themselves.

6 Implications and conclusions, limitations and outlook

The contribution of this paper is twofold: on the one hand, by presenting and validating an IL model, we aim at creating a basis for pedagogical concepts to promote IL in order to close a gap in the empirical curriculum and teaching research. On the other hand, the development of a standardised measuring instrument for IL in the context of upper secondary education represents a novelty in IL research. With the help of this instrument, the effectiveness and goal achievement of pedagogical interventions can be evaluated. The validated instruments for capturing IL can be used as a sound basis for the development of cross-school educational standards. Both instruments are available from the authors upon request. Furthermore, we emphasise the notion that the concept of 'digital natives' (Prensky, 2001) should be discarded as a normative basis for integrating educational technology in schools.

As an alternative, the concept of 'digital residents' and 'digital visitors' coined by White and Le Cornu (2011) could be revealing. This approach suggests a range of behaviours to provide a more accurate and useful framework for describing and analysing people's motivations for using technology and information. Phipps and Lanclos (2017) also take up this concept and expand it with implementation notes to contrast with the reductionist model of digital natives and immigrants. This is also an interesting approach for information competences, as it is not about labelling people with their fundamental identities, but about identifying different behavioural patterns. We plan to integrate this idea into our conceptional work in order to establish a solid normative orientation about how to understand and develop IL in education.

Furthermore, as a practical implication, we can systematically and continually collaborate together with our partner schools in order to update the competence facets. For example, the phenomenon of fake news has become a permanent fact with new technological progress.

Needless to say, our research faces limitations. SIL items are no longer included for the second competence facet 'information sources'. This is critical in terms of content validity. Hence, new items have to be constructed and validated. These items should be more difficult compared to the available items. Interpretations at the level of the seven IL facets are currently not possible due to insufficient reliability. A longitudinal design using a latent growth model may be suitable for investigating OIL development. On this basis, various development paths could be described and the influence of personal and context variables investigated.

Most importantly, instructional measures are necessary to foster IL systematically. Our 7i model may act as a basis for creating instructional designs. For instance, Seufert et al. (2019) relied on the 7i model to develop a massive open online course ('iMOOC': https://i-mooc.ch/). Since the OIL and SIL instruments also draw on the 7i model, an alignment between curriculum, instruction, and assessment might be achieved and learning gains could be diagnosed in a pre-post design.

Along with instruction, we may also address students' poor ability to assess themselves. This might foster student willingness to study IL-related content. Moreover, self-reflection is regarded as an important prerequisite for the lifelong learning process (Eyler, 2002).

References

Adams, R.J. (2005) 'Reliability as a measurement design effect', *Studies in Educational Evaluation*, Vol. 31, Nos. 2/3, pp.162–172.

American Library Association (ALA) (1989) *Presidential Committee on Information Literacy. Final Report*, American Library Association, Chicago.

Andersen, E.B. (1973) 'A goodness of fit test for the Rasch model', *Psychometrika*, Vol. 38, No. 1, pp.123–140.

Argelagós, E. and Pifarré, M. (2012) 'Improving information problem solving skills in secondary education through embedded instruction', *Computers in Human Behavior*, Vol. 28, No. 2, pp.515–526.

Association of College and Research Libraries (ACRL) (2000) *Information Literacy Competency Standards for Higher Education* [online] http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf (accessed 27 November 2014).

Baethge, M., Achtenhagen, F., Arends, L., Babic, E., Baethge-Kinsky, V. and Weber, S. (2006) *PISA-VET. A Feasibility Study*, Stuttgart, Steiner.

Balceris, M. (2011) *Medien- und Informationskompetenz: Modellierung und Messung von Informationskompetenz bei Schülern*, Dissertation, Freie Universität Berlin.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M. and Rumble, M. (2012) 'Defining twenty-first century skills', in Griffin, P., McGaw, B. and Care, E. (Eds.): *Assessment and Teaching of 21st Century Skills*, pp.17–66, Springer, New York.

Braun, J. (1997) 'Past, possibilities, and potholes on the information superhighway', *Social Education*, Vol. 61, No. 1, pp.149–153.

Bruce, C.S. (1992) *Developing Students' Library Research Skills*, HERDSA, Kensington, New South Wales.

Bruce, C.S. (1999) 'Workplace experiences of information literacy', *International Journal of Information Management*, Vol. 19, No. 1, pp.33–47.

Bühner, M. (2011) *Einführung in die Test- und Fragebogenkonstruktion*, 3rd ed., Pearson Studium, Munich.

Chaka, C. (2009) 'The changing faces of literacies: information and media literacy in the context of higher education', in Leaning, M. (Ed.): *Education, Practice and Pedagogy*, pp.31–50, Informing Science Press, California.

Cope, B. and Kalantzis, M. (2016) 'Big data comes to school: implications for learning, assessment, and research', *AERA Open*, Vol. 2, No. 2, pp.1–19.

Eyler, J. (2002) 'Reflection: linking service and learning – linking students and communities', *Journal of Social Issues*, Vol. 58, No. 3, pp.517–534.

Finch, H. (2008) 'Estimation of item response theory parameters in the presence of missing data', *Journal of Educational Measurement*, Vol. 45, No. 3, pp.225–245.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T. and Duckworth, D. (2019) *Preparing for Life in a Digital World. International Report. Findings from the IEA's International Computer and Information Literacy Study (ICILS) 2018*, International Association for the Evaluation of Educational Achievement (IEA), Amsterdam.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T. and Gebhardt, E. (2014) *Preparing for Life in a Digital Age. The IEA International Computer and Information Literacy Study. International Report*, Springer Open, Dordrecht.

Fraillon, J., Schulz, W., Friedman, T., Ainley, J. and Gebhardt, E. (2015) *ICILS 2013 Technical Report*, International Association for the Evaluation of Educational Achievement (IEA), Amsterdam.

Gapski, H. and Tekster, T. (2009) *Informationskompetenz in Deutschland – Überblick zum Stand der Fachdiskussion und Zusammenstellung von Literaturangaben, Projekten und Materialien zu einzelnen Zielgruppen*, Landesanstalt für Medien Nordrhein-Westfalen, Düsseldorf.

Glas, C. and Verhelst, N.D. (1995) 'Testing the Rasch model', in Fischer, G.H. and Molenaar, I.W. (Eds.): *Rasch Models: Foundations, Recent Developments, and Applications*, pp.69–95, Springer, New York, New York.

Hartig, J. and Frey, A. (2013) 'Sind Modelle der Item-Response-Theorie (IRT) das 'Mittel der Wahl' für die Modellierung von Kompetenzen?', *Zeitschrift für Erziehungswissenschaft*, Vol. 16, No. 18, pp.47–51.

Jones-Kavalier, B.R. and Flannigan, S.L. (2006) 'Connecting the digital. Literacy of the 21st century', *Educause Quarterly*, Vol. 29, No. 2, pp.8–10.

Kirschner, P.A. and van Merriënboer, J.J.G. (2013) 'Do learners really know best? Urban legends in education', *Educational Psychologist*, Vol. 48, No. 3, pp.169–183.

Koller, I., Alexandrowicz, R. and Hatzinger, R. (2012) *Das Rasch Modell in der Praxis*, UTB, Stuttgart.

Large, A., Beheshti, J. and Rahman, T. (2002) 'Gender differences in collaborative web searching behaviour. An elementary school study', *Information Processing and Management*, Vol. 38, No. 3, pp.427–443.

Livingstone, S., Van Couvering, E. and Thumim, N. (2005) *Adult Media Literacy: A Review of the Research Literature*, Office of Communications, London.

Lundeberg, M.A., Fox, P.W. and Punćcohaŕ, J. (1994) 'Highly confident but wrong: gender differences and similarities in confidence judgments', *Journal of Educational Psychology*, Vol. 86, No. 1, pp.114–121.

Mair, P., Hatzinger, R., Maier, M.J. and Rusch, T. (2018) *Package 'eRm'* [online] https://cran.r-project.org/web/packages/eRm/eRm.pdf (accessed 8 December 2018).

Marshall, R.K. (2006) 'An instrument to measure information competency', *Journal of Literacy and Technology*, Vol. 6, No. 1, pp.1–27.

Marzano, R.J. and Kendall, J.S. (2007) *The New Taxonomy of Educational Objectives*, Corwin Press, Thousand Oaks.

Marzano, R.J. and Kendall, J.S. (2008) *Designing and Assessing Educational Objectives: Applying the New Taxonomy*, Corwin Press, Thousand Oaks.

Masters, G.N. (1982) 'A Rasch model for partial credit scoring', *Psychometrika*, Vol. 47, No. 2, pp.149–174.

McDaniel, M.A. and Nguyen, N.T. (2001) 'Situational judgment tests: a review of practice and constructs assessed', *International Journal of Selection*, Vol. 9, Nos. 1/2, pp.103–113.

McGrew, S. (2020) 'Learning to evaluate: an intervention in civic online reasoning', *Computers & Education*, Vol. 145, No. 2 [online] https://doi.org/10.1016/j.compedu.2019.103711.

McNeish, D., Stapleton, L.M. and Silverman, R.D. (2017) 'On the unnecessary ubiquity of hierarchical linear modeling', *Psychological Methods*, Vol. 22, No. 1, pp.114–140.

Miller, C. and Bartlett, J. (2012) ''Digital fluency': towards young people's critical use of the internet', *Journal of Information Literacy*, Vol. 6, No. 2, pp.35–55.

Naumann, A., Rieser, S., Musow, S., Hochweber, J. and Hartig, J. (2019) 'Sensitivity of test items to teaching quality', *Learning and Instruction*, Vol. 60, pp.41–53.

OECD (2014) *PISA 2012. Technical Report*, OECD Publishing, Paris.

OECD (2015) *Students, Computers and Learning. Making the Connection*, OECD Publishing, Paris.

OECD (2017) *PISA 2015. Technical Report*, OECD Publishing, Paris.

Palfrey, J. and Gasser, U. (2008) *Born Digital. Understanding the First Generation of Digital Natives*, Basic Books, New York.

Pellegrino, J.W. (2010) *The Design of an Assessment System for the Race to the Top: A Learning Sciences Perspective on Issues of Growth and Measurement*, Educational Testing Service, Princeton.

Phipps, L. and Lanclos, D. (2017) 'Leading with digital in an age of supercomplexity', *Irish Journal of Technology Enhanced Learning*, Vol. 2, No. 1, DOI: 10.22554/ijtel.v2i2.22.

Prensky, M. (2001) 'Digital natives, digital immigrants', *On the Horizon*, Vol. 9, No. 5, pp.1–6.

Pustejovsky, J.E. and Tipton, E. (2018) 'Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models', *Journal of Business & Economic Statistics*, Vol. 36, No. 4, pp.672–683.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna.

Rader, H. (2003) 'Information literacy – a global perspective', in Martin, A. and Rader, H. (Eds.): *Information and IT Literacy: Enabling Learning in the 21st Century*, pp.24–42, Facet Publishing, London.

Reise, S.P. (2012) 'The rediscovery of bifactor measurement models', *Multivariate Behavioral Research*, Vol. 47, No. 5, pp.667–696.

Robitzsch, A., Kiefer, T. and Wu, M. (2018) *TAM: Test Analysis Modules (Version 3.0-21)* [online] http://cran.r-project.org/web/packages/TAM/index.html (accessed 11 November 2018).

Rosman, T., Mayer, A-K. and Krampen, G. (2015) 'Combining self-assessments and achievement tests in information literacy assessment. Empirical results and recommendations for practice', *Assessment and Evaluation in Higher Education*, Vol. 40, No. 5, pp.740–754.

Rosman, T., Mayer, A-K. and Krampen, G. (2016) 'Measuring psychology students' information-seeking skills in a situational judgment test format', *European Journal of Psychological Assessment*, Vol. 32, No. 3, pp.220–229.

Rost, J. (2004) *Lehrbuch Testtheorie – Testkonstruktion*, 2nd ed., Huber, Bern.

Roy, M. and Chi, M.T.H. (2003) 'Gender differences in patterns of searching the web', *Journal of Educational Computing Research*, Vol. 29, No. 3, pp.335–348.

Seufert, S., Guggemos, J., Moser, L. and Sonderegger, S. (2019) 'Developing a MOOC to foster information literacy (IL) by means of a conjecture map', in Uden, L., Liberona, D., Sanchez, G. and Rodríguez-González, S. (Eds.): *Communications in Computer and Information Science. Learning Technology for Education Challenges*, pp.202–213, Springer, Cham.

Stanoevska-Slabeva, K., Müller, S., Seufert, S. and Scheffler, N. (2015) 'Towards modeling and measuring information literacy in secondary education', *63rd International Conference on Information Systems*, Fort Worth.

State Secretariat for Education, Research and Innovation (SERI) (2017) *Herausforderungen der Digitalisierung für Bildung und Forschung in der Schweiz*, SERI, Bern.

Tapscott, D. (1997) *Growing Up Digital. The Rise of the Net Generation*, McGraw Hill, New York.

Taylor, A. and Dalal, H.A. (2017) 'Gender and information literacy. Evaluation of gender differences in a student survey of information sources', *College & Research Libraries*, Vol. 78, No. 1, pp.90–113.

Tsai, M. (2009) 'Online information searching strategy inventory (OISSI). A quick version and a complete version', *Computers and Education*, Vol. 53, No. 2, pp.473–483.

UNESCO (2003) *The Prague Declaration: Towards an Information Society* [online] http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/PragueDeclaration.pdf (accessed 8 December 2018).

Vrieze, S.I. (2012) 'Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)', *Psychological Methods*, Vol. 17, No. 2, pp.228–243.

White, D.S. and Le Cornu, A. (2011) 'Visitors and residents: a new typology for online engagement', *First Monday*, Vol. 16, No. 9 [online] https://firstmonday.org/ojs/index.php/fm/article/view/3171/3049.

Wilson, M. (2005) *Constructing Measures: An Item Response Modelling Approach*, Lawrence Erlbaum, Mahwah.

Wise, S.L. and DeMars, C.E. (2005) 'Low examinee effort in low-stakes assessment: problems and potential solutions', *Educational Assessment*, Vol. 10, No. 1, pp.1–17.

Wise, S.L. and Kong, X. (2005) 'Response time effort: a new measure of examinee motivation in computer-based tests', *Applied Measurement in Education*, Vol. 18, No. 2, pp.163–183.

Notes

1 Only the year of birth was requested.