Original Manuscript

Can Overclaiming Technique Improve Self-Assessment Tools for Digital Competence? The Case of DigCompSat

Social Science Computer Review
2022, Vol. 0(0) 1–24

DOI: 10.1177/08944393221117269
journals.sagepub.com/home/ssc
SAGE

Marek Muszyński[1], Artur Pokropek[1], Jonatan Castaño-Muñoz[2], and Riina Vuorikari[3]

Abstract

Digital competence is crucial for living, working and participating in current societies. Despite its huge importance, objective measurement tools for it are scarce due to its developmental difficulty. Self-assessment of digital competence seems a promising proxy of objective tests, and it additionally offers the possibility for surveying otherwise unmeasurable constructs such as attitudes and beliefs. However, self-assessment tools are burdened with validity problems, most notably response biases such as overly positive descriptions, overclaiming or careless and insufficient effort responding. In this paper, we investigate how these problems can be mitigated by using the overclaiming technique, a technique that identifies and corrects the bias variance in self-assessments. Our main result was that the use of the overclaiming technique can lead to higher reliability and validity of digital competence self-assessment tools, especially for short scales. Moreover, it allows for correcting additional spurious variance in comparison with careless responding indexes, which allows the use of both these techniques in parallel to increase the quality of data. Our results are important in providing advances in enhanced information on digital competence that can result in better lifelong learning decisions when used at the individual level and in better policy-making decisions when used at the aggregate level.

Keywords

overclaiming, overclaiming technique, digital competence, digital skills, self-assessment, DigCompSat

[1]Institute of Philosophy and Sociology, Polish Academy of Sciences, Warsaw, Poland
[2]Department of Teaching and Educational Organization, Faculty of Education, University of Seville, Seville, Spain
[3]European Commission, Joint Research Centre (JRC-Seville), Sevilla, Spain

Corresponding Author:
Marek Muszyński, Institute of Philosophy and Sociology, Polish Academy of Sciences, Nowy Świat 72, Warsaw 00-330, Poland.
Email: marek.a.muszynski@gmail.com

The confident and responsible use of digital technologies is becoming increasingly important. Digital competence is considered a key element for working and living in contemporary societies (European Council, 2018). Having access to digital technologies and the competence to use them well provide a series of benefits related to a range of online learning, employment, networking and informational opportunities (van Deursen et al., 2014). The European Council has included digital competence as one of the key competence for personal fulfilment and social inclusion (European Council, 2018; Helsper & Reisdorf, 2017), and many European Union (EU) educational systems are already promoting its early acquisition (European Commission/EACEA/Eurydice, 2019).

Digital competence assessment in the form of objective tests is a difficult and costly endeavour. Measuring online information searching skills seems easy, but in fact it requires building from scratch a dedicated information technology (IT) system that can track test takers' behaviours, record outputs and grade performances (Hargittai & Hsieh, 2012). This should be done in a closed digital environment that allows for results standardisation and does not favour particular operational systems, browsers, search engines or devices. This seemingly simple task is so difficult that even though we have elaborate testing frameworks today (Deursen & Van Dijk, 2010), actual operating tests that can measure online search skills in a valid and reliable way are still scarce (Haddon et al., 2020; Hargittai & Hsieh, 2012; Helsper et al., 2020; ITU, 2020).

One common and approachable way of measuring digital competence is to simply ask participants to self-assess themselves (Hargittai, 2005, 2009; ITU, 2020; Monoi et al., 2005). However, many studies, including some in other domains, show that competence self-assessment might be problematic (Litt, 2013; Paulhus, 2002). In some settings, self-assessment was a reliable indicator of objective skills (Johansson, 2013), but in others, it was more a measure of what level of competence respondents would like to have rather than a reliable source of information on their actual competence (Baxter & Norman, 2011). Brown and Harris (2014) provided a good summary, showing that correlations between student self-ratings and cognitive tests varied significantly from $r = .2$ to $r = .8$, indicating that some of the self-assessments could be reasonably valid, whereas others provided only a vague sense of objective abilities.

The literature on information and communications technology (ICT) skills self-assessment shows that this domain is not free of these problems either (e.g. Litt, 2013; Palczyńska & Rynko, 2020; Vonkova et al., 2021) and that respondents are likely to yield biased responses, that is, they indicate what they think they should know (social desirability bias) or what they think they know rather than what they actually do know (positivity bias). Overall, there is evidence that approximately 30% of participants overestimate their performance and 10%–15% underestimate it (John & Robins, 1994; Randall & Fernandes, 1991; Robins & Beer, 2001). These biases remain a key concern for measurements based on self-reports because they can have direct consequences for respondents themselves (e.g. when self-assessments are to be used as formative assessments [at the individual level] or as valid indicators needed to develop evidence-based policies [at the aggregate/collective level]).

A solution to these problems could be the overclaiming technique (OCT). It was initially used by Phillips and Clancy (1972) and further elaborated by Paulhus et al. (2003), who proposed to treat overclaiming bias as a general indicator of self-favouring (positivity) bias. A typical overclaiming scale entails rating the familiarity of a list of items (e.g. computer skills, software and devices) with existent terms (*targets* or *reals*) that are complemented (in various proportions, but most often 20%) by nonexistent terms (*lures* or *foils*). Controlling for overclaiming aims to account for motivated misrepresentation in self-reports (Bing et al., 2011; Paulhus, 2012).

In this paper, we explored the possibility of improving the reliability and validity of digital competence self-assessments by using the OCT. We used the empirical data collected by the European Commission's Joint Research Centre to pilot and implement a self-assessment online instrument, the DigCompSat (Digital Competence Self-Assessment Tool; Clifford et al., 2020), which is based on The Digital Competence Framework for Citizens (DigComp; Carretero et al., 2017; Vuorikari et al., 2022). The DigCompSat is a good example of a tool that provides information on digital competence that would be very hard to collect using skill tests. This tool also serves the goal of achieving lifelong learning, having a better understanding of one's own digital competence level and, eventually, taking actions to improve the level.

The Aim of the Study

The overall aim of the study was to explore the adequacy of the OCT in improving the interpretation of the information collected by digital competence self-evaluation tools. First, we evaluated the internal consistency of the overclaiming scale. We assumed that the tendency to overclaim was constant within the questionnaire and therefore should result in a reliable, essentially one-dimensional scale. Second, we investigated the relations between the overclaiming scale, careless responding indices and responding time. The aim of this analysis was to check if the OCT scale accounted for spurious variance over and above the careless responding indexes (e.g. by measuring the self-enhancement tendency [positivity bias and self-favouring bias]). Third, we checked to see whether controlling for overclaiming could increase the reliability and validity of the target scale, and we compared its performance against the careless indices and response time as other indices of spurious variance. Finally, we looked at the correlates of overclaiming, thus checking its convergent validity and identifying groups more prone to provide biased responses when assessing their digital competence.

Literature Review: The Positivity Bias Problem in Digital Competence Self-Assessment

One of the main problems in self-assessment data is positivity bias. Participants not only assess themselves too favourably but also claim to have traits they do not have or perform actions they do not perform (e.g. Helsper et al., 2020). For example, Hargittai (2009) showed that approximately 20% of participants claimed some understanding of made-up digital competence terms (i.e. foils) such as *proxypod* or *filtibly*. Similarly, in the Polish edition of the Programme for the International Assessment of Adult Competencies (PIAAC) study, approximately 20% of participants declared that they frequently used computers in a self-descriptive survey but then failed a very simple computer proficiency test (Burski et al., 2013; Palczyńska & Rynko, 2020). In general, a number of studies employing ICT skills self-assessment in student samples reported significant overestimations in various samples (Grant et al., 2009; Maderick et al., 2016;McCourt Larres et al., 2003; Merritt et al., 2005; Vonkova et al., 2021; see also Mabe & West, 1982; Zell & Krizan, 2014).

These and other results (Helsper et al., 2020: pp. 17–22) show that self-reports on digital competence cannot be fully trusted based on face value only and need additional indices that gauge their validity. In the next section, we present how OCT can be used to boost the quality of data in tools similar to the DigCompSat.

Overclaiming Technique and Data Quality Improvement

Research results point to the usefulness of OCT in controlling for spurious variance in cross-country research (He & van de Vijver, 2016; Vonkova et al., 2018) or in acting as a suppressor or moderator in criterion-related validity studies (e.g. Anderson et al., 1984; Muszyński, 2020: p. 145). Despite the fact that the boost in criterion-related validity was much larger for cross-country coefficients than within-country coefficients, the difference between the uncorrected (r = .40) and overclaiming-corrected (r = .49) correlation was deemed practically significant (Vonkova et al., 2018; see similar results in He & Van de Vijver, 2016). In the field of digital competence, optimistic results were recently presented by Vonkova et al. (2021), who reported a change from $r = .37$ (unadjusted) to $r = .53$ (adjusted) for the criterion-related validity coefficient; this was a correlation between self-reported ICT skills and skills measured by objective assessment (cognitive test) for a representative sample of Czech high school students. Muszyński (2020, pp. 145–148) also showed that OCT adjustment can lead to increased regression coefficients and boosted R^2 indices in a math skills self-assessment.

The above results suggest that the OCT can enhance the criterion-related validity of self-reports. However, this technique can have one more advantage, namely, it can also serve as a measure in addition to a careless responding bias that also lowers the measurement quality in self-reports and controlling which can lead to similar effects as controlling overclaiming bias.

Other Techniques for Data Quality Improvement

Careless/insufficient effort responding (C/IER) is a common framework under which biased data can be identified and corrected (Curran, 2016). There is ample evidence that invalid data provided by unmotivated, careless, inattentive or otherwise incompetent participants (Johnson, 2005) can lead to a serious distortion of the most popular psychometric indices and can also affect analytic conclusions (Maniaci & Rogge, 2014).

Arias et al. (2020) presented a comprehensive series of analyses showing how only a small percentage of careless respondents (<10% of the sample) can lead to lowered reliability, a distorted factorial structure (i.e. spurious cross-loadings, emergence of method factors and missing substantial factors), lowered factor loadings and a model misfit. These results are widely confirmed by other studies (e.g. DeSimone et al., 2018; Kam, 2019; Kam & Meyer, 2015). Moreover, the presence of C/IER participants in the data can lead to distorted convergence validity coefficients and lowered effect sizes in experiments (Brühlmann et al., 2020; Credé, 2010; Huang et al., 2015). Response times are considered one of the useful C/IER indicators, and typically, very fast responses indicate insufficient effort (Curran, 2016).

Overclaiming, Careless Responding and Speed of Responding

The prediction that careless/insufficient effort responding (C/IER) may be related to the emergence of overclaiming was voiced by many researchers (e.g. Bensch et al., 2019; Bing et al., 2011; Dunlop et al., 2020), but the empirical support for such claims is rather scarce to date. However, Barber et al. (2013) showed a substantial correlation between OCT bias and C/IER ($r = .45$ and $r = .51$, respectively). Similarly, Ludeke and Makransky (2016) found that OCT bias was related to the Mahalanobis distance ($r = .15$) and person fit measures ($r = .45$) and that it was unrelated to contralogical survey errors. All C/IER measures used in this study were negatively correlated with OCT accuracy (i.e. a preference for endorsing reals over foils), further corroborating the hypothesis of relationship between C/IER and OCT scores. Calsyn and Winter (1999) used a similar measure of contralogical survey errors and also failed to find any relation between it and OCT bias, but they found a small negative correlation between survey errors and OCT accuracy ($r = -.16$). Muszyński (2020) evidenced that part of the overclaiming variance is in fact related to the C/IER but that careless responding alone cannot explain the overclaiming scores and their relations to self-reports (pp. 161–166).

In the present study, a range of C/IER indices representing straightlining (DeSimone et al., 2018) were used to investigate their relation with OCT scores and to check whether the DigCompSat data quality could be enhanced by controlling for both OCT and C/IER. Moreover, we used the survey completion time (CT) as a measure of careless responding, in line with findings linking (too) fast responding with this response bias (e.g. Curran, 2016; Leiner, 2019). To the best of the authors' knowledge, survey paradata (e.g. response times) were not used to date in scales using overclaiming. We introduced this analysis in order to better understand overclaiming scores and their relation to response speed.

Overclaiming and Sociodemographic Covariates

In order to broaden our understanding of self-enhancement in reporting ICT skills, we analysed the relationship between OCT scores and a set of key sociodemographic covariates.

Gender. Many studies reported no gender differences in OCT (Calsyn & Winter, 1999; Feeney & Goffin, 2015; Franzen & Mader, 2019; Mesmer-Magnus et al., 2006; Paulhus & Dubois, 2014; Paulhus et al., 2003). However, other studies showed that male participants yield more better-than-average bias regarding intelligence self-assessment, they claimed opinions on fictitious issues more often than women and were more prone to scholastic cheating than female students (Bishop et al., 1986; Zhang et al., 2019).

This seemingly discrepant pattern of results could be expected if the specific content used to construct OCT items is taken into consideration, namely, the items' relation to agentic and communal values and their gender-desirability, importance and identity relation (Paulhus, 2002). Paulhus and John (1998) stated that women should be more susceptible to communal bias, whereas men should be more prone to agentic bias. The self-assessment of academic skills and knowledge is definitely a type of agentic content (Paulhus & Trapnell, 2008; Paunonen, 2016); hence, it should predict a higher amount of overclaiming by male participants. However, not all academic skills are considered 'boy things' as is the case with math abilities (Cipora et al., 2018). Palczyńska and Rynko (2020), who analysed PIAAC data, found more overclaiming among men in comparison with women in math-related tasks but only very small amounts of overclaiming with ICT-related tasks. Vonkova et al. (2021) showed that male participants overclaimed their digital competence more than females. In brief, gender was predicted to be related to overclaiming only in certain domains related to gender stereotypes and social norms in a given society (Borgonovi & Pokropek, 2019) or in situations where the construct measured differed significantly in desirability or importance between the genders (Paulhus, 2002; Paulhus & John, 1998).

Age. Ludeke and Makransky (2016) found no relation between a respondent's age and overclaiming for general cognitive abilities (the group was restricted to 15- to 30-year-olds). Similar results were obtained by Mesmer-Magnus et al. (2006). It is important to note that most of the studies in the field were based on college or high school student samples, so the age range in the samples was usually very limited. Probably the only study that included a wide range of ages (30–70+ years) was the one by Calsyn and Winter (1999), who reported a small negative correlation between age and OCT bias ($r = -.11$). Thus, there is only very limited evidence of age-related differences in OCT. This variable is, however, important in our research because the social desirability of digital competence can be higher among younger people (see, e.g. Palczyńska & Rynko, 2020).

Socioeconomic Status

Calsyn et al. (2001) and Muszyński (2020) found that self-reported income was not related to overclaiming and obtained only a small and negative relation between educational level and overclaiming, unlike Jerrim et al. (2019) who found higher overclaiming in students of a higher socioeconomic status.

In this study, gender, age and socioeconomic status were correlated with overclaiming indices in order to explore their nomological network. The self-reported educational level was used as a proxy for socioeconomic status. Of course, it is to be remembered that socioeconomic status and level of abilities are most often correlated (Marks & Pokropek, 2019).

Method

Materials

In this study, we used data from the pilot study of the DigCompSat that was conducted in Ireland. DigCompSat is a self-assessment tool based on the Digital Competence Framework for Citizens (Carretero et al., 2017; Vuorikari et al., 2022). The aim of the framework was to become a reference to improve citizens' digital competence and to help plan educational and training initiatives to improve the digital competence of specific target groups. DigComp also supported policymakers in formulating policies for digital competence building. The DigComp reference framework is composed of 21 competence grouped under five main areas: (1) information and data literacy, (2) communication and collaboration, (3) digital content creation, (4) safety and (5) problem solving. The DigCompSat tool has the same areas.

Questionnaire

The DigCompSat questionnaire was administered via an online platform. Respondents first filled in a sampling questionnaire that identified individuals with the demographic features required for participation in the DigCompSat pilot. The sampling questionnaire included questions about the respondent's gender, age and educational background. Moreover, it included an adapted set of questions similar to those used in the Digital Skills Indicator (DG Connect and Eurostat, 2014; 2015) intended to provide information about recent digital experiences based on which scoring thresholds were defined for three levels (basic, intermediate and advanced). In our study, the scoring data were used to analyse our main instrument criterion-related validity. Respondents who were qualified to participate then answered 107 questions on digital competencies, including five questions intended to measure overclaiming bias (foils). All responses were given on a fully labelled, five-point scale ranging from 1 *(no mastery)* to 4 *(full mastery of a topic)*; additionally, participants could choose 0 *(I do not understand this question)*. The questionnaire covered the five DigComp areas. The participants were asked to respond to statements such as: *I know how to restrict or refuse access to my geographical location*; *I can detect when digital content is made available illegally (e.g. software, movies, music, books, TV)*, etc. (see Clifford et al., 2020, for the full set of questions and response category labels).

OCT Items

The same group of experts developed five foil items designed to be absurd or refer to nonexistent tools and services and purposely built to detect and control for overclaiming:

(1) I know how to access media apps to update personal search strategies.
(2) I know how to use a spelling checker to speed up software execution.
(3) I know I have to keep the windows closed when I enter the password to access my personal computer.

(4) I know how to use the ProblemSolver app that has a solution for all technical problems with digital devices.
(5) I can manage my online reputation using the SmartR(r) application.

Participants

The participants were selected to proportionally represent the demographic structure in Ireland, the country chosen for the pilot study. All 157 of the participants answered sociodemographic questions asking for the following information:

- Gender (male, 52.87%; female, 47.13%)
- Age group (16–24, 22.29%; 25–54, 61.78%; 55–65, 15.92%)
- Educational level (low = primary or lower secondary school, 3.82%; medium = upper secondary school and college (nonuniversity), 44.59%; high = university graduate, master or equivalent and doctor or equivalent, 51.59%)
- And Digital Skills Indicator (continuous variable).

The total interview duration was collected by the surveying system, but no data on item-level or screen-level response times were available.

Procedure

Data collection took place over a 3-week period in January 2020. Participants responded to the questionnaire on an online platform in a self-paced mode and were supported by local coordinators/observers face-to-face. Coordinators did not help with interpreting the items and could give only technical support at the request of the participants (none was required). The proctors were also to observe if candidates had any issues with particular items and were to record them on a structured note-taking sheet.

Results

Internal Consistency of the Overclaiming Scale

Before using the OCT scale, we needed to guarantee that the items and the scale as a whole were adequate. Because of our relatively small sample size ($n = 157$), we employed psychometric tools based on the Classical Test Theory (CTT) to do this. The item-total analysis includes point-biserial correlation coefficients between the individual items and the total score, and the item-rest correlation contains the point-biserial correlations between each item and the total score with that item omitted (Lord & Novick, 1968). The difficulty is the mean of the responses on a scale of 0–4. The discrimination is a difference in the average answer to an item between the top third and the bottom third of the respondents' distribution.

To assess the dimensionality of the scale, we used parallel analysis (Horn, 1965) based on polychoric correlations. Parallel analysis is a simulation-based technique that relies on principal component analysis (PCA). The method is based on comparing the eigenvalues of the correlation matrix of the actual database with the eigenvalues from simulated datasets (in our case, 400) that are equal to the number of items, sample size and distribution of the responses, but the responses are simulated assuming no correlations between the items. The eigenvalues from the actual data are compared with the averages of the eigenvalues from the simulated data. By noting the number of eigenvalues from the observed data that are larger than the eigenvalues from the simulated data, one can gather a possible number of dimensions. Simulation studies showed that parallel analysis based on polychoric correlations is a much more reliable tool than the classical PCA approach based on the Kaiser criterion (eigenvalue>1) or scree tests because it effectively protects against the capitalisation on chance in small samples and against large eigenvalues that can be produced by random data (Van der Eijk & Rose, 2015). Finally, we used the Cronbach's alpha and McDonald's omega coefficients as reliability measures for the OCT scale.

Relation Between OCT Scale and Careless/Insufficient Effort Responding Indices and Completion Time

Of the C/IER indices described in the literature (Curran, 2016; Meade & Craig, 2012), we used three in this study. The first one, the intraindividual response variability (IRV), is a within-person SD calculated for a given set of data. Low values of this index may indicate straightlining, and very high values may suggest (pseudo)random responding (Dunn et al., 2018; Marjanovic et al., 2015). The second one, a long-string measure, is a number of identical responses in a row (i.e. an uninterrupted string of identical responses). If few such strings are present in a given participant's response vector, the highest number of them is given as this measure's value. The third one was the average string length index, which indicates the average number of identical responses in a row for a given response vector. Finally, because the careless indices were strongly correlated, we computed the PCA score based on the three indices (Meade & Craig, 2012). The score was based on the first component that explains 77% of the total variability of the three indices. All of these measures are often used as indicators of straightlining, a survey-responding behaviour represented by choosing only one answer throughout the whole survey (e.g. Reuning & Plutzer, 2020; Zhang & Conrad, 2014).

In addition to the C/IER indexes, we also correlated the OCT scale with the respondent time because this is another way to measure the quality of the responses provided in self-assessments and questionnaires (Meade & Craig, 2012). Because the CT was right-skewed, we used the inverse hyperbolic sine (IHS) transformation to normalise this variable.

To investigate the relation between OCT scores and indices describing the response process, we used correlation coefficients. The C/IER indices were calculated using the 'careless' R package (Yentes and Wilhelm, 2018/2021).

Impact of Overclaiming on Reliability

To see how overclaiming affects the reliability of the scores, we performed a series of estimates of the omega coefficient for the five areas of the tool and the combined scale for the full sample (n = 157). Then we excluded observations with high OCT scores starting with the observation of respondents with a maximum overclaiming score of 20 (n = 7) and ending with a score of more than 6 (n = 31).

Validity of the Target Scale Controlling for OCT

In order to explore the role that controlling for overclaiming can play in increasing the validity of a self-assessment instrument, we benefited from the fact that respondents completed two different self-assessment tools. Therefore, we compared the relation between the digital skills level measured by the questionnaire inspired by Eurostat Digital Skills Indicator (DSI) and the DigCompSat (for each area and for the combined score). We presented the standardized results of this relation without any controls (raw) and with controlling for (1) OCT scores (OCT), (2) the combined careless index (C/IER) and (3) the CT. Additionally, the R^2 indices were computed together with the 95% confidence interval (CI) based on percentiles from the bootstrap procedure, where 1000 EIV models were computed using 1000 bootstrap samples (Efron, 1982). We expected that the regression coefficients and R^2 values depicting the relation between the questionnaire inspired by DSI and the DigCompSat would be higher once they were controlled for variables that accounted for irrelevant response behaviours.

Correlates of Overclaiming

To investigate the correlates of the OCT scores, we used information on gender, age, educational levels and digital skills. For the digital skills, we disposed of three categories, foundational, intermediate and advanced, that denoted the skill level and that were derived from the questionnaire inspired by the Eurostat index. To simplify the modelling, we treated the educational level and digital skills as approximations of continuous variables. In the first step, we compared the correlations of prediction variables and OCT scores with the correlations between predictors and other indices describing the response process. In the second step, we used four nested ordinary least squares regression models to investigate the conditional relations between the predictors and the OCT. In model 1, we used demographic variables (age, gender and education); in model 2, we added digital skills as a proxy of the digital competence level; in model 3, we added the CT and in model 4, we introduced the combined careless index.

Results

Internal Consistency of the Overclaiming Scale

Figure 1 presents the distributions of responses to the overclaiming items sorted from the most overclaimed (Q2) to the least overclaimed (Q5). It appears that items based on fake things were the most robust to overclaiming (Franzen & Mader, 2019; Hargittai, 2009). For item Q5, 20% of respondents claimed that they could use a nonexistent application in a good or very good way. Sixty-eight percent of respondents agreed that they could perform nonsense actions such as ‘use spelling checker to speed up software execution’ (Q2) in a good or very good way. Such results might indicate a high level of overclaiming and overconfidence and/or a strong effect of social desirability bias. However, one needs to remember that questions could be subject to a logical,

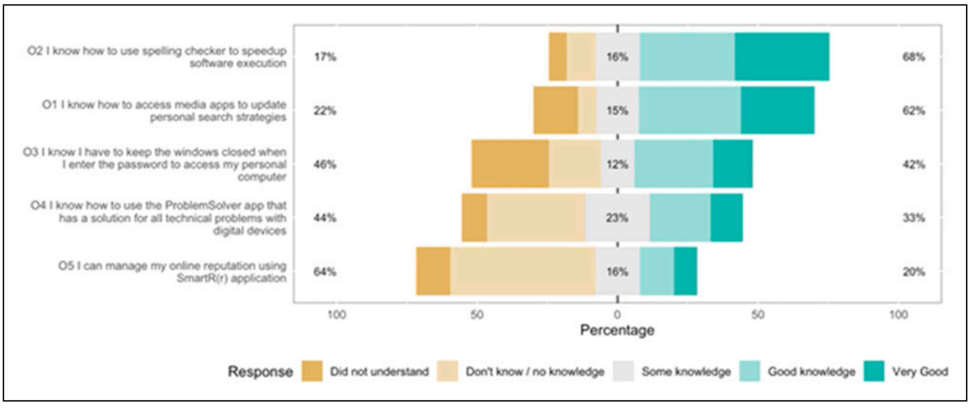

Figure 1. Distribution of overclaiming responses.

sense-making interpretation, and in this way, respondents could give a meaning even to nonsensical actions and objects. This could have been the case for Q1 because media applications could be a part of personal search strategies or for Q2 because in some situations switching off the spelling checker could speed up the text editor. Similarly, in some specific office settings, security regulations could mention that covering windows would prevent password stealing (Q3). In the same vein, the name of the 'ProblemSolver' application could be interpreted as a general name for this type of software (Q4). Therefore, overclaiming could not be judged based on single responses, but it could be judged on a consistent pattern of responses to a set of items using a scale with adequate internal consistency.

Table 1 depicts the basic item analysis performed in the Classical Test Theory framework. High item-total and item-rest correlations indicated that items measured similar traits. Item 2 was the one least associated with the general trait (item-total of .64, item-rest of .44 and discrimination of 1.60), but still the relation was reasonably high because the overall reliability of the scale based on five items was higher than with this item excluded. The total reliability measured by Cronbach's alpha was .77 with 95% CI intervals of .71–.82; it is exactly the same as when measured by McDonald's omega (.77; 95% CI 0.71–.82).

The argument of the unidimensionality of the scale was strengthened by parallel analysis (depicted in Figure 2), which confirmed that there was only one dimension in the overclaiming scale.

Once the good psychometric properties of the items were shown, an OCT scale was constructed as a simple sum of responses to the overclaiming items and went from 0 to 20. The scale's distribution was approximately normal, although it had some overrepresentation of highly overclaiming respondents (scores above 15; see Figure 3).

Relations Between Overclaiming Scale and Careless Responding

In Table 2, correlations between the OCT scores, C/IER measures and CT are presented. Less diverse response vectors were associated with more overclaiming. Moreover, the OCT scores were negatively correlated with the CT, indicating that fast respondents were scoring high on overclaiming items. Interestingly, OCT scores had a higher correlation with the response time than any of the careless responding measures. The careless indices were highly correlated with each other (especially the long string and average string indices; r = .81). Overclaiming was negatively related to IRV and positively related to the long string and average string indices, but the sizes of those correlations were smaller than the correlations among the careless indices. The correlation between the OCT and C/IER PCA index was higher than the correlations between the straightlining indices and the OCT, but they were roughly the same size as the correlation between the IRV and the OCT.

Table 1. Item-Descriptive Statistics.

Item number	Item-total correlation	Item-rest correlation	Difficulty	Discrimination	Alpha when excluded
OItem1	.72	.51	2.50	2.19	.73
OItem2	.64	.44	2.78	1.60	.75
OItem3	.77	.58	1.83	2.60	.71
OItem4	.71	.55	1.92	1.87	.72
OItem5	.75	.61	1.53	1.77	.70

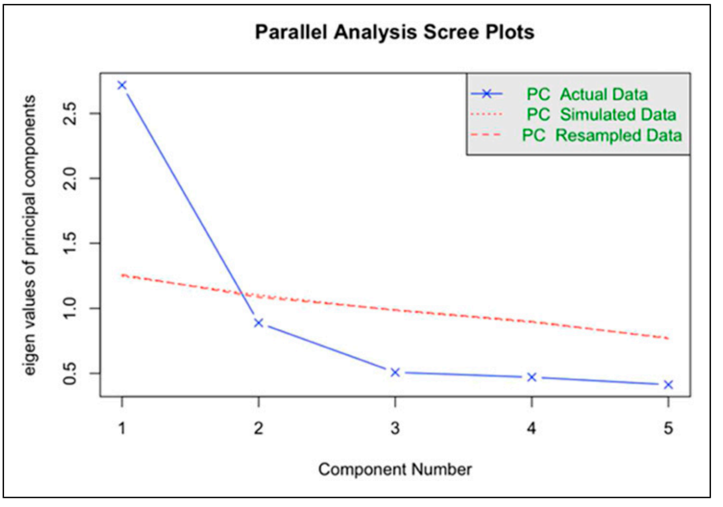

Figure 2. Parallel analysis.

Reliability Analysis of the Target Scale Controlling for OCT

The results presented in Figure 4 show that including participants with higher overclaiming scores led to a slight decrease in reliability. The drop was more pronounced in each of the content domains than in the combined scale, probably due to the impressive length of the whole scale (107 items). The drop in reliability was not large, but it showed some tendency to indicate that overclaiming was related to the scale's reliability.

Validity of the Target Scale Controlling for OCT

Results show that in all models, the digital competence as measured by the DigCompSat were strongly related to the digital skills (as measured by the questionnaire inspired by the Eurostat index). Controlling for DigCompSat scores and OCT scores was negatively and statistically significant related to digital skills, meaning that overclaimers that had the same score on the DigCompSat had, on average, lower digital skills. This result was consistent across all models. Interestingly, the effect of the careless index and response time on digital skills was not significant, and this was also consistent across all models. These relations are strictly related to the conditional effect of the DigCompSat score for digital skills. For instance, the score for the combined scale not adjusted for the standardized coefficient for DigCompSat was .54 (95% CI .40–.68), whereas when controlling for the OCT score, it increased to .66 (95% CI .51–.80), yielding a statistically significant increases of conditional relationship compared with unconditional relationship. It suggests that the relation between the DigCompSat score and digital skills is biased downward if not controlled for overclaiming (Tables 3 and 4).

Controlling for overclaiming brought a larger increase of the variance accounted for than the C/IER index and survey CT. The increase was not of the same value in all areas: in some, the R^2 value tripled (area 1), and in others, it was hardly noted (areas 2 and 5). Again, it seems that controlling for overclaiming (and to a lesser extent for C/IER) increases the measurement quality of self-assessment tools.

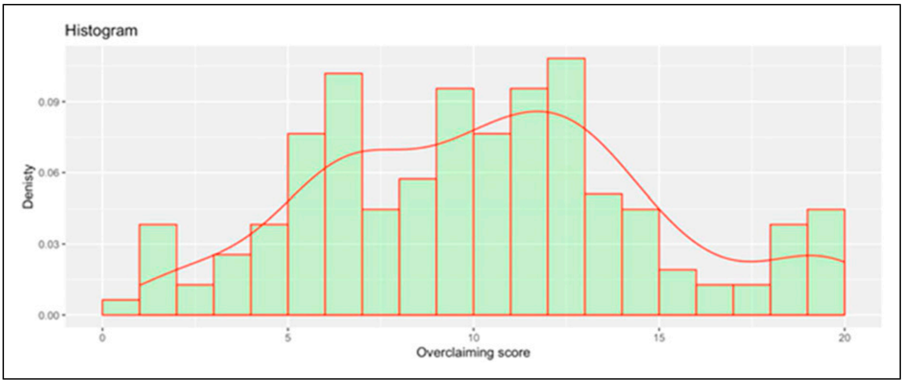

Figure 3. Distribution of overclaiming scores.

Table 2. Pearson's Correlation Coefficients Between OCT Scores and Careless Responding Measures.

	OCT	IRV	Long string	Average string	Completion time
OCT	1				
IRV	–.68***	1			
Long string	.47***	–.58***	1		
Average string	.47***	–.56***	.81***	1	
Completion time (CT)	–.34***	.27***	–.32**	–.26**	
Straightlining index (SL)	.61***	–.80***	.92***	.91***	–.32***

Note. ***p < .001 **p < .01 *p < .05.

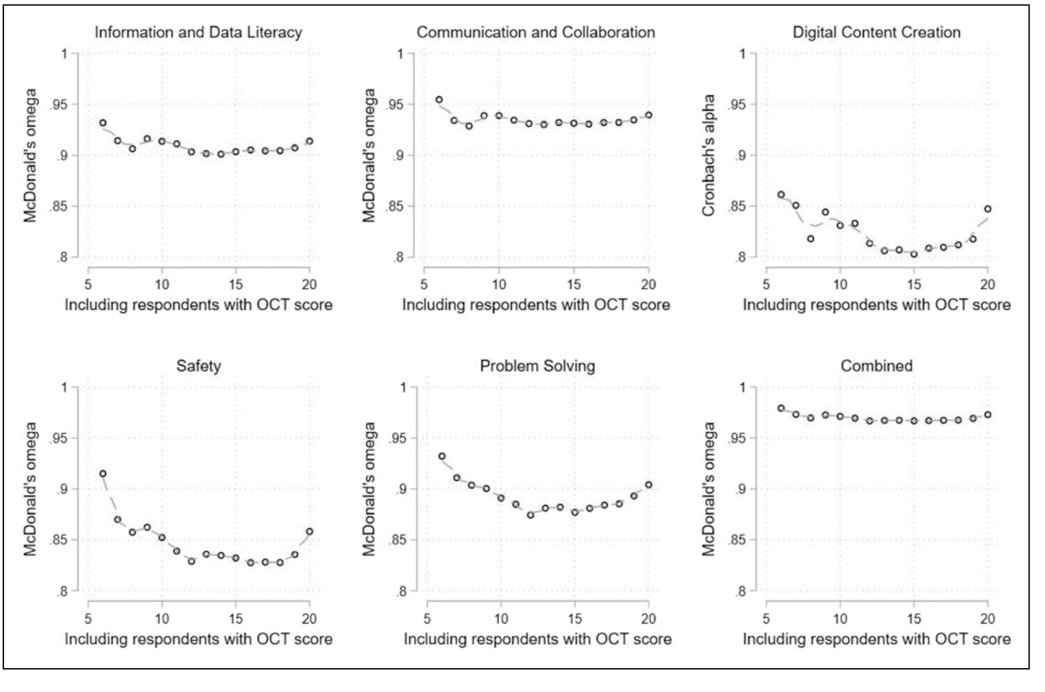

Figure 4. Relation of reliability to the number of overclaimers in the sample.

Table 3. Regression Coefficients for the Relation Between Digital Skills (Eurostat-Inspired Measure) and DigCompSat. p-values for Testing the Coefficient Against the Raw Model in Parenthesis.

Area/coefficient	Raw	Controlled for overclaiming (OCT)	Controlled for careless responding (C/IER)	Controlled for completion time (CT)
(1) Information and data literacy	.59	.65* (.02)	.60 (.50)	.59 (.85)
(2) Communication and collaboration	.46	.52 (.07)	.45 (.44)	.46 (.97)
(3) Digital content creation	.41	.50* (.02)	.44 (.30)	.42 (.69)
(4) Safety	.48	.56** (.01)	.49 (.51)	.48 (.94)
(5) Problem solving	.42	.52** (.01)	.43 (.72)	.42 (.85)
Areas 1–5 combined	.54	.66** (.002)	.56 (.17)	.54 (.61)

Note. ***p < .001 **p < .01 *p < .05.

Table 4. R^2 Values for Models Presented in Figure 5 With 95% Bootstrapped CI in Parentheses.

Area/coefficient	Raw	Controlled for overclaiming (OCT)	Controlled for careless responding (C/IER)	Controlled for completion time (CT)
(1) Information and data literacy	.34 (0.23–0.46)	.37 (0.26–0.48)	.34 (0.23–0.47)	.34 (0.23–0.46)
(2) Communication and collaboration	.21 (0.10–0.34)	.23 (0.12–0.35)	.21 (0.11–0.33)	.21 (0.11–0.35)
(3) Digital content creation	.17 (0.08–0.28)	.19 (0.10–0.31)	.17 (0.08–0.30)	.17 (0.08–0.29)
(4) Safety	.23 (0.12–0.36)	.26 (0.15–0.38)	.23 (0.12–0.37)	.23 (0.13–0.36)
(5) Problem solving	.18 (0.08–0.30)	.20 (0.12–0.32)	.18 (0.09–0.30)	.18 (0.09–0.30)
Areas 1–5 combined	.29 (0.18–0.41)	.33 (0.23–0.45)	.29 (0.17–0.41)	.29 (0.18–0.42)

Covariates of Overclaiming, Careless Responding and Completion Time

Being male was related to a higher amount of overclaiming and a lower amount of differentiation in responding, as measured by the IRV. However, gender did not correlate with the other two straightlining indices.

Age was not related to overclaiming in either of the straightlining indices (Table 5). It turned out to be related only to the IRV, with lower IRV scores (less differentiation) in older participants. Neither educational level nor self-reported digital skills (inspired by the Eurostat index) were correlated with overclaiming scores or any of the C/IER measures. The CT was negatively related to all of them (save the IRV), with faster responses linked with

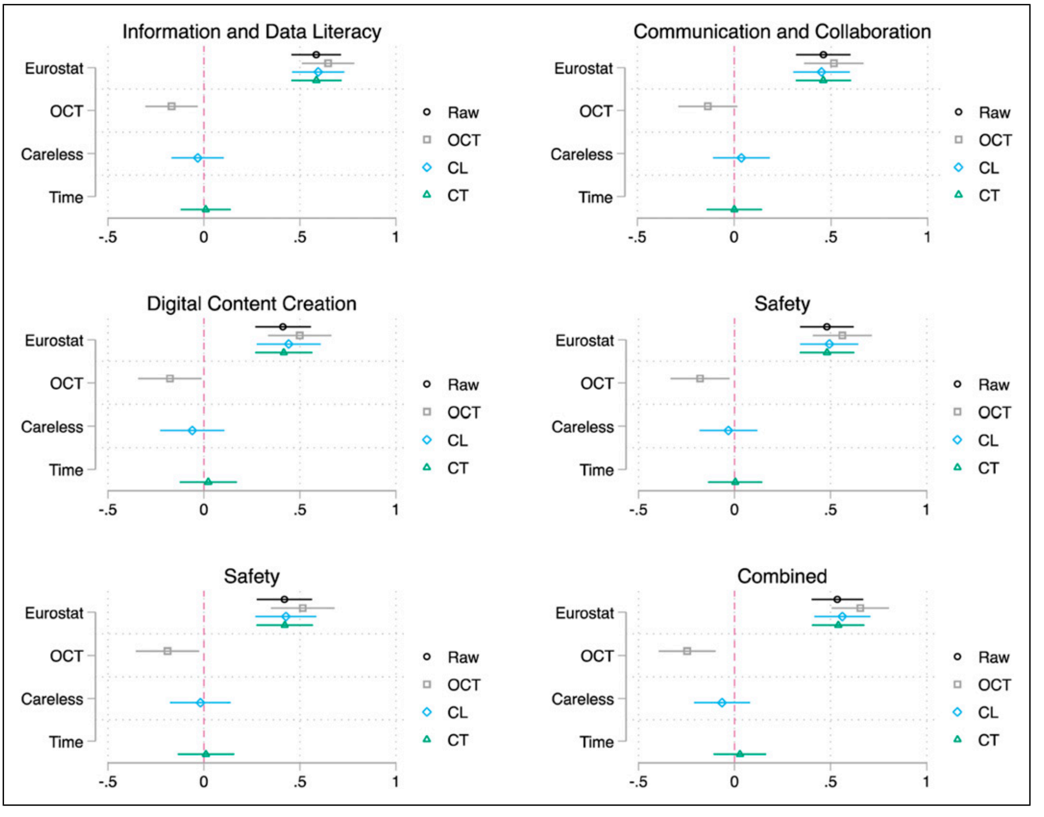

Figure 5. Standardized regression coefficients with 95% CI where digital skills were predicted by DigCompSat scores without adjustments and when controlled for OCT scores, careless responding index and completion time. *Note.* Raw - models without any controls DigComp scores regressed on digital skill level; OCT - model with overclaiming scores as an additional predictor; C/IER - model with the combined careless index as an additional predictor and CT - model with completion time as an additional predictor.

Table 5. Pearson's Correlation Coefficients Between Sociodemographic Variables and OCT Scores, Careless Responding Measures and Completion Time.

Socio-demographic variable	Overclaiming	IRV	Long string	Average string	Completion time
Gender (male)	.22**	–.33***	.14	.14	–.10
Age	.02	–.21**	.10	.13	.03
Educational level	.01	–.06	.00	.00	.03
Digital skills (Eurostat)	.07	–.10	.14	.18*	–.06

Note. ***$p < .001$ **$p < .01$ *$p < .05$.

higher rates of overclaiming and straightlining. This last result is a replication of many results, indicating that participants who were rushing to complete the survey were often responding carelessly (e.g. Leiner, 2019; Meade & Craig, 2012; Niessen et al., 2017; Wood et al., 2017).

Table 6. Standardized Regression Coefficients for the OCT Score–Dependent Variable.

	(1)	(2)	(3)	(4)
Male	.194*	.187*	.156	.077
	(.020)	(.028)	(.054)	(.253)
Age	–.028	−.029	−.013	−.091
	(.750)	(.741)	(.880)	(.197)
Educational level	.010	.003	.011	.034
	(.909)	(.971)	(.897)	(.616)
Digital skills		.033	.016	−.039
		(.693)	(.844)	(.556)
Completion time			−.325***	−.153*
			(.000)	(.022)
Straightlining index				.568***
				(.000)
N	157	157	157	157
adj. R^2	.017	.011	.112	.387
BIC	459.064	463.959	451.049	397.056

Note. *p < .05, **p < .01, ***p < .001; BIC – Bayesian Information Criterion.

Results presented in Table 6 focus on overclaiming. Models 1 and 2 show that there was no significant relationship between overclaiming and the background variables. Overclaiming was related to the completion time (model 3), and this could explain approximately 10% of the variance in the overclaiming scores and straightlining index (model 4), which added another 28% of the explained variance.

Discussion

In this paper, we have examined the adequacy of the OCT to improve the information provided by self-assessment digital competence tools. First, we proposed and empirically validated a set of overclaiming items using a pilot test of the DigCompSat, a self-assessment tool developed by the European Commission based on the Digital Competence Framework. The items form an adequate one-dimensional scale with good psychometric properties. We moved next to analyze an under-studied topic, the relationship between overclaiming and careless responses, and whether both these frameworks, used together, could help in increasing a scale's quality over and above a scale that used only one of the frameworks. We provided results showing that the C/IER, CT and overclaiming were correlated but that correlation was not strong enough to indicate that overclaiming was totally caused by careless or rushing respondents. It meant that only a part of the overclaiming variance was related to response biases but that the rest needed further delineation, with memory, self-monitoring biases and self-enhancement being the most plausible explanations (Goecke et al., 2020).

Then we moved forward to assess how important it was to control overclaiming in order to improve the quality of the data provided by DigCompSat. First, focusing on the reliability of the tool, the results showed that the whole tool, composed of around 100 items, was only slightly improved when individuals with higher overclaiming scores were excluded. However, when we analysed each area individually, there was an increase in reliability in all of them when overclaimers were dropped out. This result showed the importance of controlling for overclaiming in self-assessment tools, especially when scales are short or only subscales (e.g. content areas) of longer measurement tools are used. This last point is of utmost importance because the use of data at this level is a condition *sine qua non* to identify weak areas in the use of digital technologies and can guide effective lifelong learning decisions and policies towards remediation of the identified weaknesses. We also showed that controlling for overclaiming was important in terms of the (convergent) validity of self-assessments, which was here represented by a correlation with the DSI inspired by Eurostat. Moreover, the elimination of individuals with high overclaiming scores tended to increase the relationship between the two self-reporting digital competence tools.

Finally, our data showed that education and age were not related to overclaiming and that men were more prone to overclaim than women. However, this difference was no longer significant when controlling for C/IER. This result points to an important alternative explanation of gender differences in overclaiming and similar biases (Vonkova et al., 2021). Traditionally, social desirability, importance and gender-related stereotypes of the studied domain were seen as factors explaining the precision of self-assessment in sociodemographic groups (e.g. Palczyńska & Rynko, 2020) Our results point out that any sociodemographic differences, such as more overclaiming in male participants, should be first controlled for methodological factors, such as C/IER, before a search is made for substantial (socio-psychological) explanations. The topic of sociodemographic covariates of overclaiming warrants further research, especially in situations where groups that highly value a given domain are predicted to yield more overly positive self-assessments in light of self-enhancement theories.

Conclusions

Based on the results presented in this paper, we can conclude that the OCT can be used to enhance the data quality of self-assessment scales and surveys aiming to capture information on digital competence, at least when the aggregated level is in question. Use of the OCT can increase a scale's reliability and convergent validity, with the increase in reliability being potentially more important for shorter scales or facets. These results reinforce recent findings from the digital competence domain (Vonkova et al., 2021). Moreover, our data support a combined approach of overclaiming and careless responding analysis because these two frameworks capture partially separate spurious variances and can account for reducing the measurement error over and above each other. Finally, we show that identification of more overclaiming sociodemographic groups should be backed up by controlling for other response biases, such as careless responding, because between-group differences in biases can cause spurious between-group differences in overclaiming scores.

Limitations

The data used in this study were based on a rather small sample that was not representative of any population. Moreover, the evidence presented here came from only a single country, Ireland; this means that the findings cannot be used to measure any cross-country cultural differences in overclaiming tendencies (e.g. Vonkova et al., 2018). Finally, it is important to note that although the OCT scale is useful for identifying overclaiming, the qualitative evidence obtained by the observers during the measurement shows that respondents can feel confused and even a bit annoyed when responding to OCT foils. More evidence on best ways to construct foils is needed.

Moreover, using foils rises some ethical concerns that should be addressed by research ethic committees as well. It seems that at least participants debriefing on the use of OCT foils and its purpose is needed. All these limitations should be addressed in future studies.

Directions for Future Studies

Self-reports are widely used due to their advantages over other research methods; these are mainly their cost-efficiency, versatility and ease of use (Brückner, 2009). However, they can be used not only to measure a given construct but also to measure and develop self-assessment competencies as such (Andrade & Valtcheva, 2009). Self-assessment tools are widely tested in formative and classroom-based assessments and can be used to present and clarify all the dimensions of a domain, but they can also enhance lifelong learning skills by teaching participants how to validly self-assess and self-monitor their own competence (Panadero et al., 2016, 2017, 2018). It is for future studies to establish how introducing the OCT into formative assessment tools could account for their problems and contribute to enhancing such tools' accuracy and validity (Brown et al., 2015). One can wonder, however, whether OCT is an efficient form of increasing assessments' psychometric qualities – a modest gain in reliability and validity requires adding additional items. In our opinion, it is still worthy to add them, especially as few OCT foils take less than 1 minute to answer (Huang et al., 2012).

We have shown that the OCT can increase the quality of aggregated data by eliminating data from individuals flagged as overclaimers. However, the extrapolation of this approach to correct self-assessments on an individual level can be a very promising research avenue. In this way, individuals that engage in overclaiming unintentionally and persons, organisations or systems providing training recommendations on digital competence on the basis of self-assessment data would be able to make better decisions and configure their own effective learning paths in order to improve the quality of their lives and work. Using OCT as a self-reflection tool with lifelong learning, upskilling and reskilling aims is one of the most interesting directions for future research.

The OCT can also be used to further enhance the quality of survey data when these data are needed for a new and exciting purpose, the linking with digital traces data. Many studies point to stark discrepancies between survey and digital traces data (e.g. log-based measures, social media and search engine data); this hampers the development of methodological and theoretical frameworks of digital big data and limits the possibilities of their validation (Al Baghal et al., 2020; Hargittai, 2020; Parry et al., 2021; Stier et al., 2020). Enhancing the quality of survey data is an important goal in the social sciences because it is the most available source method for comparisons with digital data.

Using paradata to improve the quality of self-reports is on the rise, but the OCT has been to date only sporadically linked with such data. As evidenced by this paper, developing process indicators is a worthwhile goal (Kroehne & Goldhammer, 2018). This is also the case for OCT scales, which would not only serve to increase our interpretation of overclaiming scores but would also improve further convergence between OCT and C/IER. This research not only has methodological importance but also theoretical importance because it can refine the current interpretation of overclaiming as a result of motivated self-enhancement (Paulhus et al., 2003) and can also include other explanations, such as memory biases, interpretation errors (Goecke, et al., 2020; Müller & Moshagen, 2018) and measurement errors (e.g. careless responding; Dunlop et al., 2020; Ludeke & Makransky, 2016).

Finally, more elaborate procedures for developing better OCT scales are needed because the rules for generating foils and the piloting procedures for OCT scales are still only vaguely suggested in the field of social sciences (see Goecke et al., 2020, for some good ideas) and many methodological issues, such as the ideal proportion of foils to reals or the most desirable OCT scale length, have still not been tested empirically.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is financed by the National Science Centre (NCN) research grant (2019/33/B/HS6/00937) *Understanding response styles in self-report data: Consequences, remedies and sources*, awarded to Artur Pokropek. Work of Jonatan Castano Munoz was funded by grant RYC2020-030157 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future". Data collection for this study sponsored by JRC

Data Collection

Data collection was sponsored by European Commission's Joint Research Centre, Seville site.

ORCID iD

Marek Muszynski https://orcid.org/0000-0002-9306-1237

References

Al Baghal, T., Sloan, L., Jessop, C., Williams, M. L., & Burnap, P. (2020). Linking Twitter and survey data: The impact of survey mode and demographics on consent rates across three UK studies. *Social Science Computer Review*, *38*(5), 517–532. https://doi.org/10.1177/0894439319828011

Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. *Journal of Applied Psychology*, *69*(4), 574–580. https://doi.org/10.1037/0021-9010.69.4.574

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, *48*(1), 12–19. https://doi.org/10.1080/00405840802577544

Arias, V. B., Garrido, L. E., Jenaro, C., Martínez-Molina, A., & Arias, B. (2020). A little garbage in, lots of garbage out: Assessing the impact of careless responding in personality survey data. *Behavior Research Methods*, *52*(6), 1–17. https://doi.org/10.3758/s13428-020-01401-8

Barber, L. K., Barnes, C. M., & Carlson, K. D. (2013). Random and systematic error effects of insomnia on survey behavior. *Organizational Research Methods*, *16*(4), 616–649. https://doi.org/10.1177/1094428113493120

Baxter, P., & Norman, G. (2011). Self-assessment or self deception? A lack of association between nursing students' self-assessment and performance. *Journal of Advanced Nursing*, *67*(11), 2406–2413. https://doi.org/10.1111/j.1365-2648.2011.05658.x

Bensch, D., Paulhus, D. L., Stankov, L., & Ziegler, M. (2019). Teasing apart overclaiming, overconfidence, and socially desirable responding. *Assessment*, *26*(3), 351–363. https://doi.org/10.1177/1073191117700268

Bing, M. N., Kluemper, D., Kristl Davison, H., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, *116*(1), 148–162. https://doi.org/10.1016/j.obhdp.2011.05.006

Bishop, G. F., Tuchfarber, A. J., & Oldendick, R. W. (1986). Opinions on fictitious issues: The pressure to answer survey questions. *Public Opinion Quarterly*, *50*(2), 240–250. https://doi.org/10.1086/268978

Borgonovi, F., & Pokropek, A. (2019). Seeing is believing: Task-exposure specificity and the development of mathematics self-efficacy evaluations. *Journal of Educational Psychology*, *111*(2), 268–283. https://doi.org/10.1037/edu0000280

Brown, G. T., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444–457. https://doi.org/10.1080/0969594x.2014.996523

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, *2*(1), 22–30. https://doi.org/10.14786/flr.v2i1.24

Brückner, H. (2009). Surveys. In P. Hedström & P. Bearman (Eds.), *The oxford handbook of analytical sociology* (pp. 666–687). Oxford University Press.

Brühlmann, F., Petralito, S., Aeschbach, L. F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology*, *2*, 100022. https://doi.org/10.1016/j.metip.2020.100022.

Burski, J., Chłoń-Domińczak, A., Palczyńska, M., Rynko, M., & Śpiewanowski, P. (2013). *Umiejętności Polaków – Wyniki Międzynarodowego Badania Kompetencji Osób Dorosłych (PIAAC)*. Instytut Badań Edukacyjnych.

Calsyn, R. J., Kelemen, W. L., Jones, E. T., & Winter, J. P. (2001). Reducing Overclaiming in Needs Assessment Studies an Experimental Comparison. *Evaluation Review*, *25*(6), 583–604. https://doi.org/10.1177/0193841X0102500601.

Calsyn, R. J., & Winter, J. P. (1999). Understanding and controlling response bias in needs assessment studies. *Evaluation Review*, *23*(4), 399–417. https://doi.org/10.1177/0193841x9902300403

Carretero, S., Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use (No. JRC106281)*. European Commission: Joint Research Centre. (Seville site).

Cipora, K., Willmes, K., Szwarc, A., & Nuerk, H. C. (2018). Norms and validation of the online and paper-and-pencil versions of the abbreviated math anxiety scale (AMAS) for polish adolescents and adults. *Journal of Numerical Cognition*, *3*(3), 667–693. https://doi.org/10.5964/jnc.v3i3.121

Clifford, I., Kluzer, S., Troia, S., Jakobsone, M., & Zandbergs, U. (2020). *DigCompSat. A self-reflection Tool for the European digital Framework for citizens (No. JRC123226)*. Joint Research Centre. (Seville site).

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement*, *70*(4), 596–612. https://doi.org/10.1177/0013164410366686

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, *66*, 4–19. https://doi.org/10.1016/j.jesp.2015.07.006.

DeSimone, J. A., DeSimone, A. J., Harms, P. D., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology*, *67*(2), 309–338. https://doi.org/10.1111/apps.12117

Deursen, A. V., & Van Dijk, J. A. (2010). Measuring internet skills. *International Journal of Human-Computer Interaction*, *26*(10), 891–916. https://doi.org/10.1080/10447318.2010.496338

Dunlop, P. D., Bourdage, J. S., de Vries, R. E., McNeill, I. M., Jorritsma, K., Orchard, M., Austen, T., Baines, T., & Choe, W. K. (2020). Liar! Liar!(when stakes are higher): Understanding how the overclaiming technique can be used to measure faking in personnel selection. *Journal of Applied Psychology*, *105*(8), 784–799. https://doi.org/10.1037/apl0000463

Dunn, A. M., Heggestad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: Comparison to other indicators and relationships with individual differences. *Journal of Business and Psychology*, *33*(1), 105–121. https://doi.org/10.1007/s10869-016-9479-0

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Society for industrial and applied mathematics (SIAM).

European Commission (2018). *Proposal for a Council recommendation on key competences in lifelong learning* (p. 24). COMPublications Office of the European Union.

European Commission/EACEA/Eurydice (2019). *Digital education at school in Europe*. Publications Office of the European Union.

Feeney, J. R., & Goffin, R. D. (2015). The overclaiming questionnaire: A good way to measure faking? *Personality and Individual Differences*, *82*(August), 248–252. https://doi.org/10.1016/j.paid.2015.03.038

Franzen, A., & Mader, S. (2019). Do phantom questions measure social desirability? *Methods, Data, Analyses*, *13*, 37–57. https://doi.org/10.12758/mda.2019.01.

Goecke, B., Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). Testing competing claims about overclaiming. *Intelligence*, *81*, 101470. https://doi.org/10.1016/j.intell.2020.101470.

Grant, D. M., Malloy, A. D., & Murphy, M. C. (2009). A comparison of student perceptions of their computer skills to their actual abilities. *Journal of Information Technology Education: Research*, *8*(1), 141–160. https://doi.org/10.28945/164

Haddon, L., Cino, D., Doyle, M.-A., Livingstone, S., Mascheroni, G., & Stoilova, M. (2020). *Children's and young people's digital skills: A systematic evidence review*. KU LeuvenySKILLS.

Hargittai, E. (2005). Survey measures of web-oriented digital literacy. *Social Science Computer Review*, *23*(3), 371–379. https://doi.org/10.1177/0894439305275911

Hargittai, E. (2009). An update on survey measures of web-oriented digital literacy. *Social Science Computer Review*, *27*(1), 130–137. https://doi.org/10.1177/0894439308318213

Hargittai, E. (2020). Potential biases in big data: Omitted voices on social media. *Social Science Computer Review*, *38*(1), 10–24. https://doi.org/10.1177/0894439318788322

Hargittai, E., & Hsieh, Y. P. (2012). Succinct survey measures of web-use skills. *Social Science Computer Review*, *30*(1), 95–107. https://doi.org/10.1177/0894439310397146

He, J., & Van de Vijver, F. J. R. (2016). The motivation-achievement paradox in international educational achievement tests: Toward a better understanding. In R. King & A. Bernardo (Eds.), *The psychology of asian learners*. Springer.

Helsper, E. J., & Reisdorf, B. C. (2017). The emergence of a "digital underclass" in Great Britain and Sweden: Changing reasons for digital exclusion. *New Media & Society*, *19*(8), 1253–1270. https://doi.org/10.1177/1461444816634676

Helsper, E. J., Schneider, L. S., van Deursen, A. J. A. M., & van Laar, E. (2020). *The youth digital Skills Indicator: Report on the conceptualisation and development of the ySKILLS digital skills measure*. KU LeuvenySKILLS.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/bf02289447

Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, *27*(1), 99–114. https://doi.org/10.1007/s10869-011-9231-8

Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology*, *100*(3), 828–845. https://doi.org/10.1037/a0038510

ITU (2020). *Digital skills assessment Guidebook*. ITU Publications. https://www.itu.int/myitu/-/media/Publications/2020-Publications/Digital-Skills-Assessment-Guidebook.pdf

Jerrim, J., Parker, P., & Shure, N. (2019). Bullshitters. Who are they and what do we know about their lives? In *IZA discussion papers, No. 12282*. Institute of Labor Economics (IZA). https://docs.iza.org/dp12282.pdf.

Johansson, S. (2013). The relationship between students' self-assessed reading skills and other measures of achievement. *Large-scale Assessments in Education*, *1*(1), 1–17. https://doi.org/10.1186/2196-0739-1-3

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*(1), 206–219. https://doi.org/10.1037/0022-3514.66.1.206

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Kam, C. C. S. (2019). Careless responding threatens factorial analytic results and construct validity of personality measure. *Frontiers in Psychology*, *10*, 1258. https://doi.org/10.3389/fpsyg.2019.01258.

Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: The case of job satisfaction. *Organizational Research Methods*, *18*(3), 512–541. https://doi.org/10.1177/1094428115571894

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika*, *45*(2), 527-563. https://doi.org/10.1007/s41237-018-0063-y.

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403.

Litt, E. (2013). Measuring users' internet skills: A review of past assessments and a look toward the future. *New Media & Society*, *15*(4), 612–630. https://doi.org/10.1177/1461444813475424

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Ludeke, S. G., & Makransky, G. (2016). Does the over-claiming questionnaire measure overclaiming? Absent convergent validity in a large community sample. *Psychological Assessment*, *28*(6), 765–774. https://doi.org/10.1037/pas0000211

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, *67*(3), 280–296. https://doi.org/10.1037/0021-9010.67.3.280

Maderick, J. A., Zhang, S., Hartley, K., & Marchand, G. (2016). Preservice teachers and self-assessing digital competence. *Journal of Educational Computing Research*, *54*(3), 326–351. https://doi.org/10.1177/0735633115620432

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008.

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, *84*, 79–83. https://doi.org/10.1016/j.paid.2014.08.021.

Marks, G. N., & Pokropek, A. (2019). Family income effects on mathematics achievement: Their relative magnitude and causal pathways. *Oxford Review of Education*, *45*(6), 769–785. https://doi.org/10.1080/03054985.2019.1620717

McCourt Larres, P., Ballantine, J., & Whittington, M. (2003). Evaluating the validity of self-assessment: Measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities. *Accounting Education*, *12*(2), 97–112. https://doi.org/10.1080/0963928032000091729

Meade, & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Merritt, K., Smith, D., & Renzo, J. C. D. (2005). An investigation of self-reported computer literacy: Is it reliable. *Issues in Information Systems*, *6*(1), 289–295.

Mesmer-Magnus, J., Viswesvaran, C., Deshpande, S., & Joseph, J. (2006). Social desirability: The role of over-claiming, self-esteem, and emotional intelligence. *Psychology Science*, *48*(3), 336–356.

Monoi, S., O'Hanlon, N., & Diaz, K. R. (2005). Online searching skills: Development of an inventory to assess self-efficacy. *Journal of Academic Librarianship*, *31*(2), 98–105. https://doi.org/10.1016/j.acalib.2004.12.005

Müller, S., & Moshagen, M. (2018). Overclaiming shares processes with the hindsight bias. *Personality and Individual Differences*, *134*, 298–300. https://doi.org/10.1016/j.paid.2018.06.035

Muszyński, M. (2020). *Validity of the overclaiming technique as a method to account for response bias in self-assessment questions. Analysis on the basis of the PISA 2012 data*. Jagiellonian University. Unpublished doctoral thesis https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/251494/muszynski_validity_of_the_overclaiming_technique_as_a_method_2020.pdf?sequence=1&isAllowed=y

Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2017). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality*, *63*, 1–11. https://doi.org/10.1016/j.jrp.2016.04.010.

Palczyńska, M., & Rynko, M. (2020). ICT skills measurement in social surveys: Can we trust self-reports? *Quality & Quantity*, *55*(3), 1–27. https://doi.org/10.1007/s11135-020-01031-4

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: A roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher*, *45*(1), 13–31. https://doi.org/10.1007/s13384-018-0258-y

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*(4), 803–830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E.., Jönson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004.

Parry, D. A., Davidson, B. I., Sewall, C. J., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1–13. https://doi.org/10.1038/s41562-021-01117-5

Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Erlbaum.

Paulhus, D. L. (2012). Overclaiming on personality questionnaires. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 151–164). Oxford University Press.

Paulhus, D. L., & Dubois, P. J. (2014). Application of the overclaiming technique to scholastic assessment. *Educational and Psychological Measurement*, *74*(6), 975–990. https://doi.org/10.1177/0013164414536184

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*(4), 890–904. https://doi.org/10.1037/0022-3514.84.4.890

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*(6), 1025–1060. https://doi.org/10.1111/1467-6494.00041

Paulhus, D. L., & Trapnell, P. D. (2008). Self-presentation of personality: An agency-communion framework. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 492–517). The Guilford Press.

Paunonen, S. V. (2016). Sex differences in judgments of social desirability. *Journal of Personality*, *84*(4), 423–432. https://doi.org/10.1111/jopy.12169

Phillips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology*, *77*(5), 921–940. https://doi.org/10.1086/225231

Randall, D. M., & Fernandes, M. F. (1991). The social desirability response bias in ethics research. *Journal of Business Ethics*, *10*(11), 805–817. https://doi.org/10.1007/bf00383696

Reuning, K., & Plutzer, E. (2020). Valid vs. invalid straightlining: The complex relationship between straightlining and data quality. *Survey Research Methods*, *14*(5), 439–459. https://doi.org/10.18148/srm/2020.v14i5.7641.

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*(2), 340–352. https://doi.org/10.1037/0022-3514.80.2.340

Stier, S., Breuer, J., Siegers, P., & Thorson, K. (2020). Integrating survey data and digital trace data: Key issues in developing an emerging field. *Social Science Computer Review*, *38*(5), 503–516. https://doi.org/10.1177/0894439319843669

Van der Eijk, C., & Rose, J. (2015). Risky business: Factor analysis of survey data–assessing the probability of incorrect dimensionalisation. *Plos One*, *10*(3), e0118900. https://doi.org/10.1371/journal.pone.0118900.

van Deursen, A. J., Helsper, E., & Eynon, R. (2014). *Measuring digital skills: From digital skills to tangible outcomes project report*. London School of Economics, University of Twente. https://www.researchgate.net/publication/267037582_Measuring_Digital_skills_From_Digital_Skills_to_Tangible_Outcomes_project_report

Vonkova, H., Papajoanu, O., & Stipek, J. (2018). Enhancing the cross-cultural comparability of self-reports using the overclaiming technique: An analysis of accuracy and exaggeration in 64 cultures. *Journal of Cross-Cultural Psychology*, *49*(8), 1247–1268. https://doi.org/10.1177/0022022118787042

Vonkova, H., Papajoanu, O., Stipek, J., & Kralova, K. (2021). Identifying the accuracy of and exaggeration in self-reports of ICT knowledge among different groups of students: The use of the overclaiming technique. *Computers & Education*, *164*, 104112. https://doi.org/10.1016/j.compedu.2020.104112.

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes* (p. ISBN 978-92-76-48882-8 (online)). Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/115376.

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science*, *8*(4), 454–464. https://doi.org/10.1177/1948550617703168

Yentes, R., & Wilhelm, F. (2018/2021). Careless: Procedures for computing indices of careless responding. R package version 1.2.*1*, 2018/2021.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, *9*(2), 111–125. https://doi.org/10.1177/1745691613518075

Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey research methods*, *8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453.

Zhang, J., Paulhus, D. L., & Ziegler, M. (2019). Personality predictors of scholastic cheating in a Chinese sample. *Educational Psychology*, *39*(5), 572–590. https://doi.org/10.1080/01443410.2018.1502414

Author Biographies

Marek Muszyński (MA, psychology; PhD, sociology) is a post-doc researcher at the Institute of Philosophy and Sociology, Polish Academy of Sciences. His main research investigations concentrate around response biases (response styles, careless responding, socially desirable responding) in self-report data with a special focus on web surveys. He is also interested in psychometrics and psychological/educational assessment.

Artur Pokropek is a professor at the Institute of Philosophy and Sociology of the Polish Academy of Sciences. His main areas of research are statistics, psychometry and research methodology. He had developed several methodological and statistical approaches for analysing survey data. He gained knowledge as a visiting scholar at the Educational Testing Service (Princeton, USA) and associate researcher at the EC Joint Research Centre (Ispra, Italy) from 2016 to 2018. He holds a habilitation degree in social sciences, Ph.D. in Educational Science, and MA degrees in Educational Sciences and in Sociology.

Jonatan Castaño Muñoz is a senior researcher (Ramón y Cajal programme) at the Department of Teaching and Educational Organization in the Faculty of Education of the University of Seville. His main research topic is the digitalisation of educational systems and its pedagogical and socioeconomic implications. Jonatan holds a Ph.D in Information and Knowledge Society and a MA degree in Research Methods for Social Sciences.

Riina Vuorikari is a senior research fellow at the Joint Research Centre of the European Commission (2013-2022). Her work focuses on citizens' digital competence, recently she led the DigComp 2.2 update and contributed to the Digital Skills Indicator 2.0. Vuorikari has degrees in education (M.Ed in 1998 in Finland), hypermedia (DEA in 1999 in France) and her PhD is from the Dutch research school for Information and Knowledge Systems (2009). Since 1999, she has worked in the field of digital education.