

The challenges of conducting systematic evidence reviews: a case study of factors shaping children’s digital skills

Leslie Haddon

Department of Social Policy, London School of Economics, London, UK.

Davide Cino,

Department of Communication, Università Cattolica del Sacro Cuore, Milan, Italy.

Mary-Alice Doyle

Department of Social Policy, London School of Economics, London, United Kingdom.

Corresponding author

Leslie Haddon, Department of Social Policy, London School of Economics, London, United Kingdom.

L.G.Haddon@lse.ac.uk

Résumé

Les défis de la conduite d’une revue de littérature systématique : une étude de cas des déterminants des compétences numériques des enfants.

Les revues de littérature systématiques rassemblent les résultats de plusieurs études, aidant les chercheurs et les décideurs à comprendre les tendances de recherche et identifier les résultats issus d’une pluralité de contextes et méthodes de recherche. Ils sont devenus plus populaires au cours des vingt dernières années, avec divers guides traitant des différentes manières dont ils peuvent être menés et des problèmes soulevés pour leur réalisation. Cette étude de cas, sur les facteurs qui façonnent les compétences numériques des enfants, explore les défis, les risques et les stratégies potentielles pour réaliser une revue de littérature systématique. Elle revient sur comment les chercheurs réfléchissent aux jugements impliqués dans le choix des critères d’inclusion, le filtrage, le codage des études et la synthèse du matériel collecté.

Abstract

Systematic evidence reviews draw together findings from multiple studies, helping researchers and decision makers to understand patterns of research and findings across varying contexts and research methodologies. They have become more popular over the last twenty years, with various guides discussing the different ways in which they can be conducted and the issues arising in this process. This case study of a systematic review of the factors shaping children’s digital skills explores the challenges, risks and potential strategies in this process, as those involved in that review reflect upon the various judgements involved in choosing inclusion criteria, filtering and coding studies and synthesising the material collected.

Mots-clés

Revue de littérature systématique, étude de cas, compétences numériques, enfants et jeunes.

Keywords

Systematic review, case study, digital skills, children and young people`

Introduction

In his classic book ‘*What is history?*’ Carr (1961) compares the study of history to fishing:

The (historical) facts...are like fish swimming about in the vast and sometimes inaccessible ocean: and what the historian catches will depend, partly on chance, but mainly on what part of the ocean he chooses to fish in and what tackle he chooses to use – these two factors being, of course, determined by the kind of fish he wants to catch (pp.19-20).

In fact, Carr is not criticising, but instead defending and clarifying the historian’s role, drawing attention to how history always reflects researchers’ decisions about their goals and about how they operationalise the historical search process. In the social sciences, this has been discussed in terms of the degrees of freedom that researchers have when deciding how to implement research, for example, how to an experiment to test a hypothesis (Wicherts et al. 2016). There are equivalent decision-making processes when conducting a systematic evidence review (Vrieze, 2018). This article provides a case study of such a review, going beyond outlining the steps involved in that process to draw attention to the nature of the judgements that, inevitably, have to be made. The main intended audience is researchers who are considering conducting a systematic evidence review, providing the reflections of insiders who have implemented this procedure.

The approach taken here is first to outline the nature of systematic evidence reviews. The next step is to describe the systematic evidence review of children’s digital skills that was conducted as one part of the wider EU-funded ySKILLS project (Haddon et al., 2020). This article next reflects upon the initial decisions involved in searching databases. It then examines the subsequent filtering and coding of the studies found before moving on to chart decisions influencing how this material is organised and ‘mapped’. In so doing, it indicates numerous choices about how to transform an *ontological* object (i.e., a phenomenon out there) into an *epistemic* one (i.e. something worth being studied) (Caronia, 2011).

Systematic evidence reviews

Systematic evidence reviews originated in health research, as documented in the Oakly’s account of the early work of the EPPI¹-Centre, the founding institution in this field (Oakly, 2017). But in the last 20 years they have expanded into other disciplines, especially the social sciences. For example, and of relevance to the project documented in this article, Van Laar et al. (2017) and Schreeder, et al (2017) are reviews looking at digital skills more generally. Other examples from social science research are Nef et al. (2013) looking at social networking sites (SNSs) and older users, and Williams, (2019), examining SNSs and social capital.

As the term ‘systematic’ suggests, this mapping involves a more rigorous approach than a ‘narrative’ literature review that relies on the researchers’ prior knowledge of the field (Grant and Booth, 2009; Gough et al, 2017a). It entails developing a strategy to search research databases for relevant material, and then to select for detailed analysis those items most likely

¹ Evidence for Policy and Practical Information and Coordinating Centre.

to address the research question under consideration, in order to build up a comprehensive understanding and evaluation of the field². The search words, search strategies and all criteria used in filtering out inappropriate material are made explicit for the reader to follow and, in principle, reproduce (re-analyse the same data) or replicate the exercise (use the same process to analyse different data³). Thus transparency, rigour and comprehensiveness are often cited as virtues of the systemic review.

In describing a typology of reviews, Grant and Booth (2009) note that different types of reviews may have different specific goals, but that the general objectives of a systematic review are to summarise what is known in a field in order to draw recommendations for practice, and to identify gaps in the literature as directions for future research. That said, one useful distinction made by Gough et al. (2012) is between aggregative and configurative reviews. In the former the emphasis is on making empirical statements about the findings in a field of study. The authors refer to adding up and averaging findings when seeking evidence to inform decisions. Configurative reviews focus on how to interpret a field, for example, how to make sense of variation within the studies and how these fit together. These authors note that some studies contain elements of both these processes, and indeed the project reported in this article covered both a review of the evidence and a mapping of the field.

The particular systematic evidence review discussed in this article was undertaken as part of the ySKILLS project funded by the EU's Horizon 2020 programme. That broader project involved 15 partners from 13 countries and aimed to enhance and maximise the long-term positive impact of the information communications technology (ICT) environment on multiple aspects of wellbeing for children aged 12-17. The review focused on both factors that influenced the development of children's digital skills (antecedents) and the consequences of those skills. It was primarily intended to inform the development of a survey for that project, as well as being a research output in its own right. Although that review covered both antecedents and consequences of children's digital skills, this article will focus on the antecedents part of the review, i.e. the mapping and evaluation of which aspects of children's lives might have a bearing upon their digital skills. The team that carried out the review were native English and Italian speakers, the latter having a very high standard of English, so language was not an issue. They had varying degrees of familiarity with this literature and at various points showed their planning and intermediate findings to those within the project who had greater expertise as a check. Two of the research team had conducted several systematic reviews prior to this one, and so had developed relevant practical competencies (as identified by Oliver et al, 2017).

² An example guide to how to conduct such reviews is Atkinson and Cipriani (2018).

³ For example, to see if the field of study has changed by doing a systematic review of the same topic at a later time period.

The database search process

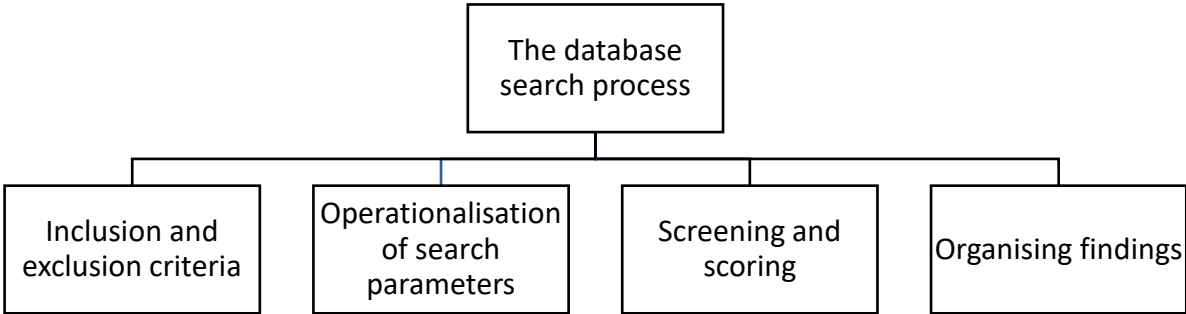


Figure 1. The database search process

All systematic evidence reviews involve decisions about criteria to be used for including and excluding material (Brunton et al., 2017), and in this case study, most of those parameters were decided at the very start of the review process. It concentrated on children aged 12-17, not all children, reflecting the focus of the wider EU project. It selected only quantitative studies since the aim was to inform both what questions should be in the survey and particular conceptual models that were under development within that project. The review only considered published material. While this involved a form of quality control because of the process of peer review in publication, it ran the risk of missing relevant material in non-published reports, book chapters and theses⁴ (Brunton et al., 2017). One particular concern of the literature on publication bias, is that results that are statistically significant are more likely to be published (Dwan et al., 2013).

Like many other systematic reviews, and reflecting the language competencies of the reviewing team, only studies where the full text was available in English were considered – even though this runs the risk of bias, if, for example, researchers prefer to publish positive results in English-language journals (Brunton et al., 2017). Full details of the search process were pre-registered with PROSPERO, with record ID CRD42020172272. That process involved another form of quality control since most of the steps and precautions taken had to be described in some detail in order to make sure the review reached a recognised standard.⁵

However, review parameters can develop even during the review process (ibid), especially in an exploratory study where it is difficult to make all judgements in advance when the nature of the field is not yet known. In the case study, the original proposal covered the last 30 years of research, until it became clear that there was virtually no research until 20 years ago, and the vast majority of even these studies took place in the last 10 years. The review team

⁴⁴ Although some of these would not have appeared in our search anyway because they were not in the electronic databases that were searched.

⁵ Not only does this encourage transparency, but it forces researchers to consider more decisions in advance (Lakens, 2019).

judged that these latter studies were also likely to be more relevant for the empirical component of the project given the rapid change in the nature of the digital world over time. Hence, the review eventually focused only on the period 2010-January 2020, the search being conducted at the end of that final month.

This decision also reflects the fact that the staffing and time frames available for conducting any review can itself have a bearing upon a review's parameters (Gough and Thomas, 2017). In fact, the review described here was relatively well resourced, with six researchers working on the project after the search stage, collaborating over a total period of ten months. That is not to say that reviews with fewer resources are less systematic, but that they are necessarily more focussed in order to make the review process more manageable within time limitations and resources available for the analysis.

A next step was to decide how to operationalise these search parameters. To do this, one aggregate database, *Web of Science*, was used as a testing ground to check the effectiveness of employing different search words concerning children, technologies and skills. A sample of abstracts of the items that were found by different experimental searches were evaluated to decide if these search terms did indeed produce results that were of interest and, equally significant, the extent to which they led to results that were irrelevant, that might be considered to be 'noise.' For example, searching for the words 'digital' and 'skills' as separate terms produced an enormous amount such noise. Hence, the best strategy to overcome this problem was to search for phrases like 'digital skill*'. The value of this experimental step was to avoid the later potentially time-consuming task of reading thousands of irrelevant abstracts, given the fixed resources of even this project.

Eight databases were searched both for more comprehensive coverage and to overcome the problem of disciplinary bias associated with any one database (Brunton et al., 2017). Some of these databases allowed searches of the whole text, but it became clear that this increased the chance of finding irrelevant material. One of the search terms may be present in the main text, but be mentioned just in passing rather than being a central element of that publication: e.g. 'future research might look at...' (followed by one of search words). Hence, it was decided to search only by title, keywords and abstract.

As indicated in Figure 2 above, that original search process produced 4,811 items. A majority of these still failed to meet the criteria of the project, and so a first step was to check ('screen') the titles and abstracts according to the three criteria so far discussed, plus an assessment of quality (whether the study was robust enough and had a suitable method). Before applying those criteria (e.g. about whether a study dealt with the right age range, whether a study covered children's skills), the team trained to make sure that all team members were making judgements in the same way, and established a process for adjudicating grey areas. Intercoder reliability was checked after this training. Nevertheless, a different project with a different team could have taken all these same precautions, but made different decisions about how to apply the criteria. In principle, this same observation could apply to a content analysis of, say, newspapers. In any one study there could be agreement about the particular way to operationalise the process of classifying content, and there could be tests afterwards to check

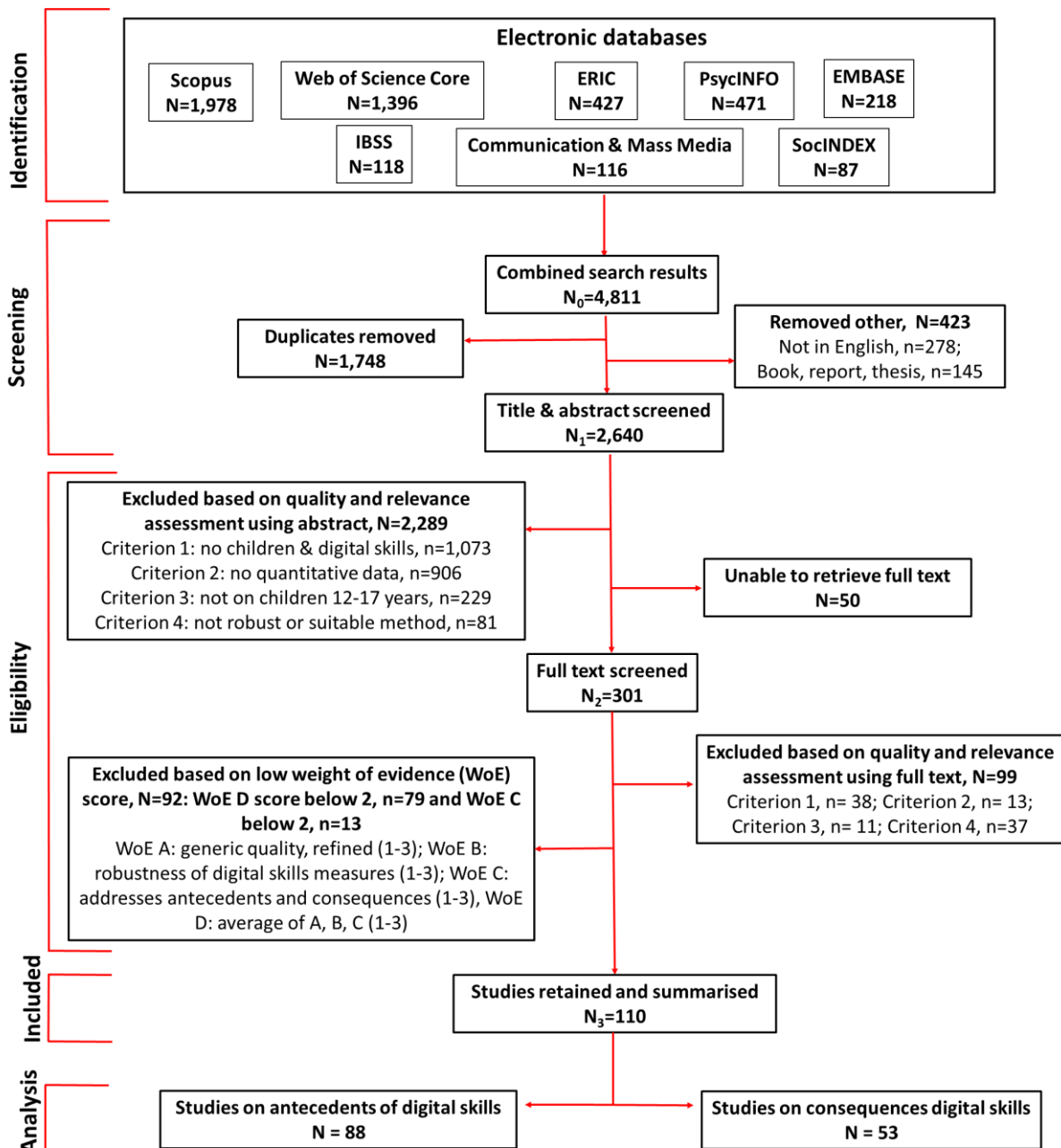


Figure 2. Flow diagram of the search and filtering processes in the mapping process

that coders coded in the same way – but a different project might have operationalised this process differently.

This stage in the screening process produced a sample of 301 items. Here, as in every stage in the process, it became clear that we had to make yet further decisions about what to include and exclude that we had not considered at the outset of the review process. For example, small-sample experiments were excluded because of the wider policy in the review of excluding studies with very small samples, even though the methodology might be robust and the risk is of losing findings that may only occur in these types of studies.

The next step involved reviewing all 301 full texts in order to score them on three criteria derived from a ‘weight of evidence’ framework (Gough, 2007). In effect, this was another quality control decision, where the team developed detailed checklists of considerations to take

into account when valuating (a) the overall quality of the methods used⁶, (b) how the study discussed and measured digital skills⁷ and (c) how the study discussed and measured antecedents (and consequences) of digital skills⁸. Each of these three criteria had a score of 1-3 (poor, fair, good), which was used to produce an average, where studies with scores above 2 overall were retained. Given that each criterion had its own list of guidelines on how to score it, this stage involved the most complex judgements by the team members in the whole filtering process. The team trained first together to make sure they scored on the same basis, set up a process of adjudication, and checked intercoder reliability, as in the previous stages. But for the same reasons noted above, there is always scope for another team to make other evaluations of which studies to include and exclude at this stage, for instance, by virtue of drawing up different guidelines for scoring the three criteria.

The final sample size after this step was 110 publications. Within the final sample, 88 studies analysed at least one antecedent of digital skills.⁹ However, even the decision about which findings represented antecedents, and which represented consequences of digital skills often involved some judgement. For instance, there were six studies that analysed learning outcomes as an antecedent of digital skills, and nine that analysed learning outcomes as a consequence of digital skills. Of those, the vast majority (85%) were based on analysis of correlations from survey data. It is therefore a matter of judgement whether to interpret these correlations as the effect of digital skills on education, the effect of education on digital skills, or some other more complicated relationship. This evidence review relied on the original study authors' interpretation of whether a particular finding represents an antecedent or consequence, but other reviewer teams may justifiably treat these cases differently.

In sum, there are many decisions shaping what material is found in the original search and then what material is subsequently selected for analysis through the filtering process. Clearly, these decisions in part reflect the goals of the particular project and the resulting inclusion criteria, they could be specified in advance and hence considered deductive modes of reasoning. But there were also unanticipated situations that arose and required solutions during the course of the research, a more inductive approach¹⁰. Finally, in filtering the material from the search process, the teams' decisions were also influenced by the practical consideration of finding a 'manageable' number of studies, that can be examined by the team of a certain size operating within certain time limits.

The coding and mapping process

Following on from data collection, this last section explains the processes at work in the analysis stages. First, it outlines how the original coding frame was developed but why and how this was adjusted once a first stage of coding had taken place. Then there are issues related to the diverse ways in which digital skills are defined and measured in the various studies: while such skills are particular to this review, the equivalent situation could emerge in other systematic reviews on different topics. Finally, the section describes the challenges that arose for a variety

⁶ Example questions included whether the study used controls, randomised representative sampling and how the data were used to test hypotheses.

⁷ For example, the clarity and complexity of the definition of skills, whether there was a theoretical explanation of skills.

⁸ For example, how well antecedents were defined, how much discussion or modelling there was to show how and why they influenced skills.

⁹ The number of papers analysing antecedents and the number analysing consequences added up to more than 110 because some papers analysed both antecedents and consequences.

¹⁰ Armat M et al. (2018), in a discussion of qualitative content analysis, argued that often research contains both inductive and deductive elements. This would apply in this case study, including in the later discussion in this article of how coding frames were initially developed but then revised.

of reasons when trying to make sense of and characterise the results for the different antecedents.

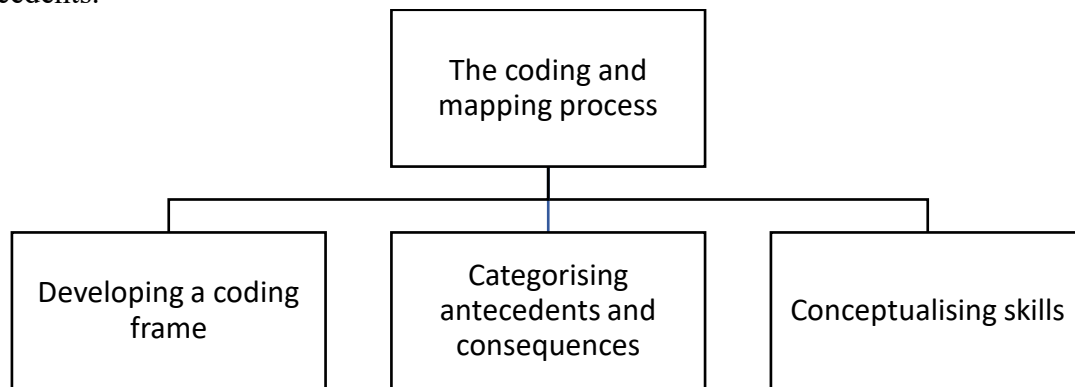


Figure 3. The coding and mapping process

Developing and completing the coding frame

Returning to the historian Carr, after providing his fish metaphor he deals with the challenge that all histories involve some interpretation of ‘the facts’. That is also reflected in the categories used by the historian, for example, dividing history into periods and then deciding their boundaries (for example, when the Middle Ages starts and ends) or dividing histories geographically (leading to questions of whether Russia should be included in Europe, for example – p.56). Carr implies that we need to be reflexive about such decisions, but regards such categorisation as a necessary process, a ‘tool of thought’ (p.55) that can be illuminating. This section reflects upon the equivalent processes of categorisation in the case of systematic evidence reviews, which can also be seen as the part of configuring process noted by Gough et al. (2012).

As a first step to reviewing the material collected, the details of each study were entered into a ‘coding frame’. This coding frame was used to record details of the study itself (e.g., the country to which it related, the method and the sample size) and details of how the study measured digital skills (e.g. were children asked for their own assessment of their skills, or were they tested?). Which antecedent was being examined in each study was also coded¹¹ as were the details of any results relating to antecedents of digital skills (e.g., whether the effect was statistically significant, and if so, the direction of the effect).

This was the stage of devising an initial tentative framework based on a field of study that was relatively well developed, in what has been called a ‘framework synthesis’ (Thomas et al., 2017). It involved making judgements about the broader categories to be employed as antecedents to digital skills (e.g., the ‘personal attributes’, ‘social context’, ‘ICT environment’) but also decisions about the particular antecedents to be included in these categories (age, gender, etc.). Examples of these antecedents were at this stage provided to the coders so that the research team members would have some idea about what each broad category covered.

Some of these elements were straightforward to code, for example, all studies clearly stated the country of the study participants. However, there were two types of challenges involved in this coding process.

First, some papers contained incomplete information about their data and methods. For instance, many of the studies in our sample did not specify the ages of participants, instead providing information only on participants’ school year, or stating that the children were in

¹¹ Some studies were more focused, dealing with a particular antecedent while others explored a wide range of factors that might influence digital skills.

‘elementary school’, ‘middle school’ or ‘secondary school’. The reviewing team followed a policy of estimating ages based on the typical ages of children within each school year in the relevant country. A different project team might have treated this information as missing¹², which could have made a difference to conclusions about the age of children, given that about a third of studies reported school rather than age per se.

Second, in coding analytical results on the associations between antecedents and digital skills, there were some practical, conceptual, and methodological challenges. For example, it was decided not to code associations between antecedents and digital skills that were evident in tables (e.g., in tables of correlation coefficients between all variables used in analysis), but which were not discussed in the actual text of the article reporting the study, although a different team may have chosen to do otherwise.

As another example, when completing the coding frame, the team chose to only record whether the direction of the effect was ‘positive or negative’ if the estimated effect was statistically significant. Some systematic reviews, particularly those that also include a meta-analysis, report the direction, precision and magnitude of findings that are not statistically significant (see, e.g., Gewandter et al., 2017).

These challenges are present in all types of systematic reviews – including in medical sciences, where systematic reviews are used more regularly. However, while many medical reviews would typically limit their scope to articles analysing the impacts of an intervention, for which most articles will have a single preferred estimate, research articles in the social sciences may report many estimated associations within a paper - as was the case in this review. It has already been noted that at the filtering stage this required the review team to make many and diverse judgements when deciding which studies were relevant to the research question. Now, at the coding stage, it was because of this diversity of frameworks and methods used in this social sciences study that the review team was unable to fully pre-define the coding framework that would be included in the review, as would be more typical in a medical systematic review (Higgins et al., 2021).

Mapping antecedents into categories

After coding each individual study, a next step in any mapping exercise involves grouping antecedents into broader categories, also known as descriptive themes (Thomas et al., 2017), in order to synthesise the studies and make statements such as ‘there are more studies about antecedent x compared with studies about antecedent y’. This was an iterative process, because it became increasingly clear that the categories used in the initial coding of the studies, as described above, did not fully align with the content of the actual studies that had been assembled – some new categories emerged that had not been anticipated, while others became less relevant. While some studies based on the original coding frame (e.g. gender) proved unproblematic, other labels, such as ‘education attainment’ were emergent, added during the next stage of the coding (or ‘re-coding’). This involved making further choices, to produce a revised, emergent framework (ibid). For example, instead of just ‘personal attributes’ the decision was made to distinguish between ‘ascribed personal attributes’, ‘achieved personal attributes’ and ‘digital personal attributes’. Then there was some re-labelling of individual antecedents in order to describe better the material actually collected (e.g., ‘vulnerabilities’ became ‘mental and physical health problems’).

Specifically, in the first round of coding there were quite a few studies where the topic did not fit into the pre-defined categories and so were classified as ‘other’. Hence, one next step

¹² Also discussed in Lakens, 2019.

in the process of developing the emergent thematic framework was to try to reduce the number of studies labelled ‘other’. One strategy was to see whether they could fit under some existing antecedent label. Another was to see whether a new antecedent could be added that would group together a few of these ‘other’ studies: for example, ‘perceptions and attitudes’ was one such new category. Yet another strategy was to see if some of the existing antecedents could be re-named (i.e., re-characterised) in order to include some of these ‘other’ studies: for example, ‘frequency of use’ of digital devices became ‘frequency and amount of use’. In other words, there was a process of developing a better, revised map of antecedents and that can be shown by comparing the original coding frame (Table 1) with the final coding frame (Table 2). This was arguably the clearest case of (re-) configuring the data. Or, returning to Carr’s fishing metaphor, this was that stage of rethinking how to classify (and then re-classify) the fish found in the ocean.

Skills issues

Before moving on to a more detailed description of the review process, it is worth noting a particular challenge arising from the nature of specific this field of study: the sheer diversity in terms of how digital skills are measured, how skill levels are defined, and which types of skills are considered to be ‘digital’.

In different studies, digital skills were measured in various ways: for example, self-efficacy measures (self-confidence in one’s ability to achieve different goals), measures of particular knowledge claims (‘I can do X’) or action taken (‘Sometimes I use an online account with a different name, so that other people believe I am a different person’) and performance tests where children are asked to demonstrate their skills. And within each of these measures there was further variation. Self-efficacy ranged from measuring narrower skills (e.g. confidence in ability to remove a virus from your computer) or general ones (e.g. being comfortable using digital devices). Performance tests had diverse formats, including requiring the child to achieve a specific goal in a simulation test or demonstrate knowledge by answering factual multiple choice questions.

Even skills could not be neatly classified as ‘basic’, ‘intermediate’ or ‘advanced’. This reflects both the age range of children (‘advanced’ skills for a 12-year old may be considered ‘basic’ or ‘intermediate’ for a 17-year old), and the skills that are tested. Instead of classifying skills into levels, some researchers focus on ‘Functional skills’, such as ability to open an email attachment or search for information online. Others focus on ‘Critical skills’, such as ability to critically evaluate information found online. While ‘functional skills’ are generally basic ones,

Table 1: Original coding frame

<i>Personal attributes</i>	Gender
	Age
	Ethnicity (including migrant background)
	Personality type
	Vulnerabilities (SEND – special educational needs and disabilities, mental health problems)
	Approach to learning (e.g. motivation, learning style)
	Interests (e.g. in science, news)
	Attitudes to computers/internet
	Digital self-efficacy (i.e. confidence in one's skills).
<i>Social context</i>	SES (including proxies like parent's education and income)
	Parental mediation (including active support)
	Other parent variables (e.g. parental attitudes to ICTs/internet, ICT competence, children's general relationship with parents, whether parents informally teach children about ICTs)
	Teacher variables (e.g. ICT competence, attitudes to ICTs, teaching methods, amount of teacher support)
	Pupil experience in the school (e.g. what ICTs are used for in class, number of lessons when children use computers, having a personalised learning curriculum, enrolment in technology related classes)
	School variables (e.g. policy, ICT support, Technological Literacy component in the curriculum)
	Peer variables (e.g. informal teaching of ICTs skills)
	Urban-rural residence
	Other community variables (e.g. after/outside school clubs for teaching ICTs)
<i>ICT environment</i>	ICT availability at home (e.g. no internet vs. shared internet; having access to a computer)
	Frequency of use of ICTs (e.g. computer, internet)
	Age of first use of ICTs (e.g. computer, internet)
	Number and type of devices used
<i>Online activities</i>	Gaming
	Use of social media/SNS
	Other activities using ICTs (learning, community participation, civic participation, creative participation, social relationship, entertainment, personal (health, support), commercial, communication)
<i>System-level</i>	Country/cultural environment (e.g. Coronavirus rates, Hofstede's cultural categories)
	Media systems

Table 2: Final coding frame

<i>Personal attributes</i>	Personal attributes with no agency	
		Age
		Gender
		Ethnicity
		Personality type
		Mental and physical health problems
		Cognitive abilities (e.g. cognitive style, reading ability)
	Personal attributes with agency	
		Educational attainment (e.g. grades)
		Leisure activities (e.g. time spent reading, religious activities)
		Approach to learning
		Interests
		Past experiences (e.g. exposure to media violence)
		Perceptions and Attitudes (e.g. attitudes to how important credible news is, perceptions of the reliability of information on the internet)
	Digitally specific attributes	
		Attitudes to computers/internet and understanding of them
		Digital self-efficacy
		Information literacy
		Evaluation method (types of sources and arguments children use to evaluate online information)
		ICT-related social engagement (e.g. “I like to talk to my friends about the current progress on computers”)
		Smartphone skills
	Other	
		Prior knowledge (Students read something relating to what they were evaluating)
<i>Social context</i>		SES
	Parental/home variables	
		Parental mediation
		Other parent variables
	School/education variables	
		Teacher variables
		Pupil experience in the school

		School variables
	Other social context variables	
		Peer variables
		Urban-rural residence
		Other community variables
		Context of acquisition of skills (e.g. learning at school, learning on their own, private lessons)
		Media literacy education
<i>ICT environment</i>		ICT availability at home
		Amount of use of ICTs (e.g. frequency of computer/internet use and amount of different types of use)
		Age of first use of ICTs
		Number, location and type of devices used
<i>Online activities</i>		Gaming
		Use of social media/SNS, use for social communication
		Other activities using ICTs
		Negative online experiences (e.g. cyber-victimisation, problematic internet use)
<i>System level</i>		Country/cultural environment
		Media systems

some may require intermediate or advanced skills, in the sense that a beginner could not achieve this goal. ‘Critical skills’ is often a version of ‘advanced’ (in the sense of multi-stage) but implying some interpretation is taking place more akin to media literacy.

Meanwhile the domain addressed by skills covered broader headings like ‘informational skills’ (e.g. searching for information), ‘social interaction skills’ (e.g. having an awareness of the conventions of social communication), ‘content creation skills’ (e.g. design and editing skills) and ‘programming/coding skills’. But some studies focused on very specific ‘skills’ such as ethical behaviour online and digital safety skills.

Sometimes, as will be demonstrated below, it is the very variation in the skills studied or how they are measured that can be used to explain differences in findings, for example, when certain measures lead to one conclusion but different measures lead to the opposite. That very diversity illustrates the value of a systematic review. But it also provides a challenge in the review stage, when seeking to make statements about the balance of findings relating to a particular antecedent. To use Carr’s fishing metaphor, it became clear that fish which initially looked the same were, in fact, different species.

The synthesis processes

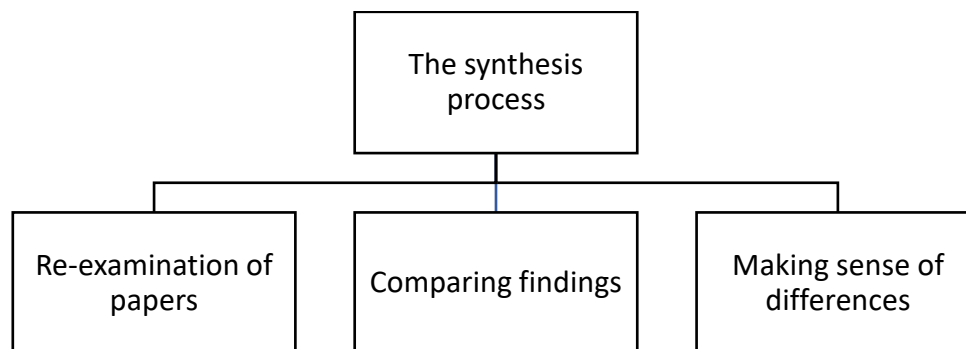


Figure 4. The synthesis process

As part of the process of synthesising findings from the 88 studies covered in this review, the original papers relating to a number of studies of antecedents were re-examined in order to understand why those researchers had looked at a particular topic. The reasons were diverse. For example, in relation to ethnicity, some looked at skills and ethnicity simply because ethnicity had been found important in other walks of life (e.g. Duarte et al., 2013). Others based their studies on previous skills research on that particular antecedent (e.g. Hatlevik et al., 2015). The case of ethnicity also exemplifies a common distinction, whereby some studies aim to test multiple possible antecedents of digital skills in order to determine which were more influential,¹³ while others focused on just one antecedent, providing a literature review to support either one or a few hypotheses.¹⁴ This exploration of the rationales of various studies helped to provide a sense of how the map of the field (here, studies of antecedents of digital skills) emerged from a combination of very different research strategies.

Since no research takes place in a cultural vacuum a second step for each antecedent was to note the pattern of countries where the research took place. Realities, values, and even epistemological stances regarding what to consider a variable of interest and how to explore it can vary across countries. It became evident that most studies from our sample were conducted in the Global North or middle- and upper-income countries, meaning that the findings from the systematic review primarily reflect Western or at least economically advanced parts of the world.

Turning now to the aggregative process of analysing results, sometimes comparing studies was relatively straightforward in spite of the heterogeneity of approaches and measures employed. For example, in the case for age the vast majority of the research indicated a positive relationship between age and digital skills, with children becoming more skilled as they grow older. Even when looking at the minority of studies of age pointing towards the opposite direction, it was easy to detect common methodological details that explained away this difference, such as the sampling strategies or the types of digital skills being measured. Other examples of antecedents where the analysis was less problematic were ‘educational attainment’, ‘learning’ (including learning motivation and learning styles) and ‘leisure activities’, where all studies found a positive correlation.

One of the dilemmas when trying to make an overall assessment of the role of a particular antecedent was that studies originated from very different analytical frameworks. For example, in the case of the antecedent ‘health’, some studies emanated from the digital divide literature (e.g. Helsper and Eynon, 2013) while other research came from addiction studies (e.g.

¹³ For example, Aesaert and van Braak, 2014.

¹⁴ For example, Macedo-Rouet et al., 2020.

Williams-Diehm et al., 2018). This is in part a product of the fact that the databases searched covered such a broad range of material. In practice, all health studies found a correlation between health and digital skills, but in order to draw conclusions and identify gaps in the research, it was still necessary to consider the different frameworks and conceptualisations, as well as the different aspects of health being studied (e.g. physical vs. mental health, learning disabilities).

Another example showing the difficulty in comparing such diverse studies is provided by the antecedent ‘personality traits’. One study came from the literature on personality tests used by employers and in career guidance,¹⁵ one came from the identity formation literature¹⁶ and a third study came from the literature on online risky activities, looking at personality traits such as self-image and risk perception.¹⁷ Given that all three studies found that certain traits were associated with greater digital skills, it is possible to say that overall personality is an influential antecedent. Yet the actual typologies of traits in these studies were totally different from each other and hence it was difficult to develop any further conclusions, apart from noting reasons for this diversity. In other words, apart from the different frameworks, the challenge for comparative analysis was that different research traditions operationalised ‘personality’ in incompatible ways.

The antecedent ‘gender’ showed how the way in which skills were measured could lead to contradictory findings. According to popular beliefs, boys outperform girls in terms of digital skills. While it is true that boys tended to score higher than girls in self-report studies, no important differences emerged when considering studies based on performance test results, where sometimes girls even outperformed boys. This suggests that a social desirability bias may be at work in studies using self-reported skill measures¹⁸, reflecting broader cultural discourses on what it means to be a boy or a girl in relation to digital technology.

At times it was not the way skills were measured but the way the antecedent was measured that produced differences in results. This can be illustrated by ‘socio-economic status’ (SES). Generally speaking, SES that was measured by parents’ education was statistically significant and positively associated with digital skills, while studies using income as the measure of SES generated mixed results. It is unclear why this difference should exist, but at least by conducting a systematic evidence review and noting this pattern, this raised a question to be addressed in further research.

It was also necessary to focus on the specificities of different studies relating to an antecedent. For example, when looking at the effect of children’s ‘personal interests’ on their digital skills, certain interests, such as an interest in science, predicted digital skills. Meanwhile other interests, like an interest in politics, did not (Lin et al., 2019). This shows how variables that were gathered under the same umbrella in the coding process because they seemed conceptually close can actually lead to different and even contrasting results. A similar effect arose from the process of trying to organise ‘other’ studies under the same antecedent headings. For example, the emergent category ‘perceptions and attitudes’ contained very diverse

¹⁵ Colvin-Sterling, 2016.

¹⁶ In this study personality types are conceptualised as identity statuses, based on different approaches by which young people evaluate and commit to future life paths (Mannerström, et al., 2018).

¹⁷ Vandoninck et al, 2010.

¹⁸ In general, since performance tests are considered more objective, their use was noted in reporting studies of different antecedents. In many cases, there was not notable differences in results from performance tests and self-reports. Even within self-report studies the way questions are framed can make a difference: it is one thing to ask children about their knowledge or ability and another thing to ask them about their ‘digital confidence’, with the latter being more likely to be overestimated.

examples.¹⁹ Unsurprisingly, the findings as to whether this antecedent correlated with digital skills were mixed depending on which perceptions and which attitudes were being examined.

A similar need to look at details arose in the case of ‘ethnicity’. While some research indicated white children (in majority-white populations) were more skilled than their non-white peers, other studies found that ethnic minority children reported more skills with respect to some specific areas (such as social entertainment skills, games and social network access) but also in the more ‘critical’ skill of evaluating the credibility of online information. In other words, by focusing on the specific skills it is possible to go beyond asking whether a particular antecedent was influential for digital skills overall, instead asking why some children may develop more of certain digital skills and less of others.

In sum, even before making comparisons between studies it is possible to contextualise the research on different antecedents by exploring the origins of research interests and the geographical location of studies. While some comparisons of studies were then relatively unproblematic, others required additional analysis to take into account the bodies of literature from which they stemmed (and hence their conceptualisations of antecedents), the operationalisation of those antecedents, and the consequences of how both skills and antecedents were measured. Sometimes the insights from this process came from the team’s own judgements based on their reading of the details of how both antecedents and skills were treated in different studies, rather than a direct interpretation of the results of our systematic coding framework, which reported more crudely simply on whether a statistically significant relationship existed.

Overall, when trying to understand the patterns of studies when making so many comparisons, especially in the light of the sheer range of studies returned by this search strategy, a certain amount of reflexivity was required. That diversity of approaches employed in the studies meant that it was possible to go beyond charting which antecedents were important to appreciate a more complex picture. This led to a more nuanced mapping of the field and indicated some lines for future enquiries.

Conclusion

This case study of a particular systematic evidence review provides a worked example. This article elaborates in some detail for a wider audience on the challenges faced and strategies used, in what is itself a research process (Gough et al., 2012). Carr’s fishing metaphor, initially intended to encourage reflection upon historical research, was explored here, and extended, in order to stand back from and reflect up the various steps in this exercise.

Although systematic reviews can vary in their nature (as demonstrated in the collection of Gough et al., 2017b), common guidelines were followed, at each point examining the nature of the judgement being made and its possible consequences. For example, inclusion and exclusion criteria were mainly derived from the goals of the wide ySKILLS project, but included decisions about language and types of publication, as well as re-setting date parameters in the light of data from the exploratory phase. And within the process of selecting the studies to be reviewed, the numerous micro-decisions may well provide some limitations to how much reviews can be reproduced or replicated even when the reasons for all the main choices are made explicit.

The process of setting up the initial coding frame and theme and then the basis upon which it was revised was explained. And the issues arose when trying to conduct the synthesis

¹⁹ Helsper and Eynon (2013) looked at how much children feel in control of their lives; Kiili et al.(2018) examined children’s beliefs and prior knowledge about a commercialised item; Nygren and Guath (2019) considered if the credibility of news online was important to children; Jean, Subramaniam, Taylor et al (2015) focused on children’s pre-existing knowledge about a topic; Metzger et al (2015) covered children’s ‘comparative optimism’.

– requiring further reflection on the variation in definitions and measurement of both the main object being studied (digital skills) and variables affecting it. Relevant factors to consider included how the motivations behind the original research affected the map that arose, and how the construction and definition of key variables varied depending on the analytical frameworks used in the original research.

This process of reflection was not meant to imply this particular piece of research was somehow flawed. Within the wider ySKILLS project the sheer detail generated by this review was appreciated. Instead, the goal of this article is more generic, to provide examples of the different decisions that have to be made throughout this type of research process, with an eye to increasing reflexivity about all the steps that lead to a finalised product. By providing specific examples of the types of challenges, judgements required, and strategies employed in this case study, this article highlights some of the factors that teams conducting a systematic review for the first time may wish to consider.

Declaration of Conflicting Interests

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

Funding

The research reported here has received funding from the European Union's Horizon 2020 Research & Innovation Programme under Grant Agreement no. 870612.

Acknowledgements

We thank our colleagues and our reviewers.

References

Aesaert K and van Braak J (2014) Exploring factors related to primary school pupils' ICT self-efficacy: A multilevel approach. *Computers in Human Behavior* 41: 327–341.

Armat M, Assaroudi A, Rad M, Hassan S and Heydari A (2018) Inductive and deductive: Ambiguous labels in qualitative content analysis. *The Qualitative Report*, 23(1): 219-221.

Atkinson LZ and Cipriani A (2018) How to carry out a literature search for a systematic review: A practical guide. *BJPsych Advances*, 24, 78-82.

Brunton G, Stansfield C, Caird J and Thomas J (2017). Finding relevant studies. In: Gough D, Oliver S and Thomas J (eds.) *An introduction to systematic reviews*, Second Edition. London: Sage, 93-122.

Caronia L (2011) *Fenomenologia dell'educazione: intenzionalità, cultura e conoscenza in pedagogia* (Phenomenology of education: intentionality, culture and knowledge in pedagogy). Milano: Franco Angeli.

Carr E (1987) *What is history?* (2nd ed.). London: Penguin.

Clifford J and Marcus GE (eds) (1986) *Writing Culture. The poetics and politics of ethnography*. Oakland: University of California Press.

Colvin-Sterling S (2016) The correlation between temperament and technology preference and proficiency in middle school students. *Journal of Information Technology Education: Research* 15: 1–18.

Duarte R, Cazelli S, Migliora R and Coimbra C (2013) Computer skills and digital media uses among young students in Rio de Janeiro, *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas* 21: 1–29.

Dwan K, Gamble C, Williamson PR and Kirkham JJ (2013) Systematic review of the empirical evidence of study publication bias and outcome reporting bias: An updated review. *PLoS ONE*, 8(7): e66844

Gewandter JS, McDermott MP, Kitt RA, Chaudari J, Koch JG, Evans SR, Gross RA, Markman JD, Turk DC and Dworkin RH (2017) Interpretation of CIs in clinical trials with non-significant results: systematic review and recommendations. *BMJ Open* 2017, 7:e017288.

Gough D (2007) Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education* 22(2): 213-228.

Gough D and Thomas J (2017) Commonality and diversity in reviews. In: Gough D, Oliver S and Thomas J (eds.) *An introduction to systematic reviews*, Second Edition. London: Sage. 43-70.

Gough D, Oliver S and Thomas J (2017a) Introducing systematic reviews. In: Gough D, Oliver S and Thomas (eds) *An introduction to systematic reviews*, Second Edition. London: Sage. 1-18.

Gough D, Oliver S and Thomas, J. (eds) (2017b) *An introduction to systematic reviews*, Second Edition. London: Sage.

Gough D, Thomas J and Oliver S (2012) Clarifying differences between review designs and methods. *Systematic Reviews* (28): 1-9.

Grant MJ and Booth A (2009) A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91-108.

Haddon L, Cino D, Doyle MA, Livingstone L, Mascheroni G and Stoilova M (2020) *Children and young people's digital skills: A systematic literature review*, ySKILLS, LSE, London.

Helsper EJ and Eynon R (2013) Distinct skill pathways to digital engagement. *European Journal of Communication* 28(6): 696–713.

Hatlevik OE, Guðmundsdóttir GB and Loi M (2015) Examining factors predicting students' digital competence. *Journal of Information Technology Education: Research* 14(14): 123–37.

Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (eds) (2021) *Cochrane Handbook for Systematic Reviews of Interventions version 6.2*. Available from www.training.cochrane.org/handbook

Jean BS, Subramaniam M, Taylor NG, Follman R, Kodama C and Casciotti D (2015) The influence of positive hypothesis testing on youths' online health-related information seeking. *New Library World* 116(3/4): 136–154.

Kiili C, Leu DJ, Marttunen M, Hautala J and Leppänen PH (2018) Exploring early adolescents' evaluation of academic and commercial online resources related to health. *Reading and Writing* 31(3): 533–557.

Lakens D (2019) The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review* 62(3): 221-230.

Macedo-Rouet M, Salmeron L, Ros C, Pérez A, Stadler M and Rouet JF (2020) Are frequent users of social network sites good information evaluators? An investigation of adolescents' sourcing abilities. *Journal for the Study of Education and Development* 43(1): 101–138.

Mannerström R, Hietajarvi L, Muotka J and Salmela-Aro K (2018) Identity profiles and digital engagement among Finnish high school students. *Cyberpsychology*, 12(1): 1-15.

Metzger M, Flanagin A, and Nekmat E (2015) Comparative optimism in online credibility evaluation among parents and children. *Journal of Broadcasting & Electronic Media* 59(3): 509–529.

Nef T, Ganea R, Muñiri R and Mosimann U (2013) Social networking sites and older users – a systematic review. *International Psychogeriatrics* 25(7): 1041–1053.

Nygren T and Guath M (2019) Swedish teenagers' difficulties and abilities to determine digital news credibility. *Nordicom Review*, 40(1): 23–42.

Oakly A (2017) Foreword. In: Gough D, Oliver S and Thomas J (eds) *An introduction to systematic reviews*, Second Edition. London: London. xii-xvi.

Oliver S, Dickson K, Bangpan Mand Newman M (2017). Getting started with a review. In: Gough D, Oliver S and Thomas J (eds.) *An introduction to systematic reviews*, Second Edition. London: Sage. 71-92

Schreuder A, van Deursen A and van Dijk J (2017) Determinants of Internet skills, uses and outcomes: a systematic review of the second- and third-level digital divide. *Telematics and Informatics* 34(8): 1607–1624.

Thomas J, O'Mara-Eves A, Kneale D and Shemilt I (2017) Synthesis methods for combining and configuring textual or mixed methods data. In: Gough D, Oliver S and Thomas J (eds.) *An introduction to systematic reviews*, Second Edition. London: Sage. 181-210

van Laar E, van Deursen A, van Dijk J and de Haan J (2017) The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior* 72: 577-588.

Vandoninck S, d'Haenens L and Donoso V (2010) Digital literacy of Flemish youth: How do they handle online content risks? *Communications* 35(4): 397–416.

Vrieze J (2018) Meta-analyses were supposed to end scientific debates. Often, they only cause more controversy. *Science* 18: article 4515.

Williams J (2019) The use of online social networking sites to nurture and cultivate bonding social capital: A systematic review of the literature from 1997 to 2018. *New Media and Society* 21(11-12): 2710–2729.

Williams-Diehm K, Miller CR, Sinclair TE and Wronowski ML (2018) Technology-based employability curriculum and culturally diverse learners with disabilities. *Journal of Special Education Technology* 33(3): 159–170.

Wicherts J, Veldcamp C, Augusteijn H, Bakker M, van Aert R and van Assen M (2016) Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid P-hacking. *Frontiers in Psychology* 7: Article 1832.