



A five-step guide for validating umbrella reviews with bibliometric and network indicators

Klaudijo Klaser¹ · Paolo Casari² · Roberta Cuel¹

Received: 4 September 2024 / Accepted: 23 December 2025
© The Author(s) 2026

Abstract

This paper presents a simple five-step guide for validating umbrella reviews—or literature reviews of existing literature reviews—using bibliometric and network indicators. The recent and unprecedented proliferation of scholarly output entails a need for robust review protocols to synthesize and systematize scientific publications, as well as tools that are suitable for assessing review quality, such as the one presented in this paper. Adopting our five-step guide as a final robustness check will improve the rigor, transparency, and objectivity of umbrella reviews and, therefore, strengthen their representativeness of the knowledge base within a specific research domain. We apply the developed five-step guide to an existing umbrella review on flexible work arrangements as a case study. Finally, we discuss the role of artificial intelligence in the expanding scientific publication landscape. We highlight how human judgment will remain a pivotal element when using algorithmic tools to support literature reviews.

Keywords Artificial intelligence · Bibliometric analysis · Literature review · Network analysis · Methodology · Umbrella review

Communicated by Leif Brändle

✉ Klaudijo Klaser
klaudijo.klaser@unitn.it

¹ Dipartimento di Economia e Management, Università degli Studi di Trento, via Inama 5, 38122 Trento, Italy

² Dipartimento di Ingegneria e Scienza dell'Informazione, Università degli Studi di Trento, via Sommarive 9, 38123 Trento, Italy

1 Introduction

According to some recent estimates, more than 64 million scholarly texts have been published between 1996 and the end of 2022. In 2022 alone, over 5 million scientific articles were published—an increase of about 23% compared to 2018 and 47% compared to 2016 (Curcic 2023; Hanson et al. 2024).

These few figures demonstrate the recent exponential expansion of scientific publications (Thelwall and Sud 2022) without yet fully capturing the effects of artificial intelligence, which is expected to further accelerate the trend (Kobak et al. 2025). Therefore, it has become increasingly necessary to develop effective practices, such as reviews, to synthesize and systematize the vast body of knowledge represented in scientific publications. This need arises not only from the imperative to consolidate existing theories, but also the opportunity to identify potential research gaps and cross-fertilization areas across scientific domains (Ooms et al. 2023).

Given these developments, it is unsurprising that the practice of writing literature reviews—together with efforts to develop methodological contributions to guide and improve their preparation—is gaining momentum as a standalone and independent research practice¹ (Kraus et al. 2022; Snyder 2019). Scholars have developed a broad spectrum of methodologies and techniques to synthesize and systematize scientific publications (Fan et al. 2022; Onwuegbuzie and Frels 2016; Snyder 2019; Sutton et al. 2019). As a general descriptive scheme, Table 13 in the Appendix summarizes some widely recognized and embraced literature review methods.

As a direct consequence of these developments, umbrella reviews—or literature reviews of existing literature reviews—have become a widespread practice. The proliferation of literature reviews, especially in the health domain, led to the development of specific guidelines and protocols to ensure rigorously conducted umbrella reviews (Abdellatif et al. 2025; Fernandez et al. 2025; Smith et al. 2011). Despite these advancements, to the best of our knowledge to date, neither practices nor tools exist for umbrella reviews that can objectively validate the transition from a formal protocol to a substantial result in terms of knowledge representativeness. In other words, no comprehensive instruments currently exist to assess the overall quality of umbrella reviews—an equivalent of the AMSTAR tool developed for systematic reviews (Shea et al. 2007, 2017).

Therefore, in this rapidly evolving publication landscape, umbrella reviews' capacity to accurately capture and represent the underlying knowledge base emerges as a critical research gap and an equally fundamental research question arises: Is there any way to assess the final quality of an umbrella review aside from its adherence to a formal protocol?² We address this research question and contribute to closing the

¹ According to the Scopus database, in 2022, more than 10,000 literature reviews were published (Kraus et al. 2022).

² When conducting a systematic (umbrella) review, rigorous adherence to the formal criteria included in the protocol is essential, and any deviation should be explicitly explained and addressed (Page et al. 2021; Smith et al. 2011). Complementarily, in substantial terms, a systematic (umbrella) review must accurately reflect the knowledge base derived from the examined primary sources under examination. By construction, adherence to the former is often assumed to guarantee the latter. In other words, the quality guarantee of systematic (umbrella) reviews usually lies in their protocol itself being developed to avoid

related research gap by proposing a method we have operationalized into a five-step guide for validating umbrella reviews from a substantive perspective.

At the core of our method lies the idea of leveraging the strengths of various existing review practices and combining them to assess an umbrella review's final quality. Therefore, as here designed, the guide we propose does not constitute an expanded and integrated functional protocol of different review methods that contributes to the review content³ (Linnenluecke et al. 2020; Marzi et al. 2025). Our guide has no explicit purpose for content addition, other than residually. Instead, our guide constitutes a corollary tool that follows the umbrella review and serves to assess its final quality in terms of knowledge representativeness.

As a paradigmatic case study of how to apply the five-step guide, we build on an existing umbrella review about flexible work arrangements (FWAs) conducted by Toscano and Zappalà (2020). To validate this umbrella review, we performed a comparative analysis using two distinct sets of publications that are typically available when an umbrella review is conducted. On the one hand, we used the formal criteria (e.g., keywords, time span, subject area, etc.) necessary to construct the umbrella review. By leveraging these criteria, we can directly identify a parallel set of primary scientific publications that serve as a control group (direct identification method). On the other hand, we have used the complete list of publications (i.e., references) embedded in the literature reviews that are included in the umbrella review (indirect identification method). We quantitatively and qualitatively compared (Tantardini et al. 2019) these two datasets via bibliometric and network indicators obtained from the software packages VOSviewer (Van Eck and Waltman 2007; Eck and Waltman 2010, 2014) and Gephi, respectively.

The remainder of this article is divided into three main sections. Section 2 outlines the five-step guide for validating umbrella reviews. This includes a descriptive overview of the bibliometric and network indicators used for the comparative analysis between the direct and indirect identification methods. Section 3 presents the umbrella review selected as a case study (Toscano and Zappalà 2020) and applies the five-step guide to it. The results of this section are summarized in the final synoptic Table 12. Finally, Sect. 4 summarizes our main contributions and examines some limitations of our method. This section also includes a broader discussion of how algorithmic tools adopted to support literature reviews still require a pivotal human component.

arbitrariness and ensure rigor, transparency, and replicability, and, consequently, produce objective and unbiased results in terms of knowledge representativeness.

³ Similarly, we do not question or scrutinize the protocols of building umbrella reviews per se—this aspect has been addressed in other studies, including Lunny et al. (2017, 2018).

2 Five-step guide for validating umbrella reviews

In this section, we outline the five-step guide for validating umbrella reviews through bibliometric and network indicators. In Sect. 3, we apply this five-step guide to the selected case study (Toscano and Zappalà 2020) and, in that context, describe and comment on each step in detail.

2.1 Step 1: umbrella review selection or writing

Step 1 entails either selecting an existing umbrella review or identifying a research domain for conducting an umbrella review to validate. Although this step may seem superfluous, explicitly specifying the need to select or produce an umbrella review is essential for at least two reasons. First, it is important to verify whether an umbrella review on the topic already exists to determine whether we must conduct one. Second, we must clarify the motivations underpinning the umbrella review, such as the rationale for selecting a specific research domain or topic, particularly for validating a published umbrella review.

2.2 Step 2: dataset construction

Step 2 requires the construction of two distinct publication datasets that will be compared:

1. *Direct identification method.* This dataset is created by searching for primary publications that meet the same criteria (e.g., keywords, time span, subject area, etc.) used to select the literature reviews included in the umbrella review. This search should exclude reviews (to avoid loops), and the resulting dataset will serve as the control group.
2. *Indirect identification method.* This dataset includes all the references cited in the literature reviews that the umbrella review encompasses. It constitutes the set of publications indirectly selected for the umbrella review's knowledge synthesis.

Both datasets must be cleaned (e.g., removing duplicates, fixing structural errors, etc.). At this stage, other filters can also be applied, such as excluding publications that are not peer-reviewed or that are not research articles. The construction of the datasets requires rigor, as any oversight in this step will be reflected in biased results. Moreover, different umbrella reviews might require specific adjustments, which must be explicitly documented. These choices will mostly depend on the research domain under investigation.

2.3 Step 3: bibliometric and network indicators and software selection

In *Step 3*, a set of bibliometric and network indicators necessary for comparing the two publication datasets constructed in *Step 2* must be identified. Concurrently, the software programs needed to calculate the selected indicators must be selected. In the absence of standard protocols, researchers retain full discretion in choosing both

the sets of indicators and the software, especially as these choices may depend on the specific characteristics of the umbrella review under examination.

2.4 Step 4: indicator calculation and preliminary similarity assessment

Step 4 entails actively processing the two datasets and calculating the outputs or values of the selected indicators for each dataset. In this step, an initial assessment of the degree of similarity between the indicators is conducted, and both convergences (which indicate quality) and divergences (which may reflect scope, expertise, or other biases in the umbrella review) are noted for each indicator.

2.5 Step 5: discussion of the overall similarity between the two datasets and final assessment

Step 5 is a final, overall assessment of the quality of the umbrella review. Broadly, the basis of the comparison in this step is: If the majority of indicators calculated on the list of publications identified indirectly coincide with the indicators based on the list of publications identified directly (i.e., the control group), the umbrella review can be considered valid and, therefore, representative of the knowledge base underpinning the literature under investigation. In our case, this would mean that Toscano and Zappalà (2020) have provided an accurate overarching representation of FWAs (until 2017), with no significant omissions of fundamental literature. Conversely, if the two datasets do not produce similar indicators, major discrepancies may exist, and the umbrella review should be considered ineffective in illustrating the key elements of the literature under consideration. In the latter case, it might be appropriate to assess whether the differences originate from methodological flaws or deliberate scope choices, as well as to consider using the information from the two datasets to complement each other and integrate the umbrella review.

3 Application of the five-step guide to a case study

3.1 Step 1: umbrella review selection or writing

In 2020 and 2021, the global COVID-19 pandemic and resulting lockdown measures disrupted daily routines and generated unprecedented changes in working tasks. Given the magnitude of the phenomenon—with millions of individuals suddenly and involuntarily involved in flexible work arrangements (FWAs)—researchers quickly and widely explored this historical shift (Klaser et al. 2023; Leonardi 2020). The number of publications on the topic of FWAs rapidly increased⁴ (Bailey and Breslin 2021; Cuel et al. 2022; Lamovšek and Černe 2023). Contextually, a synthesis and systematization of the existing (fragmented) literature was necessary to support the

⁴ According to an elementary search on the electronic Scopus database (December 1, 2025), in 2019, 35 articles with the terms *telework*, *telecommuting*, *remote work*, or *hybrid work* in the title were published. The same search produces 97 results in 2020, 218 in 2021, 338 in 2022, and over 400 in 2023 and 2024.

new wave of scholarly attention on the subject. Intercepting this demand, Toscano and Zappalà (2020) conducted an umbrella review.

As the guide indicates, the motivations underpinning the choice of umbrella review are relevant. First, Toscano and Zappalà (2020) reacted swiftly to the need to systematize the existing literature in the growing domain of FWAs research. Second, theirs is among the first umbrella reviews outside of the health sciences.⁵ Because of this (forced) celerity and the (new) research domain, validating their umbrella review is of the utmost importance.

3.2 Step 2: dataset construction

For their umbrella review, Toscano and Zappalà (2020, p. 206) adopted three filters to identify relevant literature reviews on FWAs. They considered reviews:

- Published between 2000 and 2018.
- In the fields of psychology or management.
- Including, in the title or⁶ abstract, the keywords *telecommuting*, *telework*, *remote work*, *flexible work*, or *flexplace*, and any of these accompanied by the term *review*.

For their umbrella review, Toscano and Zappalà (2020) identified a total of 16 relevant literature reviews (Table 1). The right column of Table 1 reports the total number of primary references for each literature review according to Scopus at the moment of our analysis.⁷ For, Allen et al. (2015) considered 193 primary publications in their literature review, while De Lorenzi Cancelier et al. (2017) utilized a comparatively lower number of references (26). An important aspect to consider during this validation step is that some primary references might not be strictly functional for the main goal of reviewing the existing literature. Instead, they might be methodological or seminal publications that play a contextual role.

To assess the quality of Toscano and Zappalà's (2020) umbrella review, we constructed and then compared two different datasets of primary publications:

- *Direct identification method.* On Scopus, we applied the same filters originally adopted by Toscano and Zappalà (2020, p. 206), excluding the word *review*. In other words, we looked for all the publications on Scopus that contained the keywords *telecommuting* or *telework* or *remote work* or *flexible work* or *flexplace* within their titles or abstracts. Drawing on the original criteria, we limited the

⁵ Toscano and Zappalà's (2020) early attempt—together with a few other recent examples (Johnson et al. 2022; Pahlevan-Sharif et al. 2019; Shehata et al. 2023)—could mark an emblematic passage to the broader adoption of umbrella reviews outside of the health sciences (Kitchenham et al. 2009), extending this review practice to other research domains, including the social sciences (Zupic and Čater 2015).

⁶ Toscano and Zappalà (2020, p. 206) technically write “in the title and in the abstract,” but the Boolean connector “and” seems to be incorrect, as the (key)word *review* often fails to appear in the titles of the review articles they used. Instead, it appears only in the abstracts. Therefore, we interpreted this as a typo and deduced that the correct Boolean connector was “or.”

⁷ In the meantime, further references may have been indexed on the Scopus database.

Table 1 Literature reviews based on Toscano and Zappalà (2020)

Literature review article	Total primary references
Allen et al. (2015)	193
Bailey and Kurland (2002)	103
Baker et al. (2006)	26
Baruch (2001)	111
Bélanger et al. (2013)	84
Kowalski and Swanson (2005)	60
Bessa and Tomlinson (2017)	53
De Lorenzi Cancelier et al. (2017)	26
De Menezes and Kelliher (2011)	206
Golden (2009)	60
Martin and MacDonnell (2012)	58
Ng (2016)	48
Pyöriä (2011)	53
Shin, El Sawy, Sheng, and Higa (2000)	65
Siha and Monroe (2006)	163
Spreitzer, Cameron, and Garrett (2017)	153

search to the two subfields of psychology and management. Unlike the original filters, we omitted the lower bound year. Moreover, we introduced an upper time bound, restricting our search to contributions published until 2017, the year of the last systematic reviews identified by Toscano and Zappalà (2020).⁸ Through this search, after cleaning some duplicates (7), we identified 858 unique publications, of which 659 (77%) were classified as research articles published across 318 different journals. Chess and Lyman (1969) produced the oldest article in this dataset, cited only eight times.⁹ We define this procedure as the “direct identification method” because it identifies a set of primary publications directly through the original keywords used in the original umbrella review. The resulting dataset constitutes the control group for our comparative analysis.

- *Indirect identification method.* As indicated in the guide, we created a parallel dataset with all the references retrieved from the 16 literature reviews (Table 1) selected by Toscano and Zappalà (2020). We downloaded this list of references from the Scopus database, as all 16 original reviews were indexed on it. In principle, this step requires no other filters. With this procedure, we identified an initial set of 1,462 publications. From these, we discarded the duplicates (269). There were substantially more duplicates than those identified through the direct method. However, this reflects the different search processes construction. The references of multiple literature reviews on the same topic are very likely to feature the same set of primary publications. Cleaning duplicates reduced the dataset to 1,193 single records, of which 573 (48%) were classified as research articles and published across 230 different journals. Trist and Bamforth (1951)

⁸ De Lorenzi Cancelier et al. (2017) and Spreitzer et al. (2017). These reviews cannot refer to contributions published in 2018 or later. As mentioned in the guide, specific adjustments might be needed to make the two datasets fully comparable. In this case, the time filter used for the umbrella review (from 2000 to 2018) and the criterion used to circumscribe the direct set of primary references for our validation (upper bound set to 2017) needed to be distinct.

⁹ We always consider the citations recorded by the Scopus database.



Fig. 1 Flowchart of the construction of the direct and indirect datasets for comparison

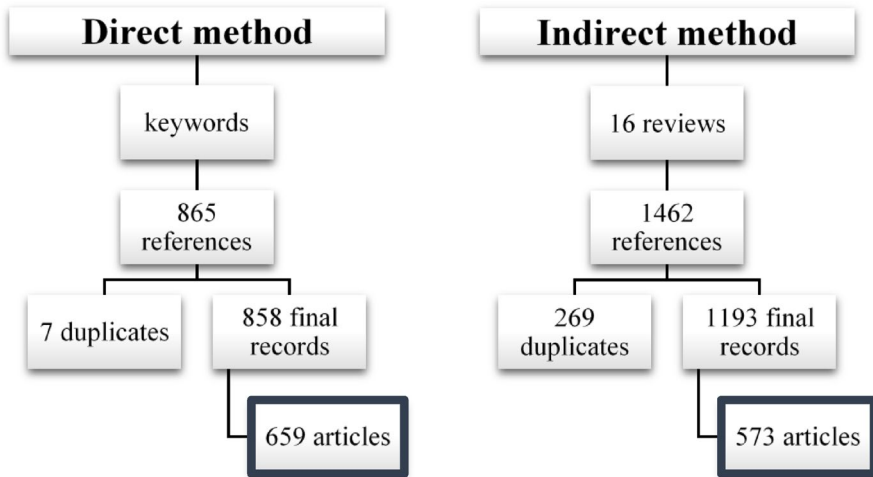


Fig. 2 Summary of the identification of our datasets

produced the oldest article in this dataset, which was cited 1,421 times. We define this procedure as the “indirect identification method” as it identifies a publication set through the reference lists of the literature reviews included in the umbrella review.

As indicated in the guide, further specific adjustments or methodological choices might be possible or necessary at this point. For example, we decided to focus on the subset of publications classified by Scopus as research articles. This choice offered advantages and disadvantages: On the one hand, excluding other types of publications reduces the risk of spurious comparisons (e.g., comparing articles identified via the direct method with books identified via the indirect method). On the other hand, the comparative picture might be incomplete because it might lack the fundamental shades added by various types of publications. Such concerns must be addressed and balanced in each umbrella review, depending on the rationale for the validation. This time we opted for a prudential approach for our validation. Therefore, to guarantee as much homogeneity as possible between our datasets, we considered only research articles as units of analysis. In total, 659 and 573 publications were identified for the direct and indirect datasets, respectively.

The main sub-steps of the two methods for our analysis are presented in Figs. 1 and 2.

3.3 Step 3: bibliometric and network indicators and software selection

To assess the quality of Toscano and Zappalà's (2020) umbrella review, we must compare the two publication datasets through bibliometric and network indicators. As previously mentioned in the guide, so far, there is no standardized validation protocol, and the specific sets of indicators and software might be context-dependent. For our case study, we propose some possible indicators, gathering them into three main categories. For the sake of synthesis and clarity, we only list the relevant indicators here. We describe them in detail in *Step 4*.

Performance indicators (Ayoko et al. 2022):

- Number of publications.
- Number of shared publications between datasets.
- Number of publications per year.
- Total and average number of citations.
- Number of citations per year.
- The 20 most cited publications.
- The 20 most time-weighted cited publications.

*Bibliometric indicators and science mapping*¹⁰ (Caputo et al. 2021; Donthu et al. 2021; Ferreira 2018; Torres-Pruñonosa et al. 2020):

- Bibliographic coupling (articles, journals, and authors).
- Total link strength (articles and authors).
- Co-citation (articles and journals).
- Keyword co-occurrence.

Network indicators:

- Number of connected nodes.
- Number of edges.
- Number of clusters.
- Average degree (the ratio between the number of edges and the number of nodes).
- Graph density (the ratio between the actual number of edges and the maximum number of edges that a network can potentially contain).
- Modularity (within-cluster node connection with respect to between-cluster node connection).
- Average clustering coefficient (average degree to which nodes in a network tend to cluster together).
- Average shortest path length (average number of steps along the shortest paths for all possible pairs of network nodes).
- Betweenness centrality (number of times a node is on the path between two other

¹⁰ Bibliometric analysis permits the conversion of hundreds or even thousands of publications into a few indicators or network charts that summarize the performance and map the evolution of a given research domain (Donthu et al. 2021).

nodes).

- Closeness centrality (average of the farness—that is, inverse distance—of a node to all other nodes).

From the various available software for bibliometric analysis (Aria and Cuccurullo 2017; Moral-Muñoz et al. 2020), we selected VOSviewer (Van Eck and Waltman 2007; Eck and Waltman 2010, 2014), which provides simplified visualizations, or maps (i.e., clusters), of the relationships between the constituents under investigation. For network metrics, we adopted Gephi (Bastian et al. 2009).

3.4 Step 4: indicator calculation and preliminary similarity assessment

In this step, we present the bibliometric and network descriptive results obtained from the two datasets. Contextually, we provide a preliminary assessment of the similarity and dissimilarity of the datasets. We provide an overall assessment of the quality of the umbrella review in *Step 5*.

3.4.1 Performance indicators

Through the direct and indirect methods, we identified 659 and 573 research articles, respectively. At first glance, these absolute numbers are comparable.

The two datasets have 156 articles in common. This means that only 15% of the total number of unique research articles (1,076) belong to both datasets. Although this overlap is not negligible, there is relatively little shared material.

Figure 3 presents the total number of publications per year for each dataset. The paths of the two datasets follow a broadly similar pattern until 2012. After this year, the number of publications diverges, with a noticeable decrease in the number of articles identified indirectly. However, this divergence should be considered a structural time lag, rather than a methodological bias. By construction, literature reviews are published after primary publications, so the review authors have no access to newly

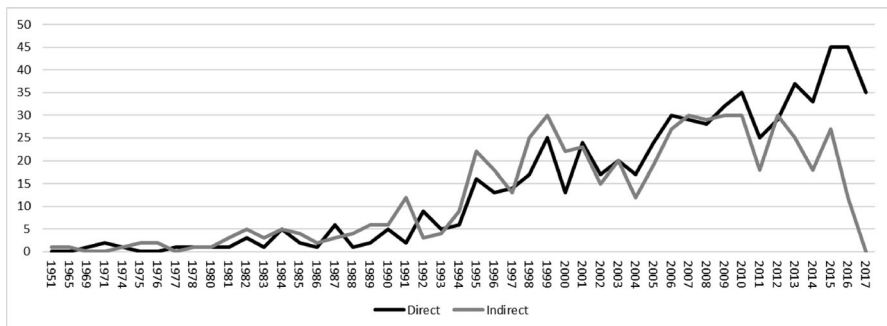


Fig. 3 Number of publications per year (1951–2017). Visualization. Number of publications per year (1951–2017) for the direct (black line) and indirect (gray line) datasets. By construction, the number of publications per year is fairly similar until 2012, when the direct dataset begins to surpass the number of publications included in the indirect dataset. There is a discontinuity in the x-axis of the years before 1980 because years when there were no publications in either dataset were excluded from the figure

(or subsequently) published research articles. The comparison might be improved by artificially truncating the examined time period at the year in which the two datasets begin to diverge.

The total (average) number of citations¹¹ of the articles identified directly is 26,785 (41). The total (average) number of citations of the articles identified indirectly is 114,851 (200). The citation count of the indirect dataset is four to five times greater than that of the direct dataset, indicating potential differences between them.

Figure 4 plots the absolute number of citations per year for each dataset. A different citation path is evident for the two datasets, with the direct dataset following a more linear path. Moreover, two notable citation peaks for the indirect dataset emerge in 1995 and 1998. Again, this divergence in citations might reflect a structural difference and should not immediately be interpreted as indicating bias in the methodology: literature reviews often rely on methodological or seminal publications that are not strictly linked to the domain of interest—FWAs, in this case—but which may be repeatedly cited (see Table 2).

Table 2 ranks the 20 most-cited articles according to the Scopus database. The most-cited direct article is Olson and Olson (2000), cited 1,119 times. The most-cited indirect article is Baumeister and Leary (1995), with 10,727 citations. The data presented in Table 2 suggests three empirical regularities: (1) the 20 most relevant articles of the indirect method accrued about ten times more citations than the equivalently ranked articles identified through the direct method; (2) the most-cited research article from the direct dataset has almost the same number of citations as the twentieth most-cited article of the indirect dataset, and the most-cited indirect article achieved more citations than the total of the whole direct dataset; (3) the two rankings have no articles in common. The other citation patterns might be explained by the fact that literature reviews likely include in their references (indirect method) methodological contributions or seminal articles in their references (indirect method) that are not identified through the direct method using subject-specific keywords.

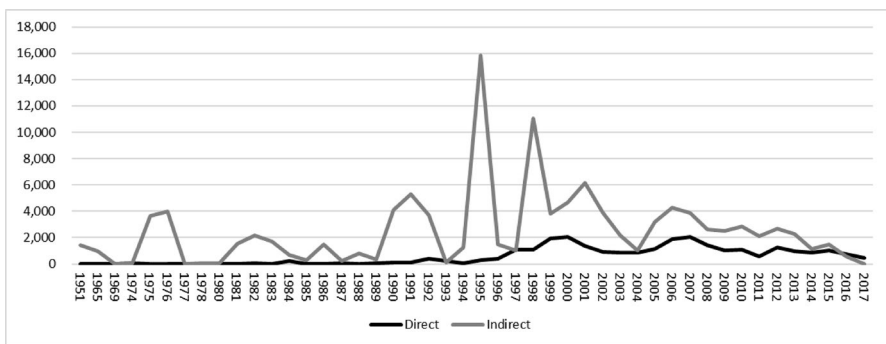


Fig. 4 Total number of citations per year (1951–2017). Visualization. Number of citations per year (1951–2017) for the direct (black line) and indirect (gray line) datasets. The indirect dataset shows peaks in 1995 and 1998. There is a discontinuity in the x-axis of the years before 1980 because years when there were no publications in either dataset were excluded from the figure

¹¹ All citation metrics reflect those recorded by the Scopus database until March 16, 2022.

Table 2 The 20 most relevant articles ranked by absolute number of citations

N°	Direct method	Citations	Indirect method	Citations
1	Olson and Olson (2000)	1,199	Baumeister and Leary (1995)	10,727
2	Gajendran and Harrison (2007)	706	Nahapiet and Ghoshal (1998)	9,633
3	Gilmore and Pine (1997)	600	Meyer and Alien (1991)	4,226
4	Baltes et al. (1999)	485	Hackman and Oldham (1976)	3,971
5	Kossek et al. (2006)	448	Davis et al. (1992)	3,611
6	Cascio (2000)	351	Hackman and Oldham (1975)	3,554
7	Lapierre and Allen (2006)	335	Goldberg (1990)	3,497
8	Allen et al. (2013)	304	Goodhue and Thompson (1995)	2,908
9	Malone and Laubacher (1998)	267	Harter et al. (2002)	2,101
10	Straub and Karahanna (1998)	254	Quinn and Rohrbaugh (1983)	1,608
11	Hill et al. (1998)	251	Kimberly and Evanisko (1981)	1,478
12	Hornung et al. (2008)	234	Sproull and Kiesler (1986)	1,446
13	Staples et al. (1999)	232	Akerlof (1982)	1,425
14	Cooper and Kurland (2002)	230	Trist and Bamforth (1951)	1,421
15	Rau and Hyland (2002)	223	Ashforth et al. (2000)	1,285
16	Golden et al. (2008)	218	Morgeson and Humphrey (2006)	1,236
17	Shockley and Allen (2007)	202	Cramton (2001)	1,232
18	Leslie et al. (2012)	198	Faraj and Sproull (2000)	1,116
19	Gittell (2000)	194	Thomas and Ganster (1995)	1,099
20	Sullivan and Lewis (2001)	191	Allen (2001)	1,074

One further factor to consider is that the number of citations can be biased due to publication year. Indeed, *ceteris paribus*, long-published articles are likely to have more citations than recently published articles recently published. A common practice to mitigate this time effect is to take the ratio between the absolute number of citations of each article and the total number of citations of all the articles published in the same year within a dataset, producing a time-weighted number of citations. Table 3 lists the 20 most relevant articles ranked according to the time-weighted number of citations. This reveals: (1) several articles published during the second half of the 20th century have scores equal to 1.00 because they are the only articles

Table 3 The 20 most relevant articles ranked by time-weighted number of citations

N°	Direct method	Weighted citation	Indirect	Weighted citation
1	Golembiewski et al. (1974)	1.00	Golembiewski et al. (1974)	1.00
2	Metzger and Von Glinow (1988)	1.00	Golembiewski and Proehl (1978)	1.00
3	Winett and Neale (1981)	1.00	Trist and Bamforth (1951)	1.00
4	Chess and Lyman (1969)	1.00	Lodahl and Kejnar (1965)	1.00
5	Gordon and Van Arsdale (1986)	1.00	Pierce and News-trom (1980)	1.00
6	Cooper and Davidson (1983)	1.00	Hackman and Oldham (1976)	0.99
7	Gray et al. (1977)	1.00	Sproull and Kiesler (1986)	0.98
8	Olson (1989)	0.98	Davis et al. (1992)	0.97
9	O'Conaill et al. (1993)	0.72	Hackman and Oldham (1975)	0.97
10	Katz (1987)	0.59	Kimberly and Evanisko (1981)	0.95
11	Hartman et al. (1991)	0.58	Quinn and Rohrbaugh (1983)	0.93
12	Olson and Olson (2000)	0.58	Nahapiet and Ghoshal (1998)	0.87
13	Pierce and News-trom (1982)	0.55	Harris and Schaubroeck (1988)	0.86
14	Yap and Tng (1990)	0.55	Goldberg (1990)	0.85
15	Gilmore and Pine (1997)	0.54	Meyer and Alien (1991)	0.79
16	Thompson and Scalpone (1985)	0.50	Baumeister and Leary (1995)	0.68
17	Elling (1985)	0.50	Akerlof (1982)	0.66
18	Hirschhorn and Gilmore (1992)	0.47	Wiley (1987)	0.57
19	Winett et al. (1982)	0.43	Gilmore and Pine (1997)	0.57
20	Olszewski and Mokhtarian (1994)	0.42	Huber (1984)	0.55

in the dataset published in that specific year; (2) there are two articles in common across datasets (Gilmore and Pine 1997; Golembiewski et al. 1974), showing some convergence compared to the Table 2 rankings. A refinement of the methodology might consider excluding the stand-alone items from the indicator to avoid biasing the results in these terms.

3.4.2 Bibliometric indicators, science mapping, and network indicators

In the following part of the analysis for each bibliometric indicator, considering each dataset separately, we present two contextual results: (1) a reduced visual map of the investigated constituents using VOSviewer software; (2) the full network properties calculated on the whole set of connected constituents using Gephi software. This double perspective—zoom-in and overview, respectively—provides adequate information from which we can draw some conclusions on the general similarity or dissimilarity between the direct and indirect datasets.

Bibliographic coupling (Kessler 1963) produces a measure of similarity between articles, authors, or journals based on the references they have in common, and considering that related and similar constituents draw upon the same knowledge basis. Essentially, the greater the number of common references (i.e., publications) between constituents, the stronger their similarity in terms of content. This indicator is a static metric because the relationship between two constituents cannot change over time.

Figure 5 visually compares the networks of the 20 most relevant articles (according to the bibliographic coupling indicator) of the direct and the indirect dataset. The central node of the two networks coincides (Gajendran and Harrison 2007). This result is affirmed by the betweenness centrality and closeness centrality reported in Table 4. The combined evidence from Fig. 5; Table 4 suggests that the articles (nodes) identified through the direct method have greater interconnectivity than those identified through the indirect method. This is plausible if we consider that all the articles identified directly share some topical commonalities, as they were all retrieved through the same string of keywords, whereas the indirect dataset contains methodological and seminal articles that might not be strictly related to FWAs.

The total link strength measures how much an article is embedded in the knowledge structure of the considered research domain (of FWAs). The higher the link strength value, the more relevant that single article is within the network. Table 5 shows the 20 most relevant articles according to the total link strength calculated on the whole network with VOSviewer.

The total link strength analysis reveals two results: (1) the two highest ranked articles in the direct and indirect datasets are the same, and they have a comparable

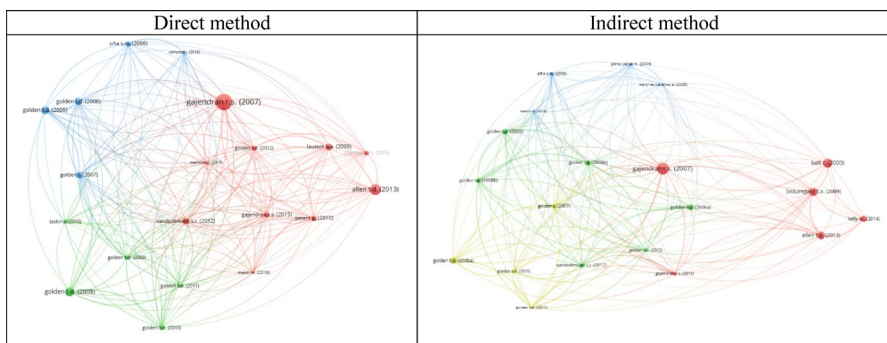


Fig. 5 Bibliographic coupling, network map of the 20 most relevant articles. Any color similarity between the two networks does not imply substantial similarities between them

Table 4 Bibliographic coupling of articles, full network properties

Indicator	Direct	Indirect
Number of connected nodes	534 (out of 659)	493 (out of 573)
Number of edges	20,789	16,858
Number of clusters	14	10
Average degree	77.86	68.38
Graph density	0.146	0.139
Modularity	0.293	0.348
Average clustering coefficient	0.564	0.507
Average shortest path length	2.129	2.039
Betweenness centrality	3,585.40 Stavrou and Iero- diakonou (2011)	2,886.53 Gajendran and Harrison (2007)
Closeness centrality	0.659 Gajendran and Har- rison (2007)	0.658 Gajendran and Harrison (2007)

Second in order (2,639.99) was Gajendran and Harrison (2007).

total link strength (Gajendran and Harrison 2007; Golden 2007); (2) out of the 20 top-ranked articles, 14 publications are common to both datasets, suggesting strong similarity in terms of core literature.

Following Ayoko et al. (2022), we clustered the lists of publications obtained through the three ranking methods (Tables 2, 3, and 5). The direct method generated a list of 55 unique articles (five duplicates) and the indirect method generated 49 unique articles (11 duplicates). Comparing these two lists shows that 15 articles, or 17% of the total number of the core publications (89), belong to both sub-datasets. This percentage is comparable to that of the shared number of articles presented above, suggesting the same type of preliminary assessment.

The bibliographic coupling indicator is also suitable for relating the research domains, encapsulated in the journals in which articles are published. In the direct dataset, there are 318 total outlets. In the indirect dataset, there are 230 total outlets. Only the journals containing at least five published articles were selected for visual mapping (Fig. 6).

From Fig. 6; Table 6, we can derive the following.

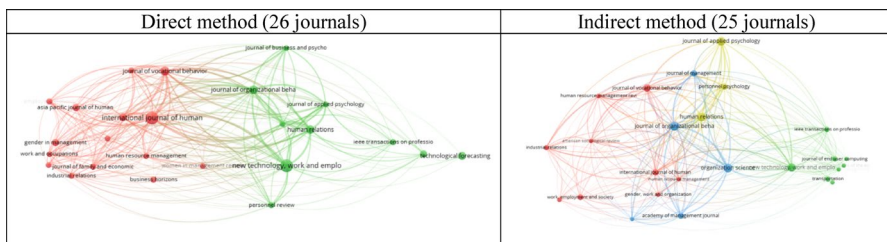
Out of 39 unique journals, 12 (31%) are shared between datasets.

The 26 core journals of the direct method are the outlets for 203 publications (31% of the total direct articles). Out of 13,249 total citations received by the journals meeting the threshold of containing at least five publications, the *Journal of Applied Psychology* scored the highest (1,831). *New Technology, Work and Employment* (24 publications and 1,338 citations) and the *International Journal of Human Resource Management* (24 publications and 769 citations), represent the center(s) of the network.

The 25 core journals of the indirect method are the outlet for 245 publications (43% of the total indirect articles). Out of 57,927 total citations received by the journals meeting the threshold of containing at least five publications, the *Journal of Applied Psychology* scored the highest (12,397). The *Journal of Organizational Behavior* (22 publications and 1,945 citations), followed by *Organization Science* (20 publications

Table 5 20 most relevant articles ranked by total link strength (tls)

	N ^o	Direct method	tls	Indirect method	tls
	1	Gajendran and Harrison (2007)	962	Gajendran and Harrison (2007)	883
	2	Golden (2007)	805	Golden (2007)	758
	3	Golden and Fromen (2011)	766	Beauregard and Henry (2009)	718
	4	Gajendran et al. (2015)	744	Golden and Fromen (2011)	671
	5	Allen et al. (2013)	731	Siha and Monroe (2006)	667
	6	Martinez and Gómez (2013)	723	Golden (2006)	649
	7	Golden (2012)	721	Golden and Veiga (2008)	604
	8	Taskin and Bridoux (2010)	715	Gajendran et al. (2015)	593
	9	Sardeshmukh et al. (2012)	713	Taskin and Bridoux (2010)	591
	10	Siha and Monroe (2006)	699	Golden and Veiga (2005)	587
	11	Golden (2006)	664	Golden (2012)	582
	12	Golden (2009)	650	Allen et al. (2013)	555
	13	Peters and Heusinkveld (2010)	628	Golden and Raghuram (2010)	550
	14	Golden and Veiga (2005)	628	Golden et al. (2008)	547
	15	Thompson et al. (2015)	614	Sardeshmukh et al. (2012)	539
	16	Lautsch et al. (2009)	601	Batt and Valcour (2003)	505
	17	Golden and Raghuram (2010)	600	Golden et al. (2006)	505
	18	Campbell and Heales (2016)	581	Martinez and Gómez (2013)	487
	19	Mayo et al. (2016)	559	Kelly et al. (2014)	479
	20	Golden et al. (2008)	555	Pérez Pérez et al. (2004)	463

**Fig. 6** Bibliographic coupling, network map of journals with at least five published articles. Any color similarity between the two networks does not imply substantial similarities between them

and 5,161 citations) and *Human Relations* (20 publications and 3,036 citations), represent the center(s) of the network.

An initial comparison of the journal networks indicates that their underlying structure is quite different. Specifically, the network of the indirect method seems to be less

Table 6 Bibliographic coupling of journals, full network properties

Indicator	Direct	Indirect
Number of connected nodes	260 (out of 318)	197 (out of 230)
Number of edges	7,768	5,743
Number of clusters	13	8
Average degree	59.75	58.30
Graph density	0.231	0.297
Modularity	0.170	0.175
Average clustering coefficient	0.698	0.692
Average shortest path length	1.913	1.776
Betweenness centrality	1,951.54 <i>New Technology, Work and Employment</i>	925.02 <i>Journal of Organizational Behavior</i>
Closeness centrality	0.799 <i>The International Journal of Human Resource Management</i>	0.841 <i>Journal of Organizational Behavior</i>

Second in order (0.789) was *New Technology, Work and Employment*.

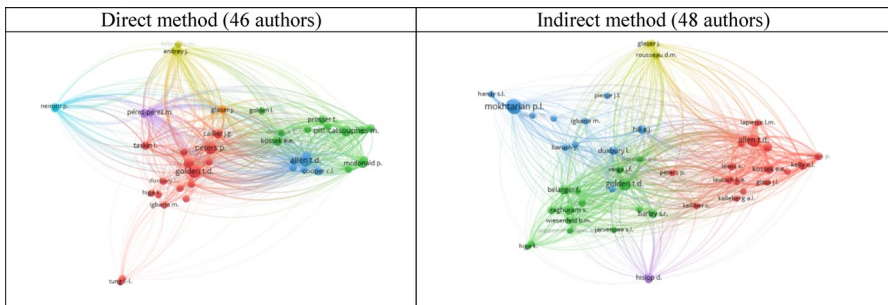


Fig. 7 Bibliographic coupling, network map of authors with at least three published articles. Any color similarity between the two networks does not imply substantial similarities between them

differentiated and more compact. This result is somewhat expected because authors, when writing a review article, define clear boundaries for their research question and domain, and, consequently, for the possible article outlets.

Bibliographic coupling can also produce a network relationship among authors¹² based on the references they share across articles. This analysis can reveal authors who, despite not directly collaborating, address the same research areas by referencing the same knowledge base. Only authors with at least three published articles were selected for visual mapping (Fig. 7). This filter restricted the sample to 46 authors (out of 1,281) in the direct dataset and 48 (out of 1,063) in the indirect dataset.

¹² This is distinct from a co-authorship analysis, which was also performed but did not provide relevant results.

Table 7 Bibliographic coupling of authors, full network properties

Indicator	Direct	Indirect
Number of connected nodes	1,077 (out of 1,281)	945 (out of 1,063)
Number of edges	93,343	78,557
Number of clusters	29	19
Average degree	173.39	166.25
Graph density	0.161	0.176
Modularity	0.367	0.426
Average clustering coefficient	0.645	0.596
Average shortest path length	2.058	1.981
Betweenness centrality	13,068.02	12,402.87
	Golden TD	Golden TD
Closeness centrality	0.739	0.758
	Golden TD	Golden TD

Table 8 Most relevant authors ranked by total link strength (tls)

N°	Direct				Indirect			
	Author	Docs	Citations	tls	Author	Docs	Citations	tls
1	Golden TD	10	1,123	14,117	Golden TD	10	1,196	13,344
2	Allen TD	8	1,149	9,896	Allen TD	11	2,255	11,280
3	Peters P	8	300	6,437	Kelly EL	6	961	7,513
4	Masuda AD	3	318	5,628	Kossek EE	6	992	6,422
5	Shockley KM	4	611	5,421	Moen P	4	740	6,092

In the restricted datasets, only 18% (14 out of 80) of the authors appear in both lists. Moreover, both Fig. 7; Table 7 underscore the centrality of Timothy Golden in FWAs research. This centrality is further confirmed by the data presented in Table 8, which ranks the five most relevant authors based on the VOSviewer total link strength. However, although these are relevant aspects regarding influential authors, the two networks exhibit significant heterogeneity in terms of authors and their connections.

Co-citation occurs when two references are cited together by a third article (Small 1973). This indicator essentially measures the frequency with which two articles are cited simultaneously by other articles. Focusing on the similarities in the references (third-level sources in our case)¹³ of the articles included in the two datasets, co-citation analysis reveals the structure of the background knowledge of the main set of publications being analyzed. This indicator is dynamic, as the co-citation relation can change over time as new contributions are published that might connect new articles or areas of research. Only references cited at least 10 times were included in the network visualization (Fig. 8).

Among 18 identified unique articles, six (33%) were common to both datasets, indicating a relatively significant overlap. Bailey and Kurland (2002) emerged as a cornerstone article of both networks. Table 9 illustrates very similar full network properties, including Bailey and Kurland (2002) as a central node.

Regarding the journals in which articles are published, the co-citation network represents the structure of the relationship between journals that are common land-

¹³ For our analysis, these are the references of the articles we selected (Fig. 2).

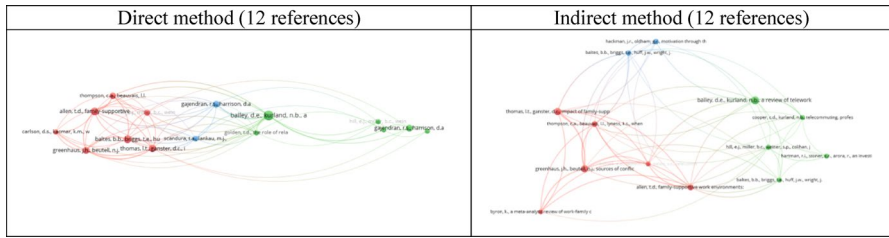


Fig. 8 Co-citation, network map of articles with at least 10 citations. Any color similarity between the two networks does not imply substantial similarities between them

Table 9 Co-citation of articles, full network properties

Indicator	Direct	Indirect
Number of connected nodes	20,456 (out of 27,170)	22,717 (out of 27,118)
Number of edges	755,119	988,176
Number of clusters	128	127
Average degree	73.82	86.99
Graph density	0.004	0.004
Modularity	0.878	0.901
Average clustering coefficient	0.968	0.969
Average shortest path length	3.681	3.962
Betweenness centrality	17,933,175.35 Bailey and Kurland (2002)	20,992,324.32 Bailey and Kurland (2002)
Closeness centrality	0.426 Bailey and Kurland (2002)	0.401 Bailey and Kurland (2002)

marks for the references of the two main datasets. This analysis provides a deeper view of the underlying structure of the two sets of articles identified through the direct and indirect methods. For the network visualization (Fig. 9), we set a minimum threshold of 50 citations for a journal to be included. Out of 10,878 and 9,848 journals represented in the references of the datasets identified through the direct and indirect methods, respectively, only 55 and 54 surpassed the selected threshold.

In both core networks, the *Journal of Applied Psychology* and the *Academy of Management Journal* play a substantial role (Fig. 9; Table 10). Perhaps even more relevant is the share of common journals: 43 journals out of 66 unique outlets (65%). In the references of the articles identified through the direct method, the *Journal of Applied Psychology* is cited 827 times, the *Journal of Vocational Behavior*, 535 times, and the *Academy of Management Journal*, 460 times. In the references of the articles identified through the indirect method, the *Journal of Applied Psychology* is cited 1,009 times, the *Academy of Management Journal*, 610 times, and the *Journal of Vocational Behavior*, 482 times. These results suggest that both datasets are deeply rooted in the same journals set in terms of their references.

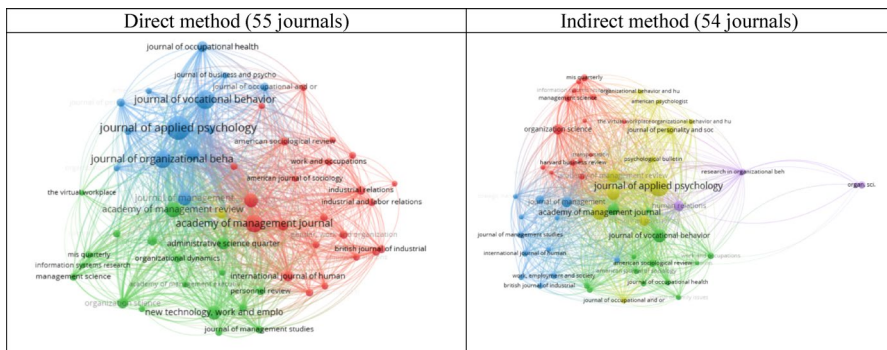


Fig. 9 Co-citation, network map of journals with at least 50 citations. Any color similarity between the two networks does not imply substantial similarities between them

Table 10 Co-citation of journals, full network properties

Indicator	Direct	Indirect
Number of connected nodes	10,747 (out of 10,878)	9,810 (out of 9,848)
Number of edges	407,157	445,090
Number of clusters	81	70
Average degree	75.77	90.74
Graph density	0.007	0.009
Modularity	0.390	0.440
Average clustering coefficient	0.907	0.897
Average shortest path length	2.473	2.424
Betweenness centrality	3,640,633.29 <i>Journal of Applied Psychology</i>	2,991,250.54 <i>Academy of Management Journal</i>
Closeness centrality	0.619 <i>Journal of Applied Psychology</i>	0.640 <i>Academy of Management Journal</i>

A final indicator to establish a relationship between articles is their textual content. The analysis of the co-occurrence of (key)words (Caputo et al. 2021) is based on the idea that content similarity can be identified by the associations between articles' (key)words (López-Fernández et al. 2016). This semi-content analysis is also useful for identifying and mapping thematic areas, or clusters (Manesh et al. 2020).

Here, we consider only the authors' keywords. For the network visualization (Fig. 10), we set a minimum threshold of three times that a keyword must appear among all the articles of each dataset. In the direct dataset, 108 out of 1,416 keywords met the threshold. In the indirect dataset, 77 out of 1,033 keywords met the threshold. This is somewhat counterintuitive as the direct articles were identified through specific words and should be more specific.

In the network of the direct dataset, *telework* appears 86 times; *telecommuting*, 74 times; *gender*, 26 times; *flexibility*, 23 times; and *teleworking*, 20 times. In the network of the indirect dataset, *telecommuting* appears 69 times; *telework*, 41 times;

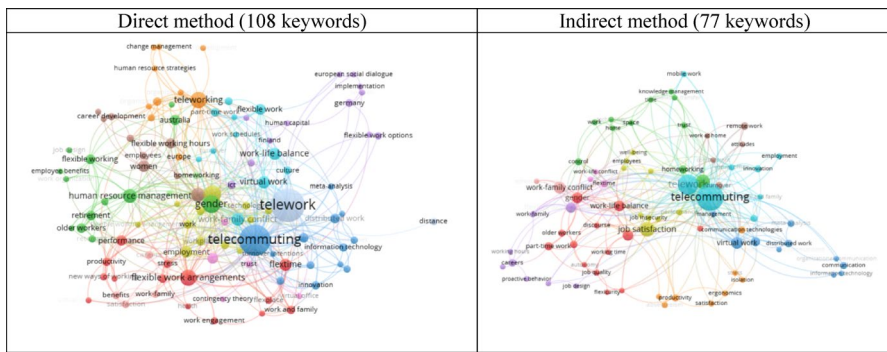


Fig. 10 Co-occurrences of author keywords, network of words appearing at least three times. Any color similarity between the two networks does not imply substantial similarities between them

job satisfaction, 22 times; *teleworking*, 20 times; and *work-family conflict*, 14 times. Out of 141 unique keywords, a total of 44 (31%) were shared between datasets. As the full network properties demonstrate, the keywords *telework* and *telecommuting* serve as central nodes in the direct and indirect networks, respectively. Overall, the two datasets seem to converge in terms of content, as represented by keywords.

3.5 Step 5: discussion of the overall similarity between the two datasets and final assessment

Before presenting the final assessment, it is important to recall the primary goal of this paper. We did not seek to track research trends or map the state of knowledge in a specific research domain. Instead, using bibliometric and network indicators, the main objective was to propose a procedure for validating umbrella reviews, taking Toscano and Zappalà's (2020) umbrella review as a case study. Furthermore, it is important to remark that a bibliometric or network analysis can be performed using several types of indicators and software, of which we demonstrate only some. As mentioned in the guide, as long as no standardized validation protocol exists, the specific sets of indicators and software might be context-dependent. Nevertheless, whatever the chosen indicator, a single value or result may not offer a comprehensive answer to the research question. Therefore, these indicators must be collectively interpreted to formulate an overall assessment from a global and complementary perspective (Ayoko et al. 2022; Donthu et al. 2021; Kleminski et al. 2020).

From the calculations in *Step 4*, we can see visual and numerical similarities between the two datasets, as well as notable differences. Table 11 summarizes these similarities and differences for each calculated indicator.

As far as similarities are concerned, the number of publications (peer-reviewed articles) per year squarely coincides across the two datasets. Only toward the end of the considered period do the datasets diverge. However, this is exclusively because of structural differences, as literature reviews are published after primary publications. The total link strength indicator revealed comparable results. Moreover, for most of the indicators concerning articles, journals, authors, and keywords, we identified many overlaps in the core networks calculated after introducing reasonable

Table 11 Co-occurrence of author keywords, full network properties

Indicator	Direct	Indirect
Number of connected nodes	1,252 (out of 1,416)	852 (out of 1,033)
Number of edges	4,429	3,150
Number of clusters	71	54
Average degree	7.07	7.39
Graph density	0.006	0.009
Modularity	0.692	0.698
Average clustering coefficient	0.886	0.882
Average shortest path length	3.655	3.540
Betweenness centrality	307,526.21	158,054.68
	Telework	Telecommuting
Closeness centrality	0.474	0.491
	Telework	Telecommuting

thresholds. This was especially true for the co-citation analysis of journals. Finally, although we did not extensively comment on this aspect in the results, most of the full network indicators provided comparable values, especially net of the structural differences between datasets.

However, at the same time, some clear differences emerged. First, a low percentage of articles is shared between the two datasets, even when we consider some sub-datasets defined via the selected indicators (e.g., the most-cited articles). Second, the total number of citations and the average number of citations are substantially different between the direct and indirect datasets. However, this divergence might also be explained by the structural difference between the two datasets. Specifically, the indirect dataset relies on articles retrieved from the literature reviews underpinning the umbrella review and encompasses some seminal and methodological articles that are highly popular, that is, frequently connected and cited, but are not identified through the direct method, which relies on specific keywords. Third, when considering the bibliographic coupling of articles, the direct dataset network is more compact than the indirect dataset network, while the opposite holds for authors and journals.

Therefore, many of the observed divergences between the indicators used to compare the two datasets can likely be attributed to structural differences in the datasets. These must be considered and discounted when formulating an assessment of the datasets overall similarity: this aspect was not initially considered in the five-step guide. Following this reasoning and combining the information obtained from the indicators into a comprehensive judgment allows us to produce a positive final assessment about Toscano and Zappalà's (2020) umbrella review: It is accurate and reliable, that is, it successfully illustrates the key elements of the literature and knowledge concerning FWAs up to 2017, the year of the last included literature review.

However, it is important to note that, in this case, we come to this final assessment by proof of contradiction, that is, we did not observe any substantive differences that would disprove a positive assessment. This umbrella review, given its broad scope, succeeds in retaining the main constituents—articles, authors, journals, and concepts—of the FWAs domain. These are substantially present in the umbrella review compared to the articles comprising the control group (direct method).

Table 12 Synoptic table

	Caption	Direct	Indirect	Similarities	Differences
Performance indicators					
Total number of publications	Figure 2	Comparable (659 articles)	Comparable (576 articles)	The number of published articles in the two datasets is comparable	The number of publications identified through the indirect method is lagged compared to the direct method, and this might explain the slightly lower number of identified articles
Shared publications		156 articles	156 articles	The two datasets share only a non-negligible fraction of the unique articles (15%)	The two methods identify a largely heterogeneous set of articles
Total number of publications per year	Figure 3	Comparable	Comparable	The publication path of the two datasets is similar until 2012	After 2012, there is a structural time lag as literature reviews are published after primary articles are published
Total number of citations		26,785	114,851		The magnitude is different, possibly due to structural differences as literature review articles rely more on methodological or seminal publications
Average number of citations		41	200		The magnitude is different, possibly due to structural differences as literature review articles rely more on methodological or seminal publications
Total number of citations per year		Constant	Fluctuating		The indirect citation chart shows citation peaks
Bibliometric indicators, science mapping, and network indicators					
20 most-cited publications (absolute number)	Table 2	More recent	Older		There are no articles in common between the datasets because the indirect method literature includes less-recent seminal and methodological articles correlated to the main topic but not centered on it

Table 12 (continued)

	Caption	Direct	Indirect	Similarities	Differences
Performance indicators					
20 most-cited publications (time-weighted indicator)	Table 3	Old	Old	There are only two articles in common and, in both rankings, the most-cited articles (weighted indicator) were published before 2000	The rankings identify two different sets of articles because the indirect dataset includes seminal and methodological articles that unbalance the absolute number of citations per year (denominator of the weighted ratio)
Bibliographic coupling of articles	Figure 5; Table 4	More compact network	More scattered network	Same importance of Gajendran and Harrison (2007) in the network	The distribution of the nodes is different: the nodes (articles) of the direct dataset tend to cluster more closely and be more connected than those of the indirect dataset
Total link strength (bibliographic coupling of articles)	Table 5	Relevant overlap	Relevant overlap	The two subsets almost coincide (14 references are shared between the two datasets), showing that the core of the references is the same	
Joint rankings	Tables 2, 3 and 5	55 unique articles	49 unique articles	15 articles (17%) of the total number of unique research articles (89) belong to both sets obtained by joining the ranking methods	The two joint rankings identify a largely heterogeneous set of articles
Bibliographic coupling of journals	Figure 6; Table 6	More scattered network	More compact network	One-third of the core journals is shared	The underlying structure of the networks is quite different in terms of central nodes (journals), with the network of the indirect dataset being less differentiated and more compact

Table 12 (continued)

	Caption	Direct	Indirect	Similarities	Differences
	Performance indicators				
Bibliographic coupling of authors	Figure 7, Table 7, Table 8	More scattered network	More compact network	Timothy Golden is a reference point in both datasets	The total number of authors is almost the same (about 1,000) but only 14 authors out of 80 (18%) in the core network are common to both datasets
Co-citation of articles	Figure 8, Table 9	Same network properties	Same network properties	In both networks, Bailey and Kurland (2002) represent the cornerstone article	
Co-citation of journals	Figure 9, Table 10	Same network properties	Same network properties		In the direct dataset, the <i>Journal of Applied Psychology</i> is more central, whereas in the indirect dataset, the <i>Academy of Management Review</i> is more central
Keyword co-occurrence	Figure 10, Table 11	Larger set of keywords	Smaller set of keywords	The central nodes (keywords) are comparable	A greater variety of keywords occurs in the direct dataset, which is somehow counterintuitive

4 Discussion and conclusions

4.1 Methodological contributions

The number of scientific publications is growing exponentially, and this trend does not yet fully reflect the potential acceleration that artificial intelligence may induce. Therefore, synthesizing and systematizing the existing body of knowledge and literature has become increasingly essential. In this context, umbrella reviews—or literature reviews of existing literature reviews—are becoming a widely adopted research device. Their quality is typically associated with adherence to established protocols designed to prevent arbitrariness and ensure rigor, transparency, and replicability. These features, in turn, support objectivity and reduce bias in representing the underlying knowledge base. Nevertheless, to date, no tools exist to assess the final quality of umbrella reviews beyond their adherence to and compliance with formal protocols.

We contribute a novel five-step guide for validating umbrella reviews to the literature review research practice (Kraus et al. 2022; Snyder 2019). This guide functions as a final quality seal to certify that—after following the rigorous application of a formal protocol—an umbrella review also substantially represents the knowledge base of the research area under investigation. To accomplish this, we

compared several bibliometric and network indicators calculated on two sets of publications that are usually available when an umbrella review has been conducted: the set of primary publications directly identified through keywords and other filters that constitutes the control group for the analysis, and the set of publications indirectly identified through the reference lists of the literature reviews that are included in the umbrella review.

Our five-step guide is innovative in many aspects but also offers considerable room for improvement.

First, it lays the groundwork for developing a more structured validation protocol. As currently designed, our guide operates as a corollary tool that is detached from the main content purpose of the umbrella review and provides a final qualitative assessment—although it is based on quantitative indicators and metrics. These indicators serve as the foundation for developing robust criteria to improve umbrella review protocols. Specifically, some indicators might be combined into a final objective quantitative index (e.g., a percentage of overlap) with a defined minimum quality threshold.

Second, a failed validation or a negative quality assessment does not imply that the umbrella review lacks value. Although the validation process does not contain explicit content purposes, the divergences that emerge afterward have potential either as warnings for methodological choices or as further information to complete and complement the umbrella review, making it more representative of the domain's knowledge base.

Third, our method offers new insights into the innovative and flexible use of bibliometric and network techniques, including the combined deployment of multiple software tools. Such techniques have traditionally been used to track research trends or to map cumulative knowledge within a research domain (Marzi et al. 2025). Here, they acquire a new role in validating the representativeness of knowledge synthesis.

Fourth, and related to the second and third points, our five-step guide can be used as a preliminary quality check. Before drafting an umbrella review, but after identifying the set of literature reviews to discuss, authors may apply the proposed five steps to assess the initial robustness of their research strategy.

Fifth, the method is replicable across different domains and fields, especially as umbrella reviews become widespread outside of the health sciences.

Sixth, the validation guide provides scholars with a tool to evaluate better how different methodological approaches or protocols for conducting umbrella review perform in terms of knowledge representativeness. Moreover, in a context where artificial intelligence tools for literature reviews—such as Elicit, Connected Papers, Claude, or Research Rabbit—are rapidly evolving, our method can offer an empirical benchmark to assess the quality and reliability of these emerging supportive technologies.

4.1.1 Limitations

Some important limitations of our approach should be acknowledged as well.

First, in the final assessment, it is important to distinguish the methodological limits of the five-step guide from objective differences between the two datasets.

Although the datasets are, in principle, comparable, and we adopted a conservative approach in building them, they still exhibited several structural differences that significantly affected the values of certain indicators. Further research should, therefore, explore ways to refine *Step 2*—dataset construction—to minimize these structural disanalogies and improve comparability.

Second, and related to the first point, we focused exclusively on research articles to guarantee maximum homogeneity. Although this was a deliberate and conservative choice, considering other types of publications (e.g., books, conference proceedings, etc.) could reveal additional insights. This must, however, be balanced against the risk of introducing inconsistencies or spurious comparisons, which may reduce the datasets' comparability.

Third, a successful validation assessment result might be context-dependent. We relied on a single case study to demonstrate the utility of this guide. Furthermore, we used only the Scopus database, which, while widely acknowledged as one of the most comprehensive scientific databases, does not index all publications. This issue is particularly salient as the umbrella review we analyzed concerns FWAs as discussed prior to the COVID-19 pandemic. The significant wave of new publications that followed the pandemic may have introduced some relevant changes, modifying the knowledge network's structure. However, as we focused on methodology—and the method is replicable—the guide can be applied to additional cases to strengthen the generalizability of our conclusions.

As is common in scientific research, these limitations also serve as the foundation for further developments, especially as our approach already serves to advance umbrella review practices and methods.

4.1.2 A final reflection on our guide and literature reviews in the era of artificial intelligence

In an era in which algorithmic automation is increasingly utilized for routine and labor-intensive tasks, it may be instinctive—especially in the current phase of widespread enthusiasm for artificial intelligence adoption—to assume that machines will replace human labor. Even literature reviews, usually if built on predefined steps, are not exempt from this speculation: artificial intelligence challenges traditional review processes in multiple ways, so that reviewers' work might appear obsolete, if not redundant (Block et al. 2025). However, when conducting literature and umbrella reviews and using tools such as our five-step validation guide, human judgment remains indispensable for several reasons (Block and Kuckertz 2024).

For example, the selection and motivations for the research area in which to conduct a literature or umbrella review (our *Step 1*) remain a human task that likely cannot be replaced by black-box algorithmic procedures. Although artificial intelligence can be used to scan the existing literature and automate specific tasks, scholars' experience, judgment, and interests remain an essential factor in defining the boundaries of a review. Likewise, different individual perspectives and scholarly dialogue within a research team—elements that foster flexibility and theoretical exploration—result in different ways of problematizing and theorizing aspects of the research area under

investigation. This is why, despite its apparent superfluity, we explicitly remarked the need for *Step 1* in our guide, anchoring the entire review process in human judgment from its inception.

During dataset construction in *Step 2*, human review ensures that inclusion and exclusion criteria are consistently applied and contextually appropriate. Human insight also helps to refine the search strategies and validate the completeness and relevance of the data entries, reducing the risk of omitting important sources or including irrelevant ones. Algorithms can support this step, but they cannot fully replace the contextual reasoning required and the transparency used to verify the selected pieces.

As far as *Step 3* is concerned, while artificial intelligence tools may automate parts of the process, human decision-making remains vital to choosing the most appropriate set of indicators and software that align with the goal of validating the umbrella review. As noted, these choices are highly context-dependent (e.g., the scientific domain or the broader research question). Machines can compute metrics, but humans must determine which indicators are appropriate for this purpose.

And finally, in *Steps 4* and *5*, at least in the absence of comprehensive indicators, human interpretation to contextualize metrics and assess their overall meaning remains pivotal. In our case, an algorithm might have judged the validation unsuccessful due to low overlaps and percentages—this is a hypothetical prediction but also an exercise that could be actually performed. However, as human researchers, we could consider the structural divergences between the two datasets and the limitations of the selected indicators, leading us to a more nuanced and, ultimately, positive assessment of the umbrella review's representativeness. Therefore, human judgment remains necessary for synthesizing findings, interpreting similarities and differences, and deriving practical, meaningful conclusions that go beyond numerical similarity scores.

For these reasons, we advocate for a hybrid and balanced approach to conducting review studies in which algorithmic tools and human decision-making are combined, and neither component supersedes the other.

Appendix

See Table 13.

Table 13 Main review methods

Review method	Purpose and main features
Narrative literature reviews (Green et al. 2006)	To examine a specific problem without a predetermined research question or protocol. Narrative literature reviews offer a broad overview of a topic without necessarily reaching a comprehensive understanding of the current state of the science domain related to it.
Integrative literature reviews (Kutcher and LeBaron 2022)	To offer a more comprehensive and holistic understanding of a particular problem or topic of interest by incorporating qualitative and/or quantitative evidence as a more structured and replicable version of a narrative literature review.
Scoping reviews (Peters et al. 2015)	To clarify the definitions and conceptual boundaries of a topic or field when a body of literature has not yet been comprehensively reviewed through an exploratory review process.
Systematic literature reviews (Page et al. 2021; Tranfield et al. 2003)	To critically assess a specific body of scientific literature by adopting strict protocols and predetermined criteria. The emphasis is on ensuring rigor, transparency, and replicability throughout the review process.
Bibliometric literature reviews (Donthu et al. 2021; Mukherjee et al. 2022; Zupic and Čater 2015)	To use quantitative indicators to analyze large bodies of scientific publications. This approach enables the exploration of the evolutionary nuances of a research field, illuminating emerging research areas and identifying cross-fertilization paths between scientific domains (Ooms et al. 2023).
Meta-analysis reviews (Glass 1976)	To combine quantitative findings from multiple studies that address analogous or comparable research questions. The purpose is to estimate the mean size effect between the variables under consideration, and provide a collective, comprehensive synthesis of related but independent research outcomes.
Meta-synthesis reviews (Walsh and Downe 2005)	To integrate findings from various interrelated qualitative studies. In contrast to the quantitative aggregative intent of meta-analysis reviews, a meta-synthesis review aims to synthesize and interpret qualitative data. It has a revelatory intent, seeking to derive a deeper understanding of the research domain under investigation.
Machine-learning (van de Shoot et al. 2023) or AI-enhanced literature reviews (Bolaños et al. 2024)	To semantically analyze and aggregate large bodies of scientific studies by employing algorithms based on machine-learning and deep-learning models. Specifically, these emerging techniques, while still subject to many challenges given their early development, promise to advance traditional review methods by automating labor-intensive tasks such as data extraction, article screening, and thematic coding (Block and Kuckertz 2024; de la Torre-López et al. 2023).
Umbrella reviews (Choi and Kang 2022)	To provide an overarching synthesis of a research area by comparing and summarizing the findings of existing literature reviews that address similar or related topics (Hunt et al. 2018; McKenzie and Brennan 2017).

This list represents a subset of review methods. Other types of reviews might be included as well, such as historical reviews, exploratory reviews, descriptive reviews, methodological reviews, theoretical reviews, and critical reviews (Palmatier et al. 2018; Paré et al. undefined2015). Researchers can also adopt various combinations of the aforementioned methods, thus broadening the formal spectrum (Linnenluecke et al. 2020; Marzi et al. 2025)

Author contributions Klaudio Klaser: Conceptualization, Data curation, Formal analysis, investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. Paolo Casari: Conceptualization, Project administration, Resources, Writing – original draft, Writing – review & editing. Roberta Cuel: Conceptualization, Project administration, Resources, Writing – original draft, Writing – review & editing.

Funding There is no funding to be acknowledged.

Data availability Data are freely available from external sources.

Declarations

Consent for publication All authors gave their consent for publication.

Competing interests The authors have no competing interests to declare.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Abdellatif M, Dadam MN, Vu NT, Nam NH, Hoan NQ, Taube Z, Huy NT (2025) A step-by-step guide for conducting an umbrella review. *Trop Med Health* 53(1):1–9. <https://doi.org/10.1186/s41182-025-00764-y>
- Akerlof GA (1982) Labor contracts as partial gift exchange. *Q J Econ* 97(4):543–569. <https://doi.org/10.2307/1885099>
- Allen TD (2001) Family-supportive work environments: the role of organizational perceptions. *J Vocat Behav* 58(3):414–435. <https://doi.org/10.1006/jvbe.2000.1774>
- Allen TD, Johnson RC, Kiburz KM, Shockley KM (2013) Work–family conflict and flexible work arrangements: deconstructing flexibility. *Pers Psychol* 66(2):345–376. <https://doi.org/10.1111/peps.12012>
- Allen TD, Golden TD, Shockley KM (2015) How effective is telecommuting? Assessing the status of our scientific findings. *Psychol Sci Public Interest* 16(2):40–68. <https://doi.org/10.1177/1529100615593273>
- an De Schoot R, De Bruin J, Schram R, Zahedi P, De Boer J, Weijdemans F, Oberski DL (2021) An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell* 3(2):125–133. <https://doi.org/10.1038/s42256-020-00287-7>
- Aria M, Cuccurullo C (2017) Bibliometrix: an R-tool for comprehensive science mapping analysis. *J Informetrics* 11(4):959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Ashforth BE, Kreiner GE, Fugate M (2000) All in a day's work: boundaries and micro role transitions. *Acad Manage Rev* 25(3):472–491. <https://doi.org/10.5465/amr.2000.3363315>
- Ayoko OB, Caputo A, Mendy J (2022) Management research contributions to the COVID-19: a bibliometric literature review and analysis of the contributions from the. *J Manage Organ J Manage Organ* 1–27. <https://doi.org/10.1017/jmo.2021.70>
- Bailey K, Breslin D (2021) The COVID-19 pandemic: what can we learn from past research in organizations and management? *Int J Manage Reviews* 23(1):3–6. <https://doi.org/10.1111/ijmr.12237>
- Bailey DE, Kurland NB (2002) A review of telework research: Findings, new directions, and lessons for the study of modern work. *J Organizational Behav* 23(4):383–400. <https://doi.org/10.1002/job.144>

- Baker P, Moon NW, Ward AC (2006) Virtual exclusion and telework: barriers and opportunities of technocentric workplace accommodation policy. *Work* 27(4):421–430
- Baltes BB, Briggs TE, Huff JW, Wright JA, Neuman GA (1999) Flexible and compressed workweek schedules: A meta-analysis of their effects on work-related criteria. *J Appl Psychol* 84(4):496. <https://doi.org/10.1037/0021-9010.84.4.496>
- Baruch Y (2001) The status of research on teleworking and an agenda for future research. *Int J Manage Reviews* 3(2):113–129. <https://doi.org/10.1111/1468-2370.00058>
- Bastian M, Heymann S, Jacomy M (2009), March Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media* (Vol. 3, No. 1, pp. 361–362). <https://doi.org/10.1609/icwsm.v3i1.13937>
- Batt R, Valcour PM (2003) Human resources practices as predictors of work-family outcomes and employee turnover. *Industrial Relations: J Econ Soc* 42(2):189–220. <https://doi.org/10.1111/1468-232X.00287>
- Baumeister RF, Leary MR (1995) The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Interpers Dev* 57–89. <https://doi.org/10.1037/0033-2909.117.3>
- Beauregard TA, Henry LC (2009) Making the link between work-life balance practices and organizational performance. *Hum Resource Manage Rev* 19(1):9–22. <https://doi.org/10.1016/j.hrmr.2008.09.001>
- Bélanger F, Watson-Manheim MB, Swan BR (2013) A multi-level socio-technical systems telecommuting framework. *Behav Inform Technol* 32(12):1257–1279. <https://doi.org/10.1080/0144929X.2012.705894>
- Bessa I, Tomlinson J (2017) Established, accelerated and emergent themes in flexible work research. *J Industrial Relations* 59(2):153–169. <https://doi.org/10.1177/0022185616671541>
- Block J, Kuckertz A (2024) What is the future of human-generated systematic literature reviews in an age of artificial intelligence? *Manage Rev Q* 74:1975–1980. <https://doi.org/10.1007/s11301-024-00471-8>
- Block J, Brändle L, Limbach P, Röth T, Sabel CA (2025) Half dead or more alive than ever? New, old, and (increasingly) redundant literature review formats. *Manage Rev Q* 1–16. <https://doi.org/10.1007/s11301-025-00513-9>
- Boell SK, Cecez-Kecmanovic D (2015) On being ‘systematic’ in literature reviews. *Formulating Res Methods Inform Systems: Volume 2*:48–78. https://doi.org/10.1057/9781137509888_3
- Bolaños F, Salatino A, Osborne F, Motta E (2024) Artificial intelligence for literature reviews: opportunities and challenges. *Artif Intell Rev* 57(10):259. <https://doi.org/10.1007/s10462-024-10902-3>
- Campbell J, Heales J (2016) Factor structure of individual consequences for teleworking professionals. *Australasian J Inform Syst* 20. <https://doi.org/10.3127/ajis.v20i0.906>
- Caputo A, Pizzi S, Pellegrini MM, Dabić M (2021) Digitalization and business models: where are we going? A science map of the field. *J Bus Res* 123:489–501. <https://doi.org/10.1016/j.jbusres.2020.09.053>
- Cascio WF (2000) Managing a virtual workplace. *Acad Manage Perspect* 14(3):81–90. <https://doi.org/10.5465/ame.2000.4468068>
- Chess S, Lyman MS (1969) A psychiatric unit in a general hospital pediatric clinic. *Am J Orthopsychiatry* 39(1):77. <https://doi.org/10.1111/j.1939-0025.1969.tb00622.x>
- Choi GJ, Kang H (2022) Introduction to umbrella reviews as a useful evidence-based practice. *J Lipid Atherosclerosis* 12(1):3. <https://doi.org/10.12997/jla.2023.12.1.3>
- Cooper CL, Davidson MJ (1983) The female manager—The pressures and the problems. *Long Range Plann* 16(1):10–14. [https://doi.org/10.1016/0024-6301\(83\)90130-9](https://doi.org/10.1016/0024-6301(83)90130-9)
- Cooper CD, Kurland NB (2002) Telecommuting, professional isolation, and employee development in public and private organizations. *J Organizational Behavior: Int J Industrial Occup Organizational Psychol Behav* 23(4):511–532. <https://doi.org/10.1002/job.145>
- Cramton CD (2001) The mutual knowledge problem and its consequences for dispersed collaboration. *Organ Sci* 12(3):346–371. <https://doi.org/10.1287/orsc.12.3.346.10098>
- Cuel R, Ravarini A, Varriale L (2020) *Technology in Organisation: Digital Transformation and People*. Italia: Maggioli. - ISBN: 9788891646088
- Cuel R, Cacciatore E, Ravarini A, Varriale L (2022) Do organizations need a head of remote work? In: Cuel R, Ponte D, Virili F (eds) *Exploring digital Resilience. Challenges for people and organizations*. LNISO, pp 274–286. <https://doi.org/10.1007/978-3-031-10902-7>
- Curcic D (2023) Number of academic papers published per year, WORDSRATED. <https://wordsrated.com/number-of-academic-papers-published-per-year/#:~:text=It%20is%20estimated%20that%20at,%2C%20reviews%2C%20and%20conference%20proceedings>

- Davis FD, Bagozzi RP, Warshaw PR (1992) Extrinsic and intrinsic motivation to use computers in the workplace 1. *J Appl Soc Psychol* 22(14):1111–1132. <https://doi.org/10.1111/j.1559-1816.1992.tb00945.x>
- de la Torre-López J, Ramírez A, Romero JR (2023) Artificial intelligence to automate the systematic review of scientific literature. *Computing* 105:2171–2194. <https://doi.org/10.1007/s00607-023-01181-x>
- De Lorenzi Cancelier MM, Lapolli EM, Gomes RK (2017) Definições sobre Trabalho flexível. Uma revisão sistemática Da literatura. *Revista Espacios* 38:25–39
- De Menezes LM, Kelliher C (2011) Flexible working and performance: A systematic review of the evidence for a business case. *Int J Manage Reviews* 13(4):452–474. <https://doi.org/10.1111/j.1468-2370.2011.00301.x>
- Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM (2021) How to conduct a bibliometric analysis: an overview and guidelines. *J Bus Res* 133:285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Elling M (1985) Remote work/telecommuting a means of enhancing the quality of life, or just another method of making business more brisk? *Econ Ind Democr* 6(2):239–249. <https://doi.org/10.1177/0143831X8562007>
- Fan D, Breslin D, Callahan JL, Iszatt-White M (2022) Advancing literature review methodology through rigour, generativity, scope and transparency. *Int J Manage Reviews* 24(2):171–180. <https://doi.org/10.1111/ijmr.12291>
- Faraj S, Sproull L (2000) Coordinating expertise in software development teams. *Manage Sci* 46(12):1554–1568. <https://doi.org/10.1287/mnsc.46.12.1554.12072>
- Fernandez RS, Sharifnia AM, Khalil H (2025) Umbrella reviews: a methodological guide. *Eur J Cardiovasc Nurs* 24(6):996–1002. <https://doi.org/10.1093/eurjcn/zvaf012>
- Ferreira FA (2018) Mapping the field of arts-based management: bibliographic coupling and co-citation analyses. *J Bus Res* 85:348–357. <https://doi.org/10.1016/j.jbusres.2017.03.026>
- Gajendran RS, Harrison DA (2007) The good, the bad, and the unknown about telecommuting: meta-analysis of psychological mediators and individual consequences. *J Appl Psychol* 92(6):1524. <https://doi.org/10.1037/0021-9010.92.6.1524>
- Gajendran RS, Harrison DA, Delaney-Klinger K (2015) Are telecommuters remotely good citizens? Unpacking telecommuting’s effects on performance via i-deals and job resources. *Pers Psychol* 68(2):353–393. <https://doi.org/10.1111/peps.12082>
- Gilmore JH, Pine BJ (1997) The four faces of mass customization. *Harvard Business Rev* 75(1):91–102
- Gittel JH (2000) Organizing work to support relational co-ordination. *Int J Hum Resource Manage* 11(3):517–539. <https://doi.org/10.1080/095851900339747>
- Glass GV (1976) Primary, secondary, and meta-analysis of research. *Educational Researcher* 5(10):3–8. <https://doi.org/10.3102/0013189X005010003>
- Goldberg LR (1990) An alternative description of personality: the big-five factor structure. *J Personal Soc Psychol* 59(6):1216. <https://doi.org/10.1037/0022-3514.59.6.1216>
- Golden TD (2006) The role of relationships in Understanding telecommuter satisfaction. *J Organizational Behavior: Int J Industrial Occup Organizational Psychol Behav* 27(3):319–340. <https://doi.org/10.1002/job.369>
- Golden T (2007) Co-workers who telework and the impact on those in the office: Understanding the implications of virtual work for co-worker satisfaction and turnover intentions. *Hum Relat* 60(11):1641–1667. <https://doi.org/10.1177/0018726707084303>
- Golden TD (2009) Applying technology to work: toward a better Understanding of telework. *Organ Manag J* 6(4):241–250. <https://doi.org/10.1057/omj.2009.33>
- Golden TD (2012) Altering the effects of work and family conflict on exhaustion: telework during traditional and nontraditional work hours. *J Bus Psychol* 27(3):255–269. <https://doi.org/10.1007/s10869-011-9247-0>
- Golden TD, Fromen A (2011) Does it matter where your manager works? Comparing managerial work mode (traditional, telework, virtual) across subordinate work experiences and outcomes. *Hum Relat* 64(11):1451–1475. <https://doi.org/10.1177/0018726711418387>
- Golden TD, Raghuram S (2010) Teleworker knowledge sharing and the role of altered relational and technological interactions. *J Organizational Behav* 31(8):1061–1085. <https://doi.org/10.1002/job.652>
- Golden TD, Veiga JF (2005) The impact of extent of telecommuting on job satisfaction: resolving inconsistent findings. *J Manag* 31(2):301–318. <https://doi.org/10.1177/0149206304271768>
- Golden TD, Veiga JF (2008) The impact of superior–subordinate relationships on the commitment, job satisfaction, and performance of virtual workers. *Leadersh Q* 19(1):77–88. <https://doi.org/10.1016/j.leaqua.2007.12.009>

- Golden TD, Veiga JF, Simsek Z (2006) Telecommuting's differential impact on work-family conflict: is there no place like home? *J Appl Psychol* 91(6):1340. <https://doi.org/10.1037/0021-9010.91.6.1340>
- Golden TD, Veiga JF, Dino RN (2008) The impact of professional isolation on teleworker job performance and turnover intentions: does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? *J Appl Psychol* 93(6):1412. <https://doi.org/10.1037/a0012722>
- Golembiewski RT, Proehl CW Jr (1978) A survey of the empirical literature on flexible workhours: character and consequences of a major innovation. *Acad Manage Rev* 3(4):837–853. <https://doi.org/10.5465/amr.1978.4289294>
- Golembiewski RT, Hilles R, Kagno MS (1974) A longitudinal study of flexi-time effects: some consequences of an OD structural intervention. *J Appl Behav Sci* 10(4):503–532. <https://doi.org/10.1177/002188637401000404>
- Goodhue DL, Thompson RL (1995) Task-technology fit and individual performance. *MIS quarterly*, 213–236. STABLE URL: <https://www.jstor.org/stable/249689>
- Gordon SL, Van Arsdale A (1986) Simulating some impacts of telecommunications/transportation substitution. *Technol Forecast Soc Chang* 29(3):285–294. [https://doi.org/10.1016/S0040-1625\(86\)80004-X](https://doi.org/10.1016/S0040-1625(86)80004-X)
- Gray P, Nilles JM, Lopez DA (1977) Communications for Transportation: Implications for Traffic Engineering. *Transportation Engineering*
- Green BN, Johnson CD, Adams A (2006) Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *J Chiropr Med* 5(3):101–117. [https://doi.org/10.1016/S0899-3467\(07\)60142-6](https://doi.org/10.1016/S0899-3467(07)60142-6)
- Hackman JR, Oldham GR (1975) Development of the job diagnostic survey. *J Appl Psychol* 60(2):159. <https://doi.org/10.1037/h0076546>
- Hackman JR, Oldham GR (1976) Motivation through the design of work: test of a theory. *Organizational Behav Hum Perform* 16(2):250–279. [https://doi.org/10.1016/0030-5073\(76\)90016-7](https://doi.org/10.1016/0030-5073(76)90016-7)
- Hanson MA, Barreiro PG, Crosetto P, Brockington D (2024) The strain on scientific publishing. *Quant Sci Stud* 5(4):823–843. https://doi.org/10.1162/qss_a_00327
- Harris MM, Schaubroeck J (1988) A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Pers Psychol* 41(1):43–62. <https://doi.org/10.1111/j.1744-6570.1988.tb00631.x>
- Harter JK, Schmidt FL, Hayes TL (2002) Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *J Appl Psychol* 87(2):268. <https://doi.org/10.1037/0021-9010.87.2.268>
- Hartman RI, Stoner CR, Arora R (1991) An investigation of selected variables affecting telecommuting productivity and satisfaction. *J Bus Psychol* 6(2):207–225. <https://doi.org/10.1007/BF01126709>
- Hill EJ, Miller BC, Weiner SP, Colihan J (1998) Influences of the virtual office on aspects of work and work/life balance. *Pers Psychol* 51(3):667–683. <https://doi.org/10.1111/j.1744-6570.1998.tb00256.x>
- Hirschhorn L, Gilmore T (1992) The new boundaries of the boundaryless company. *Harvard Business Rev* 70(3):104–115
- Hornung S, Rousseau DM, Glaser J (2008) Creating flexible work arrangements through idiosyncratic deals. *J Appl Psychol* 93(3):655. <https://doi.org/10.1037/0021-9010.93.3.655>
- Huber GP (1984) The nature and design of post-industrial organizations. *Manage Sci* 30(8):928–951. <https://doi.org/10.1287/mnsc.30.8.928>
- Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G (2018) An introduction to overviews of reviews: planning a relevant research question and objective for an overview. *Syst Reviews* 7:1–9. <https://doi.org/10.1186/s13643-018-0695-8>
- Johnson BA, Kumar P, Okano N, Dasgupta R, Shivakoti BR (2022) Nature-based solutions for climate change adaptation: A systematic review of systematic reviews. *Nature-Based Solutions* 100042. <https://doi.org/10.1016/j.nbsj.2022.100042>
- Katz AI (1987) The management, control, and evaluation of a telecommuting project: A case study. *Inf Manag* 13(4):179–190. [https://doi.org/10.1016/0378-7206\(87\)90055-3](https://doi.org/10.1016/0378-7206(87)90055-3)
- Kelly EL, Moen P, Oakes JM, Fan W, Okechukwu C, Davis KD, Casper LM (2014) Changing work and work-family conflict: evidence from the work, family, and health network. *Am Sociol Rev* 79(3):485–516. <https://doi.org/10.1177/0003122414531435>
- Kessler MM (1963) Bibliographic coupling between scientific papers. *Am Doc* 14(1):10–25. <https://doi.org/10.1002/asi.5090140103>
- Kimberly JR, Evanisko MJ (1981) Organizational innovation: the influence of individual, organizational, and contextual factors on hospital adoption of technological and administrative innovations. *Acad Manag J* 24(4):689–713. <https://doi.org/10.5465/256170>

- Kitchenham B, Brereton OP, Budgen D, Turner M, Bailey J, Linkman S (2009) Systematic literature reviews in software engineering—a systematic literature review. *Inf Softw Technol* 51(1):7–15. <https://doi.org/10.1016/j.infsof.2008.09.009>
- Klaser K, Cuel R, Casari P (2023) The future of hybrid work in Italy: A survey-based Socio-Technical-System analysis. *J Innov Knowl* 8(4):100426. <https://doi.org/10.1016/j.jik.2023.100426>
- Kleminski R, Kazienko P, Kajdanowicz T (2020) Analysis of direct citation, co-citation and bibliographic coupling in scientific topic identification. *J Inform Sci* 0165551520962775. <https://doi.org/10.1177/0165551520962775>
- Kobak D, González-Márquez R, Horvát EÁ, Lause J (2025) Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Sci Adv* 11(27):eadt3813. <https://doi.org/10.1126/sciadv.adt3813>
- Kossek EE, Lautsch BA, Eaton SC (2006) Telecommuting, control, and boundary management: correlates of policy use and practice, job control, and work–family effectiveness. *J Vocat Behav* 68(2):347–367. <https://doi.org/10.1016/j.jvb.2005.07.002>
- Kowalski KB, Swanson JA (2005) Critical success factors in developing teleworking programs. Benchmarking: *Int J*. <https://doi.org/10.1108/14635770510600357>
- Kraus S, Breier M, Lim WM, Dabić M, Kumar S, Kanbach D, Ferreira JJ (2022) Literature reviews as independent studies: guidelines for academic practice. *RMS* 16(8):2577–2595. <https://doi.org/10.107/s11846-022-00588-8>
- Kutcher AM, LeBaron VT (2022) A simple guide for completing an integrative review using an example Article. *J Prof Nurs* 40:13–19. <https://doi.org/10.1016/j.profnurs.2022.02.004>
- Lamovšek A, Černe M (2023) Past, present and future: A systematic multitechnique bibliometric review of the field of distributed work. *Inf Organ* 100446. <https://doi.org/10.1016/j.infoandorg.2022.100446>
- Lapierre LM, Allen TD (2006) Work-supportive family, family-supportive supervision, use of organizational benefits, and problem-focused coping: implications for work–family conflict and employee well-being. *J Occup Health Psychol* 11(2):169. <https://doi.org/10.1037/1076-8998.11.2.169>
- Lautsch BA, Kossek EE, Eaton SC (2009) Supervisory approaches and paradoxes in managing telecommuting implementation. *Hum Relat* 62(6):795–827. <https://doi.org/10.1177/0018726709104543>
- Leonardi PM (2020) COVID-19 and the new technologies of organizing: digital exhaust, digital footprints, and artificial intelligence in the wake of remote work. *J Manage Stud* 58(1):249–253. <https://doi.org/10.1111/joms.12648>
- Leslie LM, Manchester CF, Park TY, Mehng SA (2012) Flexible work practices: a source of career premiums or penalties? *Acad Manag J* 55(6):1407–1428. <https://doi.org/10.5465/amj.2010.0651>
- Linnenluecke MK, Marrone M, Singh AK (2020) Conducting systematic literature reviews and bibliometric analyses. *Aust J Manage* 45(2):175–194. <https://doi.org/10.1177/031289621987767>
- Lodahl TM, Kejnar M (1965) The definition and measurement of job involvement. *J Appl Psychol* 49(1). <https://doi.org/10.1037/h0021692>
- López-Fernández MC, Serrano-Bedia AM, Pérez-Pérez M (2016) Entrepreneurship and family firm research: A bibliometric analysis of an emerging field. *J Small Bus Manage* 54(2):622–639. <https://doi.org/10.1111/jsbm.12161>
- Lunny C, Brennan SE, McDonald S, McKenzie JE (2017) Toward a comprehensive evidence map of overview of systematic review methods: paper 1—purpose, eligibility, search and data extraction. *Syst Reviews* 6:1–27. <https://doi.org/10.1186/s13643-017-0617-1>
- Lunny C, Brennan SE, McDonald S, McKenzie JE (2018) Toward a comprehensive evidence map of overview of systematic review methods: paper 2—risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Syst Reviews* 7:1–31. <https://doi.org/10.1186/s13643-018-0784-8>
- Malone TW, Laubacher RJ (1999) The dawn of the e-lance economy. *Electronic business engineering. Physica, Heidelberg*, pp 13–24. https://doi.org/10.1007/978-3-642-58663-7_2
- Manesh MF, Pellegrini MM, Marzi G, Dabic M (2020) Knowledge management in the fourth industrial revolution: mapping the literature and scoping future avenues. *IEEE Trans Eng Manage* 68(1):289–300. <https://doi.org/10.1109/TEM.2019.2963489>
- Martin BH, MacDonnell R (2012) Is telework effective for organizations? A meta-analysis of empirical research on perceptions of telework and organizational outcomes. *Manage Res Rev*. <https://doi.org/10.1108/01409171211238820>
- Martinez P, Gómez CB (2013) Trading telecommuting flexibility for fewer training opportunities? *Manage Research: J Iberoamerican Acad Manage*. <https://doi.org/10.1108/MRJIAM-04-2012-0478>

- Marzi G, Balzano M, Caputo A, Pellegrini MM (2025) Guidelines for bibliometric-systematic literature reviews: 10 steps to combine analysis, synthesis and theory development. *Int J Manage Reviews* 27(1):81–103. <https://doi.org/10.1111/ijmr.12381>
- Mayo M, Gomez-Mejia L, Firfiray S, Berrone P, Villena VH (2016) Leader beliefs and CSR for employees: the case of telework provision. *Leadersh Organ Dev J*. <https://doi.org/10.1108/LODJ-09-2014-0177>
- McKenzie JE, Brennan SE (2017) Overviews of systematic reviews: great promise, greater challenge. *Syst Reviews* 6:1–4. <https://doi.org/10.1186/s13643-017-0582-8>
- Metzger RO, Von Glinow MA (1988) Off-site workers: at home and abroad. *Calif Manag Rev* 30(3):101–111. <https://doi.org/10.2307/41166517>
- Meyer JP, Allen NJ (1991) A three-component conceptualization of organizational commitment. *Hum Resource Manage Rev* 1(1):61–89. [https://doi.org/10.1016/1053-4822\(91\)90011-Z](https://doi.org/10.1016/1053-4822(91)90011-Z)
- Moral-Muñoz JA, Herrera-Viedma E, Santisteban-Espejo A, Cobo MJ (2020) Software tools for conducting bibliometric analysis in science: an up-to-date review. *Profesional De La Información* 29(1). <https://doi.org/10.3145/epi.2020.ene.03>
- Morgeson FP, Humphrey SE (2006) The work design questionnaire (WDQ): developing and validating a comprehensive measure for assessing job design and the nature of work. *J Appl Psychol* 91(6):1321. <https://doi.org/10.1037/0021-9010.91.6.1321>
- Mukherjee D, Lim WM, Kumar S, Donthu N (2022) Guidelines for advancing theory and practice through bibliometric research. *J Bus Res* 148:101–115. <https://doi.org/10.1016/j.jbusres.2022.04.042>
- Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital, and the organizational advantage. *Acad Manage Rev* 23(2):242–266. <https://doi.org/10.5465/amr.1998.533225>
- Ng CF (2016) Public spaces as workplace for mobile knowledge workers. *J Corp Real Estate*. <https://doi.org/10.1108/JCRE-10-2015-0030>
- O’Conaill B, Whittaker S, Wilbur S (1993) Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. *Hum Comput Interact* 8(4):389–428. https://doi.org/10.1207/s15327051hci0804_4
- Olson MH (1989) Work at home for computer professionals: current attitudes and future prospects. *ACM Trans Inform Syst (TOIS)* 7(4):317–338. <https://doi.org/10.1145/76158.76891>
- Olson GM, Olson JS (2000) Distance matters. *Human-computer Interact* 15(2–3):139–178. https://doi.org/10.1207/S15327051HCI1523_4
- Olszewski P, Mokhtarian P (1994) Telecommuting frequency and impacts for state of California employees. *Technol Forecast Soc Chang* 45(3):275–286. [https://doi.org/10.1016/0040-1625\(94\)90050-7](https://doi.org/10.1016/0040-1625(94)90050-7)
- Onwuegbuzie AJ, Frels R (2016) Seven steps to a comprehensive literature review: A multimodal and cultural approach. Sage
- Ooms T, Klaser K, Ishkanian A (2023) The role of academia practice partnerships in the well-being economy: Retracing synergies between health and social sciences using bibliometric analysis. *Health Policy* 138:104936. <https://doi.org/10.1016/j.healthpol.2023.104936>
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 88:105906. <https://doi.org/10.1016/j.ijsu.2021.105906>
- Pahlevan-Sharif S, Mura P, Wijesinghe SN (2019) A systematic review of systematic reviews in tourism. *J Hospitality Tourism Manage* 39:158–165. <https://doi.org/10.1016/j.jhtm.2019.04.001>
- Palmatier RW, Houston MB, Hullah J (2018) Review articles: purpose, process, and structure. *J Acad Mark Sci* 46:1–5. <https://doi.org/10.1007/s11747-017-0563-4>
- Paré G, Trudel MC, Jaana M, Kitsiou S (2015) Synthesizing information systems knowledge: A typology of literature reviews. *Inf Manag* 52(2):183–199. <https://doi.org/10.1016/j.im.2014.08.008>
- Pérez MP, Sánchez AM, de Luis Rincón P, Jiménez MJV (2004) A technology acceptance model of innovation adoption: the case of teleworking. *Eur J Innov Manage*. <https://doi.org/10.1108/14601060410565038>
- Peters P, Heusinkveld S (2010) Institutional explanations for managers’ attitudes towards telehomeworking. *Hum Relat* 63(1):107–135. <https://doi.org/10.1177/0018726709336025>
- Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB (2015) Guidance for conducting systematic scoping reviews. *JBI Evid Implement* 13(3):141–146. <https://doi.org/10.1097/XEB.000000000000050>
- Pierce JL, Newstrom JW (1980) Toward a conceptual clarification of employee responses to flexible working hours: A work adjustment approach. *J Manag* 6(2):117–134. <https://doi.org/10.1177/01492063800600202>

- Pierce JL, Newstrom JW (1982) Employee responses to flexible work schedules: an inter-organization, inter-system comparison. *J Manag* 8(1):9–25. <https://doi.org/10.1177/014920638200800101>
- Pyörriä P (2011) Managing telework: risks, fears and rules. *Manage Res Rev*. <https://doi.org/10.1108/01409171111117843>
- Quinn RE, Rohrbach J (1983) A Spatial model of effectiveness criteria: towards a competing values approach to organizational analysis. *Manage Sci* 29(3):363–377. <https://doi.org/10.1287/mnsc.29.3.363>
- Rau BL, Hyland MAM (2002) Role conflict and flexible work arrangements: the effects on applicant attraction. *Pers Psychol* 55(1):111–136. <https://doi.org/10.1111/j.1744-6570.2002.tb00105.x>
- Sardeshmukh SR, Sharma D, Golden TD (2012) Impact of telework on exhaustion and job engagement: A job demands and job resources model. *New Technol Work Employ* 27(3):193–207. <https://doi.org/10.1111/j.1468-005X.2012.00284.x>
- Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Bouter LM (2007) Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 7(1):10. <https://doi.org/10.1186/1471-2288-7-10>
- Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Henry DA (2017) AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 358:1–9. <https://doi.org/10.1136/bmj.j4008>
- Shehata B, Tlili A, Huang R, Adarkwah MA, Liu M, Chang T (2023) How are we doing with student-centered learning facilitated by educational technologies? A systematic review of literature reviews. *Educ Inform Technol* 1–42. <https://doi.org/10.1007/s10639-023-12112-w>
- Shin B, Sawy E, Sheng OA, O. R. L., Higa K (2000) Telework: existing research and future directions. *J Organizational Comput Electron Commer* 10(2):85–101. https://doi.org/10.1207/S15327744JOCE1002_2
- Shockley KM, Allen TD (2007) When flexibility helps: another look at the availability of flexible work arrangements and work–family conflict. *J Vocat Behav* 71(3):479–493. <https://doi.org/10.1016/j.jvb.2007.08.006>
- Siha SM, Monroe RW (2006) Telecommuting’s past and future: a literature review and research agenda. *Bus Process Manage J*. <https://doi.org/10.1108/14637150610678078>
- Small H (1973) Co-citation in the scientific literature: A new measure of the relationship between two documents. *J Am Soc Inform Sci* 24(4):265–269. <https://doi.org/10.1002/asi.4630240406>
- Smith V, Devane D, Begley CM, Clarke M (2011) Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Med Res Methodol* 11(1):1–6. <https://doi.org/10.1186/1471-2288-11-15>
- Snyder H (2019) Literature review as a research methodology: an overview and guidelines. *J Bus Res* 104:333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sproull L, Kiesler S (1986) Reducing social context cues: electronic mail in organizational communication. *Manage Sci* 32(11):1492–1512. <https://doi.org/10.1287/mnsc.32.11.1492>
- Staples DS, Hulland JS, Higgins CA (1999) A self-efficacy theory explanation for the management of remote workers in virtual organizations. *Organ Sci* 10(6):758–776. <https://doi.org/10.1287/orsc.10.6.758>
- Stavrou E, Ierodiakonou C (2011) Flexible work arrangements and intentions of unemployed women in cyprus: A planned behaviour model. *Br J Manag* 22(1):150–172. <https://doi.org/10.1111/j.1467-8551.2010.00695.x>
- Straub D, Karahanna E (1998) Knowledge worker communications and recipient availability: toward a task closure explanation of media choice. *Organ Sci* 9(2):160–175. <https://doi.org/10.1287/orsc.9.2.160>
- Sullivan C, Lewis S (2001) Home-based telework, gender, and the synchronization of work and family: perspectives of teleworkers and their co-residents. *Gend Work Organ* 8(2):123–145. <https://doi.org/10.1111/1468-0432.00125>
- Sutton A, Clowes M, Preston L, Booth A (2019) Meeting the review family: exploring review types and associated information retrieval requirements. *Health Inform Libr J* 36(3):202–222. <https://doi.org/10.1111/hir.12276>
- Tantardini M, Ieva F, Tajoli L, Piccardi C (2019) Comparing methods for comparing networks. *Sci Rep* 9(1):17557. <https://doi.org/10.1038/s41598-019-53708-y>
- Taskin L, Bridoux F (2010) Telework: A challenge to knowledge transfer in organizations. *Int J Hum Resource Manage* 21(13):2503–2520. <https://doi.org/10.1080/09585192.2010.516600>
- Thelwall M, Sud P (2022) Scopus 1900–2020: growth in articles, abstracts, countries, fields, and journals. *Quant Sci Stud* 3(1):37–50. https://doi.org/10.1162/qss_a_00177

- Thomas LT, Ganster DC (1995) Impact of family-supportive work variables on work-family conflict and strain: A control perspective. *J Appl Psychol* 80(1):6. <https://doi.org/10.1037/0021-9010.80.1.6>
- Thompson H, Scalpone R (1985) Managing the human resource in the factory of the future. *Hum Syst Manage* 5(3):221–230. <https://doi.org/10.3233/HSM-1985-5305>
- Thompson RJ, Payne SC, Taylor AB (2015) Applicant attraction to flexible work arrangements: separating the influence of flextime and flexplace. *J Occup Organizational Psychol* 88(4):726–749. <https://doi.org/10.1111/joop.12095>
- Torres-Pruñonosa J, Plaza-Navas MA, Díez-Martín F, Prado-Roman C (2020) The sources of knowledge of the economic and social value in sport industry research: A Co-citation analysis. *Front Psychol* 11:3924. <https://doi.org/10.3389/fpsyg.2020.629951>
- Toscano F, Zappalà S (2020) Smart working in italia: origine, diffusione e possibili Esiti. *Psicologia Sociale* 15(2):203–223. <https://www.rivisteweb.it/doi/> <https://doi.org/10.1482/96843>
- Tranfield D, Denyer D, Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br J Manag* 14(3):207–222. <https://doi.org/10.1111/1467-8551.00375>
- Trist EL, Bamforth KW (1951) Some social and psychological consequences of the Longwall method of coal-getting: an examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Hum Relat* 4(1):3–38. <https://doi.org/10.1177/001872675100400101>
- Trujillo CM, Long TM (2018) Document co-citation analysis to enhance transdisciplinary research. *Sci Adv* 4(1):e1701130. <https://doi.org/10.1126/sciadv.1701130>
- Van Eck NJ, Waltman L (2007) VOS: A new method for visualizing similarities between objects. In *Advances in Data Analysis* (pp. 299–306). Springer. https://doi.org/10.1007/978-3-540-70981-7_34
- Van Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2):523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Van Eck NJ, Waltman L (2014) Visualizing bibliometric networks. In *Measuring scholarly impact* (pp. 285–320). Springer. https://doi.org/10.1007/978-3-319-10377-8_13
- Walsh D, Downe S (2005) Meta-synthesis method for qualitative research: a literature review. *J Adv Nurs* 50(2):204–211. <https://doi.org/10.1111/j.1365-2648.2005.03380.x>
- Wiley DL (1987) The relationship between work/nonwork role conflict and job-related outcomes: some unanticipated findings. *J Manag* 13(3):467–472. <https://doi.org/10.1177/014920638701300303>
- Winett RA, Neale MS (1981) Flexible work schedules and family time allocation: assessment of a system change on individual behavior using self-report logs. *J Appl Behav Anal* 14(1):39–46. <https://doi.org/10.1901/jaba.1981.14-39>
- Winett RA, Neale MS, Williams KR (1982) The effects of flexible work schedules on urban families with young children: Quasi-experimental, ecological studies. *Am J Community Psychol* 10(1):49. <https://doi.org/10.1007/BF00903304>
- Yap CS, Tng H (1990) Factors associated with attitudes towards telecommuting. *Inf Manag* 19(4):227–235. [https://doi.org/10.1016/0378-7206\(90\)90032-D](https://doi.org/10.1016/0378-7206(90)90032-D)
- Zupic I, Čater T (2015) Bibliometric methods in management and organization. *Organizational Res Methods* 18(3):429–472. <https://doi.org/10.1177/1094428114562629>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.