Reporting quality in psychological interventions. Implications for cumulative science and evidence-based practice: A scoping review.

Vives, J.[1], Lorente, S. [1], Casellas, A. [2], López-González, E. [3], Marín-García, J.A. [4], Molla, C. [3], Rodrigo, M.F. [3], Viguer, P. [3], and Losilla, J.-M.[1]

[1] Universitat Autònoma de Barcelona; [2] Universitat de Vic-Universitat Central de Catalunya
[3] Universitat de València; [4] Universitat Politècnica de València

Abstract

Background: Evidence-based practice depends on cumulative science, yet replication of psychological interventions is often hindered by incomplete reporting of implementation details.

Objective: To map how reports of psychological interventions describe implementation at both the whole-intervention level and the level of recipients' behaviors or actions (BeA), and to examine whether reporting quality varies by field, guideline use, or indicators of intervention complexity.

Design: Scoping review (PRISMA-ScR).

Data sources: PsycINFO (ProQuest) and Web of Science (Clarivate).

Eligibility criteria for study selection: Publications since 2014 that described the development and/or implementation of psychological interventions in four application domains (clinical, public health, social/organizational, education).

Charting and synthesis methods: Random sampling within fields; two-stage selection with double screening and consensus; bespoke extraction instrument developed via Delphi. Descriptive statistics and χ^2 tests compared reporting across fields, guideline use, and complexity features (e.g., number of BeA).

Results: Eighty studies were included. Only 11.3% reported development-only objectives; most combined development with effectiveness (80% overall) and feasibility (48.8%). A reporting-quality tool was used in 13.8%. Interventions typically targeted a single recipient category (87.5%); multi-level targeting was rare (6.3%). Implementation sequence was often replicable (73.8%), but structured aids were underused (tables/lists 38.8%; diagrams 16.3%). Costs were scarcely reported (3.8%). At the BeA level, structural features were usually replicable (e.g., type/form of administration, periodicity, place, timing), whereas content/materials (36.3% and 47.5%), adaptation/customization (35%), and fidelity/adherence (27.5%) were frequently incomplete. Few significant between-field differences emerged. Reporting tended to improve as the number of BeA increased.

Conclusions: Reporting of psychological interventions commonly omits replicability critical details, especially provider/evaluator prerequisites, content/materials, adaptation rules, fidelity procedures, and costs, despite available guidance. Routine use of structured representations (flow diagrams, BeA-to-component tables) and established reporting tools (e.g., TIDieR, CONSORT-SPI) could close the most consequential gaps and facilitate implementation at scale.

Registration: INPLASY 2025.7.0098.

Keywords: Psychological interventions; Reporting quality; Implementation; Fidelity; Complex interventions; Scoping review.

Introduction

Evidence-based practice is grounded in cumulative science, which is usually based on systematic reviews of the results of primary studies. However, the ability of a systematic review to synthesize evidence is often reduced—or even precluded—by deficiencies in the reporting of primary studies (Losilla et al., 2018; Turner et al., 2012). Poor quality in research reporting, whether due to a lack of completeness, transparency and/or precision, is a pressing problem that has hindered the advancement of scientific knowledge across all health science disciplines for decades (Glasziou et al., 2014). Moreover, inadequate reporting of research not only may prevent its replication, but it is also a major source of research waste, which in areas such as biomedicine is estimated to account for 85% of research resources (Almaqrami et al., 2020; Blanco et al., 2020; Ioannidis et al., 2014).

The quality of reporting in primary studies focusing on an intervention —whether concerning its development, feasibility, efficacy, effectiveness or efficiency— is not expected to differ. In the case of interventions, however, another key implication arises in relation to evidence-based practice. If the characteristics related to the implementation of proposed psychological interventions are not reported with sufficient information, it is not possible to faithfully replicate them. Consequently, any practice involving the application of such interventions cannot truly be considered evidence-based.

The emergence of reporting guidelines, like CONSORT (Begg, 1996; Moher et al., 2001; Schulz et al., 2010), STROBE (Vandenbroucke et al., 2007), PRISMA (Liberati et al., 2009), SPIRIT (Chan et al., 2013; Hopewell et al., 2022) or TREND (Des Jarlais et al., 2004), has led to an improvement in the quality of research reporting. Nevertheless, the impact of these guidelines and authors' adherence to them remain in need of improvement (Bastuji-Garin et al., 2013; Clayson et al., 2019; Mannocci et al., 2014; Orduña-Malea et al., 2022). In the case of interventions, the recent publication of guidelines to develop interventions, such as those focusing on complex interventions of Medical Research Council (MRC) (Craig et al., 2008; Skivington et al., 2021; Tanner-Smith & Grant, 2018a) and the Agency for Healthcare Research and Quality (AHRQ) (Guise, Chang, et al., 2017), and the availability of intervention-specific reporting tools and guides such as CReDECI 2 (Möhler et al., 2015; Tanner-Smith & Grant, 2018b), GUIDED (Duncan et al., 2020), TIDieR (Hoffmann et al., 2014), iCAT_SR (Lewin et al., 2017a) and RIPI-f (Lopez-Alcalde et al., 2022), may have contributed to some improvement in the reporting quality of interventions.

Note that some of these guides target complex interventions and address how the factors defining their complexity should be reported to enable accurate replication and evaluation. Although several definitions of complex interventions attempt to distinguish them from non-complex ones, complexity is neither a binary characteristic nor a property of a single type of intervention (Lewin et al., 2017a; Petticrew, 2011). Rather, any intervention can be placed on a continuum ranging from a low degree to a high degree of complexity (Moore et al., 2017). Taking the factors that contribute to the complexity of an intervention into account allows for a more thorough description of the characteristics of published interventions and to explore whether these factors, which can be challenging to report appropriately, are related to the

reporting quality. While the debate surrounding the concept of 'complex intervention' is still ongoing, the most influential conceptualizations published in recent years, such as those of the MRC and the AHRQ aforementioned, agree that, in an intervention, complexity may arise, first and foremost, from the properties of the intervention itself, being the presence of multiple components, and the relations between them and the outcomes (especially moderations and mediations) the most relevant properties, along with the adaptability or customizability of the protocol.

Partitioning the intervention into components may facilitate the study of its efficacy and effectiveness (Petticrew et al., 2015; Skivington et al., 2021). However, none of the most influential definitions of complex interventions clearly defines what a component is. The only definition of component we are aware of is offered by iCAT_SR (Lewin et al., 2017a), where a component is defined as a "discrete, active element of the intervention that could be implemented independently of other elements" (op. cit., p. 3). Based on this definition, each behavior or action (BeA) promoted by the intervention in its recipients could be considered the atomic constituent of an intervention. Each BeA may have specific implementation characteristics, which may at the same time have a verifiable effect on one or more outcomes, so that its contribution—whether direct, moderated or mediated— to the change pursued by the intervention can be better identified in a detailed and replicable report.

While the presence of multiple components and the relations that may derive from them are the first source of complexity, the second source arises from the system (e.g., an education or healthcare system, or a company) in which the intervention operates (Hawe et al., 2009; Rutter et al., 2017). A system may be a multidimensional entity composed of multiple organization layers, each hosting independent yet interacting processes. When complexity is examined at the system level, an intervention can be considered an event expected to contribute to change at one or more levels and/or categories of recipients (e.g., users, professionals, managers, policymakers). Therefore, the number of levels and categories of recipients targeted by the intervention is another important factor contributing to its complexity. As Noyes et al. (2013, p. 1268) point out, "the effects of an intervention may be attenuated when it interacts with a complex system in which multiple processes within a system are occurring simultaneously".

While it is unreasonable to implement an intervention for which there is no empirical evidence of its efficacy in achieving the desired outcomes, it is also not justifiable to implement it if the report of their development does not contain the necessary information to do so in the precise manner intended by its authors. In order to improve the reporting of interventions, the MRC (Skivington et al., 2021) encourages the publication of their development separately from the assessment of its results (such as efficacy, effectiveness, efficiency or feasibility). It is to be seen the impact of this recommendation on the literature reporting interventions.

In this study, our aim was to examine the salient features of intervention reports and to explore whether some of these features are related to their quality. To ensure certain degree of homogeneity and according to our expertise, we focused on psychological interventions. According to Munn (2018), a scoping review represents the most appropriate type of systematized review to address such exploratory objectives. Specifically, our scoping review aimed to:

a) Describe the most common features of the intervention reports.

b) Analyze the replicability of the interventions.

c) Explore the presence of elements that characterize complex interventions and whether they are related to the quality and replicability of intervention reports.

d) To analyze the relationship between the quality of the intervention report and its replicability with: a) the objective of the intervention report (devoted exclusively to the description of its development and its implementation, or also including the evaluation of its feasibility, efficacy, effectiveness or efficiency); b) the research field to which the intervention belongs (clinical, public health, social and organizational, or education); and c) the use of reporting tools or guidelines.

Method

This review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) (Tricco et al., 2018). The protocol was prospectively registered at INPLASY (Vives et al., 2025).

In accordance with the principles of open science, all datasets and supplementary information referenced in the following sections are publicly available in the CORA_RDR public repository and can be accessed at https://doi.org/10.34810/data2792. The PRISMA-ScR checklist applied can be consulted in S1_PRISMA_ScR.

Search strategy

The search was conducted in PsycInfo (Proquest) and Web of Science (Clarivate), with the aim of collecting data from a representative sample of recently published studies focusing on reporting on psychological interventions in the four application domains: clinical, public health, social and organizational, and educational. The search structure was adapted to accommodate specific descriptors of each of the four areas of psychological specialty (see S2_Search_Strategy).

Eligibility criteria

Eligible studies were those:

1. Describing the development or implementation of psychological interventions, whether or not they also included empirical assessments of their feasibility, efficacy, effectiveness, or efficiency.

2. Providing information about the implementation process, even if parts of this process referenced previous publications (e.g., a previous protocol, a previously published manual, etc.).

3. Whose reported interventions belonged to the specialties of clinical, public health, social and organizational, or educational.

4. Published between 2014 and 2024.

Selection process

A random selection of 200 studies from each of the four research fields was made, followed by a sequential selection of 21 studies, one of which was used to perform the pilot test of data extraction, and the remaining 20 to extract the information of interest for our objectives, including a total of 80 studies in the review. Two reviewers (JML, JV) independently assessed

the inclusion criteria based on the titles and abstracts of the studies and made the final selection from the full text. Discrepancies during this selection process were resolved by consensus between the two reviewers, with the participation of a third reviewer specialist in the field of intervention (AC, SL, JM/MFR, CM/EL/PV) when necessary. The degree of agreement between the raters was calculated using the Kappa index.

Data collection

References identified by the search strategy were entered into Zotero bibliographic software, and duplicates were removed automatically.

Two authors (JML, JV) generated an initial list of items based on the review of: a) publications focused on establishing the characteristics that determine the complexity of interventions (Craig et al., 2008; Guise, Butler, et al., 2017; Rutter et al., 2017; Skivington et al., 2021; Tanner-Smith & Grant, 2018b); and b) tools and guides to assess the quality of intervention reports, with particular attention to those that consider the complexity of interventions and incorporate guidelines or elements in their proposals aimed at capturing the characteristics of the complexity of the interventions. The tools that served as the basis for writing the items were SPIRIT (*Standard Protocol Items: Recommendations for Interventional Trials*) (Chan et al., 2013; Hopewell et al., 2022), iCAT-SR (*Complexity Assessment Tool for Systematic Reviews*) (Lewin et al., 2017b), Consort-SPI (Montgomery et al., 2018), CReDECI 2 (*Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare*) (Möhler et al., 2015; Tanner-Smith & Grant, 2018b), RIPI-f (*Reporting Integrity of Psychological Interventions delivered face-to-face*) (Lopez-Alcalde et al., 2022), StaRI (*Standards for Reporting Implementation Studies*) (Pinnock et al., 2015, 2017), TIDieR (*Template for Intervention Description and Replication checklist*) (Hoffmann et al., 2014) and TREND (*Transparent Reporting of Evaluations with Non-Randomised designs*) (Des Jarlais et al., 2004).

The same authors developed a second version of the items to be used for data extraction, this time including response options. A two-round Delphi survey with all the research team (AC, CM, EL, JM, JML, JV, MR, PV, SL) was conducted using this second version as the starting point. The results of the Delphi allowed the development of a third version of the data extraction spreadsheet. Then, all research team members participated in the pilot test, which involved applying this third version to one published study from each of the four research areas. The results of a pilot test, together with the research team's comments, led to modifications and/or additions to items and response options that were deemed necessary, resulting in the final version of the data extraction spreadsheet (see S3_Data_Extraction_Spreadsheet) with response options protected by drop-down lists and checkboxes (see Figures 1a and 1b).

Although most elements of the data extraction spreadsheet are relatively straightforward, certain items warrant additional clarification. Items were categorized as green, which assessed characteristics of the intervention not directly related to its implementation, or blue, which evaluated whether the information reported on implementation was sufficient to enable accurate and unambiguous replication. While both green and blue items shared response options such as Yes and No, their interpretation differed: a Yes response for a green item indicated that the statement was factually true (e.g., a Yes for Item G1a signified that at least one study objective involved assessing the feasibility of the intervention), whereas a Yes response for a blue item signified that the reported information was replicable (e.g., a Yes for Item G7 indicated that the implementation sequence could be reproduced).

Figure 1a.

Data extraction. Report on the general characteristics of the intervention.

Item			
G0	Description	Specify	
G1	Objective/s of the study in addition to the development of the intervention program:	Resp	
	a. Evaluation of the **feasibility** of the intervention	Y/N	
	b. Evaluation of the expected effects (*efficacy, effectiveness*) of the intervention	Y/N	
G2	Previous publication of **intervention program** (completely o partially) or registration of the **intervention protocol**	R/P/C/N	
	a. Part of the intervention program is described in **non peer-reviewed documents** (e.g., unpublished papers, conference proceedings, grey literature)	Y/N	
	b. Part of the intervention program is described in **inaccessible documents** (e.g., broken links)	Y/N	
G3	Reporting **quality assessment tool** of the study	Y/N	Specify
		Resp	Met
G4	**Number of categories of recipients** (i.e., those whose intervention *effects* are tested; ***individual level*** -e.g. categories: patients, professionals, policy makers-, ***group/team level*** -e.g. categories: family, clinical staff, teaching staff, patient support groups-, or ***system level*** -e.g. categories: communities, health centers/hospitals, schools, organizations-)	O/M/L	☐
G5	**Facilitators or barriers** (i.e., characteristics or factors related to material resources, recipients, providers and/or context that may facilitate or hinder the implementation of the intervention)	Y/N	☐
G6	**Prerequisites for the whole intervention** (necessary to implement the whole intervention; should include a description of the previous tasks to be carried out by the providers, evaluators and/or recipients of the intervention, or specific characteristics of these or the context in which the intervention is implemented):	Resp	Met
	a. Providers (e.g., training, competencies/skills, irrelevant/unnecessary)	Y/D/I/N	☐
	b. Evaluators (e.g., training, competencies/skills, irrelevant/unnecessary)	Y/D/I/N	☐
	c. Recipients (e.g., training, competencies/skills, specific characteristics, irrelevant/unnecessary)	Y/D/I/N	☐
	d. From the context (*setting*) (e.g., social, cultural, political, economic, physical, geographical/spatial, digital, irrelevant/unnecessary)	Y/D/I/N	☐
G7	**Implementation sequence** (both the behaviors or actions -BeA- of the recipients and the tasks to be performed by providers and evaluators)	Y/I/N	☐
	a. Report includes a **diagram (flowchart or infographic)** of the intervention program	Y/N	☐
	b. Report includes **table/s (or list/s)** describing the structure of the intervention program	Y/N	☐
G8	**Intervention outcomes** (adding up the different outcomes of all BeA)	S/M	☐
G9	**Costs** (i.e., costs of the necessary resources, e.g., personnel, materials, equipment, for the implementation of the intervention)	Y/N	☐
G10	**General considerations** on the quality of the report or on problems for the replicability of the intervention	Specify	
	Extraction time:	Minutes	

Notes: Y: yes or reported and/or replicable; D: description only and not clearly stated; I: reported, but not replicable (not enough Information); N: no or not reported; R: protocol registered; P: intervention previously published (partially); C: intervention previously published (completely); O: one category; M: multi-category; L: multi-level; S: single; M: multiple; BeA: behaviors and actions; Resp: response; Met: information included in the methods section.

Figure 1b.

Data extraction. Report of each of the recipients' behaviors or actions (BeA).

Item		BeA 1		[...]	BeA n	
B0	BeA **Description**	Specify			Specify	
			Met			Met
B1	**Levels and categories of recipients** (i.e., those whose intervention *outcomes* are collected)	Resp	☐		Resp	☐
	a. Individual level (e.g. categories: patients, professionals, policy makers)	N/O/M	Specify		N/O/M	Specify
	b. Group/team level (e.g. categories: family, clinical staff, teaching staff, patient support groups)	N/O/M	Specify		N/O/M	Specify
	c. System level (e.g. categories: communities, health centers/hospitals, schools, organizations)	N/O/M	Specify		N/O/M	Specify
		Resp	Met		Resp	Met
B2	**Facilitators or barriers** (i.e., characteristics or factors related to material resources, recipients, providers and/or context that may facilitate or hinder the implementation of the BeA).	Y/N	☐		Y/N	☐
B3	Outcomes and effects of a particular BeA:	Resp	Met		Resp	Met
	a. Number of outcomes (specific, but not exclusive, to particular BeA and measured on their recipients)	N/S/M	☐		N/S/M	☐
	b. Moderation or mediation effect on the outcomes of other BeA?	Y/N/U	☐		Y/N/U	☐
B4	**Adaptation/Customization or not** (i.e., protocol admits implementation flexibility)	Y/D/N	☐		Y/D/N	☐
B5	**Evaluation of the fidelity** of the implementation (including adherence)	Y/I/N	☐		Y/I/N	☐
B6	Prerequisites for a particular BeA:	Resp	Met		Resp	Met
	a. Providers (e.g., training, competencies, skills, qualification, profession, irrelevant/unnecessary)	Y/D/I/N	☐		Y/D/I/N	☐
	b. Evaluators (e.g., training, competencies, skills, irrelevant/unnecessary).	Y/D/I/N	☐		Y/D/I/N	☐
	c. Recipients (e.g., training, competencies, skills, personal characteristics, symptomatology, motivation, access to/or specific resources, irrelevant/unnecessary)	Y/D/I/N	☐		Y/D/I/N	☐
	d. From the **context** (*setting*) (e.g., social, cultural, political, economic, physical, geographical/spatial, digital, irrelevant)	Y/D/I/N	☐		Y/D/I/N	☐
B7	Implementation:	Resp	Met		Resp	Met
	a. Type of administration (i.e., remote, face-to-face, mixed, *tailored*).	Y/I/N	☐		Y/I/N	☐
	b. Form of administration (i.e., targeting a single individual/group/system, targeting 2 or more individuals/groups simultaneously, mixed, *tailored*)	Y/I/N	☐		Y/I/N	☐
	c. Supervision/support (i.e., directed [requires concurrent intervention of providers] with one-way interaction [e.g., supervision], directed with two-way interaction [e.g., on-demand support], autonomous [not require concurrent intervention of providers], mixed, *tailored*)	Y/I/N	☐		Y/I/N	☐
	d. Materials and support tools (e.g., audio, video, text/booklet/leaflet, computer hardware and software, mobile phone apps, gaming, pharmacological, etc., or irrelevant/unnecessary)	Y/I/N	☐		Y/I/N	☐
	e. Content and development of the sessions	Y/I/N	☐		Y/I/N	☐
	f. Periodicity (i.e., time unit; e.g., once, daily, weekly, monthly, *tailored*)	Y/I/N	☐		Y/I/N	☐
	g. Frequency per time unit (e.g., once, a certain number of times, *tailored*)	Y/I/N	☐		Y/I/N	☐
	h. Total frequency (e.g., specific number, *tailored*)	Y/I/N	☐		Y/I/N	☐
	i. Duration of sessions, dose or intensity (e.g., specific, *tailored*)	Y/I/N	☐		Y/I/N	☐
	j. Geographical, spatial, physical and/or digital place (e.g., home, specific/specialized place [e.g., school, clinic, gym, *outdoor*, etc.], *tailored*, irrelevant/unnecessary)	Y/I/N	☐		Y/I/N	☐
	k. Moment in time (e.g., specific, *tailored*, irrelevant)	Y/I/N	☐		Y/I/N	☐
	l. Social, cultural, political and/or economic setting, system, environment (e.g., specific, *tailored*, irrelevant)	Y/I/N	☐		Y/I/N	☐

Notes: Y: yes or reported and/or replicable; D: description only and not clearly stated; I: reported, but not replicable (not enough Information); N: no, none or not reported; O: one category; M: multi-category; L: multi-level; S: single; M: multiple; U: unclear; BeA: behaviors and actions; Resp: response; Met: information included in the methods section.

All items referring to information that should be reported in the Methods section included a checkbox (“Met”), which was ticked if that information was indeed reported in that section.

Items were organized into two spreadsheets: one containing items that pertained to the overall intervention, and another containing the items that affected each of the behaviors and/or actions (BeA) of the intervention’s target recipients.

Items G6 and B6 include a “D” response option (“description only and not clearly stated”), which is absent in the other items. These items aim to capture information about prerequisites related to providers, evaluators, recipients, and contextual factors. During the pilot test, it was observed that some studies described the characteristics of the sample (participants), the professionals who administered or measured the results, or the characteristics of the intervention's implementation context, but did not specify whether this information referred to prerequisites for implementing the intervention or was merely a description of how the study was conducted. For these cases, response option "D" was considered.

The data extraction process was conducted similarly to the study selection process. Two reviewers performed independently the data extraction. Discrepancies during this process were resolved by consensus between the two reviewers, with the participation of a third reviewer specialist in the intervention field (AC, SL, JM/MFR, CM/EL/PV) when necessary. Prior to data extraction, the two raters identified and agreed on the BeA of each study according to the definition provided in the introduction: any constituent of an intervention with specific implementation characteristics that might also have a verifiable effect on one or more intervention outcomes was considered a potential BeA.

Data analysis

A narrative synthesis of the results was carried out using descriptive statistical tables and bar charts. The analyses were organized according to the specific objectives set for the scoping review. Thus, a first descriptive analysis of the general characteristics of the interventions and their BeA was carried out, and the possible differences between the four research fields. Secondly, possible differences in the quality of reporting (replicability) of items related to the implementation of the intervention were evaluated due to: a) the research field of the intervention; b) the use by the study authors of guidelines for writing the report; and c) the characteristics of the intervention related to its complexity, such as the number of BeA, the number of categories and levels of the intervention recipients, and the level of adaptability or customization of the implementation reported by its authors (JAMOVI v.2.6.4).

The raw data of the study can be consulted in S3_Raw_Data.

Ethics statement

This systematic review does not require approval from an ethics committee, as no data was collected directly from humans or animals.

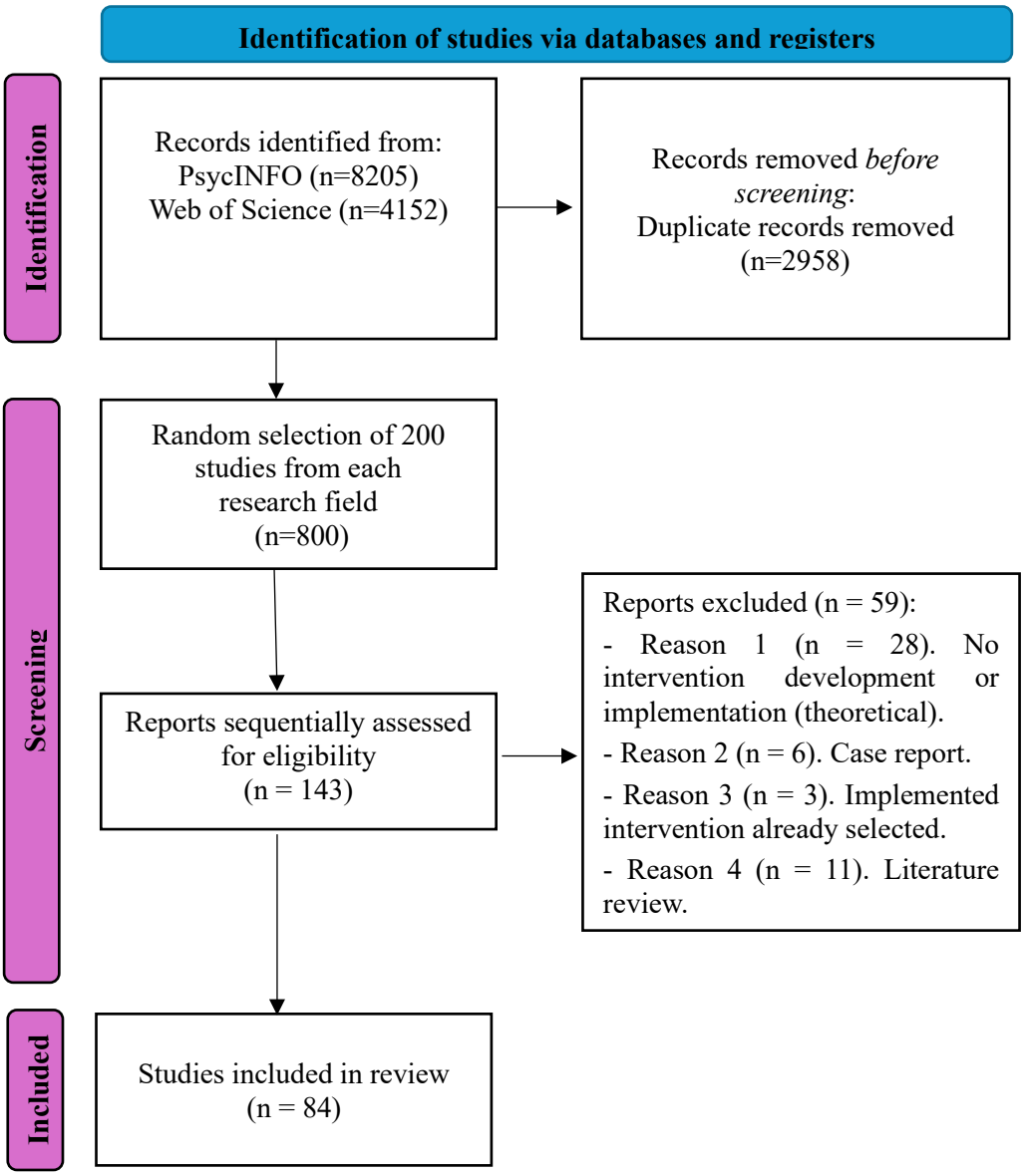
Results

Search results

Figure 2 shows the results of the search strategy, reported according to the PRISMA flowchart (Page et al., 2021). Finally, 84 studies were selected, 4 were used for testing the extraction

procedure, and the remaining 80 were included in the review. The interrater agreement for abstract and title selection was 100%, so no kappa coefficient needed to be calculated, whereas the interrater agreement for full text selection was kappa k=0.78 (CI95%: 0.59-0.96). After discussion to solve the discrepancies, the interrater agreement percentage for full text selection was 100%.

Figure 2.
PRISMA flow diagram of the selection process.



Characteristics of the whole interventions

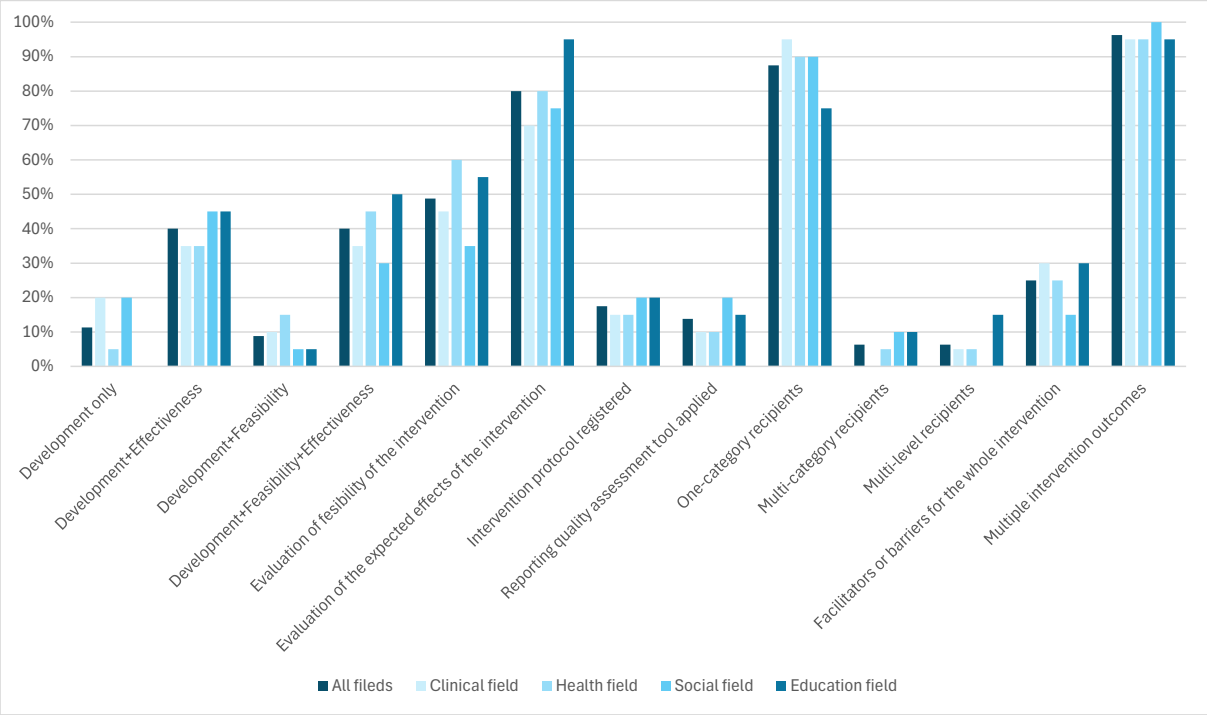
Table 1 and Figure 3 summarize the key characteristics of intervention programs in four research fields (clinical, health, social, and education). The analysis included chi-square tests to assess differences between fields. Below, we first describe the overall patterns observed for each characteristic across all fields combined, followed by a comparison across the individual fields.

Across all fields, only 11.3% of studies focused exclusively on the development of the intervention program. Most publications combined development with an additional objective, most often the assessment of effectiveness (40%) or both feasibility and effectiveness (40%). Feasibility evaluation was reported in just under half of the studies (48.8%), while the majority (80%) assessed the expected effects (efficacy or effectiveness). Most studies (61.3%) did not previously publish the intervention or register their protocol, and only 13.8% applied a reporting quality assessment tool.

In terms of recipients, most interventions targeted a single category (87.5%), while a small number addressed multiple levels (6.3%). Facilitators or barriers for the intervention were reported in only 25% of cases. Finally, nearly all interventions (96.3%) included multiple outcomes.

No statistically significant differences were found across research fields, though descriptive trends emerged. Development-only objectives appeared in clinical and social research (20% each) but were absent in education, while health studies rarely stopped at development (5%). Combined evaluations of feasibility and effectiveness were most frequent in education (50%) and least in social research (30%). Feasibility assessments dominated health (60%) and education (55%) but were less common in social studies (35%). Expected effects were almost universally assessed in education (95%) and less so in clinical research (70%). Reporting quality tools were rarely used, ranging from 10% of clinical and health to 20% of social studies. Most interventions targeted a single recipient category, especially in clinical (95%) and health (90%), whereas education showed greater diversity, with 15% of multi-level interventions compared to 0-5% in the other fields. Finally, reporting of facilitators or barriers was highest in education and clinical (30%) and lowest in social research (15%).

Figure 3.
Characteristics of the whole interventions.



Characteristics of the intervention's recipients' behaviors or actions (BeA)

Table 2 and Figure 4 outline the characteristics related to the BeA, both at global level and for each of the four specific research fields that were analyzed.

Across all fields, most interventions included more than one recipients' BeA. While 30% addressed a single BeA and 21.3% two BeA, the largest proportion (37.5%) reported 3–4 BeA, and a smaller proportion (11.3%) addressed 5–8 BeA. Most studies described recipients' behaviors at a single level (86.3%), 97,6% considered the individual level, 15.1% the group/team level, and only 1.3% the system level.

Facilitators or barriers for particular BeA were almost absent (1.3%). In terms of outcomes, just over half of the interventions (55%) reported multiple outcomes or effects for BeA, while 45% reported none. Interactive effects among BeA were infrequently explored: only 11.3% of studies reported moderating or mediating effects of the BeA on the outcomes of other BeA, and 5% were unclear.

No statistically significant differences were observed between fields. Nevertheless, some descriptive variations can be noted. Education showed a slightly higher proportion of interventions with 3–4 BeA (45%), while clinical and social fields had somewhat higher rates of single BeA (35% each). Clinical studies reported the highest proportion of single-level interventions (95%), while health interventions were slightly more likely to include two recipient levels (20 %). At individual level, education showed slightly more diversity, with 10% of studies not reporting this level, while at group/team level, education (20%) and social (15%) included it slightly more often than clinical (5%). At system level, only education (5%) reported it. Finally, clinical research led in testing moderation and/or mediation effects (20 %), followed by education (15 %), with health and social both at 5 %.

Figure 4.
Characteristics of the intervention recipients' behaviors or actions (BeA).

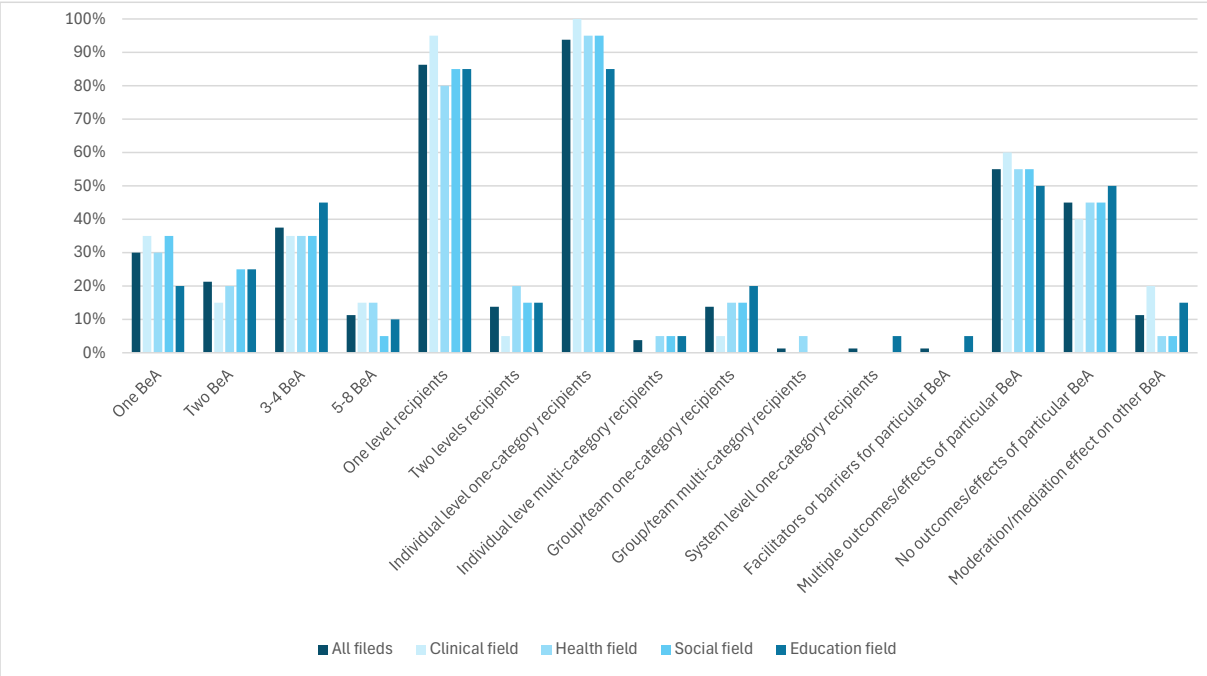


Table 1.

Characteristics of the whole intervention by field.

		All fields %	Clinical %	Health %	Social %	Education %	Between fields X^2 sig.
Additional objective/s of the study in addition to the development of the intervention program	Development only	11.3	20	5	20	0	0.438
	+Effectiveness	40	35	35	45	45	
	+Feasibility	8.8	10	15	5	5	
	+Feasibility+Effectiveness	40	35	45	30	50	
Evaluation of feasibility of the intervention	Yes	48.8	45	60	35	55	0.399
	No	51.3	55	40	65	45	
Evaluation of the expected effects (efficacy, effectiveness) of the intervention	Yes	80	70	80	75	95	0.224
	No	20	30	20	25	5	
Previous publication of intervention program (completely o partially) or registration of the intervention protocol	Intervention previously published (completely)	7.5	10	5	10	5	0.998
	Intervention previously published (partially)	13.8	15	15	10	15	
	Protocol registered	17.5	15	15	20	20	
	Without prior publication or registration	61.3	60	65	60	60	
Reporting quality assessment tool applied	Yes	13.8	10	10	20	15	0.788
	No	86.2	90	90	80	85	
Number of categories of recipients of the whole intervention	One-category	87.5	95	90	90	75	0.368
	Multi-category	6.3	0	5	10	10	
	Multi-level	6.3	5	5	0	15	
Facilitators or barriers for the whole intervention	Yes	25	30	25	15	30	0.659
	No	75	70	75	85	70	
Intervention outcomes of the whole intervention	Single	3.8	5	5	0	5	0.792
	Multiple	96.3	95	95	100	95	

Table 2.

Characteristics of the intervention's recipients' behaviors or actions (BeA) by field.

		All fields %	Clinical %	Health %	Social %	Education %	Between fields X^2 sig.
Number of BeA	One BeA	30	35	30	35	20	0.953
	Two BeA	21.3	15	20	25	25	
	3-4 BeA	37.5	35	35	35	45	
	5-8 BeA	11.3	15	15	5	10	
Number of levels of recipients of particular BeA	One level	86.3	95	80	85	85	0.572
	Two levels	13.8	5	20	15	15	
Levels & categories of recipients of particular BeA: Individual level	One-category	93.8	100	95	95	85	0.298
	Multi-category	3.8	0	5	5	5	
	No	2.5	0	0	0	10	
Levels & categories of recipients of particular BeA: Group/team level	One-category	13.8	5	15	15	20	0.534
	Multi-category	1.3	0	5	0	0	
	No	85	95	80	85	80	
Levels & categories of recipients of particular BeA: System level	One-category	1.3	0	0	0	5	0.386
	No	98.8	100	100	100	95	
Facilitators or barriers for particular BeA	Yes	1.3	0	0	0	5	0.386
	No	98.8	100	100	100	95	
Number of outcomes/effects of particular BeA	Multiple	55	60	55	55	50	0.939
	None	45	40	45	45	50	
Moderation or mediation effect of particular BeA on the outcomes of other BeA	Yes	11.3	20	5	5	15	0.523
	Unclear	5	0	10	5	5	
	No	83.8	80	85	90	80	

Table 3.

Quality reporting of implementation characteristics by field.

	All fields			Clinical field			Health field			Social field			Education field			Between fields
	RR %	RNR %	NR %	RR %	RNR %	NR %	RR %	RNR %	NR %	RR %	RNR %	NR %	RR %	RNR %	NR %	X^2 sig.
Implementation characteristics of the whole intervention																
Global prerequisites of the whole intervention	15	77.5	7.5	10	85	5	15	75	10	15	75	10	20	75	5	0.958
Providers' prerequisites	35	11.3	53.7	40	5	55	50	0	50	20	30	50	30	10	60	0.055
Evaluators' prerequisites	35	12.5	52.5	20	15	65	45	0	55	35	25	40	40	10	50	0.208
Recipients' prerequisites	57.5	12.5	30	75	5	20	60	10	30	50	25	25	45	10	45	0.256
Context prerequisites	70	7.5	22.5	65	0	35	65	15	20	75	10	15	75	5	20	0.467
Global prerequisites of the whole intervention or for BeA	13.8	78.8	7.5	10	85	5	10	80	10	15	75	10	20	75	5	0.934
Implementation sequence	73.8	17.5	8.8	80	15	5	75	15	10	60	25	15	80	15	5	0.789
Report includes a diagram of the intervention program	16.3	0	83.8	20	0	80	25	0	75	10	0	90	10	0	90	0.479
Report includes tables/lists of the structure of the intervention program	38.8	0	61.3	65	0	35	20	0	80	30	0	70	40	0	60	0.024*
Costs	3.8	0	96.3	0	0	100	5	0	95	10	0	90	0	0	100	0.283
Implementation characteristics of recipients' behaviors or actions (BeA)																
Adaptation/Customization or not	35	15	50	40	5	55	40	20	40	15	20	65	45	15	40	0.326
Evaluation of the fidelity/adherence of the implementation	27.5	7.5	65	20	0	80	40	10	50	25	5	70	25	15	60	0.533
Global prerequisites for particular BeA	32.5	61.3	6.3	30	65	5	35	60	5	35	55	10	30	65	5	0.978
Providers' prerequisites	41.3	3.8	55	45	5	50	55	5	40	35	0	65	30	5	65	0.619
Evaluators' prerequisites	43.8	0	56.3	30	0	70	55	0	45	45	0	55	45	0	55	0.459
Recipients' prerequisites	61.3	2.5	36.3	75	5	20	60	5	35	60	0	40	50	0	50	0.466
Context prerequisites	82.5	5	12.5	85	0	15	85	5	10	75	15	10	85	0	15	0.361
Global characteristics of the implementation	13.8	86.3	0	15	85	0	15	85	0	0	100	0	25	75	0	0.146
Type of administration (e.g., remote, face-to-face)	73.8	21.3	5	80	15	5	70	30	0	60	25	15	85	15	0	0.198
Form of administration (e.g., individual, group)	80	16.3	3.8	90	5	5	80	15	5	65	30	5	85	15	0	0.446
Supervision/support (e.g., directed, autonomous, mixed)	72.5	22.5	5	80	15	5	75	25	0	65	25	10	70	25	5	0.807
Materials and support tools (e.g., multimedia, text, app)	47.5	36.3	16.3	60	20	20	50	45	5	40	35	25	40	45	15	0.405
Content and development of the sessions	36.3	55	8.8	65	25	10	25	70	5	10	80	10	45	45	10	0.011*
Periodicity (time unit; e.g., once, daily, weekly, monthly)	75	11.3	13.8	90	0	10	60	15	25	75	15	10	75	15	10	0.354
Frequency per time unit (e.g., once, a number of times)	70	13.8	16.3	75	10	15	60	20	20	70	15	15	75	10	15	0.948
Total frequency (e.g., specific number, tailored)	76.3	12.5	11.3	90	0	10	70	15	15	75	15	10	70	20	10	0.591
Duration of sessions, dose or intensity (specific, tailored)	56.3	20	23.8	55	10	35	55	25	20	50	20	30	65	25	10	0.540
Geographical, spatial, physical and/or digital place	88.8	10	1.3	90	10	0	95	5	0	70	25	5	100	0	0	0.083
Moment in time (e.g., specific, tailored, irrelevant)	95	0	5	95	0	5	100	0	0	90	0	10	95	0	5	0.551
Social, cultural, political and/or economic setting	91.3	5	3.8	90	10	0	95	5	0	85	5	10	95	0	5	0.444

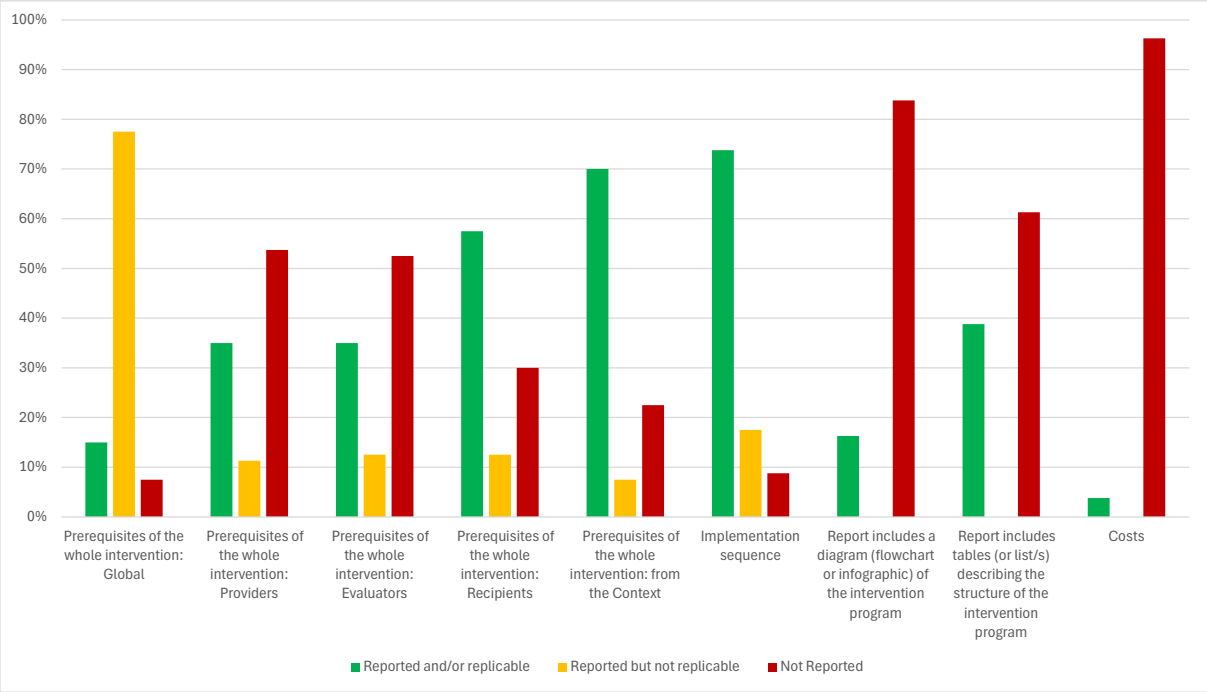
Notes: RR: reported and/or replicable; RNR: reported but not replicable (missing information); NR: not reported; *: statistically significant ($p<0.05$)

334 ***Quality reporting of implementation characteristics of the whole interventions***

335 Table 3 and Figure 5 present the quality reporting of implementation of the intervention studies
336 analyzed, categorized by reporting status: reported and/or replicable (RR), reported but not
337 replicable due to missing information (RNR), and not reported (NR).

338 Across fields, the reporting of implementation characteristics at the level of the whole
339 intervention showed a mixed picture, with notable weaknesses in several areas. Global
340 prerequisites of the intervention were seldom reported in sufficient detail: only 15% of studies
341 provided replicable information, while most either offered incomplete descriptions (77.5%) or
342 omitted them altogether (7.5%). This pattern was consistent across the different groups of
343 prerequisites. Although recipients' (57.5%) and context-related prerequisites (70%) were more
344 frequently addressed, information on providers (35%) and evaluators (35%) was less often
345 available in a replicable form. When prerequisites of the whole intervention and recipients'
346 behaviors were considered jointly, reporting was very limited, with only 13.8% adequately
347 covered. By contrast, the implementation sequence was more consistently described: nearly
348 three-quarters of studies (73.8%) included enough detail to allow replication. Nevertheless,
349 structured reporting formats that could enhance clarity and replicability were underutilized.
350 Only 16.3% of studies presented a diagram, and fewer than half (38.8%) provided tables or lists
351 outlining the intervention's structure. Costs were almost entirely neglected (3.8%).

352 Figure 5.
353 *Quality reporting of implementation characteristics of the whole interventions.*

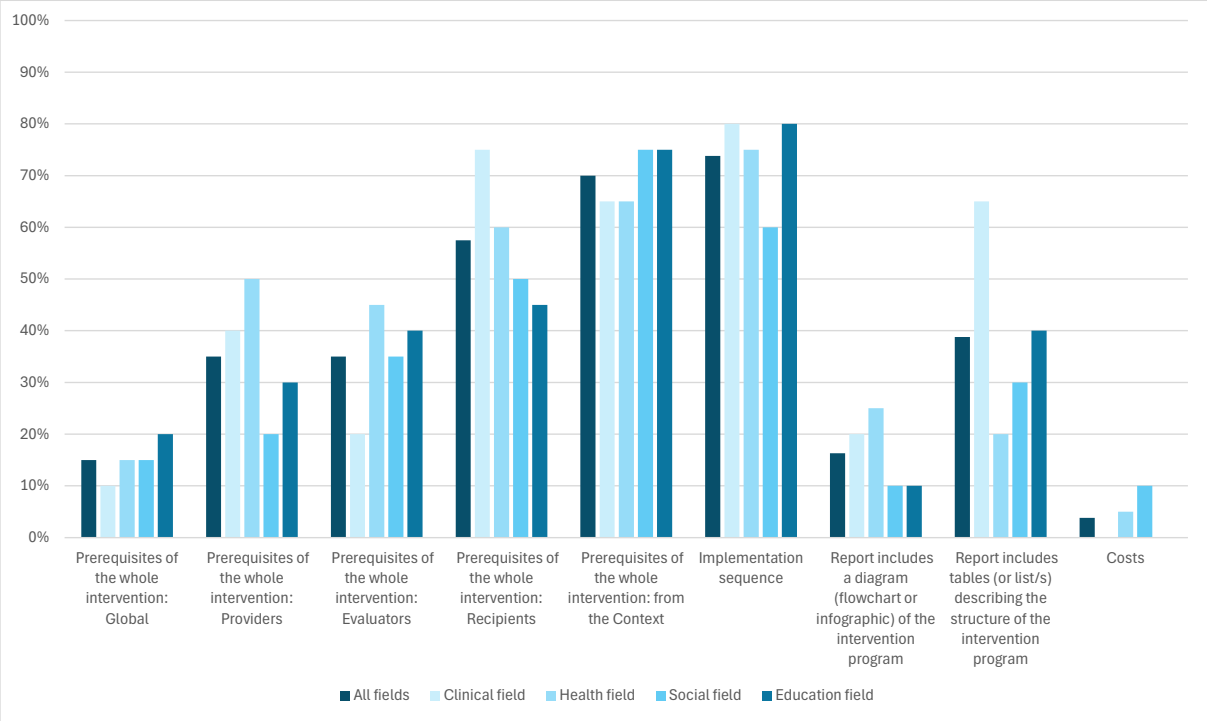


354

355 No statistically significant differences were observed between fields in the percentage of
356 characteristics well reported (see Figure6), except for the reporting of tables or lists of
357 intervention structure ($p = 0.024$), which was highest in clinical (65%) and education (40%)
358 compared to health (20%).Recipients' prerequisites were most often reported in a replicable
359 manner in clinical (75%) and health (60%) compared to education (45%, and providers'
360 prerequisites were adequately reported more often in health (50%) than in education (30%).

Reporting of tables or lists describing the structure of the intervention program was more frequent in clinical (65%) compare to other fields (20–40%).

Figure 6.
Percentage of implementation characteristics of the whole interventions reported and/or replicable by field.



Finally, it is worth noting that, regarding the section of the publications in which the previously described general characteristics of the interventions were reported, these were not included in the Method section of the studies in between 14.6% (administrator prerequisites) and 37% (evaluator prerequisites) of the cases. Moreover, in 75.8% of the studies, the reporting of contextual prerequisites (setting) was also located in sections other than the Method section.

Quality reporting of implementation characteristics of the intervention's recipients' behaviors or actions (BeA)

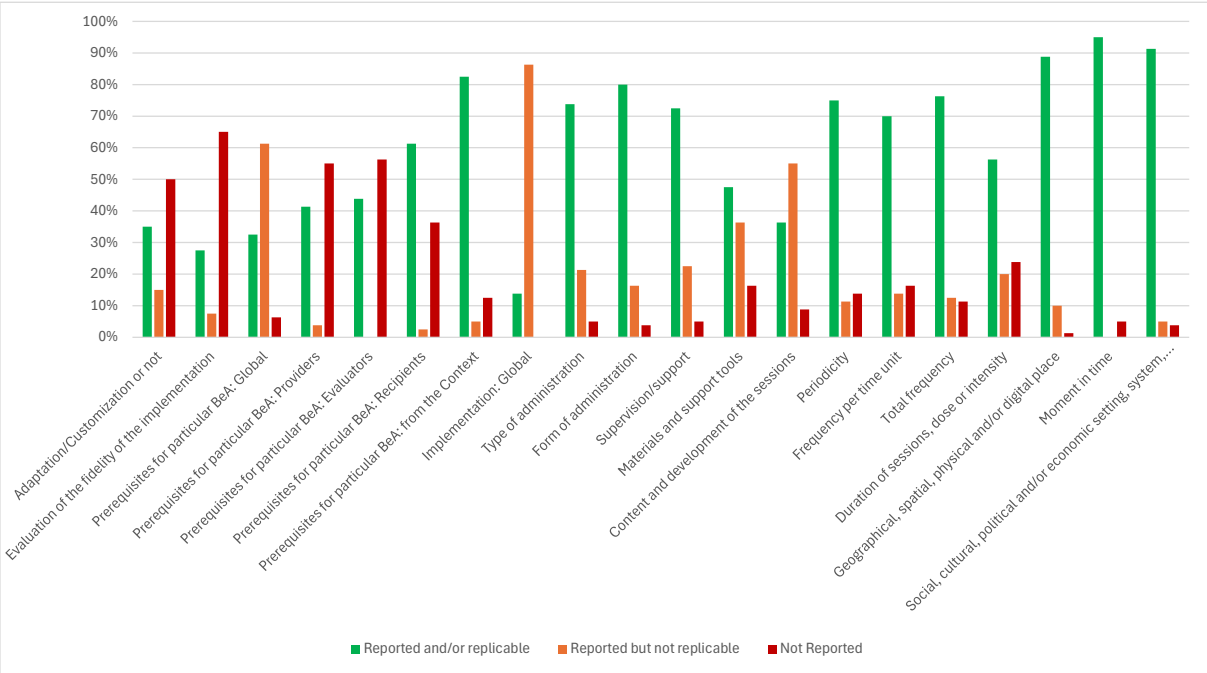
At the level of BeA, reporting quality of the implementation characteristics of the interventions also varied (Table 3 and Figure 7). The reporting of BeA adaptation or customization—entailing whether the intervention was adaptable or customizable, and, if so, what adaptations or customizations were permitted—was done in a replicable manner in just over a third of studies (35%), with half failing to address this aspect at all. Similarly, fidelity and adherence were reported by 27.5% of studies with adequate detail and nearly two-thirds (65%) omitting this implementation feature.

Global prerequisites for recipients' behaviors were also weakly documented: while 32.5% reported them with replicable detail, most studies (61.3%) provided incomplete information. When broken down by type of prerequisite, reporting was uneven. Recipients' (61.3%) and context-related prerequisites (82.5%) were more frequently specified. By contrast, providers (41.3%) and evaluators (43.8%) were less consistently addressed.

In contrast, certain global characteristics of implementation were relatively comprehensively addressed. Most studies showed adequate reporting on key structural aspects, such as type (73.8%) and form of administration (80%), supervision or support (72.5%), and periodicity (75%). Frequency (70–76.3%), location (88.8%), and timing (95%) were also commonly well reported. Moreover, the social, cultural, political, and economic setting was frequently reported and/or replicable (91.3%). However, two key elements remained comparatively underreported: the content and development of sessions (36.3%) and the materials or support tools used (47.5%). The reporting of the dose or intensity of the sessions was replicable in just over half (56.3%) of the studies. Overall, the low percentage of "not reported" responses observed for most of the features analyzed should be interpreted with caution, as while they indicated that these implementation features were universally addressed, this was rarely done in a fully replicable manner in light of the observed percentage of "reported but not replicable (missing information)" results.

Figure 7.

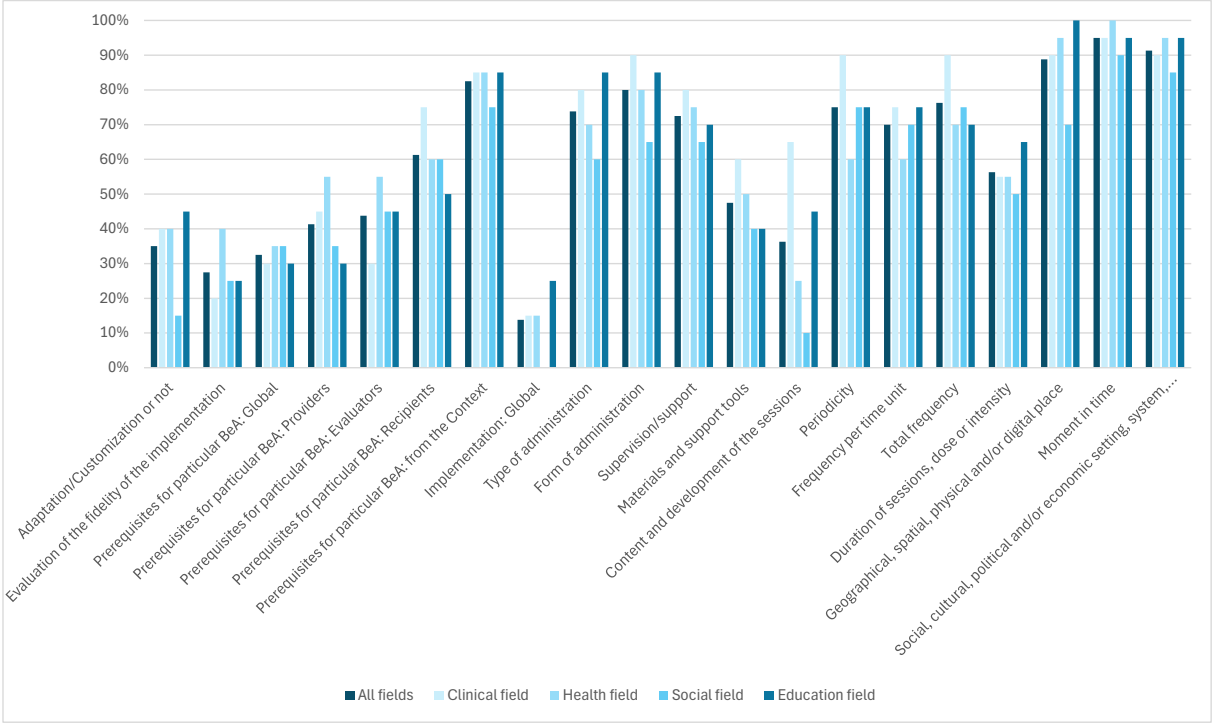
Quality reporting of implementation characteristics of the intervention's recipients' behaviors or actions (BeA).



No statistically significant differences were found across fields (Figure 8) except for content and development of sessions ($p = 0.011$), where the replicability of the reports was greater in the clinical field (65%) and lowest in health (25%) and social (10%) fields. Adaptation/customization was reported more frequently and accurately in clinical and health fields (40% each) than in social (15%). Reporting on implementation fidelity was most adequate in health (40%) and lowest in clinical (20%) field.

Figure 8.

Percentage of implementation characteristics of intervention's recipients' behaviors or actions (BeA) reported and/or replicable by field.



Finally, it should be noted that, concerning the section of the publications in which the characteristics previously described of the BeA were reported, the four groups of implementation prerequisites were not included in the Method section in between 70% (administrator prerequisites) and 86.4% (evaluator prerequisites) of the studies. The remaining implementation characteristics of the BeA were reported outside the Method section in less than 15% of the studies, except for the materials and support tools required for BeA implementation, their potential for adaptation/customization, and the procedures used to assess implementation fidelity, which were reported outside the Method section in 37.3%, 25%, and 17.9% of the studies, respectively.

Quality reporting of implementation characteristics of the interventions by number of intervention's recipients' behaviors or actions (BeA)

Table 4 and Figure 9 examine the quality of reporting for implementation characteristics in intervention studies, stratified by the number of BeA involved (1 BeA, 2 BeA, 3 BeA, 4 BeA, and 5-8 BeA). The reporting of implementation characteristics for the whole intervention varied substantially depending on the number of BeA considered. When only one BeA was addressed, replicable reporting of global prerequisites was almost absent (4.2%), and only a minority of studies included providers' (16.7%) or evaluators' (12.5%) prerequisites, and recipients' prerequisites were reported in a replicable manner somewhat more frequently (25%), and context prerequisites were well documented at a relatively high level (79.2%).

As the number of BeA increased, the completeness of reporting improved. For example, in interventions with 2 BeA, reporting of global prerequisites increased to 11.8%, providers' to 35.3%, and evaluators' to 35.3%. By the time interventions addressed 3–4 BeA, global prerequisites were well described in 22.2% of studies, with evaluators' (44.4%) and especially

Table 4.

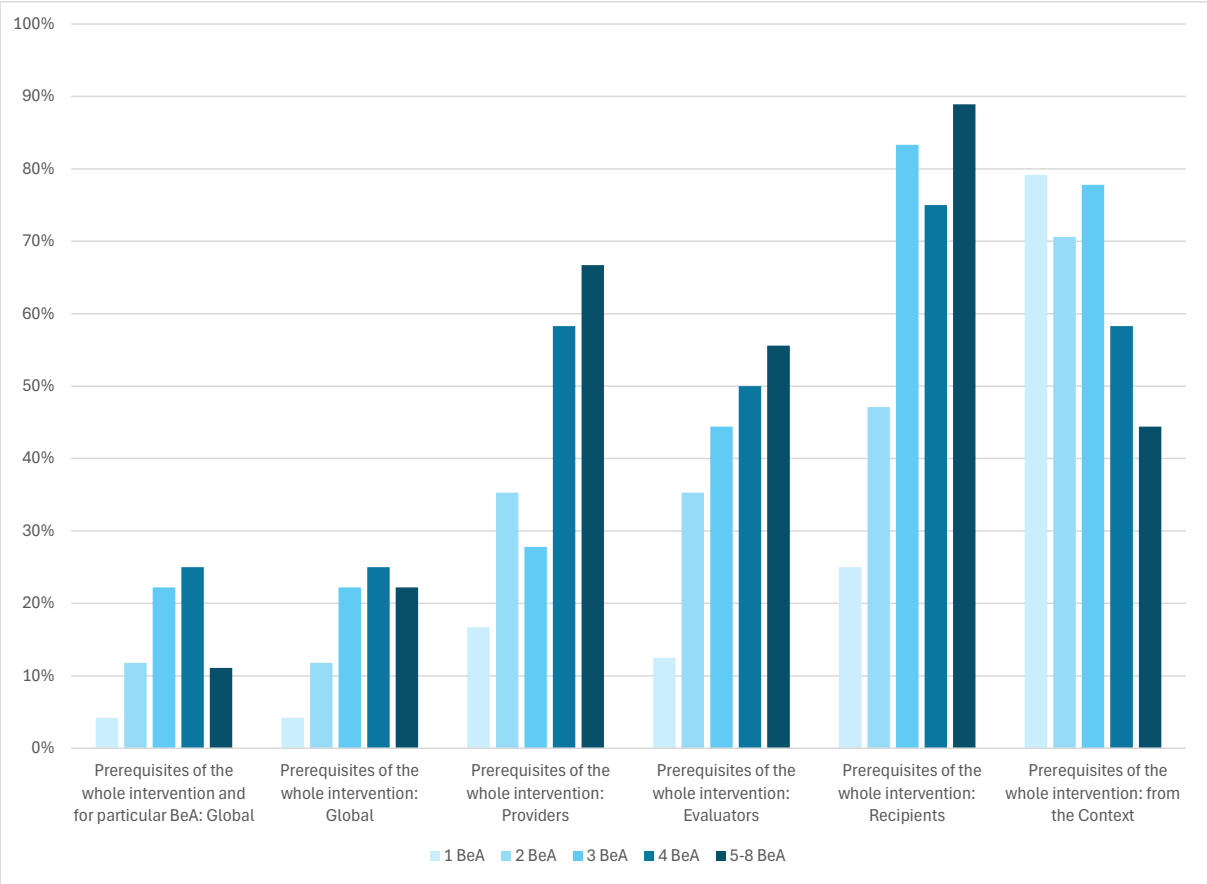
Quality reporting of implementation characteristics by number of BeA.

	1 BeA			2 BeA			3 BeA			4 BeA			5-8 BeA			Between fields
	RR %	RNR %	NR %	RR %	RNR %	NR %	RR %	RNR %	NR %	RR %	RNR %	NR %	RR %	RNR %	NR %	X^2 sig.
Implementation characteristics of the whole intervention																
Global prerequisites of the whole intervention	4.2	83.3	12.5	11.8	76.5	11.8	22.2	77.8	0	25	75	0	22.2	66.7	11.1	0.473
Providers' prerequisites	16.7	20.8	62.5	35.3	0	64.7	27.8	11.1	61.1	58.3	16.7	25	66.7	0	33.3	0.047*
Evaluators' prerequisites	12.5	8.3	79.2	35.3	17.6	47.1	44.4	5.6	50	50	33.3	16.7	55.6	0	44.4	0.016*
Recipients' prerequisites	25	20.8	54.2	47.1	11.8	41.2	83.3	5.6	11.1	75	16.7	8.3	88.9	0	11.1	0.004*
Context's prerequisites	79.2	0	20.8	70.6	5.9	23.5	77.8	11.1	11.1	58.3	16.7	25	44.4	11.1	44.4	0.412
Global prerequisites of the whole intervention or for BeA	4.2	83.3	12.5	11.8	76.5	11.8	22.2	77.8	0	25	75	0	11.1	77.8	11.1	0.488
Implementation characteristics of recipients' behaviors or actions (BeA)																
Global prerequisites for particular BeA	4.2	83.3	12.5	29.4	58.8	11.8	38.9	61.1	0	58.3	41.7	0	66.7	33.3	0	0.009
Providers' prerequisites	20.8	4.2	75	35.3	5.9	58.8	38.9	0	61.1	75	0	25	66.7	11.1	22.2	0.049
Evaluators' prerequisites	12.5	70.8	87.5	41.2	52.9	58.8	50	11.1	50	75	8.3	25	77.8	0	22.2	<.001*
Recipients' prerequisites	25	4.2	70.8	47.1	0	52.9	83.3	5.6	11.1	91.7	0	8.3	100	0	0	<.001*
Context prerequisites	75	4.2	20.8	88.2	0	11.8	88.9	5.6	5.6	75	8.3	16.7	88.9	11.1	0	0.691
Global characteristics of the implementation	12.5	87.5	0	17.6	82.4	0	11.1	88.9	0	16.7	83.3	0	11.1	88.9	0	0.974
Type of administration (e.g., remote, face-to-face)	79.2	8.3	12.5	88.2	11.8	0	55.6	38.9	5.6	66.7	33.3	0	77.8	22.2	0	0.144
Form of administration (e.g., individual, group)	83.3	12.5	4.2	88.2	5.9	5.9	72.2	22.2	5.6	75	25	0	77.8	22.2	0	0.856
Supervision/support (e.g., directed, autonomous, mixed)	79.2	16.7	4.2	88.2	5.9	5.9	55.6	44.4	0	75	25	0	55.6	22.2	22.2	0.058
Materials and support tools (e.g., multimedia, text, app)	50	29.2	20.8	47.1	23.5	29.4	38.9	50	11.1	66.7	25	8.3	33.3	66.7	0	0.219
Content and development of the sessions	45.8	37.5	16.7	41.2	47.1	11.8	27.8	72.2	0	33.3	58.3	8.3	22.2	77.8	0	0.323
Periodicity (time unit; e.g., once, daily, weekly, monthly)	66.7	4.2	29.2	70.6	17.6	11.8	83.3	5.6	11.1	91.7	8.3	0	66.7	33.3	0	0.073
Frequency per time unit (e.g., once, a number of times)	70.8	4.2	25	64.7	17.6	17.6	72.2	16.7	11.1	83.3	8.3	8.3	55.6	33.3	11.1	0.497
Total frequency (e.g., specific number, tailored)	70.8	4.2	25	70.6	17.6	11.8	88.9	11.1	0	83.3	8.3	8.3	66.7	33.3	0	0.115
Duration of sessions, dose or intensity (specific, tailored)	62.5	4.2	33.3	58.8	35.3	5.9	50	27.8	22.2	58.3	8.3	33.3	44.4	33.3	22.2	0.180
Geographical, spatial, physical and/or digital place	91.7	4.2	4.2	82.4	17.6	0	83.3	16.7	0	91.7	8.3	0	100	0	0	0.627
Moment in time (e.g., specific, tailored, irrelevant)	91.7	0	8.3	100	0	0	100	0	0	83.3	0	16.7	100	0	0	0.177
Social, cultural, political and/or economic setting	91.7	4.2	4.2	94.1	5.9	0	94.4	5.6	0	75	8.3	16.7	100	0	0	0.410

Notes: RR: reported and/or replicable; RNR: reported but not replicable (missing information); NR: not reported; *: statistically significant ($p<0.05$)

recipients' prerequisites (83.3%) reported more often. This trend culminated in interventions with 4 BeA, where reporting of prerequisites was strongest: 25% for global prerequisites, 58.3% for providers, 50% for evaluators, and 75% for recipients. Context prerequisites remained relatively stable across groups (58–88%).

Figure 9.
Percentage of implementation characteristics of the whole interventions reported and/or replicable by number of recipients' behaviors or actions (BeA).

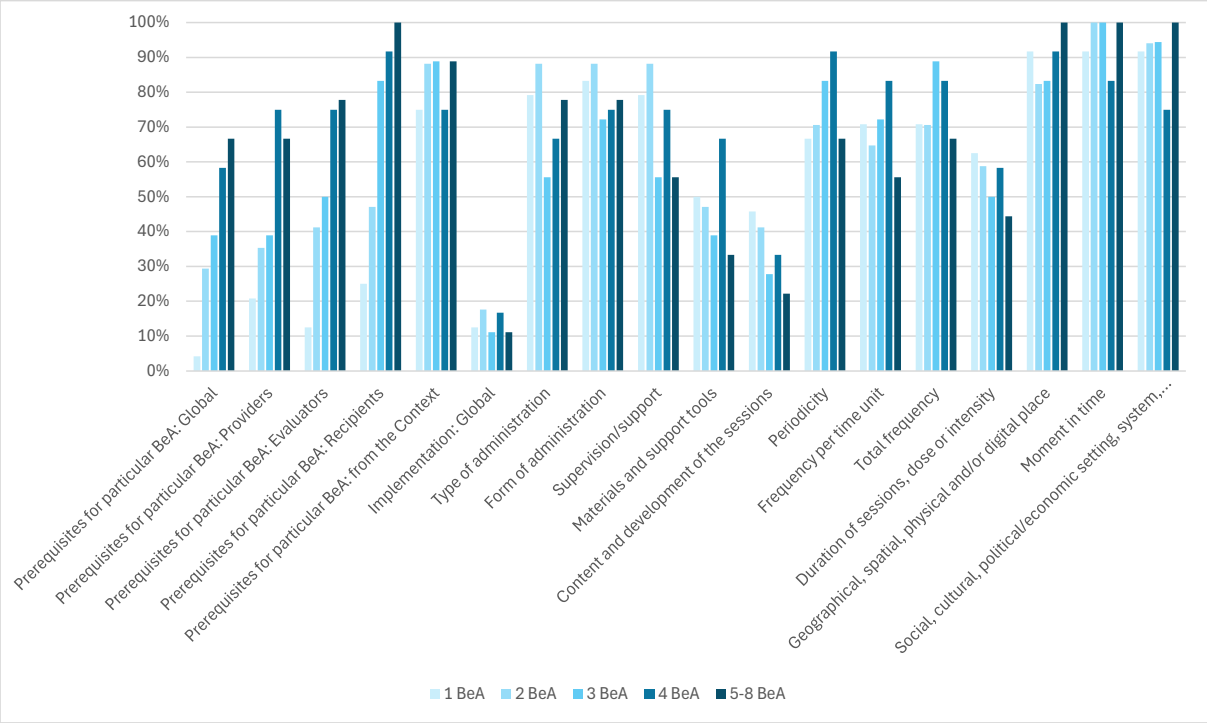


At the level of BeA (Figure 10), the same pattern of improvement with increasing complexity was observed. For interventions with only one BeA, global prerequisites were reported in a replicable manner in 4.2% of cases, while providers', evaluators', and recipients' prerequisites were accurately reported in 20.8%, 12.5%, and 25%, respectively. Again, context prerequisites were relatively well covered even in the simplest interventions (75%).

As the number of BeA increased, reporting became progressively more complete. With 2 BeA, global prerequisites well reported reached 29.4%, providers' 35.3%, evaluators' 41.2%, and recipients' 47.1%, while context prerequisites climbed to 88.2%. For 3–4 BeA, adequate reporting was higher: 38.9% for global prerequisites, 50% for evaluators, and 83.3% for recipients. In interventions with 4 BeA, the reporting of prerequisites reached its peak, with 58.3% replicable for global prerequisites, 75% for providers and evaluators, 91.7% for recipients, and 75% for context. For the most complex interventions (5–8 BeA), reporting levels were also high, with recipients' prerequisites accurately reported universally (100%) and evaluators' prerequisites in 77.8%. Global implementation characteristics followed a slightly different pattern, generally showing low replicability (11.1–17.6%)

Figure 10.

Percentage of implementation characteristics of the intervention's recipients' behaviors or actions (BeA) reported and/or replicable by number of BeA.



Discussion

This scoping review mapped how psychological intervention reports describe implementation at both the whole-intervention level and at the level of recipients' behaviors or actions (BeA). Across 80 included studies from clinical, health, social and educational research fields, only 11.3% reported *development-only* objectives; most combined development with evaluations of effectiveness (80% overall) and, to a lesser extent, feasibility (48.8%). Just 13.8% used a formal reporting quality tool. Most interventions targeted a single recipient category; multi-level targeting was rare (6.3%). Reporting of implementation sequence was comparatively strong (73.8% replicable), yet structured representations were underused (tables/lists 38.8%; diagrams 16.3%). Costs were almost never reported (3.8%). At the BeA level, generic structural features (type and form of administration, supervision/support, periodicity, location, timing) were often replicably reported, whereas content/materials, adaptation/customization, and especially fidelity/adherence were frequently incomplete (content 36.3% replicable; materials 47.5%; adaptation/customization 35%; fidelity 27.5%). Finally, features of complexity—multi-level recipients and inter-component relations (moderation/mediation)—were seldom analyzed (11.3%). Reporting quality tended to improve as the number of BeA increased.

This pattern suggests that authors tend to prioritize reporting on the "when" and "where" of interventions, and points to a key weakness: while interventions are usually situated within their contextual settings, they are less often described in ways that allow for faithful replication or critical appraisal of their mechanisms.

Our findings show that, despite decades of guideline development (CONSORT, STROBE, PRISMA, SPIRIT, TREND) and intervention-specific tools (TIDieR, CReDECI-2, GUIDED,

iCAT_SR, RIPI-f), adoption remains limited (only 13.8% used a reporting quality tool) and key implementation details remain missing (e.g., fidelity, costs). This aligns with evidence that endorsement alone does not guarantee adherence and complete reporting (e.g., Turner et al., 2012; Bastuji-Garin et al., 2013; Clayson et al., 2019).

Our results also resonate strongly with TIDieR's emphasis on specifying *what* was delivered (materials/content) and *how* (dose, mode, procedures). The comparatively weak reporting for content/materials (and moderate reporting for dose/intensity) indicates that, even when structural aspects are documented, the "active ingredients" of interventions are often left implicit. Using content taxonomies—such as the Behavior Change Technique Taxonomy (BCTTv1)—can sharpen the granularity and comparability of content reporting and thereby strengthen replication and synthesis (Michie et al., 2013).

Contemporary frameworks (MRC 2015 process evaluation guidance; MRC 2021 complex interventions update; iCAT_SR) ask authors to articulate *how* components are expected to work (mechanisms), *for whom*, *in what contexts*, and with *what interdependencies*. The rarity of multi-level targeting and explicit tests of mediation/moderation among BeA suggests that many reports still treat interventions largely as *packages* rather than as *systems of interacting parts*. Explicitly modeling and reporting inter-component relations—and using logic models to link BeA to outcomes—would align practice with these frameworks and support richer cumulative science (Moore et al., 2015).

Two practical routes could accelerate progress. First, for interventions operating at population or policy levels, the TIDieR-PHP (Campbell et al., 2018) extension tailors itemization to the realities of system-level actions (e.g., legal, fiscal, environmental levers) and should be considered in addition to generic templates. Second, for remote and hybrid delivery, TIDieR-Telehealth (Zachrison et al., 2022) provides a focused extension that captures technology, platforms, and delivery contingencies, which were unevenly described in our sample and particularly relevant to modern psychological services.

Two encouraging signals emerge. First, the implementation sequence was often reported with sufficient detail (73.8% replicable), and in more complex interventions (more BeA), the reporting of prerequisites (especially for recipients and evaluators) improved. Second, generic structural elements (e.g., type/form of administration, supervision, periodicity) were frequently replicable. These strengths offer a foundation to build upon. We argue that journals and authors should make structured representations—tables that map BeA to their materials, content, procedures, dose, and evaluation; and flow diagrams of sequences—*standard*. This would operationalize the spirit of TIDieR and related tools and appears feasible given our findings.

A further implication concerns publication architecture. Critical prerequisites (e.g., for providers/evaluators) and contextual details are frequently reported outside the Methods section, impeding discoverability and machine-readability for evidence synthesis. Consolidating these items in Methods (with stable supplementary files hosting manuals, checklists, and materials) would align with SPIRIT/CONSORT transparency goals and facilitate automated screening by meta-research tools.

Finally, two gaps stand out. First, implementation fidelity—central to interpreting outcomes and avoiding Type III errors—was rarely reported in a replicable way (27.5%). The conceptual framework of Carroll et al. (2007) describes fidelity as a multi-dimensional construct

(adherence, dose, quality of delivery, participant responsiveness, differentiation) that should be planned, measured, and reported systematically; the pattern we observed suggests these dimensions remain under-specified in psychological intervention reports. Moreover, meta-research consistently shows that implementation quality is tied to effect sizes, reinforcing the practical import of robust fidelity reporting (Durlak & DuPre, 2008). Second, costs were almost never reported, yet budget impact and cost-effectiveness information are essential to scale-up and policy translation. The CHEERS 2022 statement (Husereau et al., 2022) provides explicit, modernized guidance on transparent reporting of economic evaluations and is broadly applicable across intervention types; the near-absence of cost reporting in our sample signals an actionable opportunity for improvement in future reports. From a practice standpoint, insufficient reporting of fidelity, adaptation policies, and costs undermines implementation planning and weakens the claim that a delivered service is truly *evidence-based*. The taxonomy of Proctor et al. (2011) clarifies that feasibility, fidelity, penetration, sustainability, and cost are implementation outcomes in their own right; adequate reporting of these domains is therefore necessary both for replication and for implementation science to cumulate knowledge.

Limitations

This scoping review aimed to map reporting practices rather than quantify causal relationships or effectiveness. Although selection was randomized within fields, the sample size per field (n=20) limits power to detect small differences and may not capture all sub-disciplinary practices. Judgments of "replicable" versus "reported but not replicable" required coder interpretation; we mitigated this through piloting, independent double extraction, and consensus procedures, achieving substantial inter-rater agreement at full-text screening ($\kappa = 0.78$; 95% CI 0.59–0.96). We did not contact study authors, so information available only in unpublished manuals or inaccessible resources could have been missed; however, the data extraction instrument explicitly recorded references to prior publications and inaccessible documents, tempering this risk. Finally, the review window (2014–2024) may underrepresent very recent impacts of new guidance (e.g., MRC 2021 update; CHEERS 2022), and findings pertain to the four predefined application domains (clinical, public health, social/organizational, education), which may limit generalizability beyond them.

Conclusions and future directions

Reporting of psychological interventions shows encouraging coverage of structural "when/where/how" elements and implementation sequences, but persistent weaknesses in the what (content/materials), how well (fidelity), whether and how it changes (adaptation/customization), and how much it costs (economic data). These deficiencies blunt replication, weaken cumulative syntheses, and impede implementation at scale—precisely the problems the reporting-guideline ecosystem was designed to solve. Bringing routine use of TIDieR (and extensions such as TIDieR-PHP and TIDieR-Telehealth), fidelity frameworks, content taxonomies (BCTTv1), and economic reporting standards (CHEERS 2022) into psychological intervention reporting would close the most consequential gaps identified in our scoping review.

Future research and practice priorities should focus on a) standardizing the level of BeA: jointly developing a minimum set of BeA reports (e.g., a TIDieR BeA table) that explicitly links each BeA to prerequisites, materials/content (using BCTTv1 tags; Michie et al., 2013), procedures,

dosages, fidelity measures, and outcomes, including hypothetical mediation/moderation between BeAs; b) fidelity measurement and reporting: prospectively adopting multicomponent fidelity frameworks and evaluating whether improved fidelity reporting enhances effect estimates and reduces heterogeneity in meta-analyses (Carroll et al., 2007); c) economic transparency: integrate the elements of CHEERS 2022 (resource use, costs, analytical perspective) into intervention reports, even when a full economic evaluation is not conducted, to inform scaling-up decisions (Husereau et al., 2022); d) complexity-based designs: increase the routine use of logic models and process evaluations (MRC guidance) to assess component performance in different contexts, including multi-level receivers and intercomponent pathways (Moore et al., 2015); and e) editorial and automation levers: evaluate editorial interventions (e.g., mandatory checklists, structured tables/diagrams, automated report audits) that have demonstrated improvements in integrity in other domains, adapting them to the required level of BeA granularity.

Data availability statement

The data supporting the findings of this study are openly available in the Open Science Framework repository at https://doi.org/10.34810/data2792.

Funding

This work was supported by Grant PID2022-141403NB-I00 funded by the Spanish Government (MCIN/ AEI/10.13039/501100011033/FEDER, UE), and 2021SGR-00806, funded by the Government of Catalonia (Spain). The authors declare no conflict of interest, and the funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript. No additional external funding was received for this study.

Competing interests

The authors have declared that no competing interests exist.

References

Almaqrami, B. S., Hua, F., Liu, Y., & He, H. (2020). Research waste-related editorial policies of leading dental journals: Situation 2018. *Oral Diseases*, *26*(3), 696–706. https://doi.org/10.1111/odi.13257

Bastuji-Garin, S., Sbidian, E., Gaudy-Marqueste, C., Ferrat, E., Roujeau, J.-C., Richard, M.-A., Canoui-Poitrine, F., & on behalf of the European Dermatology Network (EDEN). (2013). Impact of STROBE Statement Publication on Quality of Observational Study Reporting: Interrupted Time Series versus Before-After Analysis. *PLoS ONE*, *8*(8), e64733. https://doi.org/10.1371/journal.pone.0064733

Begg, C. (1996). Improving the Quality of Reporting of Randomized Controlled Trials: The CONSORT Statement. *JAMA*, *276*(8), 637. https://doi.org/10.1001/jama.1996.03540080059030

Blanco, D., Schroter, S., Aldcroft, A., Moher, D., Boutron, I., Kirkham, J. J., & Cobo, E. (2020). Effect of an editorial intervention to improve the completeness of reporting of randomised trials: A randomised controlled trial. *BMJ Open*, *10*(5), e036799. https://doi.org/10.1136/bmjopen-2020-036799

Campbell, M., Katikireddi, S. V., Hoffmann, T., Armstrong, R., Waters, E., & Craig, P. (TIDieR-PHP Group). (2018). TIDieR-PHP: A reporting guideline for population health and policy interventions. *BMJ, 361*, k1079. https://doi.org/10.1136/bmj.k1079

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science, 2*, 40. https://doi.org/10.1186/1748-5908-2-40

Chan, A.-W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., Hróbjartsson, A., Mann, H., Dickersin, K., Berlin, J. A., Doré, C. J., Parulekar, W. R., Summerskill, W. S. M., Groves, T., Schulz, K. F., Sox, H. C., Rockhold, F. W., Rennie, D., & Moher, D. (2013). SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials. *Annals of Internal Medicine*, *158*(3), 200–207. https://doi.org/10.7326/0003-4819-158-3-201302050-00583

Clayson, P. E., Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*, *56*(11). https://doi.org/10.1111/psyp.13437

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, a1655. https://doi.org/10.1136/bmj.a1655

Des Jarlais, D. C., Lyles, C., & Crepaz, N. (2004). Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement. *American Journal of Public Health*, *94*(3), 361–366. https://doi.org/10.2105/AJPH.94.3.361

Duncan, E., O'Cathain, A., Rousseau, N., Croot, L., Sworn, K., Turner, K. M., Yardley, L., & Hoddinott, P. (2020). Guidance for reporting intervention development studies in health research (GUIDED): An evidence-based consensus study. *BMJ Open*, *10*(4), e033516. https://doi.org/10.1136/bmjopen-2019-033516

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., Michie, S., Moher, D., & Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, *383*(9913), 267–276. https://doi.org/10.1016/S0140-6736(13)62228-X

Guise, J.-M., Butler, M. E., Chang, C., Viswanathan, M., Pigott, T., & Tugwell, P. (2017). AHRQ series on complex intervention systematic reviews—paper 6: PRISMA-CI extension statement and checklist. *Journal of Clinical Epidemiology*, *90*, 43–50. https://doi.org/10.1016/j.jclinepi.2017.06.016

Guise, J.-M., Chang, C., Butler, M., Viswanathan, M., & Tugwell, P. (2017). AHRQ series on complex intervention systematic reviews—paper 1: An introduction to a series of articles that provide guidance and tools for reviews of complex interventions. *Journal of Clinical Epidemiology*, *90*, 6–10. https://doi.org/10.1016/j.jclinepi.2017.06.011

Hawe, P., Shiell, A., & Riley, T. (2009). Theorising Interventions as Events in Systems. *American Journal of Community Psychology*, *43*(3–4), 267–276. https://doi.org/10.1007/s10464-009-9229-9

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, *348*(mar07 3), g1687–g1687. https://doi.org/10.1136/bmj.g1687

Hopewell, S., Boutron, I., Chan, A.-W., Collins, G. S., de Beyer, J. A., Hróbjartsson, A., Nejstgaard, C. H., Østengaard, L., Schulz, K. F., Tunn, R., & Moher, D. (2022). An update to SPIRIT and CONSORT reporting guidelines to enhance transparency in randomized trials. *Nature Medicine*, *28*(9), 1740–1743. https://doi.org/10.1038/s41591-022-01989-8

Husereau, D., Drummond, M., Augustovski, F., de Bekker-Grob, E., Briggs, A. H., Carswell, C., Caulley, L., Chaiyakunapruk, N., Greenberg, D., Loder, E., Mauskopf, J., Mullins, C. D., Petrou, S., Pwu, R.-F., & Staniszewska, S. (2022). Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) statement: Updated reporting guidance for health economic evaluations. *BMC Medicine, 20*, 23. https://doi.org/10.1186/s12916-021-02204-0

Ioannidis, J. P. A., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., Schulz, K. F., & Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*, *383*(9912), 166–175. https://doi.org/10.1016/S0140-6736(13)62227-8

Lewin, S., Hendry, M., Chandler, J., Oxman, A. D., Michie, S., Shepperd, S., Reeves, B. C., Tugwell, P., Hannes, K., Rehfuess, E. A., Welch, V., Mckenzie, J. E., Burford, B., Petkovic, J., Anderson, L. M., Harris, J., & Noyes, J. (2017a). Assessing the complexity of interventions within systematic reviews: Development, content and use of a new tool (iCAT_SR). *BMC Medical Research Methodology*, *17*(1), 76. https://doi.org/10.1186/s12874-017-0349-x

Lewin, S., Hendry, M., Chandler, J., Oxman, A. D., Michie, S., Shepperd, S., Reeves, B. C., Tugwell, P., Hannes, K., Rehfuess, E. A., Welch, V., Mckenzie, J. E., Burford, B., Petkovic, J., Anderson, L. M., Harris, J., & Noyes, J. (2017b). Assessing the complexity of interventions within systematic reviews: Development, content and use of a new tool (iCAT_SR). *BMC Medical Research Methodology*, *17*(1). https://doi.org/10.1186/s12874-017-0349-x

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, *151*(4), W65–W94. https://doi.org/10.1136/bmj.b2700

Lopez-Alcalde, J., Yakoub, N., Wolf, M., Munder, T., Von Elm, E., Flückiger, C., Steinert, C., Liebherz, S., Rosendahl, J., Witt, C. M., & Barth, J. (2022). The RIPI-f (Reporting Integrity of Psychological Interventions delivered face-to-face) checklist was developed to guide reporting of treatment integrity in face-to-face psychological interventions. *Journal of Clinical Epidemiology*, *151*, 65–74. https://doi.org/10.1016/j.jclinepi.2022.07.013

Losilla, J.-M., Oliveras, I., Marin-Garcia, J. A., & Vives, J. (2018). Three risk of bias tools lead to opposite conclusions in observational research synthesis. *Journal of Clinical Epidemiology*, *101*, 61–72. https://doi.org/10.1016/j.jclinepi.2018.05.021

Mannocci, A., Saulle, R., Colamesta, V., D'Aguanno, S., Giraldi, G., Maffongelli, E., Meggiolaro, A., Semyonov, L., Unim, B., & La Torre, G. (2014). What is the impact of reporting guidelines on Public Health journals in Europe? The case of STROBE, CONSORT and PRISMA. *Journal of Public Health*, fdu108. https://doi.org/10.1093/pubmed/fdu108

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., Eccles, M. P., Cane, J., & Wood, C. E. (2013). The Behavior Change Technique Taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine, 46*(1), 81–95. https://doi.org/10.1007/s12160-013-9486-6

Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet*, *357*(9263), 1191–1194. https://doi.org/10.1016/S0140-6736(00)04337-3

Möhler, R., Köpke, S., & Meyer, G. (2015). Criteria for Reporting the Development and Evaluation of Complex Interventions in healthcare: Revised guideline (CReDECI 2). *Trials*, *16*, 204. https://doi.org/10.1186/s13063-015-0709-y

Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., & Moher, D. (2018). Reporting randomised trials of social and psychological interventions: The CONSORT-SPI 2018 Extension. *Trials*, *19*(1), 407. https://doi.org/10.1186/s13063-018-2733-1

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ, 350*, h1258. https://doi.org/10.1136/bmj.h1258

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1). https://doi.org/10.1186/s12874-018-0611-x

Noyes, J., Gough, D., Lewin, S., Mayhew, A., Michie, S., Pantoja, T., Petticrew, M., Pottie, K., Rehfuess, E., Shemilt, I., Shepperd, S., Sowden, A., Tugwell, P., & Welch, V. (2013). A research and development agenda for systematic reviews that ask complex questions about complex interventions. *Journal of Clinical Epidemiology*, *66*(11), 1262–1270. https://doi.org/10.1016/j.jclinepi.2013.07.003

Orduña-Malea, E., Alonso-Arroyo, A., Ontalba-Ruipérez, J.-A., & Catalá-López, F. (2022). Evaluating the online impact of reporting guidelines for randomised trial reports and protocols: A cross-sectional web-based data analysis of CONSORT and SPIRIT initiatives. *Scientometrics*. https://doi.org/10.1007/s11192-022-04542-z

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Petticrew, M., Anderson, L., Elder, R., Grimshaw, J., Hopkins, D., Hahn, R., Krause, L., Kristjansson, E., Mercer, S., Sipe, T., Tugwell, P., Ueffing, E., Waters, E., & Welch, V. (2015). Complex interventions and their implications for systematic reviews: A pragmatic approach. *International Journal of Nursing Studies*, *52*(7), 1211–1216. APA PsycInfo®. https://doi.org/10.1016/j.ijnurstu.2015.01.004

Pinnock, H., Barwick, M., Carpenter, C. R., Eldridge, S., Grandes, G., Griffiths, C. J., Rycroft-Malone, J., Meissner, P., Murray, E., Patel, A., Sheikh, A., & Taylor, S. J. C. (2017). Standards for Reporting Implementation Studies (StaRI) Statement. *BMJ*, i6795. https://doi.org/10.1136/bmj.i6795

Pinnock, H., Epiphaniou, E., Sheikh, A., Griffiths, C., Eldridge, S., Craig, P., & Taylor, S. J. (2015). Developing standards for reporting implementation studies of complex interventions (StaRI): A systematic review and e-Delphi. *Implementation Science*, *10*(1). https://doi.org/10.1186/s13012-015-0235-z

Proctor, E. K., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Finegood, D. T., Greaves, F., Harper, L., Hawe, P., Moore, L., Petticrew, M., Rehfuess, E., Shiell, A., Thomas, J., & White, M. (2017). The need for a complex systems model of evidence for public health. *The Lancet*, *390*(10112), 2602–2604. https://doi.org/10.1016/S0140-6736(17)31267-9

Schulz, K. F., Altman, D. G., Moher, D., & for the CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*(mar23 1), c332–c332. https://doi.org/10.1136/bmj.c332

Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M., & Moore, L. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, n2061. https://doi.org/10.1136/bmj.n2061

Tanner-Smith, E. E., & Grant, S. (2018a). Meta-Analysis of Complex Interventions. *Annual Review of Public Health*, *39*(1), 135–151. https://doi.org/10.1146/annurev-publhealth-040617-014112

Tanner-Smith, E. E., & Grant, S. (2018b). Meta-analysis of complex interventions. *Annual Review of Public Health*, *39*, 135–151. APA PsycInfo®. https://doi.org/10.1146/annurev-publhealth-040617-014112

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Turner, L., Shamseer, L., Altman, D. G., Schulz, K. F., & Moher, D. (2012). Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Systematic Reviews*, *1*(1), 60. https://doi.org/10.1186/2046-4053-1-60

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., Poole, C., Schlesselman, J. J., & Egger, M. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). *Epidemiology*, *18*(6), 805–835. https://doi.org/10.1097/EDE.0b013e3181577511

Vives, J., Lorente, S., Casellas-Grau, A., López-González, E., Marin-Garcia, J. A., Molla, C., Rodrigo, M. F., Viguer, P., & Losilla, J.-M. (2025). *Reporting quality in psychological interventions. Implications for cumulative science and evidence-based practice. A scoping review protocol*. INPLASY - International Platform of Registered Systematic Review and Meta-analysis Protocols. https://doi.org/10.37766/inplasy2025.7.0098

Zachrison, K. S., Yan, Z., Marsh, E. B., McCoy, C., Song, S., Schwamm, L. H., & Bianchi, M. T. (2022). TIDieR-telehealth: Precision in reporting of telehealth interventions used in clinical trials—Unique considerations for the Template for Intervention Description and Replication (TIDieR) checklist. *BMC Medical Research Methodology, 22*, 144. https://doi.org/10.1186/s12874-022-01640-7