



RESEARCH ARTICLE

How common are explicit research questions in journal articles?

Mike Thelwall  and Amalia Mas-Bleda 

Statistical Cybermetrics Research Group, University of Wolverhampton, Wulfruna Street, Wolverhampton WV4 4ST, UK

an open access  journal



Citation: Thelwall, M. & Mas-Bleda, A. (2020). How common are explicit research questions in journal articles?. *Quantitative Science Studies*, Advance publication. https://doi.org/10.1162/qss_a_00041

DOI: https://doi.org/10.1162/qss_a_00041

Supporting Information: https://doi.org/10.1162/qss_a_00041

Received: 08 November 2019
Accepted: 19 January 2020

Corresponding Author:
Mike Thelwall
m.thelwall@wlv.ac.uk

Handling Editor:
Ludo Waltman

Keywords: research aims, research article structures, research goals, research purpose statements, research questions

ABSTRACT

Although explicitly labeled research questions seem to be central to some fields, others do not need them. This may confuse authors, editors, readers, and reviewers of multidisciplinary research. This article assesses the extent to which research questions are explicitly mentioned in 17 out of 22 areas of scholarship from 2000 to 2018 by searching over a million full-text open access journal articles. Research questions were almost never explicitly mentioned (under 2%) by articles in engineering and physical, life, and medical sciences, and were the exception (always under 20%) for the broad fields in which they were least rare: computing, philosophy, theology, and social sciences. Nevertheless, research questions were increasingly mentioned explicitly in all fields investigated, despite a rate of 1.8% overall (1.1% after correcting for irrelevant matches). Other terminology for an article's purpose may be more widely used instead, including aims, objectives, goals, hypotheses, and purposes, although no terminology occurs in a majority of articles in any broad field tested. Authors, editors, readers, and reviewers should therefore be aware that the use of explicitly labeled research questions or other explicit research purpose terminology is non-standard in most or all broad fields, although it is becoming less rare.

1. INTRODUCTION

Academic research is increasingly multidisciplinary, partly due to team research addressing practical problems. There are also now large multidisciplinary journals, such as *PLOS ONE* and *Nature Scientific Reports*, with editorial teams that manage papers written by people from diverse disciplinary backgrounds. There is therefore an increasing need for researchers to understand disciplinary norms in writing styles and paradigms. The authors of a research paper need to know how to frame its central contribution so that it is understood by multidisciplinary audiences. One strategy for this is to base an article around a set of explicitly named research questions that address gaps in prior research. Employing the standard phrase “research question” gives an unambiguous signpost for the purpose of an article and may therefore aid clarity. Other strategies include stating hypotheses, goals, or aims, or describing an objective without calling it an objective (e.g., “this paper investigates X”). Similarly, structured abstracts are believed to help readers understand a paper (Hartley, 2004), perhaps partly by having an explicit aim, objective, or goal section. A paper that does not recognize or value the way in which the central contribution is conveyed may be rejected by a reviewer or editor if they are unfamiliar with the norms of the submitting field. It would therefore be helpful for authors, reviewers, and

Copyright: © 2020 Mike Thelwall and Amalia Mas-Bleda. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



editors to know which research fields employ explicitly labeled research questions or alternative standard terminology.

Purpose statements and research questions or hypotheses are interrelated elements of the research process. Research questions are interrogative statements that reflect the problem to be addressed, usually shaped by the goal or objectives of the study (Onwuegbuzie & Leech, 2006). For example, a healthcare article argued that “A good research paper addresses a specific research question. The research question—or study objective or main research hypothesis—is the central organizing principle of the paper” and “The key attributes are: (i) specificity; (ii) originality or novelty; and (iii) general relevance to a broad scientific community” (Perneger & Hudelson, 2004).

The choice of terminology to describe an article’s purpose seems to be conceptually arbitrary, with the final decision based on community norms, journal guidelines, and author style. For example, a research paper investigating issue X could phrase its purpose in the following ways: “research question 1: is X true?,” “this paper aims to investigate X,” “the aim/objective/purpose/goal is to investigate X,” or “X?” (as in the current paper). Implicit purpose statements might include “this paper investigates X” or just “x,” where the context makes clear that this is the purpose. Alternatively, the reader might deduce the purpose of a paper after reading it, with all these options achieving the same result with different linguistic strategies. Some research purposes might not be easily expressible as a research question, however. For example, a humanities paper might primarily discuss an issue (e.g., “Aspects of the monastery and monastic life in Adomnán’s *Life of Columba*”) but even these could perhaps be expressed as research questions, if necessary (e.g., “Which are the most noteworthy aspects of the monastery and monastic life in Adomnán’s *Life of Columba*?”).

Although there are some guidelines or advice for authors about article purposes, there is no systematic evidence about the prevalence of different strategies or disciplinary differences. In particular, the extent to which explicitly named research questions are used is unknown. In response, this article assesses the extent to which research questions or alternative terminology are explicitly employed in academic research publications, guided by the following.

1. In which fields are explicitly named research questions commonly used?
2. Has the use of explicitly named research questions increased over time?
3. Are research purposes addressed using alternative language in different fields?
4. Do large journals guide authors to use explicitly named research questions or other terminology for purpose statements in different fields?

2. BACKGROUND

2.1. Advice for Authors

There are some influential guidelines for reporting academic research. In the social sciences, Swales’ (1990, 2004) Create A Research Space (CARS) model structures research article introductions in three moves (establishing a territory, establishing a niche, and occupying a niche), which are subdivided into steps. Within the 1990 model, move 3 includes the steps “outlining purposes” and “announcing present research,” but research questions are not explicitly included, being similar the “question raising” step in move 2. In the updated 2004 model, move 3 includes an obligatory step named “announcing present research descriptively and/or purposively” (that joins the steps “outlining purposes” and “announcing present research” from the 1990 model), whereas “listing research questions or hypotheses” is a new optional step.

In medicine, the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) initiative is a checklist of items that should be included to improve reporting quality. One of these is a statement of objectives that “may be formulated as specific hypotheses or as questions that the study was designed to address” or may be less precise in early studies (Vandenbroucke, von Elm, et al., 2014). This description therefore includes stating research questions as one of a range of ways of specifying objectives. An informal advice article in medicine instead starts by arguing that the paper’s aim should be clearly defined (McIntyre, Nisbet, et al., 2007).

Researchers may also be guided about the language to use in papers by any ethical or other procedures that they need to follow before conducting their work. For example, clinical trials often need to be registered and declared in a standard format, which may include explicit descriptions of objectives (e.g., see “E.2.1: Main objective of the trial” at: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2015-002555-10/GB>).

2.2. Empirical Evidence

Journal article research questions and other purpose statements, such as aims, objectives, goals, and hypotheses (Shehzad, 2011), are usually included within Introduction sections or introductory phases, sometimes appearing as separate sections (Kwan, 2017; Yang & Allison, 2004). Some studies have analyzed research article introductions in different disciplines and languages based on the Swales’ (1990, 2004) CARS model. Although these studies analyze small sets of articles, they seem to agree that the research article introduction structure varies across disciplines (e.g., Joseph, Lim & Nor, 2014) and subdisciplines within a discipline, including for engineering (Kanoksilapatham, 2012; Maswana, Kanamaru, & Tajino, 2015), applied linguistics (Jalilifar, 2010; Ozturk, 2007) and environmental sciences (Samraj, 2002). Introductions in English seem to follow this pattern more closely than introductions in other languages (Ahmad & Yusof, 2012; Hirano, 2009; Loi & Evans, 2010; Rahimi & Farnia, 2017; Sheldon, 2011), reflecting cultural differences. Research questions and other purpose terminology, such as aims, objectives, goals, or hypotheses, might also reappear within the Results or Discussion sections (Amunai & Wannaruk, 2013; Brett, 1994; Hopkins & Dudley-Evans, 1988; Kanoksilapatham, 2005).

Previous research has shown that research questions and hypotheses are more common among English-language papers than non-English papers (Loi & Evans, 2010; Mur Dueñas, 2010; Omid & Farnia, 2016; Rahimi & Farnia, 2017; Sheldon, 2011), especially those written by English-native speakers (Sheldon, 2011). However, a study analyzing 119 English research article introductions from Iranian and international journals in three subdisciplines within applied linguistics found that “announcing present research” was more used in international journals whereas research questions were proclaimed explicitly more often in local journals (Jalilifar, 2010).

In some fields the verbs *examine*, *determine*, *evaluate*, *assess*, and *investigate* are associated with the research purpose (Cortés, 2013; Jalali & Moini, 2014; Kanoksilapatham, 2005) and the verbs *expect*, *anticipate*, and *estimate* are associated with hypotheses (Williams, 1999). Some computer scientists seem to prefer to write the details of the method(s) used rather than stating the purpose or describing the nature of their research and use assumptions or research questions rather than hypotheses (Shehzad, 2011). Moreover, scholars might state the hypotheses in other ways, like “it was hypothesized that” (Jalali & Moini, 2014).

A study analyzing lexical bundles (usually phrases) in medical research article introductions showed that the most frequent four-word phrases are related to the research objective,

Q2

such as “the aim of the,” “aim of the present,” and “study was to evaluate” (Jalali & Moini, 2014). Another study examined lexical bundles in a million-word corpus of research article introductions from several disciplines, showing that the main bundle used to announce the research descriptively and/or purposefully included the terms *aim*, *objective*, and *purpose* (e.g., “the aim of this paper,” “the objective of this study,” “the purpose of this paper”), but no bundles related to research questions or hypotheses were identified (Cortés, 2013).

These findings are in line with other previous studies investigating the structure of research articles, especially the introduction section, which report a much higher percentage of journal papers specifying the research purpose than the research questions or hypotheses across disciplines, regardless of the language in which they are published, with the exception of law articles (see Table 1). These studies also show that research questions and hypotheses are much more frequent among social sciences articles (see Table 1), which has also been found in other genres, such as PhD theses and Master’s theses (see Table 2).

A few studies have focused exclusively on research purposes, research questions, and hypotheses. Some have discussed the development of research questions in qualitative (Agee, 2009) or mixed method (Onwuegbuzie & Leech, 2006) studies, whereas others have examined the ways of constructing research questions or hypotheses within some fields, such as organization studies (Sandberg & Alvesson, 2011) or applied linguistics doctoral dissertations (Lim, 2014; Lim, Loi, & Hashim, 2014). Shehzad (2011) examined the strategies and styles employed by computer scientists outlining purposes and listing research questions. She found an increase in the use of research nature or purpose statements and suggested that the “listing research questions or hypotheses” step of Swales’s model was obligatory in computing. No study seems to have examined how often journal guidelines give authors explicit advice about research questions or other purpose statements, however.

3. METHODS

The PMC (Pub Med Central) Open Access subset (www.ncbi.nlm.nih.gov/pmc/tools/openftlist/) was downloaded in XML format in November 2018. This is a collection of documents from open access journals or open access articles within hybrid journals. The collection has a biomedical focus, but includes at least a few articles from all broad disciplinary areas. Although a biased subset is not ideal, this is apparently the largest open access collection. Only documents declared in their XML to be of type research-article were retained for analysis. This excludes many short contributions, such as editorials, that would not need research goals.

The XML of the body section of each article was searched for the test strings “research question,” “RESEARCH QUESTION,” “Research Question,” or “Research question,” recording whether each article contained at least one. This would miss papers exclusively using abbreviations, such as RQ1 or other abbreviations.

Full body text searches are problematic because terms could be mentioned in other contexts, depending on the part of an article. For example, the phrase “research question” in a literature review section may refer to an article reviewed. For a science-wide analysis it is not possible to be prescriptive about the sections in which a term must occur, however, because there is little uniformity in section names or orders (Thelwall, 2019). Making simplifying assumptions about the position in a text in which a term should appear, such as that a research question should be stated in the first part of an article, would also not be defensible. This is because the structure of articles varies widely between journals and fields. For example, methods can appear at the end rather than the middle, and some papers start with results, with little introduction. There are also international cultural differences in the order in which

Table 1. Studies analyzing research article introductions that report whether they present their research purpose and whether they introduce research questions (RQs) and/or hypotheses.

Study	Field	Present the research purpose	Specify the RQs or hypotheses
Mahzari & Maftoon (2007) N = 200 (100 each language)	Medicine	English: 93% Persian: 83%	English: 0% Persian: 9%
Tessuto (2015) N = 90	Law	English: 28%	English: 20%
Rahimi & Farnia (2017) N = 70 (35 each language)	Dentistry	English: 71.4% Persian: 91.4%	English: 37% Persian: 0%
Omidi & Farnia (2016) N = 60 (30 each language)	Physical education	English: 100% Persian: 86.6%	English: 73.3% Persian: 23.3%
Loi & Evans (2010) N = 40 (20 each language)	Educational psychology	English: 100% Chinese: 80%	English: 35% (RQs), 75% (H) Chinese: 10% (RQs), 0% (H)
Kanoksilapatham (2012) N = 179	Civil engineering (n = 60) Software engineering (n = 60) Biomed. engineering (n = 59)	English: 78.3%* English: 81.7%* English: 83%*	
Maswana, Kanamaru, & Tajino (2015) N = 69	Structural engineering (n = 9) Environ. engineering (n = 15) Electrical engineering (n = 21) Chemical engineering (n = 14) Computer science (n = 8)	English: 89%* English: 100%* English: 90%* English: 93%* English: 100%*	English: 22%** English: 33%** English: 14%** English: 43%** English: 13%**
Shehzad (2011) N = 56	Computer science	English: 98.21%	English: 32.14%
Posteguillo (1999) N = 40	Computer science	English: 95%	English: 22.5%
Anthony (1999) N = 12	Software engineering	English: 42%	English: 0%
Sheldon (2011) N = 36 (18 each language)	Applied linguistics	—	English: 33% Spanish: 5.5%
Del Saz Rubio (2011) N = 28	Agricultural sciences	English: 89.3%	English: 14.3%
Mur Dueñas (2010) N = 24	Business management	English: 100% Spanish: 92%	English: 100% (H) Spanish: 25% (H)
Joseph, Lim & Nor (2014) N = 20	Forestry	English: 100%	English: 50%

* Reference to a wide research purposes, without specifying if they are objectives or RQs/hypotheses.

** Restating RQs in the result section.

Note: Studies that have based their analysis on the Swales's (1990) CARS model (Anthony, 1999; Posteguillo, 1999; Mahzari & Maftoon, 2007) report the percentage related to "outlining purposes" and "announcing present research." For these studies, the column "Present the research purpose" reports the higher value. Moreover, for these studies, the value reported in the RQs/hypotheses column refers to the "Question raising" information.

sections are presented in some fields (Teufel, 1999). The current paper therefore uses full-text searches without any heuristics to restrict the results for transparency and to give an almost certain upper bound to the prevalence of terms, given the lack of a high-quality alternative.

Articles were separated into broad fields using the Science-Metrics public journal classification scheme (Archambault, Beauchesne, & Caruso, 2011), which allocates each journal into

Table 2. Studies analyzing PhD thesis and Master thesis introductions that report whether they present their research purpose and whether they introduce research questions (RQs) and/or hypotheses.

Study	Field	Present the research purpose	Specify the RQs or Hypotheses
Lim (2014) <i>N</i> = 32	Applied linguistics	–	English: 97% (RQs)
Lim, Loi, & Hashim (2014) <i>N</i> = 32	Applied linguistics	–	English: 50% (H)
Gil-Salom, Soler-Monreal, & Carbonell-Olivares (2008) <i>N</i> = 21	Computing	Spanish: 100%	Spanish: 4.8%
Soler-Monreal, Carbonell-Olivares, & Gil-Salom (2011) <i>N</i> = 20 (10 each language)	Computing	English: 60% Spanish: 80%	English: 50% Spanish: 10%
Geçikli (2013) <i>N</i> = 20 (10 each language)	English Language Teaching	English: 90% Turkish: 40%	English: 90% Turkish: 70%
Cheung (2012) <i>N</i> = 41	Applied linguistics (<i>n</i> = 11) Physical education (<i>n</i> = 10) Applied psychology (<i>n</i> = 10) Education management (<i>n</i> = 10)	English: 73% English: 90% English: 90% English: 90%	English: 82% (RQs), 9% (H) English: 30% (RQs), 30% (H) English: 50% (RQs), 10% (H) English: 70% (RQs), 30% (H)

Q4

exactly one category. This seems to be more precise than the Scopus or Web of Science schemes (Klavans & Boyack, 2017). The Science-Metrics classification was extended by adding the largest 100 journals in the PMC collection that had not been included in the original Science-Metrics classification scheme. These were classified into a Science-Metrics category by first author based on their similarity to other journals in the Science-Metrics scheme.

Five of the broad fields had too little data to be useful (Economics & Business; Visual & Performing Arts; Communication & Text Studies; General Arts, Humanities & Social Sciences; Built Environment & Design) and were removed. Years before 2000 were not included because of their age and small amount of data. Individual field/year combinations were also removed when there were fewer than 30 articles, since they might give a misleading percentage. Each of the 17 remaining categories contained at least 630 articles (Table 3), with exact numbers for each field and year available in the online supplementary material (columns AE to AW: <https://doi.org/10.6084/m9.figshare.10274012>). For all broad fields, most articles have been published in the last five years (2014–2018), with the exception of Historical Studies, Chemistry, and Enabling & Strategic Technology.

For the third research question, alternative terms for research goals were searched for in the full text of articles. These terms might all be used in different contexts, so a match is not necessarily related to the main goal of the paper (e.g., the term “question” could be part of a discussion of a questionnaire), but the rank order between disciplines may be informative and the results serve as an upper bound for valid uses. The terms searched for were “research questions,” “questions,” “hypotheses,” “aims,” “objectives,” “goals,” and “purposes” in both singular and plural forms. These have been identified above as performing similar functions in research. For this exploration, the term “question” is used in addition to “research question” to capture more general uses.

Table 3. The number of research articles in each broad field

Broad field	2000–2018	2014–2018
Clinical Medicine	446,017	267,370 (60%)
General Science & Technology	322,690	228,006 (71%)
Biomedical Research	245,375	136,466 (56%)
Enabling & Strategic Technology	72,563	35,066 (48%)
Chemistry	56,860	23,829 (42%)
Public Health & Health Services	49,528	28,469 (57%)
Biology	41,819	27,061 (65%)
Psychology & Cognitive Science	26,724	19,693 (74%)
Agriculture, Fisheries & Forestry	18,212	13,522 (74%)
Physics & Astronomy	16,653	11,963 (72%)
ICTs	6,934	4,084 (59%)
Engineering	2,728	1,609 (59%)
Historical Studies	2,339	599 (26%)
Earth & Environmental Science	2,094	1,444 (69%)
Social Sciences	1,776	1,409 (79%)
Philosophy & Theology	1,460	836 (57%)
Mathematics & Statistics	640	469 (73%)
Total	1,314,412	801,895 (61%)

Any of the queried terms could be included in an article out of context. For example, “research question” could be mentioned in a literature review rather than to describe the purpose of the new article. To check the context in which each term was used, a random sample of 100 articles (using a random number generator) matching each term (200 for each concept, counting both singular and plural, for a total of 1,400 checks) was manually examined to ascertain whether any use of the term in the article stated the purpose of the paper directly (e.g., “Our research questions were...”) or indirectly (e.g., “This answered our research questions”), unless mentioned peripherally as information to others (e.g., “The study research questions were explained to interviewees”). There did not seem to be stock phrases that could be used to eliminate a substantial proportion of the irrelevant matches (e.g., “objective function” or “microscope objective”). There also was not a set of standard phrases that collectively could unambiguously identify the vast majority of research questions (e.g., “Our research questions were” or “This article’s research question is”).

Journal guidelines given to authors were manually analyzed to check whether they give advice about research questions and other purpose statements. Three journals with the most articles in each of the 17 academic fields were selected for this (see online supplement doi.org/

10.6084/m9.figshare.10274012). This information is useful background context to help interpret the results.

4. RESULTS

4.1. RQ1 and RQ2: Articles Mentioning Research Questions

Altogether, 23,282 out of 1,314,412 articles explicitly mentioned the phrases “research question” or “research questions” (1.8%), although no field included them in more than a fifth of articles in recent years and there are substantial differences between broad fields (Figure 1). When the terms are used in an article they usually (63%, from the 1,400 manual checks) refer to the article’s main research question(s). Other uses of these terms include referring to questions raised by the findings, and a discussion of other articles’ research questions in literature review sections or as part of the selection criteria of meta-analyses. Thus, overall, only 1.1% of PMC full text research articles mention their research questions explicitly using the singular or plural form. There has been a general trend for the increasing use of these terms, however (Figure 2).

If the terms “question” or “questions” are searched for instead, there are many more matches, although for a minority of articles in most fields (Figures 3 and 4). When these terms are mentioned, they rarely (17%) refer to the hosting article’s research questions (excluding matches with the exact phrases “research question” or “research questions” to avoid overlaps with the previous figure). Common other contexts for these terms include questions in questionnaires and questions raised by the findings. Sometimes the term “question” referred to an idiomatic phrase or issue rather than a query (e.g., “considerable temperature gradients occur within the materials in question” and “these effects may vary for different medications. Future studies are needed to address this important question”). In Philosophy & Theology, the matches could be for discussions of various questions within an article, rather than a research question that is an article’s focus. Similarly for Social Sciences and Public Health & Health

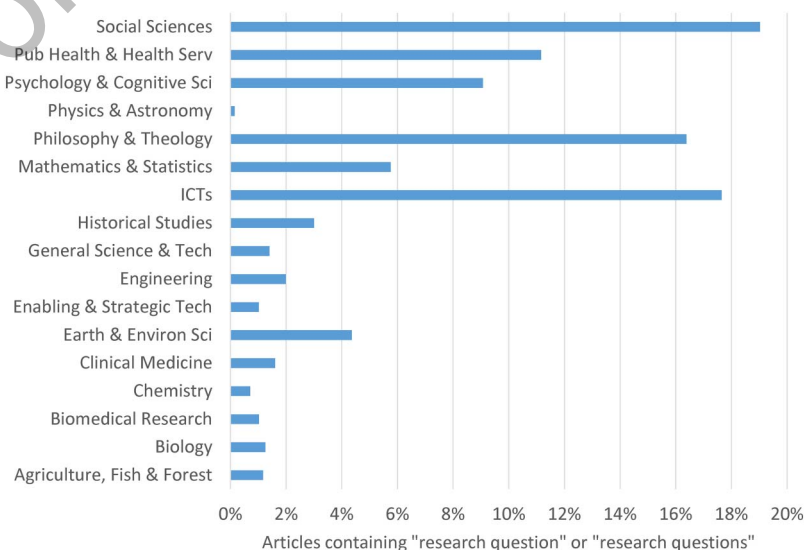


Figure 1. The percentage of full-text research articles containing the phrases “research question” or “research questions” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 63% of occurrences of these terms described the hosting article’s research question(s) ($n = 801,895$ research articles).

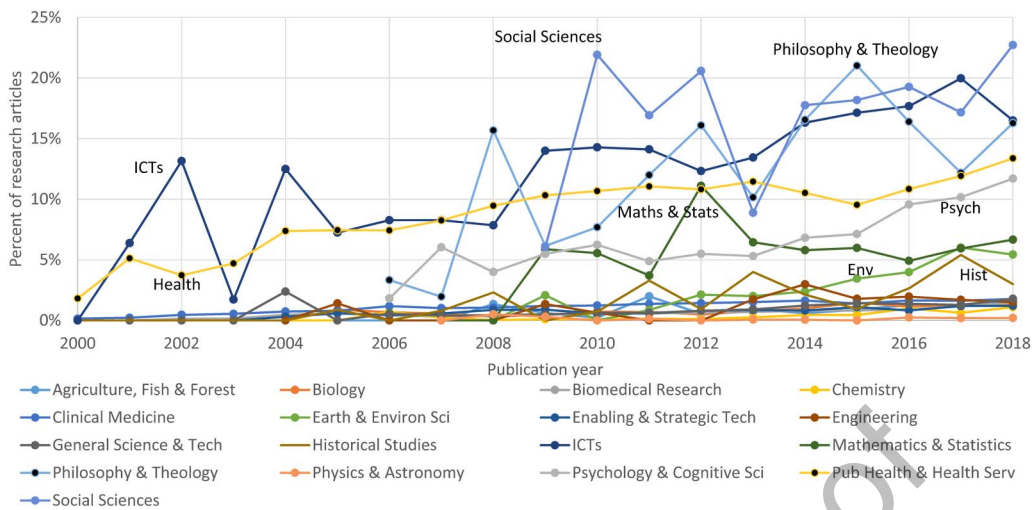


Figure 2. As for Figure 1 but covering 2000–2018 ($n = 1,314,412$ research articles). (All fields can be identified in the Excel versions of the graph within the online supplement 10.6084/m9.figshare.10274012)

Services, the question mentioned might be in questionnaires rather than being a research question. After correcting for the global irrelevant matches, which is a very rough approximation, in all broad fields less than 14% of research articles use these terms to refer to research questions. Nevertheless, this implies that the terms “question” or “questions” are used much more often than the phrases “research question” or “research questions” (1.8%) to refer to an article’s research purposes.

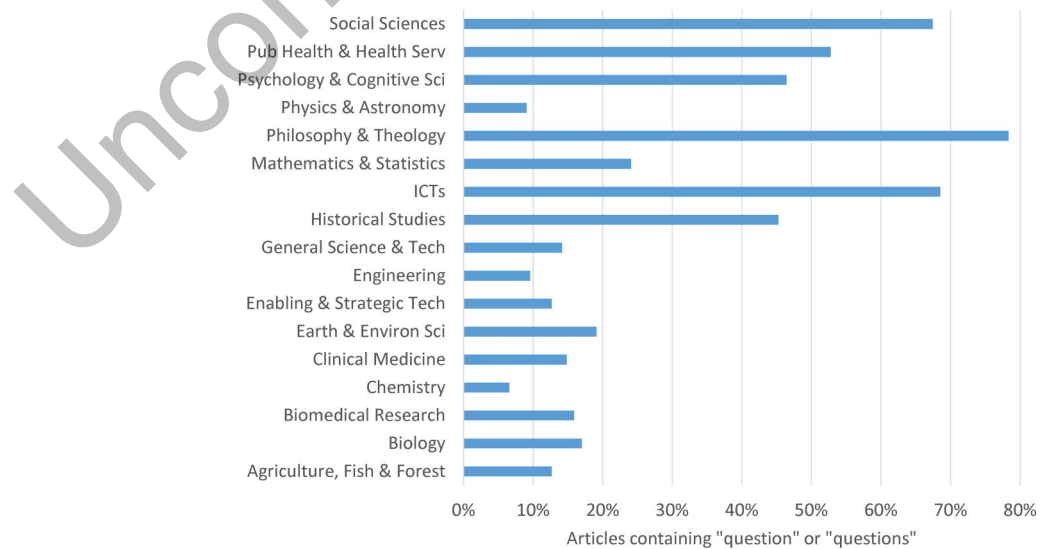


Figure 3. The percentage of full-text research articles containing the terms “question” or “questions” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 17% of occurrences of these terms described the hosting article’s main research question(s) without using the exact phrases “research question” or “research questions,” not overlapping with Figure 1(a) ($n = 801,895$ research articles).

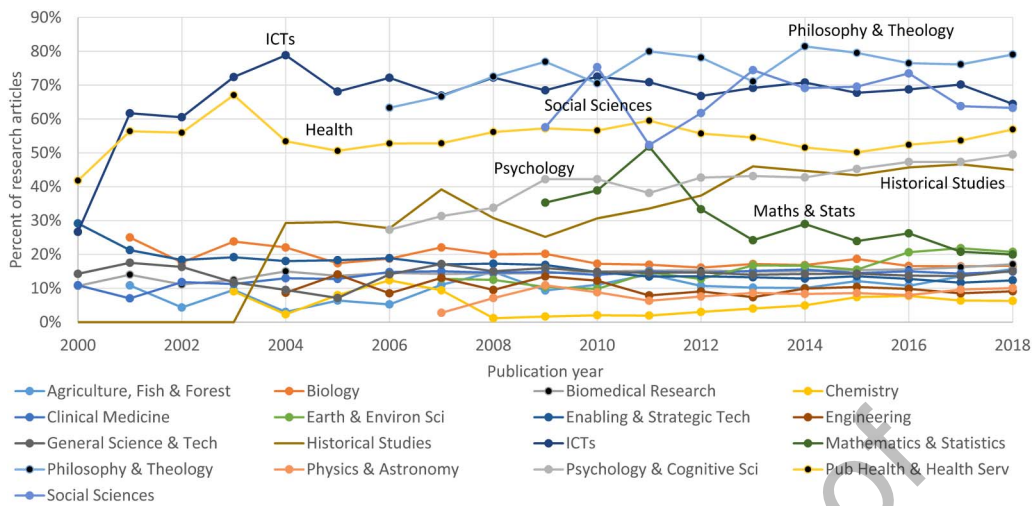


Figure 4. As for Figure 3, but covering 2000–2018 ($n = 1,314,412$ research articles).

4.2. RQ3: Other Article Purpose Terms

The terms “hypothesis” and “hypotheses” are common in Psychology and Cognitive Science as well as in Biology (Figure 5). They are used in a minority of articles in all other fields, but, by 2018 were used in at least 15% of all (or 4% after correcting for irrelevant matches). The terms can be used to discuss statistical results from other papers and in philosophy and mathematics they can be used to frame arguments, so not all matches relate to an article’s main purpose, and only 28% of the random sample checked used the terms to refer to the articles’ main hypothesis or hypotheses.

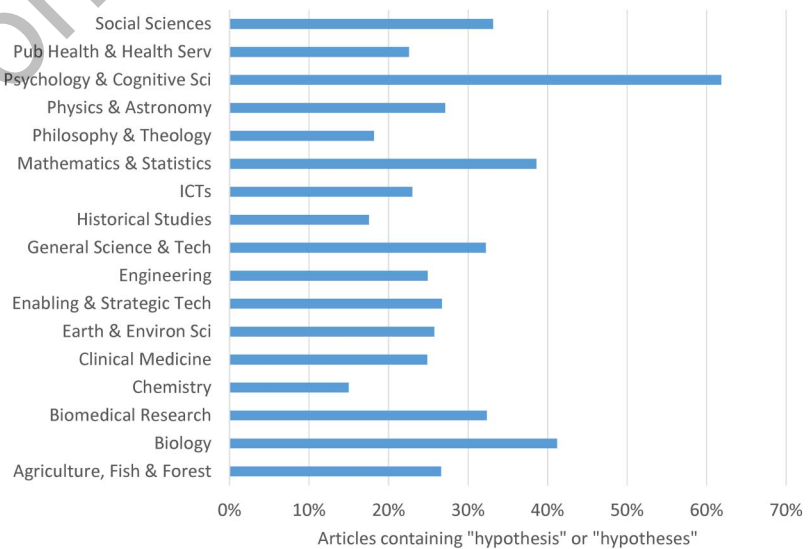


Figure 5. The percentage of full-text research articles containing the terms “hypothesis” or “hypotheses” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 28% of occurrences of these terms described the hosting article’s main hypothesis or hypotheses. A corresponding time series graph showing little change is in the online supplement ($n = 801,895$ research articles).

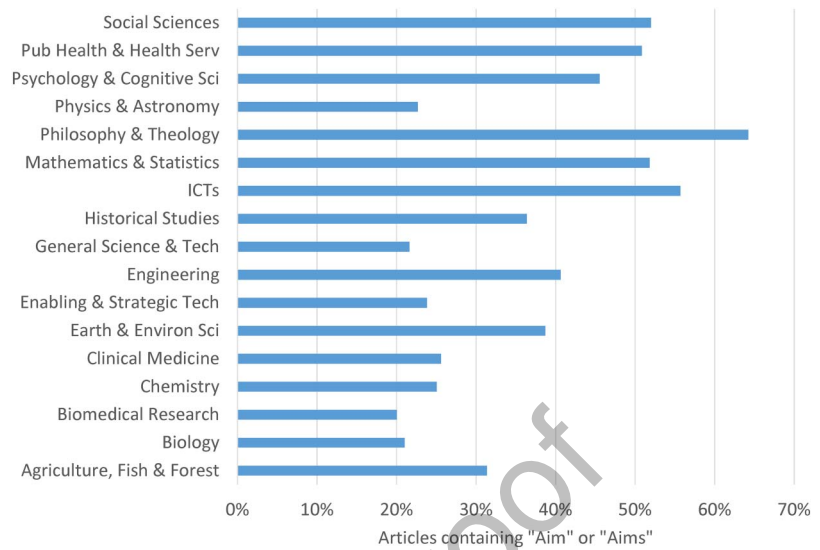


Figure 6. The percentage of full-text research articles containing the terms “aim” or “aims” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 70% of occurrences of these terms described the hosting article’s main aim(s) ($n = 801,895$ research articles).

The use of the terms “aim” and “aims” is increasing overall, possibly in all academic fields (Figures 6 and 7). Fields frequently using the term include Philosophy & Theology, Information & Communication Technologies (ICTs) and Public Health & Health Services, whereas it is used in only about 20% of Chemistry and Biomedical Research papers. Articles using the terms mostly use them (especially the singular “aim”) to describe their main aim (70%), so these are the terms most commonly used to describe the purpose of a PMC full-text article. The terms are also sometimes used to refer to wider project aims or relevant aims outside of the project (e.g., “The EU’s biodiversity protection strategy aims to preserve...”).

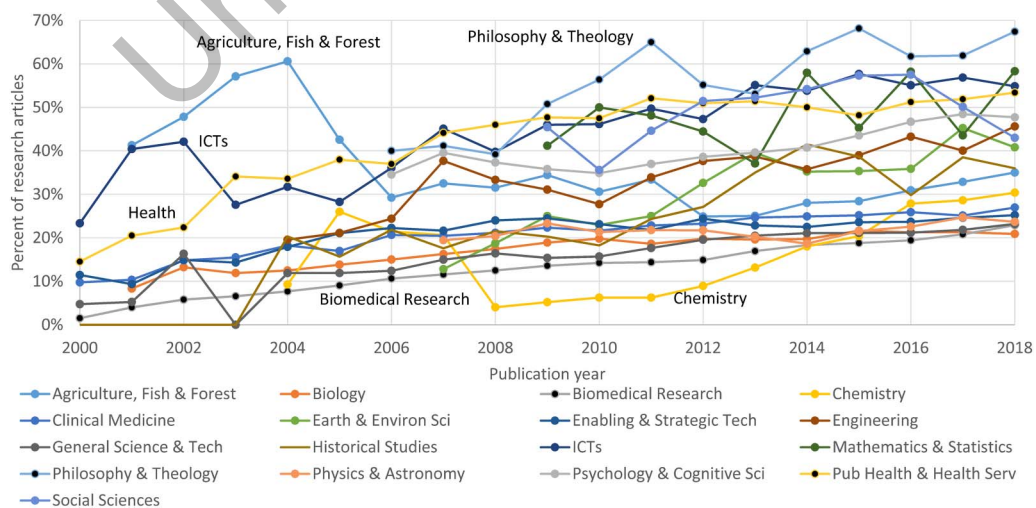


Figure 7. As for Figure 6, but covering 2000–2018 ($n = 1,314,412$ research articles).

The terms “objective” and “objectives” are reasonably common in most academic fields (Figure 8) and are used half of the time (52%) for the hosting article’s objectives. Other common uses include lenses and as an antonym of subjective (e.g., “high-frequency ultrasound allows an objective assessment...”). It is again popular within ICTs, Philosophy & Theology, and Public Health & Health Services, whereas it is used in only about 12% of Physics & Astronomy articles.

The terms “goal” and “goals” follow a similar pattern to “aim” and “objective” (Figure 9), but refer to the hosting paper’s goals in only 28% of cases. Common other uses include methods goals (“the overall goal of this protocol is...”) and field-wide goals (e.g., “over the last decades, attempts to integrate ecological and evolutionary dynamics have been the goal of many studies”).

Some articles may also use the terms “purpose” or “purposes” rather than the arguably more specific terms investigated above, and there are disciplinary differences in the extent to which they are used (Figure 10). These terms may also be employed to explain or justify aspects of an article’s methods. When used, they referred to main purposes in less than a third of articles (29%), and were often instead used to discuss methods details (e.g., “it was decided a priori that physical examination measures would not be collected for the purpose of this audit”), background information (e.g., “species are harvested through fishing or hunting, mainly for alimentary purposes”) or ethics (e.g., “Animal care was carried out in compliance with Korean regulations regarding the protection of animals used for experimental and other scientific purposes.”).

4.3. RQ4: Journal Guidelines

The “instructions for authors” from the 51 large journals analyzed (online supplement 10.6084/m9.figshare.10274012), usually (34 out of 51, or 67%) included some purpose-

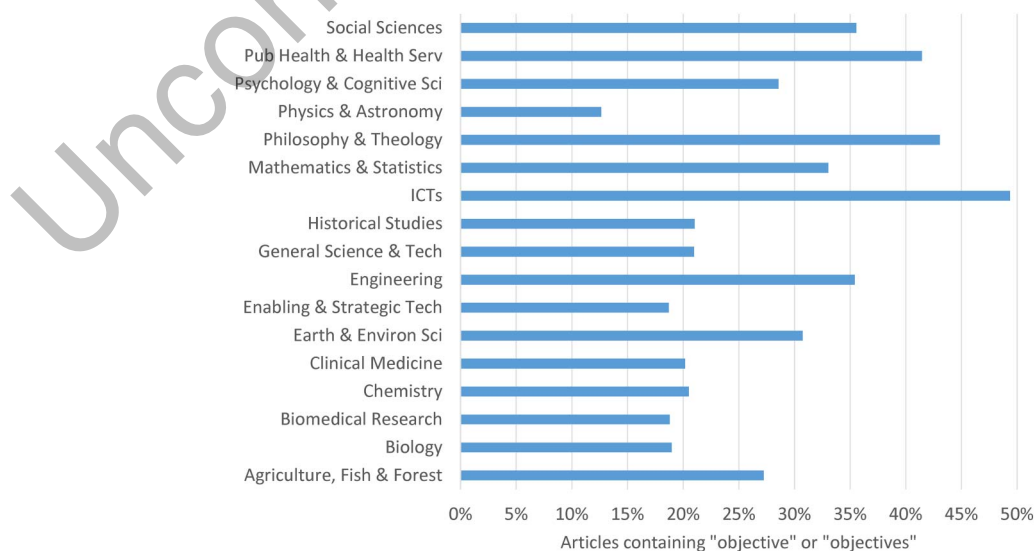


Figure 8. The percentage of full-text research articles containing the terms “objective” or “objectives” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 52% of occurrences of these terms described the hosting article’s objective(s). A corresponding time series graph showing little change is in the online supplement ($n = 801,895$ research articles).

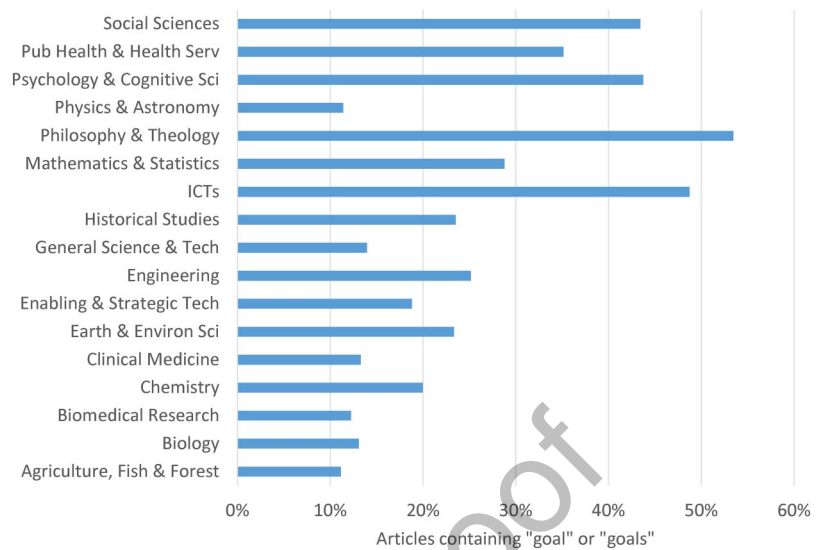


Figure 9. The percentage of full-text research articles containing the terms “goal” or “goals” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 28% of occurrences of these terms described the hosting article’s research question(s). A corresponding time series graph showing little change is in the online supplement ($n = 801,895$ research articles).

related terms (purpose, aim, objective, goal). A fifth (10) of the guidelines referred to research questions and a third (17) mentioned hypotheses. Nevertheless, none of the guidelines seemed to be prescriptive in the sense of insisting that an article contained specifically framed research questions or hypotheses. Instead, the language used was vague enough so that even guidelines

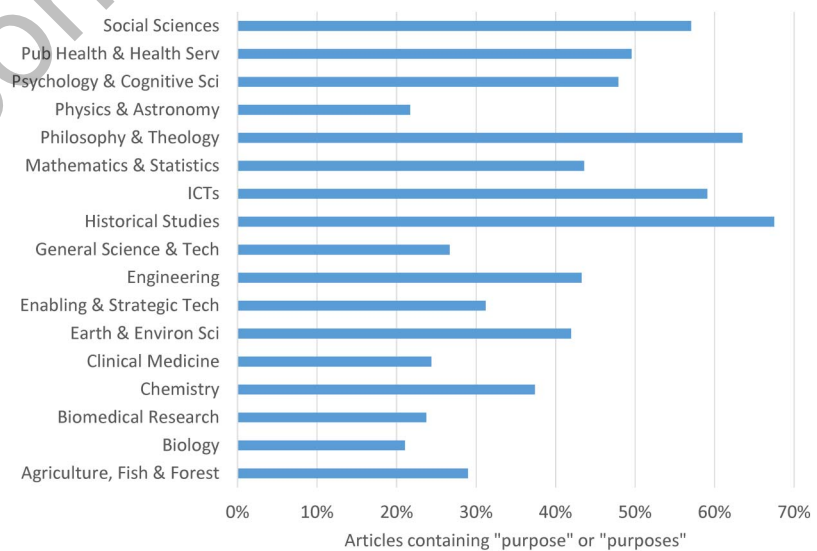


Figure 10. The percentage of full-text research articles containing the terms “purpose” or “purposes” in the body of the text, 2014–2018, for articles in the PMC Open Access collection from 17 out of 22 Science-Metrics broad fields; 29% of occurrences of these terms described the hosting article’s purpose(s). A corresponding time series graph showing little change is in the online supplement ($n = 801,895$ research articles).

mentioning these could be addressed by informal statements of research questions or hypotheses, as the following examples illustrate.

- “The motivation or purpose of your research should appear in the Introduction, where you state the questions you sought to answer” (zookeys.pensoft.net/about)
- “Define the purpose of the work and its significance, including specific hypotheses being tested” (www.mdpi.com/journal/nutrients/instructions)
- “The introduction briefly justifies the research and specifies the hypotheses to be tested” (www.ajas.info/authors/authors.php)
- “A brief outline of the question the study attempts to address” (onlinelibrary.wiley.com/page/journal/20457758/homepage/registeredreports.html)
- “Acquaint the reader with the findings of others in the field and with the problem or question that the investigation addresses.” (www.oncotarget.com)
- “State the research objective of the study, or hypothesis tested” (www.springer.com/biomed/human+physiology/journal/11517)

In the first quote above, for example, “state the questions” could be addressed literally by listing (research) questions or less literally by stating the research objectives. Thus, journal guidelines seem to leave authors the flexibility to choose how to state their research purpose, even if suggesting that research questions or hypotheses are used. This also applies to the influential American Psychological Society guidelines, such as, “In empirical studies, [explaining your approach to solving the problem] usually involves stating your hypotheses or specific question” (APA, 2009, p. 28).

5. DISCUSSION

An important limitation of the methods is that the sample contains a small and biased subset of all open access research articles. For example, the open access publishers BMC, Hindawi, and MDPI have large journals in the data set. The small fields (Table 3) can have unstable lines in the graphs because of a lack of data. Sharp changes between years for the same field are likely due to either small amounts of data or changes in the journals submitted to PubMed in those years, rather than changes in field norms. It is possible that the proportions discovered would be different for other collections. Another limitation is that although articles were searched with the text string “research question,” this may not always have signified research questions in the articles processed (e.g., if mentioned in a literature review or in a phrase like “this research questions whether”). Although the corrections reported address this, they provide a global correction figures rather than field-specific corrections. Conversely, a research question may just be described as a question (e.g., “the query of this research”) or phrased as a question without describing it as such (e.g., “To discover whether PGA implants are immunologically inert...”). Thus, the field-level results are only indicative.

RQ1: Only 23,282 (1.8%, 1.1% after correcting for irrelevant matches) out of 1,314,412 articles assessed in the current paper explicitly mentioned “research question(s),” with significant differences between fields. Although there has been a general trend for the increasing use of explicitly named research questions, they were employed in less than a quarter of articles in all fields. Research questions were mostly used by articles in Social Sciences, Philosophy & Theology, and ICTs, whereas they have been mentioned by under 2% of articles in engineering, physical, life, and medical sciences. Previous studies have shown that 73.3% of English articles in Physical Education (Omidi & Farnia, 2016), 33% of Applied Linguistics articles (Sheldon, 2011) and 32% of Computer Science articles (Shehzad, 2011) included research

questions or hypotheses. Studies focused on doctoral dissertations show that 97% of US Applied Linguistics (Lim, 2014), 90% of English Language Teaching (Geçikli, 2013), 70% of Education Management (Cheung, 2012), and 50% of computing doctoral dissertations (Soler-Monreal, Carbonell-Olivares, & Gil-Salom, 2011) listed research questions, a large difference.

The results also show that about 13% of Public Health and Health Services articles and 12% of Psychology and Cognitive Science articles use the term “research questions”. However, a study focused on Educational Psychology found that 35% of English-language papers listed research questions and 75% listed hypotheses (Loi & Evans, 2010). Thus, the current results reveal a substantially lower overall prevalence than suggested by previous research.

RQ2: There has been a substantial increase in the use of the term “research questions” in some subjects, including ICTs, Social Sciences, and Public Health and Health Services (Figure 2), as well as a general trend for increasing use of this term, but with most fields still rarely using it. This suggests that some disciplines are standardizing their terminology, either through author guidelines in journals (RQ4), formal training aided by frameworks such as Swales’ CARS model, or informal training or imitation. For example, the analysis of the “instructions for authors” given by 51 journals (online supplement doi.org/10.6084/m9.figshare.10274012) showed that the three biology journals, the three psychology journals, and two biomedical journals included in the analysis referred to both research questions and hypotheses in their author guidelines.

RQ3: Terminology for the purpose of an article seems to be quite widely used, including aims, objectives, and goals (Figures 5–9). This is in line with a study examining the lexical bundles identified in research article introductions from several disciplines, which reported the terms “aim,” “objective,” and “purpose” as the main terms used to announce the research descriptively and/or purposefully, although no phrase related to research questions or hypotheses was identified (Cortés, 2013), and with another study reporting similar terminology in medical articles (Jalali & Moini, 2014). Related to this (RQ4), the analysis of the “instructions for authors” given by 51 journals (online supplement [10.6084/m9.figshare.10274012](https://doi.org/10.6084/m9.figshare.10274012)) showed that “purpose” is the term mostly mentioned in the Abstract guidelines and “aims” is the term mainly used in the body of the text (Introduction or Background) guidelines. The term “objective” also appears in some article body guidelines, whereas the term “goal” is not mentioned in them. After correcting for irrelevant matches (e.g., articles using the term “hypothesis” but not for their main research hypotheses) using the percentages reported with the figures above, no terminology was found in a majority of articles in any field. Thus, at least from the perspective of PMC Open Access publications, there is no standardization of research terminology in any broad field.

There are substantial disciplinary differences in the terminology used. Whereas the term “research question” is relevant in Social Sciences, Philosophy & Theology, and ICTs, the term “hypothesis” is important in Psychology and Cognitive Science, used in over 60% of articles. This is in line with a study focused on Educational Psychology, which found that the 75% out of 20 English papers introduced the hypotheses, whereas 35% of them introduced the research questions (Loi & Evans, 2010). The three psychology journals with the highest frequency in the data set used for this study referred to hypotheses in their author guidelines (see online supplement [10.6084/m9.figshare.10274012](https://doi.org/10.6084/m9.figshare.10274012)).

The terms “aim,” “objective,” and “goal” are mainly used in Philosophy, Theology, ICTs, and Health. The term “aim” is also quite often used in health, mathematics, and psychological articles, whereas the term “objective” is also used in engineering and mathematics articles. The term “goal” is also used in psychology and biomedical articles. Although most articles

in all fields include a term that could be used to specify the purpose of an article (question or questions, hypothesis, aim, objective, goal), they are relatively scarce in Chemistry and Physics & Astronomy. The use of purpose-related terms has also increased over time in most academic fields. This agrees with a study about Computer Science research articles that found an increasing use of outlining purpose or stating the nature of the research (Shehzad, 2011).

An example article from Chemistry illustrates how a research purpose can be implicit. The paper, "Fluid catalytic cracking in a rotating fluidized bed in a static geometry: a CFD analysis accounting for the distribution of the catalyst coke content" has a purpose that is clear from its title but which is not described explicitly in the text. Its abstract starts by describing what the paper offers, but not why, "Computational Fluid Dynamics is used to evaluate the use of a rotating fluidized bed in a static geometry for the catalytic cracking of gas oil." The first sentence of the last paragraph of the introduction performs a similar role, "The current paper presents CFD simulations of FCC in a RFB-SG using a model that accounts for a possible non-uniform temperature and catalyst coke content distribution in the reactor." Both sentences could easily be rephrased to start with, "The purpose of this paper is to," but it is apparently a stylistic feature of chemical research not to do this. Presumably purposes are clear enough in typical chemistry research that they do not need to be flagged linguistically, but this is untrue for much social science and health research, for example, partly due to non-standard goals (i.e., task uncertainty: Whitley, 2000).

5.1. Possible Origins of the Differences Found

Although the results of this article provide evidence of field differences in the use of terminology, the methods used do not address their root causes. This is interesting from a sociology of science perspective. The following are possible causes that future research may investigate.

- Broad epistemological: Fields work with knowledge in different ways and naturally use different terminology as a result. Arts and humanities research may have the goal to critique or analyze, or may be practice-based research rather than having a more specific knowledge purpose. For this, research questions would be inappropriate. Thus, terminology variation may partly reflect the extent to which a broad field typically attempts to create knowledge.
- Narrow epistemological: Narrow fields that address similar problems may feel that they do not need to use research problem terminology to describe their work because the purpose of a paper is usually transparent from the description of the methods or outcome. For example, it would be unnecessary to formulate, "This paper investigates whether treatment x reduces death rates from disease y" as a named research question or even explain that it is the goal of a paper. This may also be relevant for fields that write short papers. It may be most relevant for papers that use statistical methods and have high standards of evidence requirement (e.g., medicine) and clearly defined problems. In contrast, many social sciences research projects are not intrinsically clearly demarcated and need an explanation to define the problem (as for the current article). Thus, describing what the problem is can be an important and non-trivial part of the research. This relates to "task uncertainty!", which varies substantially between fields (Whitley, 2000) and affects scholarly communication (Fry, 2006).
- Field or audience homogeneity: Fields with homogeneous levels and types of expertise may avoid terminology that field members would be able to deduce from the context. For example, a mixed audience paper might need to specify statistical hypotheses,

whereas a narrow audience paper might only need to specify the result, because the audience would understand the implicit null and alternative hypotheses.

- Field cultures for term choice: Academic publishing relies to some extent on imitation and reaching a consensus about the ways in which research is presented (e.g., Becher & Trowler, 2001). It might therefore become a field norm to use one term in preference to a range of synonyms, such as “aims” instead of “objectives.”
- Field cultures for term meaning: Following from the above, a field culture may evolve an informal convention that two synonyms have different specific uses. For example, “aims” could be used for wider goals and “objectives” for the narrower goals of a paper.
- Guidelines: Fields or their core journals may adopt guidelines that specify terminology, presumably because they believe that this standardization will improve overall communication clarity.

6. CONCLUSIONS

The results suggest that the explicit use of research questions, in the sense that they are named as such, is almost completely absent in some research fields, and they are at best a substantial minority (under 20%) in most others (ignoring the fields that did not meet the inclusion threshold). Although the word search approach does not give conclusive findings, the results suggest that alternative terminologies for describing the purpose of a paper are more widespread in some fields, but no single terminology is used to describe research purposes in a majority of articles in any of the broad fields examined.

The lack of standardization for purpose terminology in most or all fields may cause problems for reviewers and readers expecting to see explicit statements. It is not clear whether guidelines to standardize terminology for journals or fields would be practical or helpful, however, but this should be explored in the future. Presumably any guidelines should allow exceptions for articles that make non-standard contributions, although there are already successful journals with prescriptive guidelines, and the advantage of standardization through structured abstracts seems to be accepted (Hartley, 2004).

The disciplinary differences found may cause problems for referees, authors, editors, and readers of interdisciplinary research or research from outside of their natural field if they fail to find an article’s purpose expressed in the terminology that they expect. This issue could not reasonably be resolved by standardizing across science because of the differing nature of research. Instead, evidence in the current article of the existence of valid disciplinary differences in style may help reviewers and editors of large interdisciplinary journals to accept stylistic differences in research problem formulations.

AUTHOR CONTRIBUTIONS

Mike Thelwall: Conceptualization, Investigation, Software, Writing—original draft. Amalia Mas-Bleda: Investigation, Writing—original draft.

COMPETING INTERESTS

The authors have no competing interests to declare.

FUNDING INFORMATION

This research received no funding.

DATA AVAILABILITY

The data behind the results is available at FigShare (<https://doi.org/10.6084/m9.figshare.10274012>).

REFERENCES

- Agee, J. (2009). Developing qualitative research questions: A reflective process. *International Journal of Qualitative Studies in Education*, 22(4), 431–447.
- Ahamad, M. I., & Yusof, A. M. (2012). A genre analysis of Islamic academic research article introductions. *Procedia–Social and Behavioral Sciences*, 66, 157–168.
- Amunai, A., & Wannaruk, A. (2013). Investigating move structure of English Applied Linguistics research: Article discussions published in international and Thai journals. *English Language Teaching*, 6(2), 1–13.
- Anthony, L. (1999). Writing research article introductions in software engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication*, 42(1), 38–46.
- APA. (2009). *Publication Manual of the American Psychological Association*, 6th Ed. Washington, DC: American Psychological Association.
- Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics* (pp. 66–77). Durban, South Africa.
- Becher, T., & Trowler, P. R. (2001). *Academic Tribes and Territories*. London, UK: McGraw-Hill.
- Brett, P. (1994). A genre analysis of the results section of sociology articles. *English for Specific Purposes*, 13(1), 47–59.
- Cheung, Y. L. (2012). Understanding the writing of thesis introductions: An exploratory study. *Procedia–Social and Behavioral Sciences*, 46, 744–749.
- Cortés, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, 12(1), 33–43.
- Del Saz Rubio, M. M. (2011). A pragmatic approach to the macro-structure and metadiscourse features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes*, 30(4), 258–271.
- Fry, J. (2006). Scholarly research and information practices: a domain analytic approach. *Information Processing & Management*, 42(1), 299–316.
- Gil-Salom, L., Soler-Monreal, C., & Carbonell-Olivares, M. (2008). The move-step structure of the introductory sections on Spanish PhD theses. *RESLA*, 21, 85–106.
- Hartley, J. (2004). Current findings from research on structured abstracts. *Journal of the Medical Library Association*, 92(3), 368–371.
- Hirano, E. (2009). Research article introductions in English for specific purposes: A comparison between Brazilian Portuguese and English. *English for Specific Purposes*, 28(4), 240–250.
- Jalali, Z. S., & Moini, M. R. (2014). Structure of lexical bundles in introduction section of medical research articles. *Procedia – Social and Behavioral Sciences*, 98, 719–726.
- Jalilifar, A. R. (2010). Research article introductions: Sub-disciplinary variation in applied linguistics. *Journal of Teaching Language Skills*, 2(2), 29–55.
- Joseph, R., Lim, J. M.-H., & Nor, N.A.M. (2014). Communicative moves in forestry research introductions: Implications for the design of learning materials. *Procedia–Social and Behavioral Sciences*, 134, 53–69.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3), 269–292.
- Kanoksilapatham, B. (2012). Research article structure of research article introductions in three engineering subdisciplines. *IEEE Transactions on Professional Communication*, 55(4), 294–309.
- Klavans, R., & Boyack, K. W. (2017). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, 68(4), 984–998.
- Kwan, B. S. C. (2017). A cross-paradigm macro-structure analysis of research articles in Information Systems. *English for Specific Purposes*, 45, 14–30.
- Lim, J. M. H. (2014). Formulating research questions in experimental doctoral dissertations on Applied Linguistics. *English for Specific Purposes*, 35, 66–88.
- Lim, J. M.-H., Loi, C.-K., & Hashim, A. (2014). Postulating hypotheses in experimental doctoral dissertations on Applied Linguistics: A qualitative investigation into rhetorical shifts and linguistic mechanisms. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos*, 27, 121–141.
- Loi, C. K., & Evans, M. S. (2010). Cultural differences in the organization of research article introductions from the field of educational psychology: English and Chinese. *Journal of Pragmatics*, 42(10), 2814–2825.
- Mahzari, A., & Maftoon, P. (2007). A contrastive study of the introduction section of English and Persian medical research articles. *Iranian Journal of Language Studies*, 1(3), 201–214.
- Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, 2, 1–11. <https://doi.org/10.1016/j.amper.2014.12.002>
- McIntyre, E., Nisbet, S., Magarey, A., Keane, M., & Roeger, L. (2007). Publishing in peer review journals. *Australian Family Physician*, 36(7), 561–562.
- Mur Dueñas, P. (2010). A contrastive analysis of research article introductions in English and Spanish. *Revista Canaria de Estudios Ingleses*, 61, 119–133.
- Omid, L., & Farnia, M. (2016). Comparative generic analysis of introductions of English and Persian physical education research articles. *International Journal of Language and Applied Linguistics*, 2(2), 1–18.
- Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures 1. *The Qualitative Report*, 11(3), 474–498.
- Ozturk, I. (2007). The textual organisation of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, 26(1), 25–38.
- Perneger, T. V., & Hudelson, P. M. (2004). Writing a research article: advice to beginners. *International Journal for Quality in Health Care*, 16(3), 191–192.

- Rahimi, S., & Farnia, M. (2017). Comparative generic analysis of introductions of English and Persian dentistry research articles. *Research in English Language Pedagogy*, 5(1), 27–40.
- Samraj, B. (2002). Introductions in research articles: variations across disciplines. *English for Specific Purposes*, 21(1), 1–17.
- Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: Gap-spotting or problematization? *Organization*, 18(1), 23–44.
- Shehzad, W. (2011). Outlining purposes, stating the nature of the present research, and listing research questions or hypotheses in academic papers. *Journal of Technical Writing and Communication*, 41(2), 139–160.
- Sheldon, E. (2011). Rhetorical differences in RA introductions written by English L1 and L2 and Castilian Spanish L1 writers. *Journal of English for Academic Purposes*, 10, 238–251.
- Soler-Monreal, C., Carbonell-Olivares, M. S., & Gil-Salom, M. L. (2011). A contrastive study of the rhetorical organisation of English and Spanish PhD thesis introductions. *English for Specific Purposes*, 30(1), 4–17.
- Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge, UK: Cambridge University Press.
- Swales, J. M. (2004). *Research Genres. Exploration and Applications*. Cambridge, UK: Cambridge University Press.
- Tessuto, G. (2015). Generic structure and rhetorical moves in English-language empirical law research articles: Sites of interdisciplinary and interdiscursive cross-over. *English for Specific Purposes*, 37, 13–26.
- Teufel, S. (1999). Argumentative zoning: Information extraction from scientific text. Doctoral dissertation, University of Edinburgh.
- Thelwall, M. (2019). The rhetorical structure of science? A multidisciplinary analysis of article headings. *Journal of Informetrics*, 13(3), 555–563.
- Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., & STROBE Initiative. (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery*, 12(12), 1500–1524.
- Whitley, R. (2000). *The Intellectual and Social Organization of the Sciences*. Oxford, UK: Oxford University Press.
- Williams, I. A. (1999). Results sections of medical research articles: analysis of rhetorical categories for pedagogical purposes. *English for Specific Purposes*, 18(4), 347–366.
- Yang, R., & Allison, D. (2004). Research articles in applied linguistics: Structures from a functional perspective. *English for Specific Purposes*, 23(3), 264–279.