

Who does big team science?

Erin M. Buchanan¹ & Savannah C. Lewis²

¹ Harrisburg University of Science and Technology

² University of Alabama

Author Note

Erin M. Buchanan is a Professor of Cognitive Analytics at Harrisburg University of Science and Technology. Savannah C. Lewis is a graduate student at the University of Alabama.

Thank you to Dwayne Lieck for providing an extensive list of large scale projects for this manuscript.

The authors made the following contributions. Erin M. Buchanan: Conceptualization, Data curation, Formal Analysis, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing; Savannah C. Lewis: Conceptualization, Data curation, Methodology, Project administration, Writing – original draft, Writing – review & editing.

Correspondence concerning this article should be addressed to Erin M. Buchanan, 326 Market St., Harrisburg, PA 17101. E-mail: ebuchanan@harrisburgu.edu

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline.

Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines.

One sentence clearly stating the **general problem** being addressed by this particular study.

One sentence summarizing the main result (with the words “**here we show**” or their equivalent).

Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge.

One or two sentences to put the results into a more **general context**.

Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: big team, science, authorship, credit

Word count: X

Who does big team science?

<https://www.science.org/content/page/science-information-authors>

Libraries

Import Bib Data

In this section, we are importing the bibtex file that includes all the papers we are going to use.

Notes:

- We used the preprint manuscript for some papers because they had the entire author list, we will update their references manually for the final dataset.
- We will have to manually add the names for several papers that used consortium authors.
- The pre-reg of this document only examines a few authors/papers to ensure code can be run and develop workflow.

In this section, we will add the information for authors when they used a consortium authorship.

Import Author Data

Hand Coded Variables

Journal Information

Create Summary Statistics

Merge Information Back

Person Analysis

Career Length.

Employment Levels.**Education Levels.****Author Bean Counting.****More Authors Over Time?****Where are Authors?****CRedit and WEIRD.**

- credit contributions ? are we using de-weirding by using those people for translation and data collection only
 - grab major roles and see what the diversity is
 - look at the first 10 people and the last author
 - maybe compare to the other 10 groups (1/3rd, middle, last 1/3rd)

Journal Analysis**Which Journals.**

- make sure preprints we are using also have real journal merged in

Article Topics Global Type.

- hand coded information on what the articles are about
- Types of big team research (social, cognitive) using keywords

Article Types Over Time.

- Over time: are things becoming more varied
- are the publications more varied over time

The introduction will go here. Here's an outline:

- Big Team Science
 - one off papers

– collaborative teams

- Credibility revolution
- WEIRD
- ... more tbd, brain isn't braining

Method

Studies

We defined **big team science publications** as publications with at least 10 authors that were published in peer-reviewed journals or had posted a full paper pre-print for publication review. We specifically focused on social science research, primarily *psychology* for this manuscript. First, we added all known publications from collaborative teams, such as the PSA, Many Labs, and Many Babies. We examined journals that frequently publish registered replication reports (i.e., *Advances in Methods and Practices in Psychological Science*) for additional publications with at least 10 authors. From these manuscripts, we identified common authors who frequently participate in these studies, and examined their Google Scholar or Open Researcher and Contributor Identifier (ORCID) page for other publications. We reached out to social networks on Twitter to identify other publications. Last, we used Google Scholar and EBSCO to search for large projects using the following search terms: collaboration, multicultural, large scale, and big team science.

Using these criteria, we identified 70 articles for inclusion on this manuscript. The publication dates on these articles ranged from 2013 to 2022, and we used the pre-print last updated date as the publication date for those articles. Articles were most commonly published in *Advances in Methods and Practices in Psychological Science* ($n = 20$), *Behavior Research Methods* ($n = 1$), *Collabra: Psychology* ($n = 1$), *Computers in Human Behavior* ($n = 1$), and *International Journal of Methods in Psychiatric Research* ($n = 1$). A complete list of journals can be found on our Open Science Framework page XXX.

Data Curation

Article Information. For each publication, we coded the list of keywords into broad labels for areas of social science (i.e., Social Psychology, Cognitive Psychology). The current impact factor (i.e., 2022) for each journal was found on the journal page and included for journal statistics.

Author Information. The author list was then extracted from each publication. In the case of consortium authorship, we extracted the complete authorship from the meta-data or pre-print publication. The total number of unique authors was 3336. The number of authors on each publication ranged from 1 to 482 with an average of 68.63 authors ($SD = 89.89$).

Next, we matched each author to their Google Scholar and ORCID profile pages, if available. We originally used the *R* packages, *rorcid* [CITE] and *scholar* [CITE] to try to match published author names to profile pages. This process did not result in a large number of matches, and we therefore curated the list of profile pages manually, checking each author against the publication list. We used these two packages and profile pages to collect authorship statistics described below.

Career Length. Career length for each author was defined using multiple variables to see if results from the two data sources would converge on similar answers. Both ORCID and Google Scholar provide a list of publications for authors, and we first calculated career length as the year of first publication listed for each author. In ORCID, a researcher can enter their educational background with completion years for each degree. We defined career length for this variable as years since first degree listed. Publication years are often curated directly from meta-data provided by Crossref (ORCID) or online sources used by Google Scholar. Authors may also directly add publications and their information into both systems. The limitation to using education as a metric for career length is that the researcher must directly enter this information into ORCID.

Employment. Employment information was collected from self-entered ORCID data. These values are open text, and therefore, we coded them into coherent categories for traditional educational (graduate student, post doctoral, lecturer), tenure track (assistant, associate, full professor), and other roles (fellow, research assistant, researcher, head). Employment geopolitical region was also selected when available.

Education. As with employment information, we also collected education information from self-entered ORCID data. These values were coded into general categories of bachelor, master, and doctoral degrees. The geopolitical region of the listed education was included when available. For analyses, both employment and education levels were grouped into United Nation regions.

Types of Publications. ORCID includes information about the type of publication pulled from either researcher entered data or Crossref. We coded these publications into general categories including book, conference presentations, data-sets, journal articles, preprints, software, thesis, and other publications.

Publication Metrics. We calculated total number of publications of any type from both Google Scholar and ORCID. We additionally pulled both the h-index and i-10 index from Google Scholar. The h-index represents the highest h number of publications that have at least h citations, while the i-10 index represents the number of publications with at least 10 citations.

Data analysis

what are we gonna do to answer the research questions

Results

put stuff in here about the format/structure to answering the research questions

Discussion

References