1   Who does big team science?

2   Erin M. Buchanan[1] & Savannah C. Lewis[2]

3   [1] Harrisburg University of Science and Technology

4   [2] University of Alabama

5   Author Note

⁶   Erin M. Buchanan is a Professor of Cognitive Analytics at Harrisburg University of

⁷ Science and Technology. Savannah C. Lewis is a graduate student at the University of

⁸ Alabama.

¹⁶   Correspondence concerning this article should be addressed to Erin M. Buchanan,

¹⁷ 326 Market St., Harrisburg, PA 17101. E-mail: ebuchanan@harrisburgu.edu

¹⁸   Abstract

¹⁹ One or two sentences providing a **basic introduction** to the field, comprehensible to a

²⁰ scientist in any discipline.

²¹   Two to three sentences of **more detailed background**, comprehensible to scientists

²² in related disciplines.

²³   One sentence clearly stating the **general problem** being addressed by this particular

²⁴ study.

²⁵   One sentence summarizing the main result (with the words "**here we show**" or their

²⁶ equivalent).

²⁷   Two or three sentences explaining what the **main result** reveals in direct comparison

²⁸ to what was thought to be the case previously, or how the main result adds to previous

²⁹ knowledge.

³⁰   One or two sentences to put the results into a more **general context**.

³¹   Two or three sentences to provide a **broader perspective**, readily comprehensible to

³² a scientist in any discipline.

³³   *Keywords:* big team, science, authorship, credit

³⁴   Word count: X

35   Who does big team science?

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

37   • Big Team Science

38   – one off papers

39   – collaborative teams

40   • Credibility revolution

41   • WEIRD

42   • . . . more tbd, brain isn't braining

43   • Research Question 1: What journals publish big team science papers?

44   • Research Question 2: What are the types of articles that are being published in big

45   team science?

46   • Research Question 3: Who is involved in big team science?

47   For each of these research questions, we will examine the overall results of all big

48 team research projects, and examine for change in result trends across years of publication.

49   **Method**

50 **Studies**

51   We defined **big team science publications** as publications with at least 10 authors

52 that were published in peer-reviewed journals or had posted a full paper pre-print for

53 publication review. We specifically focused on social science research, primarily *psychology*

54 for this manuscript. First, we added all known publications from collaborative teams, such

55 as the PSA, Many Labs, and Many Babies. We examined journals that frequently publish

56 registered replication reports (i.e., *Advances in Methods and Practices in Psychological*

57 *Science*) for additional publications with at least 10 authors. From these manuscripts, we

58 identified common authors who frequently participate in these studies, and examined their

59 Google Scholar or Open Researcher and Contributor IDentifier (ORCID) page for other

60 publications. We reached out to social networks on Twitter to identify other publications.

61 Last, we used Google Scholar and EBSCO to search for large projects using the following

62 search terms: collaboration, multicultural, large scale, and big team science.

Data Curation

64 **RQ1: Journal Information.**   Using these criteria, we identified 70 articles for

65 inclusion on this manuscript. The publication dates on these articles ranged from 2013 to

66 2022, and we used the pre-print last updated date as the publication date for those articles.

67 The current impact factor (i.e., 2022) for each journal was found on the journal page and

68 included for journal statistics.

69 **RQ2: Article Information.**   For each publication, we coded the list of keywords

70 into broad labels for areas of social science (i.e., Social Psychology, Cognitive Psychology).

71 In this section, we elected to code each article with only one main research area. Research

72 is often cross-disciplinary, however, for simplicity, we applied one global label to each

73 article that represented the perceived main area of study (based on what course this

74 material might be taught in, the publication journal focus). The list of broad areas

75 includes clinical, cognitive, developmental, educational, and social psychology. Last, we

76 included a metascience category that covered articles that were detailing the science of

77 science. The Open Science Collaboration [CITE] was included in this category because the

78 aim of the paper was a focus on replication across multiple field types. Other metascience

79 categories included papers that focused on research degrees of freedom, analysis choices,

80 and types of samples.

⁸¹ **RQ3: Author Information.** The author list was then extracted from each

⁸² publication. In the case of consortium authorship, we extracted the complete authorship

⁸³ from the meta-data or pre-print publication. The total number of unique authors was 3336.

⁸⁴ The number of authors on each publication ranged from 1 to 482 with an average of 68.63

⁸⁵ authors ($SD = 89.89$).

⁸⁶ Next, we matched each author to their Google Scholar and ORCID profile pages, if

⁸⁷ available. We originally used the R packages, *rorcid* [CITE] and *scholar* [CITE] to try to

⁸⁸ match published author names to profile pages. This process did not result in a large

⁸⁹ number of matches, and we therefore curated the list of profile pages manually, checking

⁹⁰ each author against the publication list. We used these two packages and profile pages to

⁹¹ collect authorship statistics described below.

⁹² ***Career Length***. Career length for each author was defined using multiple variables

⁹³ to see if results from the two data sources would converge on similar answers. Both ORCID

⁹⁴ and Google Scholar provide a list of publications for authors, and we first calculated career

⁹⁵ length as the year of first publication listed for each author. In ORCID, a researcher can

⁹⁶ enter their educational background with completion years for each degree. We defined

⁹⁷ career length for this variable as years since first degree listed. Publication years are often

⁹⁸ curated directly from meta-data provided by Crossref (ORCID) or online sources used by

⁹⁹ Google Scholar. Authors may also directly add publications and their information into

¹⁰⁰ both systems. The limitation to using education as a metric for career length is that the

¹⁰¹ researcher must directly enter this information into ORCID.

¹⁰² ***Employment***. Employment information was collected from self-entered ORCID

¹⁰³ data. These values are open text, and therefore, we coded them into coherent categories for

¹⁰⁴ traditional educational (graduate student, post doctoral, lecturer), tenure track (assistant,

¹⁰⁵ associate, full professor), and other roles (fellow, research assistant, researcher, head).

¹⁰⁶ Employment geopolitical region was also selected when available.

107 ***Education***. As with employment information, we also collected education

108 information from self-entered ORCID data. These values were coded into general

109 categories of bachelor, master, and doctoral degrees. The geopolitical region of the listed

110 education was included when available. For analyses, both employment and education

111 levels were grouped into United Nation regions.

112 ***Types of Publications***. ORCID includes information about the type of publication

113 pulled from either researcher entered data or Crossref. We coded these publications into

114 general categories including book, conference presentations, data-sets, journal articles,

115 preprints, software, thesis, and other publications.

116 ***Publication Metrics***. We calculated total number of publications of any type from

117 both Google Scholar and ORCID. We additionally pulled both the h-index and i-10 index

118 from Google Scholar. The h-index represents the highest h number of publications that

119 have at least h citations, while the i-10 index represents the number of publications with at

120 least 10 citations.

121 **Data analysis**

122 **RQ1: Journals.** Results for types of journals will include a summary of the

123 journals that publish big-team science papers and an average of the 2-year and 5-year

124 impact factors for the journals.

125 **RQ2: Articles.** Results for types of articles will include a summary of the coded

126 areas for each article, presented overall and across the years of publication.

127 **RQ3: Authors.**

128 We will use $\alpha < .05$ for all analyses that involve hypothesis testing. We make no

129 directional predictions.

130 ***Career Length***. For each of the three variables in career length (ORCID year of

131 first pub, Scholar year of first pub, ORCID year of first degree), we will create a

132 visualization of the trend and variance of researcher career length across publication years.

133 To analyze trends over time, we will use the slope from a multilevel model (MLM) using

134 the individual as a random intercept, career length as the dependent variable, and year of

135 publication as the predictor variable. MLMs are regression models that control for the

136 correlated error due to the repeated and nested nature of the data [CITE]. In this model,

137 each individual's starting career length is arbitrary, which we allow to vary with the

138 random intercept by participant. A positive slope for year of publication would indicate

139 increasing years of first publication (i.e., more younger scholars over time), while a negative

140 slope would indicate older years of first publication (i.e., more older scholars over time). In

141 order to show variance between individuals, we will report the variance component of the

142 random intercept for individuals. Finally, to examine variance over time, we will calculate

143 the standard deviation of the individual within each publication (i.e., one standard

144 deviation for each article), and use a linear model with year predicting these standard

145 deviations. This model is a traditional simultaneous regression, as averaging the variance

146 by title eliminates the repeated measures variable (individuals). A positive slope would

147 indicate increasing variance over time (i.e., more diversity in the career lengths of scholars),

148 while a negative slope would indicate less variance and diversity in scholars over time.

149 ***Employment***. Employment information was collected from self-entered ORCID

150 data. These values are open text, and therefore, we coded them into coherent categories for

151 traditional educational (graduate student, post doctoral, lecturer), tenure track (assistant,

152 associate, full professor), and other roles (fellow, research assistant, researcher, head).

153 Employment geopolitical region was also selected when available.

154 ***Education***. As with employment information, we also collected education

155 information from self-entered ORCID data. These values were coded into general

156 categories of bachelor, master, and doctoral degrees. The geopolitical region of the listed

157 education was included when available. For analyses, both employment and education

158 levels were grouped into United Nation regions.

159   ***Types of Publications***. ORCID includes information about the type of publication

160 pulled from either researcher entered data or Crossref. We coded these publications into

161 general categories including book, conference presentations, data-sets, journal articles,

162 preprints, software, thesis, and other publications.

163   ***Publication Metrics***. We calculated total number of publications of any type from

164 both Google Scholar and ORCID. We additionally pulled both the h-index and i-10 index

165 from Google Scholar. The h-index represents the highest h number of publications that

166 have at least h citations, while the i-10 index represents the number of publications with at

167 least 10 citations.

168   **Results**

169   **RQ1: Journals.**   Articles were most commonly published in Advances in Methods

170 and Practices in Psychological Science ($n = 20$), Pre-Print ($n = 17$), Perspectives on

171 Psychological Science ($n = 7$), PLOS ONE ($n = 4$), and Nature Human Behaviour ($n = $

172 3). A complete list of journals can be found on our Open Science Framework page XXX.

173 The average 2-year impact factor for official journal publications was 8.22 ($SD = 10.67$)

174 and the average 5-year impact factor was 6.41 ($SD = 4.42$).

175   **RQ2: Articles.**   Articles were primarily Social (41.4%) and Cognitive (31.4%),

176 followed by smaller categories for Metascience (17.1%), Clinical (4.3%), Developmental

177 (4.3%), and Educational (1.4%). Here, we will talk about the figure below and how the

178 trends occur across time.

179
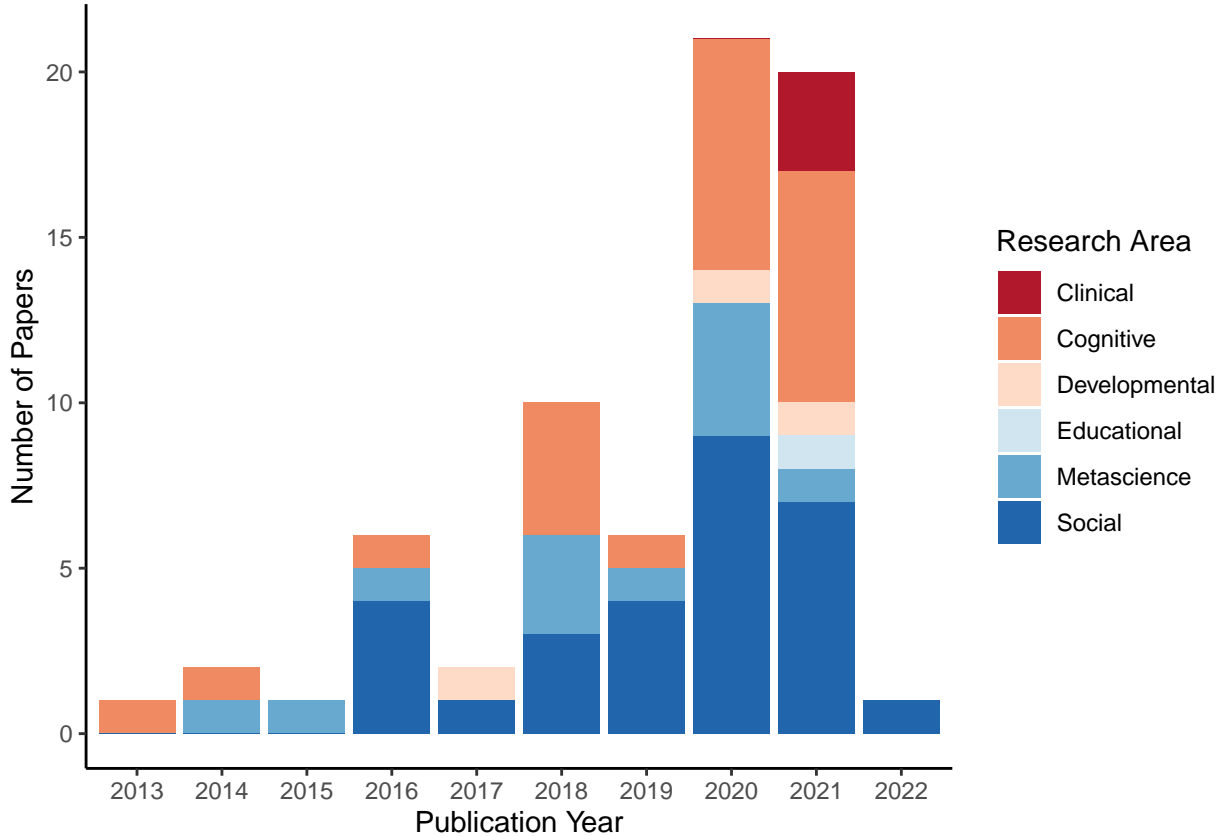
RQ3: Authors.

181 ***Career Length***. The average career length was 11.97 ($SD = 7.95$) for ORCID

182 first publication year (i.e., current year minus year of first publication), $M = 13.12$ ($SD =$

183 8.74) for Scholar year of first publication, and $M = 13.54$ ($SD = 7.77$) for ORCID year of

184 first degree. In each of the analyses, we use the actual year of first publication/education,

185 as subtracting publication and using it as a predictor would create a perfect solution for

186 each model.

187 Next, we will talk about the results from three variables:

188 1) ORCID year first publication:

189 • The slope for year of first publication was $b = 0.00$, $SE = 0.00$, $t(53) = 1.71$, $p = .092$.

190 • The variance parameter was $SD = 7.99$.

191 • The slope for variance across years was $b = 0.01$, $SE = 0.48$, $t(23) = 0.02$, $p = .983$.

2) Scholar year first publication:

- The slope for year of first publication was $b = 0.00$, $SE = 0.00$, $t(15) = 1.59$, $p = .132$.
- The variance parameter was $SD = 8.88$.
- The slope for variance across years was $b = 0.16$, $SE = 0.61$, $t(2) = 0.26$, $p = .821$.

3) ORCID year first degree:

- The slope for year of first degree was $b = 0.00$, $SE = 0.00$, $t(35) = 5.14$, $p = < .001$.
- The variance parameter was $SD = 7.77$.
- The slope for variance across years was $b = 0.03$, $SE = 0.82$, $t(14) = 0.04$, $p = .971$.

Then we will discuss how these results are represented in our graph (below). We will discuss any discrepancies in the results across the three different statistics.
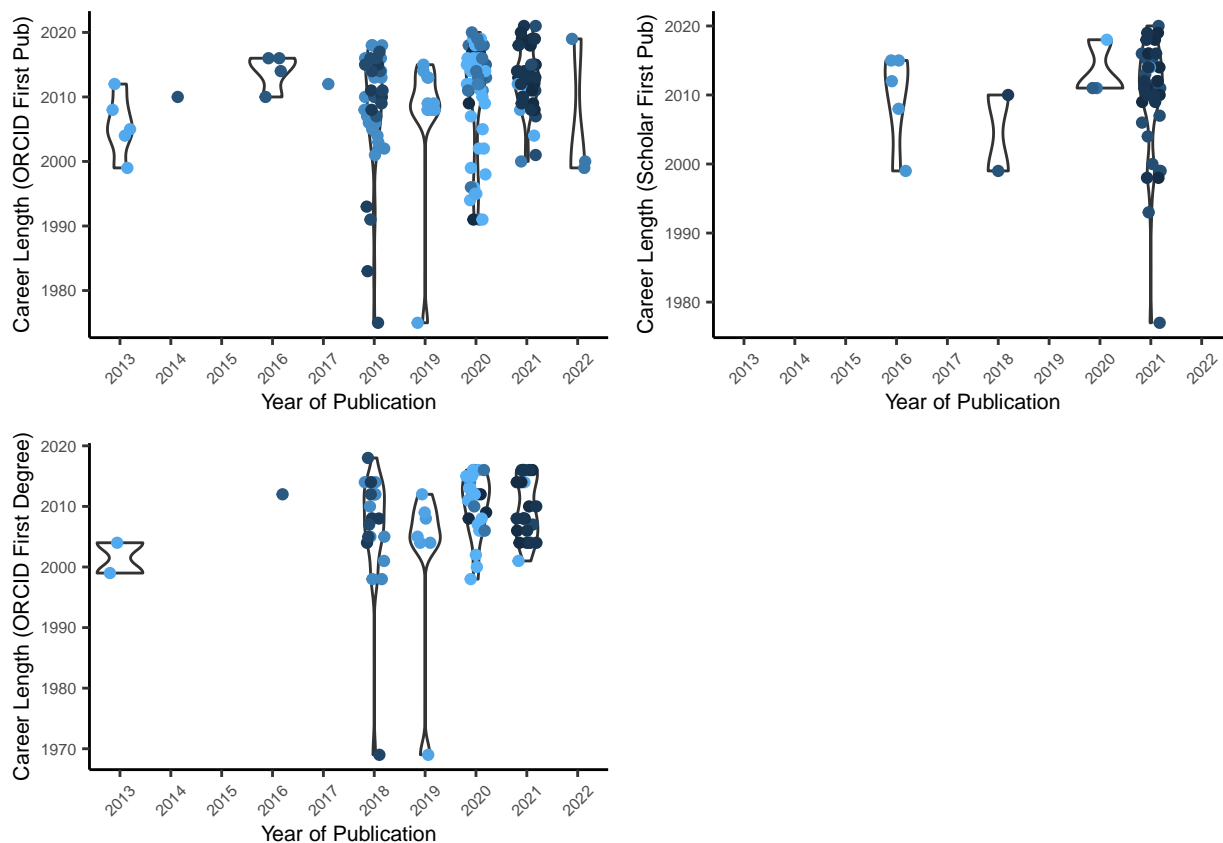
Discussion

To be included after we have completed analyses.

205

References