Success in International Science Olympiads is a Poor Predictor of Scientific Impact

Oscar Gómez[1,*,+], **Diego Kleiman**[1,*,+], **and Kenya Vázquez**[1,*,+]

[1]New York University Abu Dhabi, Division of Science, Abu Dhabi, 129188, UAE
[*]{oscar.gomez, diego.kleiman, kvn222}@nyu.edu
[+]These authors contributed equally to this work

ABSTRACT

The International Olympiads in Mathematics (IMO),[1] Physics (IPhO),[2] Biology (IBO),[3] Chemistry (IChO),[4] and Informatics (IOI)[5] are prestigious annual competitions in which high school students from around the globe solve challenging problems. This study hypothesizes that success at International Science Olympiads correlates with scientific impact. To answer this question, a database of former Olympians was constructed and matched against two bibliometric datasets (Microsoft Academic Graph and Scopus Author Search). Several factors including gender, academic age, number of coauthors, and citation inflation were controlled during the analysis. It was found that the normalized number of citations that a scientist receives is weakly correlated to their ranking in an International Science Olympiad ($r^2 = 0.009$).

Introduction

The International Olympiads in Mathematics (IMO),[1] Physics (IPhO),[2] Biology (IBO),[3] Chemistry (IChO),[4] and Informatics (IOI)[5] are prestigious annual competitions in which high school students from around the globe solve challenging theoretical and experimental tasks. Each participating country sends a team of four to six exceptional and well-prepared students to compete for prizes: gold, silver, and bronze medals, as well as honorable mentions. Such prizes are awarded based on the relative performance of the contestants at each competition.

Despite the emphasis on awards, the objectives of these academic events go beyond fostering friendly competition. In their mission statements,[1–5] steering committees often include one or more of the following aims: (1) to stimulate the interest of young students in a particular scientific field, (2) to identify scientific talent, (3) to promote international cooperation, (4) to encourage cross-cultural dialogue, and (5) to set educational and assessment standards.

International Science Olympiads (ISOs) have gained the attention of the broader scientific community due to the achievements of past competitors, who have been recipients of 16 Fields Medals, 3 Wolf Prizes, 20 EMS Prizes, 14 AMS Research Prizes, 9 Clay Awards, 3 Nevalinna Prizes, 2 Knuth Prizes, 6 Gödel Prizes, a Fundamental Physics Prize, and 2 New Horizons in Physics Prizes. ISOs alumni include notable mathematicians such as Grigori Perelman (IMO '82), computer scientists such as Peter Shor (IMO '77), and physicists such as Steven Gubser (IPhO '89). Given the anecdotal evidence of success from these individuals, it is natural to wonder whether their high scientific impact could have been predicted based on their stellar performances at IMO and IPhO.

Attempting to predict scientific impact is far from a pursuit motivated by pure curiosity. It is a widely recognized fact that scientific research is related to socioeconomic growth[6] and thus it is in the best interest of nations to invest on programs that will boost their scientific development. Consequently, evaluating the scientific impact of past competitors at ISOs is relevant to understand the influence that such competitions may have on future careers and a nation's pool of scientific talent.

Previous studies related to ISOs can be roughly divided in two groups. The first one encompasses studies that focus on the competitions themselves and not on the future prospects of Olympians. Most of these identify the factors that determine individual or country-level performance at ISOs[7–9] whereas others try to tackle the ubiquitous gender imbalance found in these events.[10] Studies on the second group look beyond the competition to observe the effects of ISOs on the career choices[11–16] and scientific impact[17–19] of these programs' alumni. Studies from both groups rely with few exceptions[18] on survey-based methodologies or personal interviews and case studies. They offer country-level or highly individualized perspectives that highlight experiences of top performers rather than entire cohorts. A detailed literature review is available in the supplementary material.

The present study utilized a large scale, data-driven approach to systematically investigate the scientific impact of ISOs past competitors. Through a bibliometric analysis, this work provides a global perspective on the academic careers of Olympians beyond the country- or individual-specific insights offered in previous studies. The principal hypothesis tested by this study is the supposition that individuals who attain a higher ranking at an ISO will have higher scientific impact later in their careers. However, while the main goal of this work is to investigate the relation between success at ISOs and success in academia, the data gathered was analyzed from multiple point of views to contextualize the work within the broader field of Science of Science. Namely, gender differences[20] and migration patterns[21] were also measured to accompany the analysis.

Data

To identify Olympians, the official websites for each competition[1,3] or reliable, community sustained unofficial websites[2,4,5] were accessed. For each ISO, tables with each Olympian's name, ranking, prize awarded (gold medal, silver medal, bronze medal, honorable mention, or none), nationality, gender –if available–, and year of competition were extracted by parsing the website's HTML source code. In the case of the IBO, the data was obtained manually since the results were kept in PDF files. In order to implement analyses on the performance and subsequent academic success of each Olympian depending on the discipline they competed, an additional column with the ISO ID was added and, based on this, columns with markers that stated whether they had participated in the same ISO for multiple years or if they have competed in multiple disciplines were added. After dropping Olympians whose names were unknown, 43,849 competitors were obtained; in the case the parsed data from the website did not include each participant's ranking, it was assumed that they had achieved the best possible ranking depending on their placement on the results table. For those competitors whose gender was not available the genderize.io API[22] was used to infer their gender based on their first names and nationalities, achieving accuracies of 94% and 90% (testing on the IMO data with gender information available) for male and female names respectively (see Supplementary Material for details about performance).

After obtaining the Olympians' data, their names were cross-referenced with Microsoft Academic Graph[23,24] with the condition that the matching researcher's first paper had to be published after the year the Olympian competed. From the database, the fields obtained were the following: number of publications, year of first publication, year of last publication, average number of coauthors, number of citations, and current affiliation (from the last published paper). Given that many competitors had quite common names, the data was validated against the Scopus Author Database[25] by looking for each author by first name and last name. Since it is common for an author to publish under different names depending on their affiliation –for instance consider renowned mathematician Terence Tao who was identified under the names T. Tao, Terence Tao, and Terrence Tao–, the cutoff for matches was set to 3 authors. This corroboration was repeated with a cutoff of 25 matches and no result from the study changed. Moreover, due to spelling variations, some authors or their publications were not properly identified; this was especially true for non-Western names –for instance, leading mathematician Nader Masmoudi had only one paper under the name *Nadir Masmoudi*, since that was the way his name was recorded for both times he participated in the IMO. Finally, in order to study migration patterns of researchers, we used Google Geocoding API in order to locate the place of an author's last affiliation. The limitations of the latter methodology are that some institutions share names –e.g. "Sofia University" in the United States and Bulgaria– and that Google Geocoding API favoured English speaking countries because the names of the affiliations are in English.

	Number of Authors
Total	12510
Male	10878
Female	1434
Undetermined	198
Current Location Known	2817

Table 2. Number of authors identified through Microsoft Academic Graph and Scopus Author Database.

Methods

It has been established that scientific citations are a good measure of scientific impact and it is common practice to use this metric in current literature.[26,27] In order to draw fair comparisons between groups of researchers spanning different fields, years of activity, and career stages the citation count was normalized by number of coauthors, citation inflation, and academic age. To account for the number of coauthors, the citations received by a given publication were equally split among all authors. To normalize by academic age, the academic lifespan was computed as *year of last publication*$-$ *year of first publication* $+1$. Then, the number of citations received by a scientist were divided by the academic lifespan. Accounting for citation inflation was more complex than the previous two data transformations. First, the yearly citation inflation rate was consulted in the literature.[28] It was then verified that this citation inflation rate (4.1%) specific to scientific fields was consistent across the entire period studied (1960–2019). From this rate, we derived a citation weighting factor using 1960 as reference year (i.e., the weight for a citation received in 1960 is 1). The functional form of this factor is an exponential decay

$$w(y;r) = \exp\{-(1+r)*(y-1960)\} \tag{1}$$

where y is the year when the citation was received and r is the inflation rate (4.1% in this study). Inspired by Candia (2019),[26] we made the assumption that nearly all citations a paper receives arrive in a 10-year window right after the publication date. Using this assumption, we computed the average weighting for such a time span as

$$W_{avg}(y;r,l) = \frac{1}{l}\int_{y}^{y+l} w(y';r)dy' \tag{2}$$

where the new parameter l is the window length (10 years in this study). Finally, the citations received by an author due to a paper published in year y where weighted

	IBO	IChO	IMO	IOI	IPhO
Years	1990–2019	1968–2019	1959–2019	1989–2019	1967–2019
Participants	4723	7855	18203	6004	7063
Male	2973	6268	16391	5474	6224
Female	1481	1182	1524	288	478
Gold Medal	516	918	1532	682	1036
Silver Medal	988	1674	3080	1323	1452
Bronze Medal	1482	2475	4540	1934	2158
Honorable Mention	222	383	3025	0	2047
Multiyear	1261	2107	5483	1594	1836
Multidiscipline	45	42	1114	858	361

Table 1. Summary of the data collected for the ISOs.

by $W_{avg}(y; r = 0.041, l = 10)$.

To study the correlation of the normalized citations to ISO performance, a linear regression was performed. The F-statistic was used to verify the statistical significance of the fit. In order to control for additional factors (gender, discipline of competition, and multiyear/multidiscipline status) a multilinear model was fitted (see Eq. 3) and the N-way ANOVA test was used to verify its statistical significance. It was not possible to control for nationality due to the high number of countries represented (which would have resulted in groups too small to perform statistics).

Linear regressions and barplots (see Supplementary Material) show 95% confidence intervals computed via bootstrapping with 10000 resampling steps. Other figures were produced using Seaborn and/or Matplotlib. World maps were plotted using Cartopy.

Results

Dataset Matching

Upon matching the ISOs dataset to the bibliometrics datasets, we found that at least 29% of participants have published scholarly work, indicating that a sizeable percentage of ISOs alumni entered academia. Figure 1 shows the nationalities distribution of all the ISOs competitors. The dataset covers most of the world, although the African continent (excluding South Africa and Morocco) is underrepresented. This is due to the fact that many African nations have not started sending teams to these competitions yet, although some countries in North Africa already participate in some ISOs. The countries where these competitions started (Eastern Europe) and countries that joined early (usually from Europe or from the Soviet sphere of influence) naturally show a higher number of competitors. However, due to the disintegration of the Soviet Union and other border-altering events, several young Eastern European countries show a low number of competitors.

Figure 2 displays the percentage of scientists from each nation that were matched to the academic dataset. We can see that for some European countries that percentage goes as high as ~60% while for some Central American and African nations it is below 10%. It should be noted that the percentages for certain countries (e.g., China) are affected by the fact that hundreds or thousands of researchers from these nations share the same name and were therefore removed from the dataset during the validation step.

A linear regression of normalized citations vs. normalized competition ranking was performed. For the fitted linear model, adjusted $r^2 = 0.003$. The resulting F-statistic was 38.6 ($p < 0.001$). The coefficient of the single predictor was 0.8577 ± 0.138 ($p < 0.001$). Given the small correlation coefficient, it is clear that performance at ISO has low (though significant) predicting power of scientific impact.

A more sophisticated multilinear model is also proposed,

$$Y = \alpha + \sum_{i=0}^{4} \beta_i(C_i) + \gamma(R) + \delta_1(G) + \delta_2(MY) + \delta_3(MD) \quad (3)$$

where C_i is a dummy variable indicating which competition the participant attended (IBO, IChO, IMO, IOI, or IPhO), R is the (continuous) normalized ranking obtained by the competitor, G is a categorical variable for gender ($f = 0$, $m = 1$), while MY and MD are also categorical variables indicating whether the participant competed in multiple years or took part in different competitions respectively. The results for this model are summarized in Table 3.

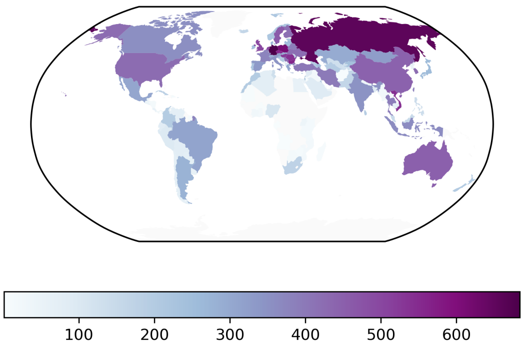

Figure 1. Geographical heat map of the number of students per country that participate in ISO

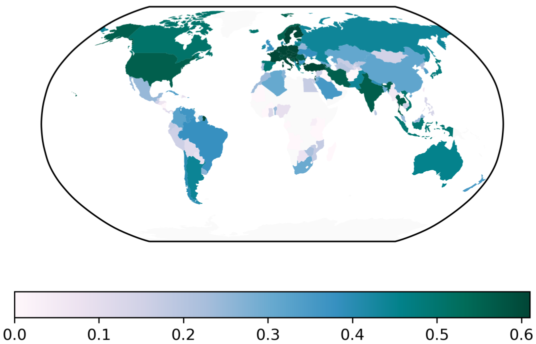

Figure 2. Geographical heat map of the percentage of Olympians per country that go on to pursue an academic career

	Normalized Citations
IBO	-0.3451*
	(0.159)
IChO	0.2088
	(0.149)
IMO	0.5154***
	(0.141)
IOI	-0.1361
	(0.177)
IPhO	0.3662*
	(0.154)
Ranking	0.8637***
	(0.139)
Gender	0.3302**
	(0.120)
Multiyear	0.3636*
	(0.109)
Multidiscipline	0.0572
	(0.217)

Table 3. Regression table for normalized citations. $F = 14.27$ (p < 0.001). Adjusted $r^2 = 0.009$.

Finally, Figure 3 shows the migration of scientists across countries as revealed by their nationality and current affiliation. Green dots indicate emigration while blue dots indicate immigration. The map reveals that the direction of movement tends to be from "source" regions (Easter Europe, Asia) towards "sink" regions (Western Europe, North America).

Discussion

Our matching results show that an impressive amount of Olympians become scholars. At around 29%, and equally so for men and women, it is above population averages. In the US, only about 13% of the population has a Master's, Professional, or Doctoral degree[29], and clearly the number that remain in academia is much lower. Even though this figures can vary significantly between countries, it is also reasonable to assume that the proportion of the population with advanced degrees is higher in more developed countries, and that the proportion of Olympians that remain in academia is higher in such countries. Our dataset is consistent with other studies that have used more reliable methods to detect ISO competitors who become researchers.[17] In other words, survey-based methods estimated that 52% of American Olympians pursued a Ph.D., which correlates well with our result for the US (see Figure 2).

Although there are clearly exceptional cases, such as those competitors who perform extremely well both in an ISO and later in academia (Fields Medal, Abel Prize, etc. winners), it is noteworthy that the correlation between ranking at an ISO and academic prowess (measured by normalized citations) is very weak. What this seems to imply is that people who perform poorly in these competitions and have the drive to become reputable scientists do not let themselves down and work to do so. We can explore such examples by looking at the interactive version of the regression plots (`https://oscargomezq.github.io/ranking_norm_cit.html`, `https://oscargomezq.github.io/ranking_tot_pubs.html`), and examining points with low ranking and high number of normalized citations. This also seems to imply that a key ISOs goal is being fulfilled, fostering interest in science, rather than competitiveness.

The correlation between rank and normalized citations, while small, was still significant. We do not know why this is the case, but we can speculate that higher ranking individuals might have more access to opportunities to enter, and remain in academia. For example, some elite universities are known to accept a high number of international applicants that have succeeded in ISOs (MIT, University of Waterloo, etc.). Furthermore, high performing individuals tend to come from more developed countries, which probably provides an advantage to enter academia.

We can also speculate the other way around however, and

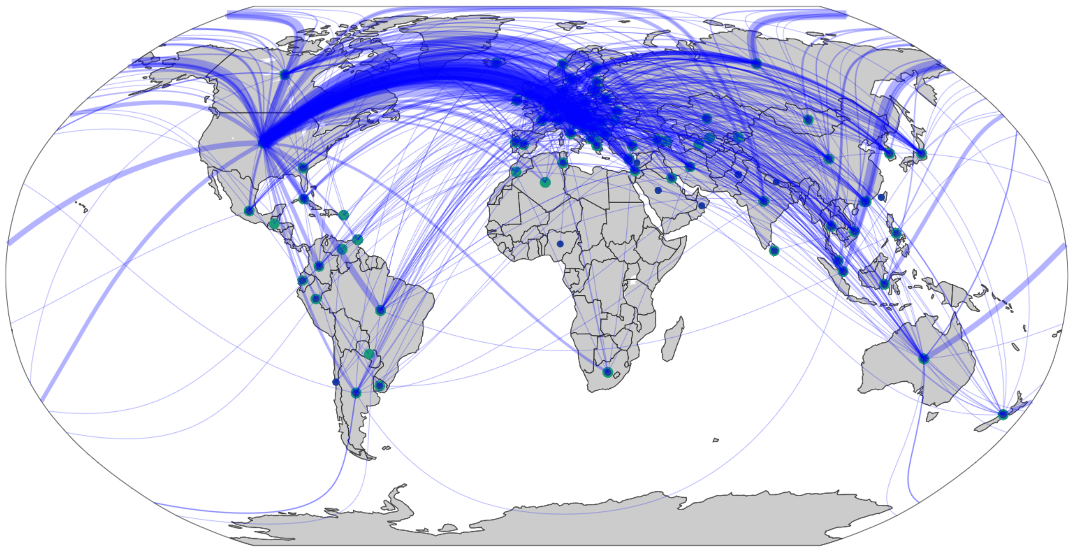

Figure 3. Patterns of migration of Olympians who pursue an academic career. The current country was obtained by taking the geolocation of the researcher's last paper

attribute the small correlation to the fact that any participant in an ISO has already been selected through a very rigorous process that ensures they are the top performers within their respective countries. At this level of high performance, it is likely that students who are driven enough to pursue a career in academia are capable of doing so. It is also worth to notice that even though high achievers in ISOs undeniable possess remarkable capabilities for complex problem solving, the problems proposed during competitions are always known to have a solution, and are meant to be solvable within a relatively short amount of time by at least a few contestants. This contrasts the very nature of scientific research, which is characterized by being open-ended; situations where scientists try to tackle unsolvable questions abound. Because of this, achieving a high ranking at an ISO does not necessarily imply one has the necessary skills and mindset for a career in academia, which someone with perhaps a lower ranking, but more perseverance, does.

It is clear from Figure 2 that the group of ISO competitors who later became scholars represents a vast number of nationalities and ethnicities. Although our results do not imply a causal effect of ISOs on the ethnic diversity of scientists, it is likely that these events are achieving their goal of promoting cultural diversity in STEM to some extent. This is relevant because it is an established fact that ethnic diversity increases scientific impact.[27]

In terms of gender differences, it was found that being male is correlated with having more citations. This agrees with a comprehensive study that measured the effect of gender on scientific careers[20]. Moreover, it was found that the academic lifespan of women who were past ISO competitors is shorter than that of men (see Supplementary Material, Figure 25). Since Huang (2020)[20] found that career length is at the center of gender differences in scientific impact, we predict that this is also the case for past ISO participants.

The multiyear indicator was also significant in our regression results. It can be vaguely interpreted as a measure of how young a participant is when they are already capable of performing at an extremely high level. We speculate that students who show high levels of achievement at a younger age are also more likely to succeed as scientists.

With respect to the migration patterns of scientists, our study is consistent with others.[21] It is likely that the patterns we observe are highly inform by political, social, and economic factors such as the disintegration of the Soviet Union, which prompted the migration of many Easter European scholars to Western Europe and North America. In conclusion, this study demonstrated that performance at ISOs is a weak indicator of future scientific success, while accounting for several variables that are known to affect scientific impact. Moreover, the agreement between this study and relevant literature was explored.

References

1. https://imof.co/about-imo/vision/.
2. https://www.ipho-new.org/statutes-syllabus/.
3. https://www.ibo-info.org/en/about/what-we-do.html.
4. https://www.ichosc.org/regulations.

5. https://ioinformatics.org/files/regulations19.pdf.
6. Stephan, P. E. The economics of science. *J. Econ. Lit.* **34**, 1199–1235 (1996).
7. Lim, S. S., Cheah, H.-M. & Hor, T. A. Science olympiads as vehicles for identifying talent in the sciences: The singapore experience. In *Communicating Science to the Public*, 195–211 (Springer, 2014).
8. Saul, M. More than a system: what we can learn from the international mathematical olympiad. *Notices AMS* **58**, 410–416 (2011).
9. Henseke, G. Country performance at the international mathematical olympiad. Tech. Rep., Thünen-Series of Applied Economic Theory-Working Paper (2009).
10. Steegh, A. M., Höffler, T. N., Keller, M. M. & Parchmann, I. Gender differences in mathematics and science competitions: A systematic review. *J. Res. Sci. Teach.* **56**, 1431–1460 (2019).
11. Sahin, A., Gulacar, O. & Stuessy, C. High school students' perceptions of the effects of international science olympiad on their stem career aspirations and twenty-first century skill development. *Res. Sci. Educ.* **45**, 785–805 (2015).
12. Nokelainen, P., Tirri, K. & Campbell, J. R. Cross-cultural predictors of mathematical talent and academic productivity. *High Abil. Stud.* **15**, 229–242 (2004).
13. Sahin, A. Stem clubs and science fair competitions: Effects on post-secondary matriculation. *J. STEM Educ. Innov. Res.* **14**, 5–11 (2013).
14. Dauber, S. L. International mathematical olympaid. *Gift. Child Today Mag.* **11**, 8–11 (1988).
15. Choi, K. M. Influences of formal schooling on international mathematical olympiad winners from korea. *Roeper Rev.* **35**, 187–196 (2013).
16. Choi, K. M. *Characteristics of Korean International Mathematical Olympiad (IMO) winners' and various developmental influences* (Columbia University, 2009).
17. Campbell, J. R. & Walberg, H. J. Olympiad studies: Competitions provide alternatives to developing talents that serve national interests. *Roeper Rev.* **33**, 8–17, DOI: 10.1080/02783193.2011.530202 (2010). https://doi.org/10.1080/02783193.2011.530202.
18. Rindermann, H., Ziegler, A. & Perleth, C. Results in the international mathematical olympiad (imo) as indicators of the intellectual classes' cognitive-ability level. *Excell. Essays Honour Kurt. A. Heller: Lit, Münster* 303–321 (2011).
19. Reiman, I. *International Mathematical Olympiad: 1976-1990*, vol. 2 (Anthem Press, 2005).
20. Huang, J., Gates, A. J., Sinatra, R. & Barabási, A.-L. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc. Natl. Acad. Sci.* **117**, 4609–4616, DOI: 10.1073/pnas.1914221117 (2020). https://www.pnas.org/content/117/9/4609.full.pdf.
21. Robinson-Garcia, N. *et al.* The many faces of mobility: Using bibliometric data to measure the movement of scientists. *J. Informetrics* **13**, 50 – 63, DOI: https://doi.org/10.1016/j.joi.2018.11.002 (2019).
22. Genderize.io. Determine the gender of a name.
23. Sinha, A. *et al.* An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web*, 243–246 (2015).
24. Wang, K. *et al.* A review of microsoft academic services for science of science studies. *Front. Big Data* **2**, 45 (2019).
25. Elsevier. Elsevier developers.
26. Candia, C., Jara-Figueroa, C., Rodriguez-Sickert, C., Barabási, A.-L. & Hidalgo, C. A. The universal decay of collective memory and attention. *Nat. human behaviour* **3**, 82–91 (2019).
27. AlShebli, B. K., Rahwan, T. & Woon, W. L. The preeminence of ethnic diversity in scientific collaboration. *Nat. communications* **9**, 1–10 (2018).
28. Petersen, A. M., Pan, R. K., Pammolli, F. & Fortunato, S. Methods to account for citation inflation in research evaluation. *Res. Policy* **48**, 1855 – 1865, DOI: https://doi.org/10.1016/j.respol.2019.04.009 (2019).
29. Bureau, U. C. Number of people with master's and doctoral degrees doubles since 2000 (2019).
30. Verhoeff, T. *The 43rd International Mathematical Olympiad: A Reflective Report on IMO 2002* (Technische Universiteit Eindhoven, Department of Mathematics and Computer . . . , 2002).

Author contributions statement

All authors were involved in the conceptualization, different process of the data curation, formal analysis, investigation, visualization, methodology. D.K. was in charge of administrating the project: coordinating the responsibilities of each author. Finally, all authors redacted and reviewed the manuscript.

Supplementary Material

S1. Literature Review

While there is literature regarding the impact of ISOs, usually the main focus of the articles is to provide evidence of how these contests encourage young students to pursue STEM careers. Others focus on comparing the performance of different countries and make connections between a country's wealth, support system for science, etc. to their team's performances at ISOs. Finally, some others provide highly individualized perspectives through interviews and case studies with top performers, who are defined by their individual results or national team results. The next following paragraphs present a short summary of each paper encountered during our literature survey.

In Sahin (2015),[11] the authors present high school students' perceptions of the factors affecting their career choices and their beliefs regarding their participation in ISO. Through a survey, this paper demonstrated that International Olympians' inspiration was mostly influenced by their teachers, personal interests, and parents while their participation in the ISO reinforced their plan to choose a career in STEM.

Similarly, Nokelainen (2004)[12] compared the factors that influenced students from Finland and from the US so that they could determine the cross-cultural and culture-specific elements that contribute to the development of the Olympians' talent. Moreover, they used the previous predictor factors–such as family support, socioeconomic class, GPA, school influences, computer literacy, and the Olympians' self-attributes and interests–to determine if any of these have an impact on their career productivity.

In Sahin (2013),[13] the impact of STEM Clubs and science fair competitions on US students' college enrollment and the majors they chose is studied. In this paper, the authors demonstrate that students who take part in extracurricular STEM activities have a higher than average college enrollment and they found a positive correlation between the number of years students were involved in STEM activities and choosing a STEM major in college.

In Rindermann (2011)[18] the authors provide a comprehensive overview of the IMO's characteristics, such as gender distribution and participant's average household income, which complements the description of the competition itself in Verhoeff (2002).[30] They also calculate correlations between countries' relative rank in IMO with GDP, Economic Freedom, and a STEM indicator (based on number of Nobel Prizes, number of scientists, patent rates, and fraction of high-technology exports). However, the claims that are most relevant to this study are the following: (1) as adults, former IMO competitors usually have academic careers, and (2) as scientists they have above-average numbers of publications and patents. In contrast, Reiman (2005),[19] emphasizes that bad performances at the IMO usually have no implications about the mathematical potential of a well prepared student.

In Steegh (2019),[10] the relation between gender and participation in ISO and other STEM-related events is explored. The authors draw a comparison between female representation in academia and female representation in ISO national teams.

Interviews with top competitors from the US' and Korea's teams are discussed in several articles.[14–16] These provide a more nuanced and subjective view of the influences that the IMO experience had on each participant, and provides an accurate, but small sized sample of the career outcomes they had: more than half had completed or were enrolled in a PhD program.

More general, country-level case studies are presented in multiple articles.[7–9] Overall they discuss strategies that have enabled successful countries such as Singapore and Peru to thrive at these competitions, and the difficulties faced by others. Overall they claim that the key strategy is building a strong support system that identifies and retains young scientific talent.

S2. Figures

Given the lack of predictive power in the linear regression of normalized citations and ranking, figures displaying the scattered data were not presented in the main part of this report. Additionally, some heat maps describing the academic success of those participants who participated in multiple disciplines or multiple years of the ISO were also produced. While the number of participants in each category might not be sufficient to pursue further analysis, these maps follow general trends of the each disciplines' pattern of citations and publications[27]. While there is still some analysis to be performed, the data seems to imply that Olympians who participate in multiple disciplines have more academic success than those who do not.

In figure 24, the percentage of female participants across competitions and years in plotted. From here it is evident that ISOs have remained a predominantly male space, with IBO being a notable exception, as it consistently has around double or triple the amount of female contestants compared to the other ISOs. Despite the stark difference in the distribution of genders, it is worth noting that female participants are just as likely as males to remain in academia and become scholars, as measured by the proportion of all ISO participants that were matched with the academic datasets. Figure 22 shows that this is clearly the case. There are some significant differences between females and males when looking at the academic lifespans (Figure 25) and the normalized citations separated by competition, especially in IMO and IPhO, (Figure 26).

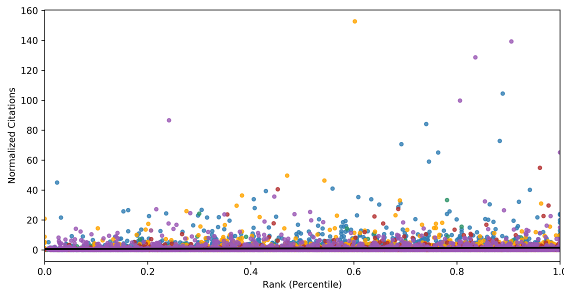

Figure 4. Linear regression of ranking against normalized citations with 95% confidence interval of all ISO Olympians matched in Microsoft Academic Graph

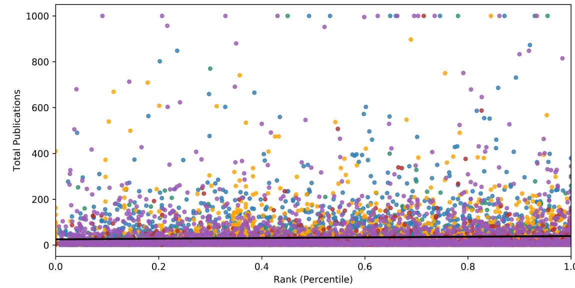

Figure 5. Linear regression of ranking against total publications with 95% interval of all ISO Olympians matched in Microsoft Academic Graph

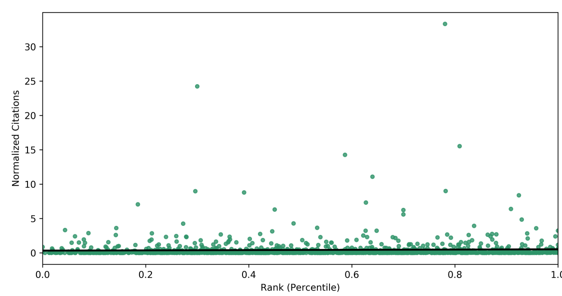

Figure 6. Linear regression of ranking against normalized citations with 95% confidence interval of all IBO Olympians matched in Microsoft Academic Graph

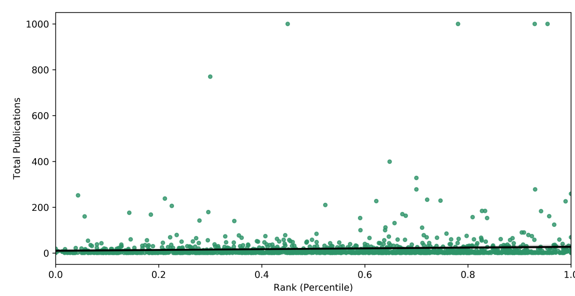

Figure 7. Linear regression of ranking against total publications with 95% interval of all IBO Olympians matched in Microsoft Academic Graph

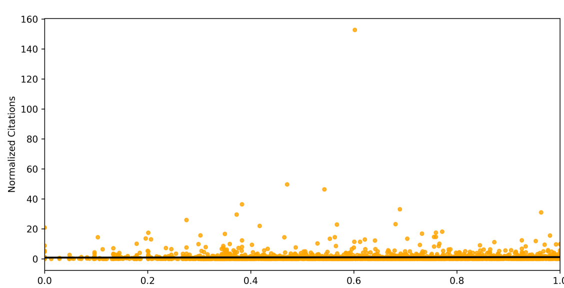

Figure 8. Linear regression of ranking against normalized citations with 95% confidence interval of all IChO Olympians matched in Microsoft Academic Graph

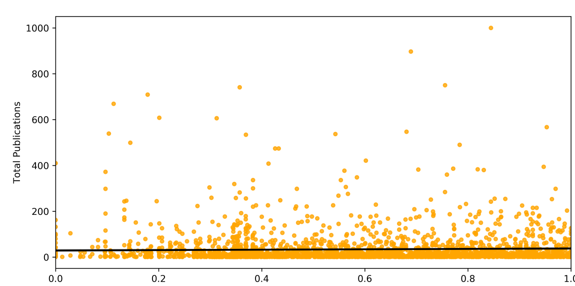

Figure 9. Linear regression of ranking against total publications with 95% interval of all IChO Olympians matched in Microsoft Academic Graph

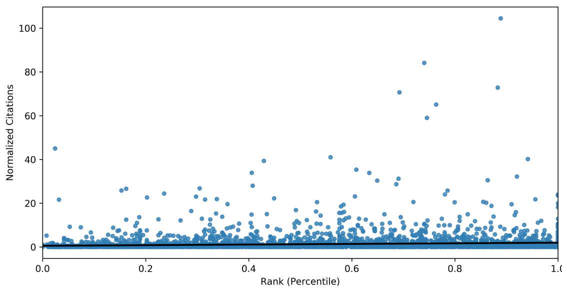

Figure 10. Linear regression of ranking against normalized citations with 95% confidence interval of all IMO Olympians matched in Microsoft Academic Graph

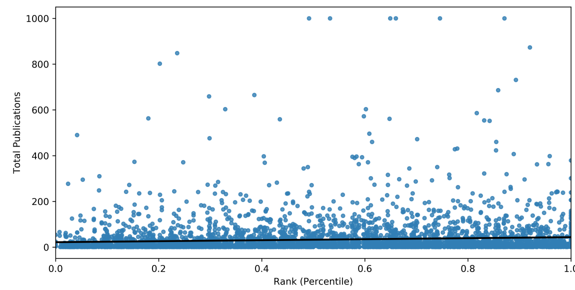

Figure 11. Linear regression of ranking against total publications with 95% interval of all IMO Olympians matched in Microsoft Academic Graph

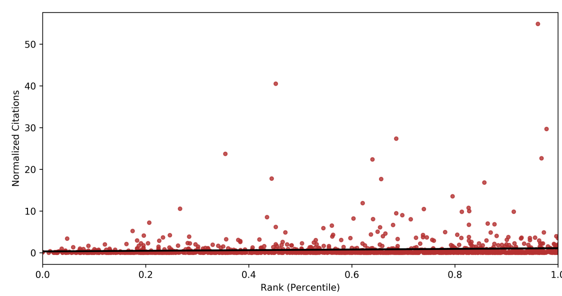

Figure 12. Linear regression of ranking against normalized citations with 95% confidence interval of all IOI Olympians matched in Microsoft Academic Graph

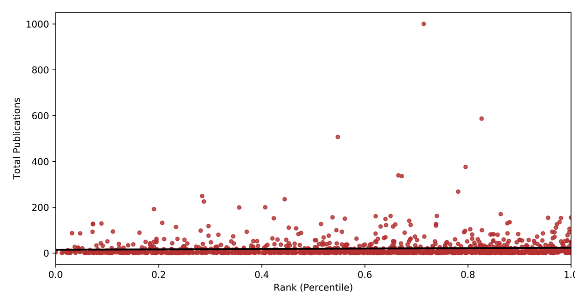

Figure 13. Linear regression of ranking against total publications with 95% interval of all IOI Olympians matched in Microsoft Academic Graph

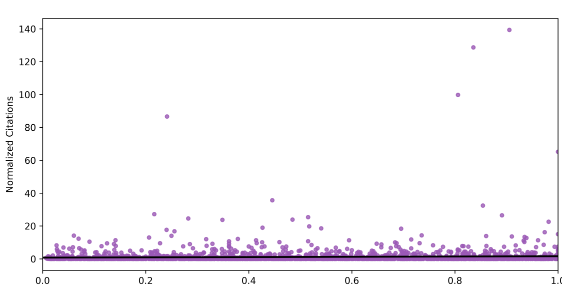

Figure 14. Linear regression of ranking against normalized citations with 95% confidence interval of all IPhO Olympians matched in Microsoft Academic Graph

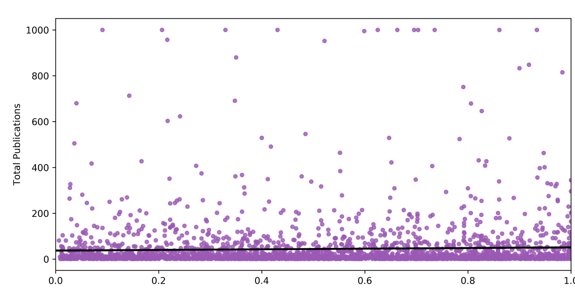

Figure 15. Linear regression of ranking against total publications with 95% interval of all IPhO Olympians matched in Microsoft Academic Graph

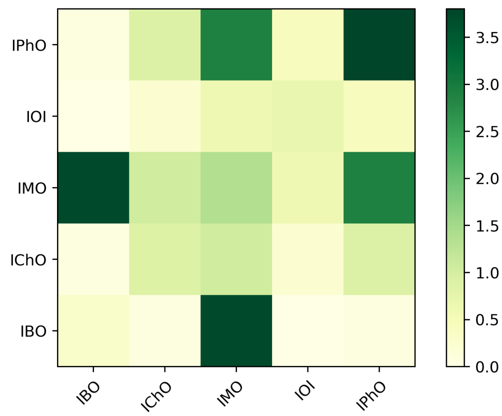

Figure 16. Heat map of the average normalized citations per each category of multi-discipline participation. Note that the categories are symmetric with respect to the main diagonal.

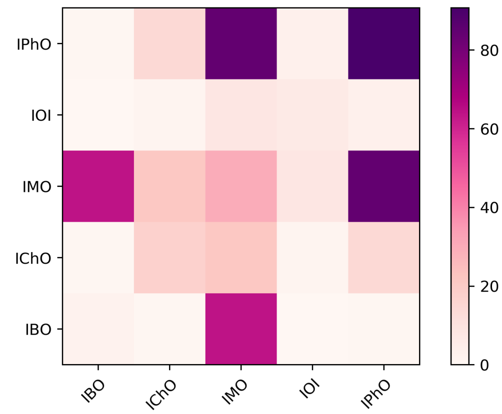

Figure 17. Heat map of the average pseudo-normalized citations per each category of multi-discipline participation. Note that the categories are symmetric with respect to the main diagonal.

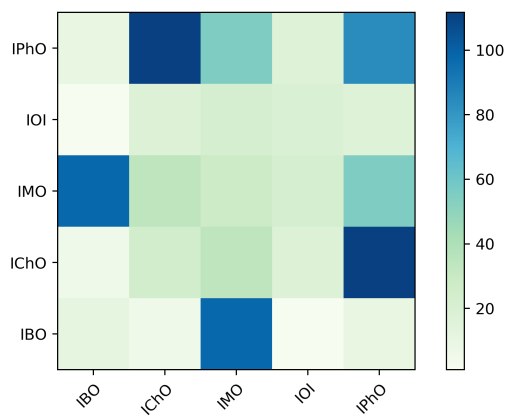

Figure 18. Heat map of the average number of publications per each category of multi-discipline participation. Note that the categories are symmetric with respect to the main diagonal.

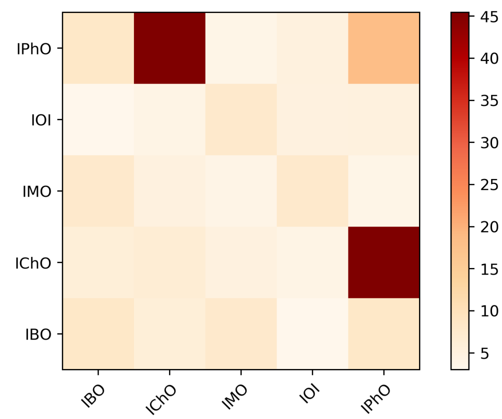

Figure 19. Heat map of the average number of coauthors per each category of multi-discipline participation. Note that the categories are symmetric with respect to the main diagonal.

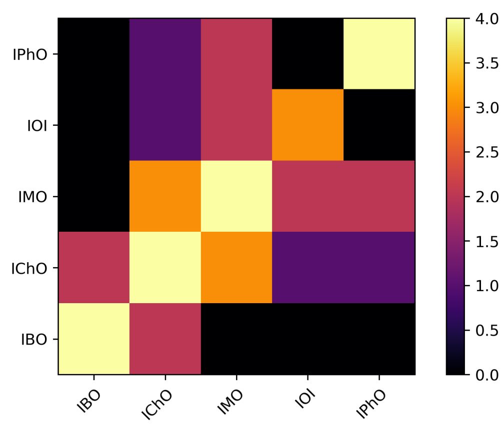

Figure 20. Heat map of the total number of female researchers who participated in multi-discipline ISO. Note that the categories are symmetric with respect to the main diagonal.

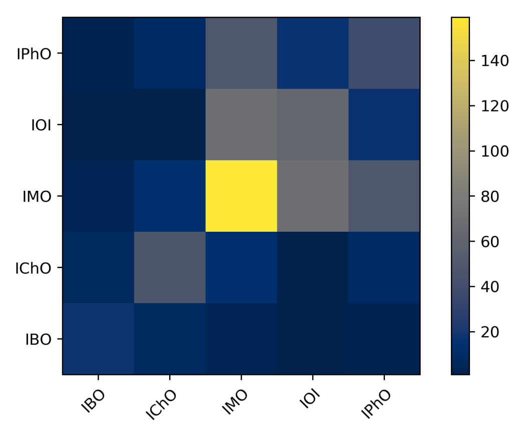

Figure 21. Heat map of the total number of male researchers who participated in multi-discipline ISO. Note that the categories are symmetric with respect to the main diagonal.

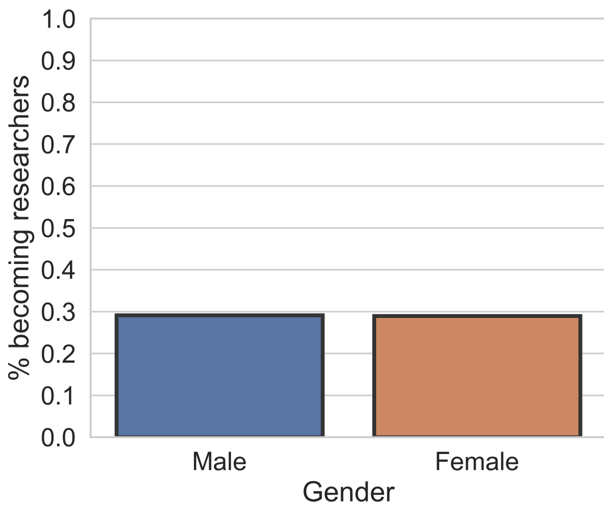

Figure 22. Percentage of contestants that were matched with the academic dataset by gender

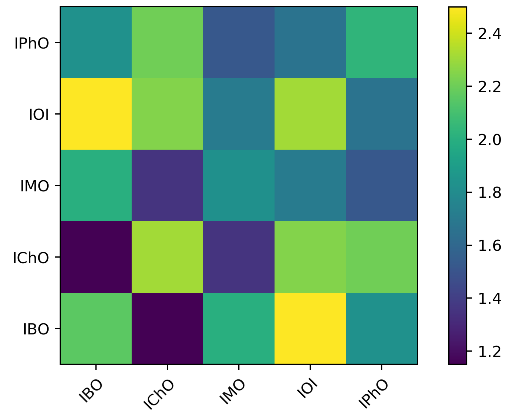

Figure 23. Heat map of the average award (where a Gold medal is 4 and participation is 0) of Olympians who participated in multi-discipline ISO. Note that the categories are symmetric with respect to the main diagonal.

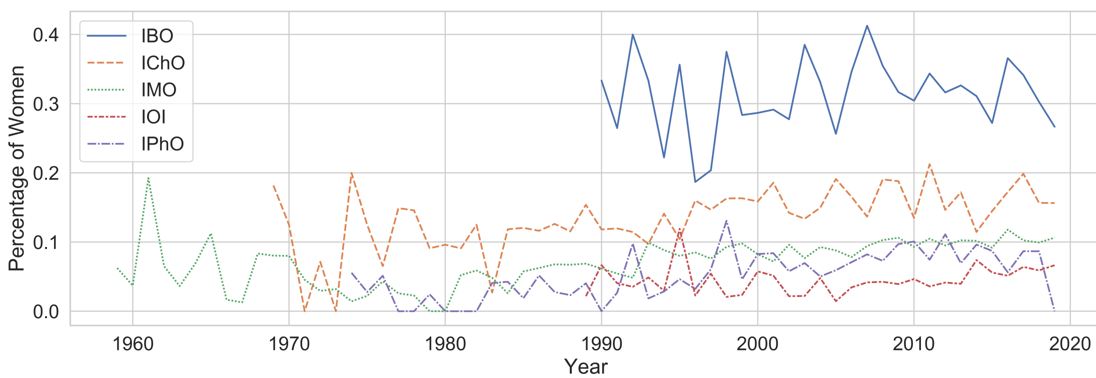

Figure 24. Percentage of female participants in each ISO across years

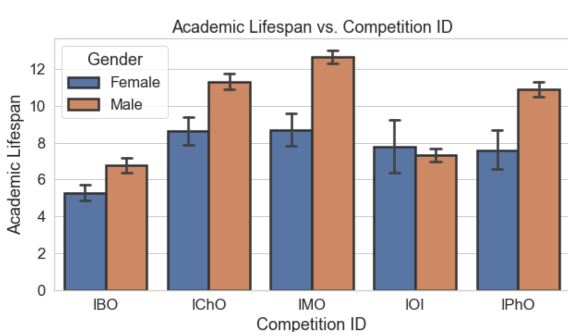

Figure 25. Average academic lifespan separated by gender

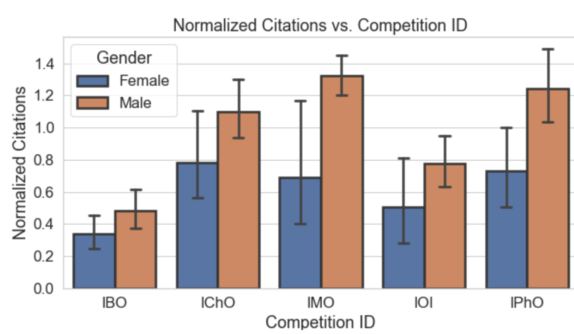

Figure 26. Average number of normalized citations separated by gender and competition