

International Journal of Education Policy & Leadership

A QUANTITATIVE ANALYSIS OF COUNTRIES' RESEARCH STRENGTHS

ANURAG SAXENA

Indira Gandhi National Open University

S. DAVID BRAZER

George Mason University

B. M. GUPTA

National Institute of Science, Technology and Development Studies

This study employed a multidimensional analysis to evaluate transnational patterns of scientific research to determine relative research strengths among widely varying nations. Findings from this study may inform national policy with regard to the most efficient use of scarce national research resources, including government and private funding. Research output from 34 countries is examined using a conceptual framework that emphasizes the ratio of research resources devoted to a particular field to research output measured by publications in peer-reviewed journals. Using cluster analysis and k-means analysis, we conclude that countries' research output (as measured by the number of published peer-reviewed articles) and their efficiency (as measured by a ratio of research output to dollars allocated to research) together indicate a comparative advantage within any given country's own menu of research choices and an absolute advantage relative to other countries. This study implies that the more countries engage in publication in areas of relative strength and consume research in areas of relative weakness, the stronger their entire research agenda will become.

Saxena, A., Brazer, S. D., & Gupta, B.M. (2009). A Quantitative Analysis of Countries' Research Strengths. *International Journal of Education Policy and Leadership* 4(3). Retrieved [DATE] from <http://www.ijepl.org>.

Introduction

When a country produces some goods and services more efficiently than others, the country achieves a comparative advantage in its more efficient economic sectors. But the advantage is compared within the country, not across countries. International trade theory, as far back as David Ricardo, has posited that international trade benefits all trading partners when countries exchange those goods and services they produce more efficiently for goods and services they produce less efficiently, despite any sort of absolute advantage they may or may not enjoy in world markets. For example, Country A may not produce anything more efficiently than Country B, yet it is in the interest of both countries to trade goods and services that each produces more efficiently relative to its production of other goods and services (Suranovic, 2006). This arti-

cle's cross-disciplinary analysis of multinational scientific research demonstrates that the international trade concept of comparative advantage is a helpful analogy for considering many countries' research production.

Data from scientific publications show that the world's scientific hierarchy has remained remarkably stable over the past decade, with the postindustrial nations of the United States, Japan, and Europe continuing to dominate in many fields (Braun, Glänzel, & Schubert, 1999). A close examination of scientific publication data helps researchers understand that there is a relationship between economic capacity and research capacity. But the relationship is not as simple as it may seem, because certain countries appear to specialize and excel in particular fields of research even though they may not be as developed as the economic powerhouses named above.



SIMON FRASER
UNIVERSITY



Poorer countries' abilities to conduct and publish research suggest that they utilize their comparative advantage in the international marketplace of peer-reviewed journals. Their relative efficiency stems from fields of higher education and targeted funding support that lead them to focus on specific types of research, even though they lack absolute advantage in any particular area.

Purpose of the Study

The purpose of this study is to demonstrate patterns of scientific research publication transnationally to determine if production patterns exist and whether or not they are consistent with economic development. This study moves beyond previous pair-wise comparisons (Doré & Ojasoo, 2001) to engage in multidimensional analysis that allows for the examination of many countries simultaneously. Our intent is to articulate specific research patterns, and then use them to make inferences about how countries might choose research priorities and develop research policies.

Research Significance

We suspect that allocation of national wealth into particular fields in higher education has implications for scientific research output. When specific patterns are made clear, then it becomes possible to suggest how countries' researchers could maximize publication in peer-reviewed journals for which they are best suited. Drawing on the comparative advantage analogy, countries should produce the scientific research that is best supported by their economies and, by extension, their educational institutions, even if other countries' universities may be better supported in those same fields. Focusing research and publication this way means that each country will maximize output, in the form of peer-reviewed journal publication, in those areas that are its comparatively greatest strengths. "Trade" takes place through journal publication and readership.

Scientific research output is evidence of the relative strength of a particular field or discipline within a country. For example, if the United States has the greatest amount of published social science research worldwide, researchers might want to know how the United States came to this position. There are several possible reasons, such as deployment of substantial resources in the form of larger amounts of government grants for social science research relative to other fields within the United States, superior post-baccalaureate education in the social sciences, relatively poor educational or monetary support

for social science research in other countries, or a combination of all of these factors that are closely related to economic development. Once relative productivity in research within a country is identified for the world's research-producing countries, then it would be possible to find out the underlying reasons why. This article takes the first steps toward understanding differentiated research output by adopting a new approach to existing data and analyzing it through the comparative advantage lens. By comparing the relative research strengths of many of the world's countries, this study demonstrates a comparative advantage that may derive from countries' allocation of resources into higher education specializations.

Research Questions

To learn more about patterns of research around the world and their relationships to economic development, we ask the following research questions:

1. Are patterns of research productivity discernable according to resources devoted to a particular discipline?
2. Are research priorities clustered according to countries' stages of economic development, such as agricultural, industrial, or postindustrial?
3. Which disciplines (e.g., chemistry and physics) show similar and different publication trends in a set of countries?

We define *research productivity* as the ratio of a country's published research in a particular discipline to the total publications by all countries in that particular discipline. Comparing a country's research productivity in multiple fields reveals a country's strengths and weaknesses in various fields. The comparisons are made within countries or within clusters (see Methodology below) to avoid restating the obvious: that wealthier countries are more productive than poorer ones.

Being published is defined as acceptance and publication in a peer-reviewed journal. Although there is wide variation in the perceived quality of peer-reviewed journals—particularly across disciplines—measuring differences is a major problem. It would be possible to look at acceptance rates, but such data is highly variable depending on how each journal engages in review (for example, a one- or two-step review process), defines a rejection (for example, the way it categorizes a "revise and resubmit" that never comes back), and handles incomplete or inappropriate submissions. Consequently, we have settled on successful publication in peer-

reviewed journals as a minimum standard to define research success.

We define research priority as the ratio of a country's published research in a particular discipline to its total publications in all disciplines. If a particular country has a ratio of 0.35 in chemistry, for example, compared to 0.17 in economics, then chemistry would be considered a higher research priority. Our conceptual framework that follows below hypothesizes that priorities result from comparative advantage. In the brief example given here, there are circumstances in the hypothetical country that make production of chemistry research more efficient than that of economics.

Conceptual Framework

We employ a relatively simple conceptual framework to address the research questions we have stipulated for this study. It stems from two key characteristics we observed in worldwide research output data prior to beginning this study: (1) research output derives from multiple factors that operate simultaneously, and (2) countries appear to have greater research output, in terms of research publications, in some fields compared to others. These phenomena suggest that a multidimensional analysis of research output would help identify production patterns, and the concept of comparative advantage provides a plausible explanation for the patterns detected.

Multidimensional Analysis

A multidimensional database is numeric data that have multiple characteristics. For example, countries' publication trends can be considered from the perspectives of discipline, country, and year. The multidimensionality of the data is lost or not considered at all if it is handled in series, as it would be in a table format.

Multidimensional analysis allows a complete and simultaneous view of the data to answer queries pertaining to the multiple dimensions, such as the interactive effect of an extended recession in a particular country and scientific output in a particular field. Examining multiple factors simultaneously is more powerful than conventional methods that account for factors serially.

Comparative Advantage

Borrowing from the international trade concept of comparative advantage is a compelling way to look at differentiated patterns of research output. Simply put, the theory of comparative advantage states that international

trade is maximized when all participating countries trade their most efficiently made products with other countries doing the same. Efficiency is relative within a country, so even if Country A produces everything more efficiently than everything Country B produces, both countries are better off when Country A purchases Country B's most efficiently produced goods, and vice versa (Suranovic, 2006).

As with trade goods, countries are likely to realize greater publication success in those fields where they have achieved the greatest degree of research efficiency. Efficiency derives from available resources and how those resources are applied. For example, it would be possible for a country with relatively low economic output to devote available funding to one area of postbaccalaureate research—possibly chemistry—where it could achieve relatively high research output per dollar spent. Similarly, a relatively wealthy country might have no systematic way of allocating dollars to research, thus funding numerous areas that are relatively unproductive. As a result, lower ratio of output to research dollar spent would be achieved; yet it might still be more efficient at producing chemistry research than the low-wealth country. Comparative advantage suggests that the low-wealth country will nevertheless produce chemistry research because it can produce that more efficiently than any other field of research. The low-wealth country will “trade” its chemistry research in peer-reviewed journals for other research produced by the high-wealth country.

We depart from the international trade analogy at this point in our conceptual framework because nations' returns from research are not monetary—at least not in the short term. There is an important return, however, and that is scholarship. By engaging in the practice of publishing in and reading peer-reviewed research journals, countries are able to improve their scholarship. Such activity may ultimately benefit countries' economies, but proving that connection is beyond the scope of this particular article.

We do not yet have the data for measuring various characteristics of efficiency, such as relative quality of educational fields or dollars devoted to various research fields. We can, however, examine research produced by field and by country and compare those to the ratio of resources allotted across research publications within a particular discipline. We hypothesize that countries at various levels will produce different kinds of research in differing numbers of fields. Low-wealth countries seem more likely to produce in fewer fields than high-wealth

countries, and they seem more likely to produce research appropriate to their economic needs.

The simple conceptual framework we employ focuses this study on identifying research output patterns across a large number of countries. It serves as a working hypothesis for the idea that such patterns exist and will organize themselves in a manner similar to comparative advantage in international trade. To test such a hypothesis requires research publication data and cluster analysis, discussed in the methodology section below.

Methodology

Data Sources

Data for this study are taken from Web of Science, which provides access to publication data from approximately 8,700 peer-reviewed research journals around the world. The database was taken from national indicators, which change from year to year.

Data Collection

These data were downloaded in the year 2003. They provide the number of publications—articles, reviews, notes, and proceedings—in 22 specific disciplines published by the 34 countries with the highest publication output over the period of 1993 to 2002.

Data Analysis

This paper uses the cluster analysis technique to answer the research questions we pose above. Cluster analysis (described in more detail below) reveals the patterns of different countries and their research productivity. The resulting analysis presents two different measures of a country's output in a particular discipline: (1) the share of the country's publications of all publications in a single discipline, and (2) the proportion a particular discipline represents of the country's total research output. We have deliberately not aggregated the data according to the Research Priorities Index (Nagpaul & Pant, 1993), which compares intercountry research priorities as follows:

$$\left[\left(\frac{n_{ij}}{n_{i0}} \right) / \left(\frac{n_{0j}}{n_{00}} \right) \right] * 100$$

where n_{ij} is the number of publications of country i in subfield j , n_{i0} is the number of publications of country i in all subfields, n_{0j} is the number of publications of all countries in subfield j , and n_{00} is the total number of publications of all countries in all subfields. This approach has an advantage over raw publication counts

in that it inherently takes into account the size of the country and the size of the subject area (Nagpaul & Sharma, 1995). Instead of combining factors as the Nagpaul and Pant method does, we wanted to see the influence of discipline and country separately. The result is that we aggregate the data in two distinct ways.

Cluster analysis.

Exploratory procedures are appropriate for understanding the nature of multivariate relationships between and among countries and their research output because distinct patterns have not been described prior to this study. We, therefore, search the research output from Web of Science to detect structure and groupings using Cluster analysis, a freely available computer program written by Michael Eisen from Stanford University (Eisen, Spellman, Brown, & Botstein, 1998). (To view the program, go to <http://rana.lbl.gov/EisenSoftware.htm>). Cluster analysis is a technique that assumes nothing about the number of groups or the group structure. It groups on the basis of similarities as indicated by distances from means. Cluster performs all necessary calculations and displays results using multiple color combinations so that cases that are clustered display similar colors.

In this article, we apply nonhierarchical clustering techniques that group cases into a collection of "k" clusters. For example, clustering disciplines means determining groups of disciplines that show similar publication trends across countries. Clustering the disciplines provides the relative position of countries within an area of world research. It brings disciplines into focus in a manner that demonstrates similar publication trends across countries (see Appendix A). Another example is clustering countries, which means finding groups of countries that show similar publication trends. Clustering the countries demonstrates the relative importance of a discipline within a particular country (see Appendix B). This also clarifies which countries place similar priority on various disciplines within the range of research each country produces.

Nonhierarchical procedures have an advantage over hierarchical procedures because a matrix of distances does not have to be determined, and thus nonhierarchical procedures can be applied to much larger data sets (Johnson & Wichern, 2001). We have applied one of the more popular nonhierarchical procedures, the k-means method, as originated by McQueen (1967). The k-means method assigns each item to the nearest mean. In other words, cluster similarity is measured in regard to the

mean value of the objects in a cluster, which can be viewed as the cluster's center of gravity (Han & Kamber, 2001). For example, Country A would be assigned to cluster Z because its value on a particular dimension is closer to Z's mean than it is to any other cluster. We have adhered to k-means clustering because we wish to highlight the country associations on the basis of country percent and subject percent.

With the help of clustering, the data can be segmented into small similar regions and thus comment on overall distribution patterns of the data (Saxena, Khare, & Garg, 2004). Clustering is done on the basis of a similarity measure—attributes or variables are used to derive the clusters so that data points in one cluster are more similar to one another than they are to data points in other clusters. In other words, observations are homogeneous within clusters and heterogeneous across clusters (Anderberg, 1973).

Results

We examined data from the 34 countries listed in the Web of Science database as of 2003. Conducting k-means clustering for research output within disciplines and across countries, we find the discipline clusters displayed in Table 1 (see page 8).

The discipline clusters in Table 1 (see Appendix A for specific discipline cluster data) reveal output similarities across countries. In other words, all countries that produce research are likely to have greater output similarity within the chemistry, materials science, and physics cluster and greater output dissimilarity compared to another cluster, such as psychiatry, social science, and economics. Yet, because the analysis captures all countries at once, there will be some variation between different countries' output within clusters. We are looking for the overall trends that generally hold true for most countries.

Some countries are outliers in many clusters, such as the United States or Malaysia (percent world contribution too high or too low), because of their much larger or much smaller economies and ability or inability to provide resources to all disciplines. Also, there are specific differences in the performance of countries in different clusters. Returning to the comparative advantage discussion above, the low-wealth country will nevertheless produce a particular type of research because it can produce that more efficiently than any other. The low-wealth country will “trade” that research in peer-reviewed journals for other research produced by the

high-wealth country. Low-wealth countries (e.g., Malaysia) seem more likely to produce in fewer fields than high-wealth countries, and they seem more likely to produce research appropriate to their economic needs (e.g., India in agricultural sciences). There are ample indicators of the differences between clusters in the table. For example, in the cluster of psychiatry, India and China have very low output. However, these countries perform much better in other clusters, such as that of chemistry, materials science, and physics. Output from Australia and the United States is relatively high in the psychiatry cluster. This suggests that each country has a unique kind of system that produces research in various disciplines. The extent of contribution depends on various factors discussed above.

Breaking out the discipline clusters, as we have in Table 1, allows us to analyze individual country's research output more easily. Applying k-means analysis to country output within discipline clusters reveals which countries produce similar amounts of research within discipline clusters. By displaying output data by country cluster and by discipline cluster in Table 2 (see page 9; see Appendix B for specific country cluster data), research capacities of individual countries are revealed.

This analysis provides an interesting classification of the countries on the basis of the share they are devoting to each discipline. In other words, this table is only indicative of the percent of total research a cluster of countries dedicated to a particular discipline. It is on this basis that countries can be clustered. As seen earlier in Table 1, the United States is producing almost 50% of the total research published in economics, but economics is only about 2–3% of the total research published in the cluster of which the United States is a part (see Table 2) because economics falls under the psychiatry cluster. Also, about 27% of published research is devoted to clinical medicine in the cluster of countries in which the United States falls, but the contribution of the United States to this discipline worldwide is only about 35%. Brazil and Argentina are left unclustered because they have different research priorities, with Brazil giving more attention to physics and Argentina placing more emphasis on clinical medicine. The benefit of this kind of analysis is that it can rate the research priorities of a cluster of countries and compare them with other countries.

Discussion

In the area of multidisciplinary output, the cluster with India, China, and Russia demonstrates major contribu-

tions to world research; India and China together have contributed about 18% of the total world research (see Appendix A). Surprisingly, Russia's contribution to this area is 21%. This makes the contribution of these three countries equivalent to 40%. More unexpected is the fact that this 40% world contribution is achieved by the cumulative research priority of 4% (see Appendix B). The United States has a research priority of 0.3% in the multidisciplinary area but contributes 30% to world research. India has done well in agriculture by making a contribution of 6% to world research, with a 5% research priority in this area. This trend is repeated in India and Japan (7.5% and 1.5% respectively). But one can say that Japan's impact on agricultural research is greater. In another example, Brazil is devoting 27% of its research priorities to physics and chemistry, but its contributions to world research are around 1.5% and 1% respectively. Similarly, Israel is devoting 23% of its research priorities to clinical medicine, but its contribution to world research is around 1%.

Table 3 (see page 10) shows the ratio of country's world contribution to research in specific disciplines to the percentage of total research output that discipline represents for the country as a whole. Argentina, for example, contributes 1.08% (from Appendix A) of the world's published research in agricultural science. This same category represents 3.93% (from Appendix B) of Argentina's total effort. Thus the ratio is $1.08/3.93 = 0.27$. The resulting figure demonstrates that, despite Argentina's relative effort in this area, it does not yield as large a contribution to world output as most other countries.

The table above demonstrates that all those countries that have a ratio greater than unity should provide research to the world market in these subjects, because even though the proportion of research effort for the country may be small, it is making a significant contribution to world research publication in that field. Those countries that have ratios less than unity are likely consumers of research in these subjects. For example, one can say that India has done well in agriculture sciences by making a contribution of 6% to world research by allotting 5% of its total research resources to the subject. Japan is performing better with a ratio of 7.5% to 1.5%, making it more efficient than India in producing agriculture sciences research—Japan gets greater research output for every dollar of effort. Nevertheless, agriculture science is a success area for India relative to its other research fields. For this reason, India's research in agriculture science should continue. Likely consumers of

this research are Malaysia, Singapore, Argentina, South Africa, and Turkey, because their ratios are among the lowest. Similarly, Australia, Russia, China, and India will likely consume research in fields such as chemistry, biology, engineering, physics, and social science.

If a country has a higher ratio of world contribution to research priority in a particular research field, this indicates a highly productive education system that is making an impact on a global level. Countries with ratios greater than unity can provide to countries with ratios less than unity in that field through publications, conferences, and direct interventions in higher education. The benefits of comparative advantage are realized when countries "trade" their specific expertise in this fashion. One result is likely to be improvement in the research base in all trading countries in all fields.

Conclusion

Definite patterns of research productivity emerge from the 2003 Web of Science publication data. Countries exhibit distinct ratios of research output to effort, as measured by research dollars, that indicate where their relative strengths—their comparative advantages—lie. Patterns are less obvious with regard to stages of economic development, indicating that research productivity is more a matter of policy priority and political will than economic imperative. Because countries specialize in different areas and achieve comparative advantage, it is in the interest of all nations to participate both as purveyors and consumers of published research.

High-wealth countries, such as the United States, France, Germany, and the United Kingdom, will continue to dominate research publications, but other countries can maintain and thrive in niche research areas. Understanding these patterns more clearly could be achieved by further analysis using either Nagpaul's Research Priority Index or some other measure to find the cumulative effect of a country's specialized research efforts. It would also be helpful to know why certain countries that appear dissimilar on other measures are clustered in specific kinds of research. Deeper examination of research priority and specialization illuminates how countries compete in the world market for research, while further examination of clusters may indicate how countries could collaborate more effectively.

References

- Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.
- Braun, T., Glänzel, W., & Schubert, A. (1999, May 1). A global snapshot of scientific trends [Electronic version]. *UNESCO Courier*, 5, 28–29. Retrieved January 24, 2005, from http://www.unesco.org/courier/1999_05/uk/dossier/txt21.htm#e1.
- Doré, J., & Ojasoo, T. (2001). How to analyze publication time trends by correspondence factor analysis: Analysis of publications by 48 countries in 18 disciplines over 12 years. *Journal of the American Society for Information Science and Technology*, 52(9), 763–769.
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, USA*, 95, 14863–14868.
- Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann Publishers.
- Johnson, R. A., & Wichern, D. W. (2001). *Applied multivariate statistical analysis* (3rd ed.). New Delhi: Prentice Hall of India.
- MacQueen, J. B. (1967). *Some methods for classification and analysis of multivariate observations*. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Vol. 1* (pp. 281–297). Berkeley, Calif.: University of California Press.
- Nagpaul, P. S., & Pant, N. (1993, June). Cross-national assessment of specialization patterns in chemistry. *Scientometrics*, 27(2) 215–235.
- Nagpaul, P. S., & Sharma, L. (1995, October). Science in the eighties: A typology of countries based on inter-field priorities. *Scientometrics*, 34(2), 263–283.
- Saxena, A., Khare, P., & Garg, S. (2004). Application of cluster analysis as a tool to analyze distance education students. *Asian Journal of Distance Education*, 2(2), 1–11.
- Suranovic, S. (2006). *The theory of comparative advantage—overview*. Retrieved November 6, 2006, from <http://internationalecon.com/v1.0/ch40/40c000.html>