Size versus performance in research

Ron Johnston
ACIIC, University of Sydney
Lyn Grigg
QUT
and
John Currie
ACIIC, University of Sydney

Abstract

Argument continues to rage about the question of whether big research groups and departments perform better than smaller groups, with obvious implications of the extent to which the national resources for research should flow to big groups, big departments and big universities.

Yet for such an apparently simple question, it appears very difficult to get any consensus. Data and analysis are disputed, and interpretations hotly contested.

A recent NBEET-commissioned report provides some answers. It indicates that in the sciences and engineering there is a threshold effect, varying in group size from 5-12, below which there are considerable inefficiencies. However beyond this size, other than in exceptional cases, there are no economies of scale.

It also raises the issue that it is group productivity, rather than per capita productivity, which is critical in the social dynamics of research (though of course not in practice). Different measures of performance are required for research groups, and universities, with different objectives.

At regular intervals there is an outburst of debate, usually marked by some acrimony, among the research community about the effects of size on research performance and output. The most recent occasion was the publication by Adey and Larkins (1994) of an evaluation of publication records of selected universities for the period 1990-92, in which it was claimed that for the fields of physics, chemistry and biological sciences, "quite clearly, a broad relationship between size of department and total publications is established".

This research was stimulated by an earlier study by Lowe (1993) based on 1990 data only, which found that "large research groups, large departments and large universities are not particularly productive by any quantitative measure of output or impact". Adey and Larkins claimed the longer time-span of their study was the principal reason for reaching an opposite conclusion.

Their findings were immediately challenged by academic staff of Wollongong University, who recalculated the data to show that "small science departments are similarly productive to large science departments" (Anon, 1994).

Why is it that the issue generates so much heat? Why does there seem to be so little consensus? And why are the data produced by one group so readily open to question and dispute? Is there no reliable evidence or analysis on which the relationships of performance to size can be based?

It was in order to address these issues that NBEET commissioned a study of the effects of resource concentration on research performance (NBEET 1993).

The policy context of resource concentration

In most industrial countries around the world, there has been a growing emphasis on the need to construct more explicit and deliberate policies for science. There are various reasons for this including the escalating costs of many areas of research, growing constraints on government spending, and political demands for greater accountability for all areas of public expenditure. The ultimate aim of these science policies has been to ensure that the limited resources which are available are used as effectively as possible. From this stems an interest in research performance, especially of groups of scientists, be they constituted in university departments, laboratories or institutes.

Evaluations of research performance, whether informal or formal, inevitably conclude that some scientific groups are more productive than others. This immediately raises the question of what accounts for performance differences. A wide range of possible determinants of group research performance have been proposed. These include the size of the groups, whether the research is closely linked to teaching, and the type of funding. However, as we shall see, the evidence that these are important determinants of research performance is extremely limited and, where it exists, is often very ambiguous.

A prominent feature of research support policy in many, though not all countries over the last twenty years has been the espousal and implementation of resource allocation processes that provide 'selectivity and concentration'.

Implicit in these policies has been the assumption that 'bigger is better'; in other words, that scientific research benefits from economies of scale. This approach has been most pronounced in the UK and to some extent other Anglo-Saxon derivative countries, but it has been the subject of consideration and experiment in many other countries as well.

In the UK, starting in 1984, the University Finance Committee (UFC) initiated a series of departmental rankings, which were shaped by a widespread view of the need to build departments of international standing, and attain a 'minimum economic size'. General assertions of critical mass, including access to equipment, technical staff and a budget to find new initiatives were used to justify a minimum size for science departments in the range 15-30, depending on discipline.

A generally similar line of argument, though a different mechanism for implementation, has emerged in Australian research and higher education policy in recent years. Thus, the White Paper on Higher Education (Dawkins, 1987) stated:

> *The application of research findings into processes of direct social or economic benefit is also crucial to the Government's objectives and must be increased. None of these areas of research can be effective if limited resources are spread thinly. Concentration and selectivity in research are needed if funding is to be fully effective.* (Dawkins, 1988, p. 90)

In contrast to the British approach, these documents emphasise that it is the universities themselves, rather than the government paymasters, that should implement a policy of selectivity and concentration. This approach was reinforced by a series of reports, the general thrust of which is captured in the statement:

> *In Australia as elsewhere throughout the developed world, the changing nature of research is creating pressures to concentrate resources for research and be more selective in resource distribution...it is necessary to ensure that the most able and effective researchers are funded in such a way that makes best use of their creativity.* (ASTEC, 1989, p. 11-12)

We note again the emphasis on concentrating resources with the best researchers, rather than institutional approaches to concentration as applied in the UK. The implementation of selectivity and concentration policies in Australia thus far have not involved any process of explicit ranking of institutions or their organisational units.

The general rationale for resource concentration is aptly captured by Ziman (1987):

> *A "critical mass" of people and instruments is thus needed, whether for a team undertaking a single large project or in a research group carrying out a programme of coordinated projects in the same field. The actual aggregate of resources required for viable research varies considerably from field to field, but even where all that individual researchers need is access to a library or a computer, advantages are seen in bringing them together into specialised groups. The intellectual environment in such a "centre of excellence" is more stimulating both for mature scientists and for graduate students requiring a thorough training in research skills.* (Ibid., p. 11)

However, there have been substantial criticisms of the assumptions of and justifications for the policy of concentration. Thus, Becher and Kogan have argued that:

> *Firstly, the logic of concentration equates all research with the 'big science' model when much is, and should remain, modest in scope; secondly in an area of commercial relevance concentration would be a liability because firms in competition will prefer to tap different sources of expertise to preserve confidentiality; and third the great diversity of types of research activity should be recognised and inappropriate policies of wholesale concentration should not be pursued.* (Becher and Kogan, 1987, pp. 8-9)

Lowe (1991) has been more scathing in his comments, identifying four problems with the 'widespread superstition' that the achievement of critical mass is necessary for high quality research:

> *The first two arise from the fact that the concentration of resources has negative effects as well as positive ones. Putting the best researchers in charge of centres turns them into research managers...*
>
> *Secondly, it is necessary to take account of what economists call opportunity costs....The third problem is that the establishment of centres or specialised research institutes create units which tend to be self-perpetuating, thereby reducing the capacity of the research funding system to respond flexibly to changing priorities....Finally, it is in the nature of the process for establishing such centres that they tend to perpetuate the traditional division of academic research into the established disciplines.* (Lowe, 1991, p. 187)

What then, is the evidence available on which to base a judgement of the relationship between size and performance?

Previous studies of the relationships between resource concentration and research performance

Despite widespread views that there was little hard data on the effects of concentration on performance (eg. Lazenby, 1992), we were able to identify more than 30 relevant studies since the early 1960s, the majority of which focussed on the relationship between group size and productivity. Unfortunately,

> *The results of this body of work can best be characterised as ambiguous and contradictory. The majority verdict is that research output is linearly related to size with no significant economies of scale apparent. Others have argued that the relationship between output and size is more complicated - for example, that there are economies of scale up to a certain group size after which diseconomies set in. One possible reason for the divergence is that different studies have focused on different units of analysis - research groups, departments or entire institutions.* (NBEET, 1993, p. xi)

The possible reasons for the wide variations between the findings of these studies are numerous, but appear to be principally a consequence of four factors:

- different organisational units, eg. departments, research centres, laboratories, research institutions, research groups;
- different measures of group or unit size, depending variably on the inclusion of students, visiting fellows, technical and support staff, and the assessment of full-time equivalents;
- different measures of production (output) and productivity; and
- confusion between individual and group productivity.

In addition, it is apparent that many other variables, themselves variously related to size, affect research productivity. It would be necessary to identify these variables, and construct methodologies which allow them to be held constant, in order to isolate effects related only to size.

It was in order to attempt to improve on the understanding of the relationships between resource concentration and research performance that a series of analyses and experiments were conducted in the UK and Australia.

The British data

These data are based on a five-year research program at the Science Policy Research Unit, University of Sussex, UK. The department has been primarily selected as the unit of analysis, given the policy climate of the focus of the UFC.

An empirical study examined the effects of size and of teaching on the published output of physics, chemistry and earth sciences departments. Simple analyses based on concentration ratios and linear regression seemed to indicate that 'productivity', measured in terms of journal articles per member of staff, does increase slightly with the size of the department in physics and chemistry, although not in earth sciences. However, even in these first two fields the effect is very small, with size explaining only a few per cent of the variation in output. Furthermore, if research students are included along with academic faculty and post-doctoral fellows in the calculation of department size, the correlation between productivity and size disappears. Furthermore the exclusion of Oxford and Cambridge University departments from the data sets eliminated any correlation between productivity and size.

Interviews with researchers in mathematics, physics, chemical engineering and biochemistry departments revealed they consider links with teaching to have a minor influence on research performance, and the effects of department size to be even less important.

Critical mass effects are regarded as important at the level of the subfield-based group rather than the department. It was considered that a researcher needs to be a member of a group of four to six staff (together with perhaps three or four research assistants and PhD students) working together in the same subfield if they are to be able to compete internationally. There are few direct economies of scale from research per se, apart perhaps from in fields like biochemistry where equipment may be shared across subfield-based groups.

A detailed statistical analysis of size effects in the published output and citation impact of British university departments revealed that the UFC rankings are highly correlated with department size but not with size-adjusted 'productivity' measures such as publications per staff or citations per paper.

For physics and chemical engineering no correlation was found between size-adjusted publication productivity indicators and size - in other words, there is no evidence of economies of scale in these two fields in relation to published output. There are moderate correlations in biochemistry suggesting that there are some benefits to be derived in this field from being located in a larger department, perhaps linked to the sharing of equipment between subfield groups. There were also small but significant correlations for mathematics which may indicate mathematicians are less closely bound into subfield groups and interact more with other colleagues.

A detailed statistical analysis of size effects in the published output and citation impact of British university departments revealed that the UFC rankings are highly correlated with department size but not with

size-adjusted 'productivity' measures such as publications per staff or citations per paper.

For physics and chemical engineering no correlation was found between size-adjusted publication productivity indicators and size - in other words, there is no evidence of economies of scale in these two fields in relation to published output.

There are moderate correlations in biochemistry suggesting that there are some benefits to be derived in this field from being located in a larger department, perhaps linked to the sharing of equipment between subfield groups.

There were also small but significant correlations for mathematics which may indicate mathematicians are less closely bound into subfield groups and interact more with other colleagues.

A detailed comparison of the characteristics and performance of British and American research groups in a tightly defined sub-field of condensed matter physics, superfluid-helium-three, provided some evidence that the type of funding mechanism may have had an impact on the relative performance of research groups.

In Britain, the decision to concentrate resources on program grants for three groups meant that UK groups tended to be larger and more experienced and to employ more technicians than their US equivalents. However, the smaller and less experienced US teams funded by shorter-term project grants seem to have been more innovative and their research earned more citations. There is no evidence from this case-study that longer-term program grants gave British researchers any appreciable advantage. However, this policy apparently had the unintended effect of 'freezing' the chosen groups, preventing the inflow of young researchers and new ideas.

The Australian data

Given the limitations of analysis based on departments, the research group was chosen as the unit of aggregation for examination in Australia. A consequent severe methodological limitation, once it was decided that it was only appropriate to compare the research productivity of groups in the same or closely-related fields, was that in Australia, the population of eligible groups was too small to allow any statistically valid conclusions to be drawn.

The measure of productivity used was one based on a weighting by both type of output and ranking of journals in which papers were published. (NBEET, 1993; Harris 1989, Pettit, 1992). This is comparable to the performance indicators proposed for use by NBEET.

An assessment of the research productivity, in terms of both output and impact, of a range of Australian research groups from five disparate disciplinary areas was unable to provide evidence of a relationship between group size and productivity mainly because of the small sample size available in the Australian context (ie there are only a relatively limited number of quality research groups in any single disciplinary research field).

A second 'experiment' sought to examine the effect of research group size on productivity by comparing the performance of researchers in a full-time research position with that achieved by matching groups in a teaching and research environment.

In selecting research groups with significant 'research only' personnel it was decided to include a number of Special Research Centres (SRCs), given their unique position within the Australian University system as 'beacons' of research concentration and excellence. The SRCs that were to be included were chosen from those that were instituted in 1988, and those of the original SRCs begun in 1982 that were re-funded in 1991. "Matching" groups active in the same or closely-related research fields as those of the SRCs, and with a strong ARC record (ie of successful grant applications), were identified.

The principal findings were:

- wide variability exists in the productivity of all research groups;
- the research output of the SRCs is more than three times that of comparator research groups;
- however, in terms of productivity per head, the larger SRCs output is little more than half that of the comparator groups;
- the SRC's published many more papers comparatively, in high impact journals;
- the number of authors per paper is similar for the SRCs and the comparator groups, indicating that the larger SRCs operate not as a corporate entity where intellectual output is concerned, but as a series of semi-autonomous, cross-linked, groups.

Interviews with Australian scientists using the same instrument as in the UK revealed a broad similarity, but with a much greater emphasis on the importance of critical mass, though opinions varied on what constitutes a 'critical mass', and how such a concentration affects performance.

The following quotes indicate the range of responses:

A critical mass exists in our discipline, but has not been reached in any one location. (Electronics Engineer, Research group).

When a critical mass is present a different type of research is undertaken. (Biochemist, Research Group).

*Probably a group of 5 or 6 is appropriate in this area. Critical mass has a downside for lesser/newer academics - in the long term they do not benefit. (*Biochemist, Research Group).

Groups are at their most productive, in specific research terms, when they are sized 6-7 and work all hours. However, to tackle the big problems, and compete with the major international groups, you need infrastructure and a breadth of capability based on 4-5 such teams. (Engineer, SRC).

The 'critical mass' effect is not necessarily a size issue, but a structure and coherence thing. To be effective, people at different levels are needed - one researcher with 50 postgraduate students does not work. (Physicist, SRC).

In addition, the SRC Directors, and other commentators, presented strong evidence of the way in which the attendant scale and continuity of funding could provide the basis not just for a higher level of research activity, but for research of a different kind, research problems which were more strategically targeted, with a higher risk, but the promise of a greater achievement. Indeed, it would appear that the SRC directors generally had a different conception of risk - less a problem of the unknown than the conscious adoption of a long-odds bet, pitting their capability against a challenge of great magnitude.

Directors adopting these high risk strategies (successfully) were able to point directly to a range of qualitative, but nonetheless objective, measures of high international scientific recognition for their achievements; measures such as invitations to provide the plenary address at international conferences, recruitment of their staff by overseas centres, invitations to review and assess competing international centres, invitations to join national and international committees of advice, and unsolicited visits by foreign experts.

Conclusions

The basis for the 'standard' argument for critical mass in research is that a minimum level of resources, particularly in the form of equipment, or library, together with a minimum level of intellectual interaction, is necessary to be able to perform research in a competitive way. This common sense argument, however, provides little guidance in the actual process of concentration, and the pursuit of economies of scale. Neither is it necessarily valid.

With regard to equipment, there is a range of relatively modestly-priced equipment which universities of a substantial size, with government funding support, can provide to research groups. At the top end of big science, the establishment of international centres is, paradoxically, reducing the importance of 'local' group support. It is in the intervening price range where access to equipment **may** be a determinant of being able to participate in a particular research field. But even here, the instrumentation requirements may have become so specialised that sharing in a large group is not practical.

With regard to intellectual interchange, we have observed that in a number of diverse fields, ranging from physics to history, researchers consider themselves members of international groups of like-minded academics, who are working on the same problems. The dramatic growth in the use of electronic mail between academics is an indication of the importance of this ready means of communication, and of the way in which it is reducing the need to establish local interaction.

The overwhelming evidence of this and other studies is that, in the natural sciences, the mode of research group size is 3-5 academic staff, possibly 2-3 postdoctoral researchers, some postgraduate students and technical/professional staff. Below this threshold, research performance is reduced. This modality does not preclude, however, strong research performance from groups of a smaller size, and in some fields, talented individuals.

A 'unit' of this size, totalling from about 5-12, apparently represents a natural maximum for effective communication. Indeed, the general picture emerging from the operation of larger units, such as research centres, is that once size exceeds this norm by any substantial level, fission will occur to re-establish the desired interaction patterns.

For fields of science where researchers follow this pattern of forming themselves into sub-field based groups of 5-12, there is little evidence that research benefits directly from economies of scale, beyond this threshold effect, either in terms of output or impact assessed through journal citation impact. However, in fields where researchers are not tightly integrated into sub-field groups, as in mathematics and the social sciences, there may be some very modest economies of scale.

Size alone, therefore is never a sufficient condition, and sometimes not even a necessary condition, of effective research performance. The only exception is at the very bottom end of the scale, where it is clearly impractical for a single researcher in a small physics department in a remote teaching-oriented university to participate in particle physics.

It is apparent that large, well-funded and well-led research groups produce more publications, of higher impact, and receive much higher international recognition than do smaller groups, when group output is the basis of comparison. At the same time, productivity per head is higher for small groups.

The issue, which does not appear to have been examined previously, is whether productivity based on group output or per head output is the most appropriate measure for the performance of a research group. Strict comparability, and opposition to resource concentration policies, favours the latter.

However, there appears to be some evidence to suggest that the traditional economic assessment of labour productivity may be less appropriate when applied to research, than a group-based analysis. Scientists themselves, when asked to identify their competitors, refer to groups:

Its the Brussels team that are our big competition
[...]'s group are the ones we have to watch.

Frequently, group achievements are signified by the name of the highly visible leader alone:

[...] has established herself as one of the top ten plant biologists. Over 9 years [...] has built a world class operation.

In research, the competition is primarily for intellectual achievement and recognition. This competition does not depend directly, or at least as much, on the ratio of outputs to inputs, which is the basis of economic productivity. Rather it is the ability to marshal resources including intellectual capability to achieve 'significant advances' ahead of the competition that counts. In this case, group productivity may be a more important than individual productivity.

There is also evidence to suggest that scientific recognition is based on group output, and the ability to capture significant attention based on quality and quantity of output:

[...]'s group made a really big impact when they produced 9 top papers in good journals in one year.

It is the ability to produce a substantial volume of sustained high quality publications in leading journals on which scientific recognition is based; not on a higher output per researcher.

These findings confirm the operation of a 'Matthew effect' - "to him who hath shall be given" (Merton, 1973). The researcher who can command the larger resources can produce the higher output which is the basis for commanding further resources. While ample resources do not alone automatically produce high output, researchers who cannot command such resources in general cannot compete with those who do.

It must also be noted that productivity may not measure all the benefits of a policy of resource concentration. Other advantages may be:

- the value of public and political visibility through a major targeted program;
- easing the problem of increasing costs of administering a competitive research grants project-oriented scheme;
- the importance of threshold size to establish a gatekeeper presence in a research field; and
- the importance of threshold size to be fully plugged into, and accepted as a member of, the international leading-edge research club.

Policy implications

A number of policy implications follow:

1. The evidence that the threshold effect supports the implicit policy of science research funding agencies to favour proposals from teams of researchers, rather than individuals.

2. There is no support for a policy of increasing resource concentration for the purposes of increasing unit research productivity. Hence policy questions such as 'what proportion of ARC funds should be devoted to concentrated forms of support, such as research centres' cannot be resolved by reference to an algorithm of the returns from resource concentration. Rather, a detailed assessment of the precise advantages of resource concentration in the particular context need to be made. However, it may be appropriate for other objectives, including achieving top international recognition for fields of research in Australia.

3. Assessment of a policy of resource concentration needs to be carried out in the context of the what might be called the different tiers of government support for research in universities, and the relative importance of and balance between these tiers.

   The first of these tiers is concerned with the maintenance of the levels of scholarship and linkages with the international research community, necessary to underpin quality higher education. The resources required at an individual level are generally modest, though collectively, for the whole "Unified National System", they amount to many millions of dollars. Support has been traditionally provided through the universities.

   The second tier can be considered to be the support of committed, cumulative research programs, of an internationally recognisable quality, with objectives of scientific or technological advance, or application to particular purposes. This activity, largely supported by the ARC, has a different relationship to economic outcome than the first category, and needs to be assessed accordingly to these different standards. It also requires a higher level of funding per researcher.

   The third tier is directed to achieving international leadership in selected research fields, or the development of new technologies of particular promise. These ambitions are much more exposed to the forces of global competition, and hence need to be resourced and managed in a way that provides the basis for international competitiveness.

   There would appear to be a strong case for a much clearer differentiation in policy of the objectives of these three tiers of research activity and support. One advance might be to establish and publicise the level of funding for each of these categories. There is also a need for an improved understanding of the form and extent of synergy between the three types of research activities.

4. The findings of this report cast some doubt on the use of performance indicators to evaluate research. While they may be appropriate , for example, for comparisons of discipline performance over a significant time period, (Bourke, 1994) the evidence of this study suggests that the scales may be too compressed to adequately distinguish between good quality, international standard research, and leading edge, world-ranked research.

References

Adey P. & Larkins F. 1993, 'Productivity Benefits from Research Concentration - Myth or Reality?', *Campus Review*, June 23-29.

Anon, 'Research Size: A Reply', Campus Review, July 7-13, 1994.

ASTEC 1989, *Profile of Australian Science*, Canberra, AGPS.

Becher B. and Kogan M. 1987, *Calling Britain's Universities to Account*, London, Education Reform Group.

Harris, G.T. 1989, 'Research Output in Australian University Research Centres in Economics', *Higher Education*, Vol.18, p. 397.

Lazenby 1992, Review of the Special Research Centres Program and the Key Centres of Teaching and Research Program: Advice of the NBEET Australian Research Council.

Lowe I. 1993, Campus Review, Oct 7-13.

Lowe I. 1991, 'Science Policy for the Future' in R. Haynes (ed), High Tech: High Cost?, Pan Macmillan, Sydney pp 177-191

Merton, Robert 1973, *The Sociology of Science*, University of Chicago Press.

p. 273, p. 416.

NBEET 1993, 'The Effects of Resource Concentration on Research Performance', Canberra, AGPS, Report No. 25,

NBEET 1994, 'Quantitative Indicators of Australian Academic Research', Canberra, AGPS, Report No. 27.

Pettit 1993, Research Performance Indicators Survey', NBEET, Canberra, AGPS, Report No. 21.

Ziman, J. 1987, *Science in a 'Steady State'*, London, Science Policy Support Group Concept Paper No.1.