Higher Education Management and Policy
Volume 23/3

Critical masses for academic research groups and consequences for higher education research policy and management

by

Ralph Kenna and Bertrand Berche
Coventry University, United Kingdom and University of Nancy, France

Smaller universities may produce research which is on a par with larger, elite establishments. This is confirmed by a recently developed mathematical model, supported by data from British and French higher education research-evaluation exercises. The detailed nature of the UK system, in particular, allows quantification of the notion of critical mass in research. It is shown that research quality increases with group quantity, but only up to a limiting size referred to as the upper critical mass. The condition for smaller universities to produce top-quality research is that they contain research groups of sizes above the upper critical masses appropriate to their respective disciplines. Policies which concentrate support into progressively fewer, larger institutions are therefore unjustified for high-quality academic research. Instead, to amplify overall research strength, support for medium-sized groups should be prioritised to help them attain upper critical mass.

Higher Education Management and Policy
Volume 23/3

Masses critiques pour les groupes de recherche académiques et conséquences pour la politique sur la recherche et la gestion de l'enseignement supérieur

par
Ralph Kenna et Bertrand Berche
Université de Coventry, Royaume-Uni et Université de Nancy, France

Les plus petites universités peuvent mener des recherches dont les résultats rivalisent avec ceux des centres d'excellence. Ceci a été confirmé par un modèle mathématique récemment développé, dont les données proviennent d'exercices d'évaluation et de recherche de l'enseignement supérieur français et britannique. La nature du système particulièrement détaillé au Royaume-Uni permet de quantifier la notion de masse critique dans la recherche. Il est démontré que la qualité de la recherche augmente proportionnellement à la dimension du groupe, mais seulement jusqu'à une taille limite appelée la masse critique supérieure. La condition permettant aux plus petites universités de mener des recherches d'excellente qualité est d'inclure des groupes de recherche dont la taille dépasse la masse critique supérieure de manière appropriée dans chacune de leurs disciplines respectives. Les politiques, qui concentrent de moins en moins leur soutien aux grands établissements, sont donc injustifiées pour ce qui est de la recherche académique de premier ordre. Au lieu de cela, donner priorité aux groupes de taille moyenne afin de leur permettre d'atteindre un niveau de masse critique supérieur permettrait de renforcer les points forts de la recherche.

Introduction

The notion of critical mass in academic research groups is an old one which, despite lacking clear definition, feeds into policy and management decisions at department, institution and government levels. The traditional notion is that there is a minimum, threshold size which a research group or department must reach in order to produce high-quality research (Harrison, 2009; Evidence, 2010). The extended notion that benefit accrues through increasing scale lies behind recent calls for a greater concentration of resources into fewer, elite universities (*e.g.* Russell Group, 2010; 1994 Group, 2010). However, recent empirical findings by Adams and Gurney (2010) provided evidence that this is not, in fact, the case. Their citation count analysis showed that the impact of smaller research-intensive universities' publications is mostly on a par with that of larger, elite ones. The purpose of this paper is to explain why this is the case.

The explanation originates from a new notion of critical mass in research. In our earlier papers (Kenna and Berche, 2010, 2011) we showed that old notions of critical mass and related or extended concepts are flawed and in need of refinement; we also proposed new, testable and robust definitions. Taking the theory a step further, we now apply it to different types of UK universities to explain the Adams-Gurney findings. In addition, this paper focuses on implications for higher education management and policy.

Peer review assessments of group or departmental quality offer opportunities to test concepts around critical mass, including the relationship between quality and quantity in research. National research evaluation systems such as the UK's Research Assessment Exercise (RAE) and the French equivalent, which is performed by the Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES) form bases on which governments, funding councils, higher education institutes and research groups formulate policies on where to focus investment. In this paper, we compare recent RAE and AERES evaluations. While they are compatible, the UK exercise is more detailed than the French one. This level of detail allows the notion of critical mass in research to be refined and gives rise to a better understanding of how the quality of group research depends on quantity.

Implicit in the notion of critical mass is the idea that interactions between individual researchers within a group are important. But the question as to how these interactions feed into emergent group quality has not been addressed until now. It is seen that such interactions lead to a linear increase of research quality with group quantity, but only up to a limiting size, beyond which meaningful collaborations between all group members cannot be sustained. Beyond this limit – denoted as the **upper critical mass** – research groups tend to fragment; increased concentration of resources does not lead to continued significant increases in research quality. This explains recent findings which showed that smaller universities' research can combine excellence with impact on a par to that of larger, elite establishments (Adams and Gurney, 2010): the condition for doing so is that they both contain research groups whose sizes tend to exceed the upper critical masses. This is the main policy-related message of this study.

This paper thus extends our earlier work (Kenna and Berche, 2010, 2011) to specific areas of concern to management and policy makers in higher education. In order to render it complete, in the next section we offer a condensed, non-technical summary of the theory. Some overlap with our earlier papers is therefore necessary and the reader is referred there for more technical expositions of the mathematics and statistical analyses. In the subsequent sections we describe research quality assessment schemes in the United Kingdom and France and give more detail of the UK higher education sector. After a discussion of causality and caveats we offer an explanation for the Adams-Gurney (2010) findings; conclusions for higher education management and policy are drawn in the final section.

The dependency of research quality on group quantity

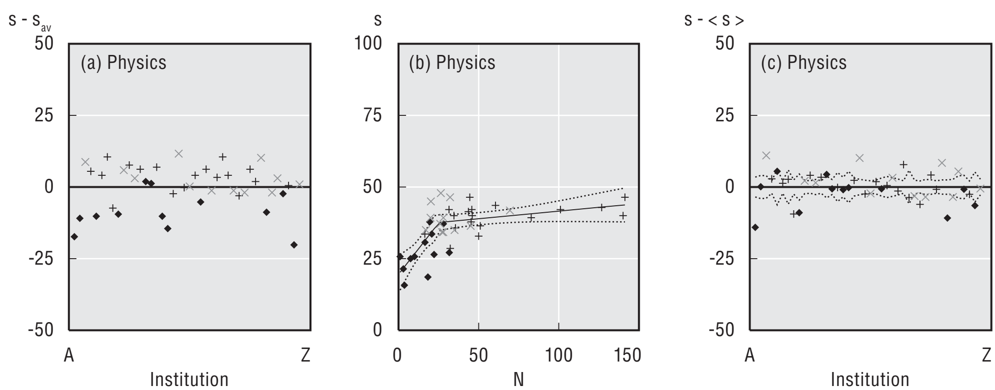
There are two viewpoints in the current debate on the nature of the relationship between research quality and group quantity (Harrison, 2009). On the one hand, there are arguments to the effect that bigger tends to be better, that benefit accrues through scale and that research support should focus on a few, elite higher education institutes which already have significant resources (Russell Group, 2010; 1994 Group, 2010). The second viewpoint is that quality and quantity are independent and that resources should be allocated to wherever research excellence is found (University Alliance, 2009; Million+, 2010). Harrison (2009) gives a qualitative discussion of the ongoing debate between the two sides of the argument in the United Kingdom.

In France, the "Opération Campus" provoked a similar debate (Ministère de l'enseignement supérieur et de la recherche, 2011). This was introduced by the government with a view to identifying specific universities into which funding would be channelled in order to enhance research and teaching infrastructure, as well as accommodation. The agenda continued with so-called "excellence initiatives" in which between five to ten facilities in the country were earmarked for greater public research funding (Agence nationale de la recherche, 2011). Both programmes have been criticised by researchers who were largely in favour of a more even distribution of public investment (Sauvons la Recherche, 2011). Thus the issue of how quality (of research) is related to quantity (of investment) is one of immediate and simultaneous concern in different countries.

A new, agent-based mathematical theory was introduced by Kenna and Berche (2010) and subjected to statistically rigorously testing (2011). This theory addresses the relationship between research quality and group quantity. It indicates that quality is indeed strongly dependent on quantity, but only up to a limit termed the upper critical mass. Beyond this limit, the relationship between quality and quantity becomes less significant or even disappears. Thus the mathematical theory offers support for aspects of both viewpoints, but each is only valid in a certain size regime. The basic reasoning behind this theory is outlined below; see Kenna and Berche (2010) for a detailed, mathematical exposition.

Ceteris paribus, it may be argued that a research group or department containing, say, ten active scientists may, on average, be expected to produce twice as many papers, train twice as many PhD students and generate twice as much income as a group of five. According to this expectation, the **strength** of a research group is approximately proportional to the number of individuals it contains. If one defines the **quality** of a research group as the average strength per head, then quality is simply given by the mean calibre of the individuals in the group. According to this argumentation, different groups would have qualities distributed around an average. Figure 1(a) illustrates such a distribution in the

Figure 1. **Quality analysis for physics research groups**



Note: (a) RAE quality scores, normalised to the overall discipline average for physics groups which are listed alphabetically, from Aberystwyth University to the University of the West of Scotland. The symbols + and × correspond to universities from the Russell and 1994 groups of large and smaller research-intensive universities, respectively, while the ◆ represent other higher-education establishments in the UK. (b) The same data plotted against the sizes *N* of research groups where the solid curve is the fit coming from the model and the dashed curves are the corresponding 95% confidence intervals. (c) Quality is renormalised to the expectation values < s > coming from the model.

case of physics groups submitted to the RAE. In the plot, the difference between group quality s and the mean quality s_{av} is compared against the institutions listed alphabetically. (The precise manner in which the quality measures s are determined and the meaning of the various symbols in the figure are explained below.) The extent to which a group's research quality is higher or lower than average is a function of many parameters such as the innate calibre, education and experience of its members, the quality of managerial support and of facilities and infrastructure, the impact of external and international collaborations, the prestige and confidence inspired by the institution and by past successes, etc. All of these parameters contribute to the quality of an individual group but, in general, research quality is size independent according to this argument. Figure 1(a) gives no indication as to the best policy to maximise the quality of a research group apart from simply recruiting members of high individual calibre in order to increase the strength of the group.

However, it would be naive to retain this viewpoint alone because it neglects the importance of interactions between group members. This is despite the concept of such interactions being vaguely linked to that of "critical mass" as advocated by proponents of this viewpoint. In this context, critical mass has been loosely described as some kind of threshold group size for effective performance, although no quantitative evidence for the existence of such a threshold has ever been presented (Evidence, 2010). In Figure 1(b), quality s is plotted against quantity N for physics. Also, there is no evidence here of a minimum group size above which there is a stepwise increase in research quality. On the contrary, group quality tends to rise continuously with quantity until a certain critical size is reached at about $N \approx 25$. Beyond this size, research quality appears to be independent of group quantity. What is the reason for this behaviour?

According to the theory developed in Kenna and Berche (2010, 2011), if there are N members of a group, the total number of pairs of individuals along which channels of communication can open is $N(N - 1)/2$. The total **strength** of the research group is then quadratically dependent on its size. The **quality** of the group (its strength per head) is therefore a linear function of group size. This explains the steady rise of research quality with group quantity in the left part of Figure 1(b), up to about $N \approx 25$. These considerations assume that a significant proportion of all potential two-way communication links are active. Beyond a certain group size, however, it is not possible for every individual to meaningfully interact with everyone else. Suppose that the average maximum number of colleagues with whom a given individual can communicate is N_c. It follows that when the actual group size N exceeds this number, the total number of interaction links is $NN_c/2$. In this case, the strength of the group is only linearly dependent on group size, so that the quality of the group is essentially constant. Such large groups can fragment into subgroups whose average size is less than N_c. There can also be connections between subgroups. Such connections give rise to a weaker quadratic dependency of strength on quantity and therefore to another linear relationship between quality and quantity.

To summarise, a simple model, developed by Kenna and Berche (2010, 2011) taking into account the importance of communication links between individuals and between sub-groups, leads to a piecewise linear dependency of group quality on quantity. Mathematically, this is expressed as follows:

$$< s > = a_1 + b_1 N \text{ if } N \leq N_c \qquad \text{Equation 1(a)}$$

or

$$< s > = a_2 + b_2 N \text{ if } N \leq N_c \qquad \text{Equation 1(b)}$$

where N represents the size or quantity of the research group, $< s >$ represents its expected research quality and the coefficients a_1, b_1, a_2 and b_2 are characteristic of the academic discipline under consideration. The terms a_1 and a_2 correspond to the average individual strengths while b_1 and b_2 measure strengths of interactions. These parameters can be estimated using least squares statistical approaches.

In Figure 1(c), the difference between **actual** and **expected** quality is plotted against the size of physics groups at RAE. The data is more closely bunched than the equivalent display in Figure 1(a). In fact, the standard deviation associated with Figure 1(c) is about 5, significantly less than that associated with Figure 1(a), where it is 8. This illustrates the superiority of the model (1) over the alternative notion that quality is independent of quantity.

Interestingly, then, while group quality increases in step with the calibre of individuals, this is not the dominant mechanism for enhancing quality. In fact, it is an order of magnitude smaller than the collaborative effect. The limit N_c, where the pattern of communications within the group changes on average, is referred to as the **upper critical mass**. This is currently the only quantitative definition of critical mass in the literature. In Kenna and Berche (2011), statistical hypothesis testing was applied to data from the United Kingdom's RAE to establish the validity of the model and critical masses were determined for a multitude of academic research disciplines. For completeness of the current exposition, the resulting critical masses are listed in Table 1.

The model allows for a definition of another critical mass in research (Kenna and Berche, 2010). The expected strength of a research group is determined by the product of its quality and quantity $< s > N$. The gradient of this strength function for $N < N_c$ exceeds that for $N > N_c$ provided $N < N_k$ where N_k is:

$$N_k = N_c/2 \qquad \text{Equation 2}$$

Table 1. **Estimates for the upper critical masses for a variety of academic disciplines, as illustrated in Kenna and Berche (2011)**

Research discipline	Upper critical mass N_c	Research discipline	Upper critical mass N_c
Applied mathematics	13 ± 2	Pure mathematics	≤ 4
Physics	25 ± 5	Biology	21 ± 4
Geography/environment	30 ± 3	Chemistry	36 ± 13
Medical sciences	41 ± 8	Nursing	18 ± 5
English (in the UK)	32 ± 3	Foreign languages	6 ± 1
Agricultural sciences	10 ± 3	Architecture/planning	14 ± 3
Law	31 ± 4	Education	29 ± 5
Economics/econometrics	11 ± 3	Business/management	48 ± 8
Politics/international	25 ± 5	Sociology	14 ± 4
History	25 ± 5	Philosophy/theology	19 ± 3
Art and design	25 ± 8	History of art, performing arts, communication and music	9 ± 2
Archaeology	17 ± 3		

This means that to maximize the strength of a given discipline, it is best to support groups which are smaller in size than N_c in order to drive them towards the upper critical mass. However, this policy only makes sense if these groups are not too small. In fact, defining a research group of size N to be:

small, if$N \leq N_k$

medium, if$N_k \leq N \leq N_c$

large, if$N \geq N_c$

then, in order to maximise the overall strength of a research discipline, it is best to prioritise support for medium-sized groups. Small groups, on the other hand, must strive to achieve the lower critical mass to survive.

While the above perspectives stem from the point of view of a group or an overall discipline, it is interesting to compare them with the viewpoint of an individual researcher joining such a research group. Clearly, from the individual's point of view, it is optimal to join a group which maximises one's opportunity to interact, *i.e.* to join a large group. In other words, it is best for an individual to join a large group whereas it is best for the discipline as a whole to boost medium-sized groups (to make them large).

The above theory, which was placed on a mathematical footing by Kenna and Berche (2010), is the only agent-based, quantitative theory in the literature explaining the relationship between quality and quantity in research. The theory is scientific, in the sense that it is falsifiable (Popper, 1959): a hypothesis is proposed which can be tested empirically. Our hypothesis is that quantity drives quality through communication links. While other factors are undoubtedly also at work, in the spirit of Occam's razor,[1] this is the primary cause-and-effect mechanism. Our empirical testing is based on the statistical analysis performed in Kenna and Berche (2011); this in turn supports the theoretical relationship between quality and quantity across a range of disciplines which have many different work methods. This universality also indicates a commonality of causal mechanism, largely independent of discipline. We have identified this causality as quantity-driving quality, which means that the theory is not only descriptive, but is also explanatory. To better understand the predictive nature of the theory, the causality arrow, and to draw specific policy and managerial conclusions for research at institutes of higher education, the theory needs to be confronted with empirical reality. National research quality assessments provide a suitable arena in which to do this. We address these in the following sections.

Research quality assessment in France and the United Kingdom

The United Kingdom's most recent RAE (2008) is considered to be the most precise evaluation of its kind to date. It was not based on citation counts, but on peer review. This is because bibliometric analyses are considered to measure **impact** rather that quality (Raan, 2005) and the one is not a proxy for the other, even for disciplines with good coverage in the Web of Science (Harnad, 2008, 2009; Mahdi *et al.*, 2008; Oppenheim and Summers, 2008; Raan, 2006; Evidence, 2009; and references therein). Another important principle underpinning the RAE is that it concerns the research quality of whole units or research groups that are put forward for assessment; it is not about individuals.

For the RAE, three aspects of group quality were considered: research outputs, research environment and research esteem. Research outputs mostly signify publications, but for some disciplines software, patents, artefacts, performances or exhibitions may be considered.

Arguments in favour of concentrating research resources go beyond the issue of quality of research outputs. For example, on the whole, larger groups and larger universities have access to more facilities and infrastructures, some of which may require a minimum level of usage to justify their purchase. The availability of expensive equipment and proximity of related groups which may share the cost burden facilitate the diversity of research direction. Such aspects of research environment were also measured at RAE and contributed to groups' final scores (RAE, 2008). Institutions were asked to provide information on research income and other funding supporting the group, on research infrastructure, the vitality of active seminar programmes, hosting workshops and conferences and on visiting researchers. Leadership of research, arrangements for supporting collaborative and interdisciplinary research, management, training and supervision of – and facilities for – postgraduate research students were also assessed. The appropriateness of accommodation and library facilities, for example, as well as hardware and software were also evaluated, as was the quality of relationships with research users. Numbers of studentships and research-degree completion rates were considered. The RAE evaluators also considered the breadth of research activity within departments and the success of funded projects in generating additional grants and activities. The extent to which researchers – at all stages of their careers – were nurtured, arrangements for research leave, coherence and research culture were all scrutinised. Departments were required to provide clear evidence in support of all claims regarding the research environment.

Finally, visibility is also scale dependent. This was considered in the RAE through the "esteem" component. Esteem indicators included prizes, awards, fellowships, honours, major research grants, conference organisation, keynote addresses and editorial roles. Significant professional services, including to industry, were also assessed. Thus, through the environment and esteem components of RAE, efficiencies of scale and visibility were components which contributed to the overall measurements of group research quality.

The precise manner in which research outputs, research environment and research esteem fed into the overall final RAE score was dependent upon discipline. For example, in pure and applied mathematics, statistics and the computer sciences these three components were weighted at 70%, 20% and 10% respectively, while in biology, pre-clinical and human biological sciences, agriculture, veterinary and food sciences, outputs, environment and esteem were weighted at 75%, 20% and 5% respectively.

Experts in various research fields scrutinised these factors in research groups in order to determine the overall proportions of quality which fell into five levels. These levels are defined as 4* (research which is considered world-leading), 3* (internationally excellent), 2* (internationally recognised), 1* (nationally recognised) and unclassified. A formula based on the resulting quality profiles was then used to determine how research funding is distributed to each higher education institute which took part. Although Scotland used a more complicated funding formula, the one used by the funding councils for England, Wales and Northern Ireland associated each rank with a weight in such a way that 4* and 3* research received, respectively, seven and three times the amount of funding allocated to 2* research. Research ranked at, or below, 1* attracts no funding. This formula may therefore be used to give a continuous measure of the quality s of a research group. Denoting the percentage of research which was evaluated as n^* by p_{n^*} we define the quality of the group by $s = p_{4^*} + 3p_{3^*}/7 + p_{2^*}/7$. In this way, the maximum theoretical possible quality score is 100, corresponding to entirely 4* research. (In fact no group achieved this score, with the top groups achieving about half this.)

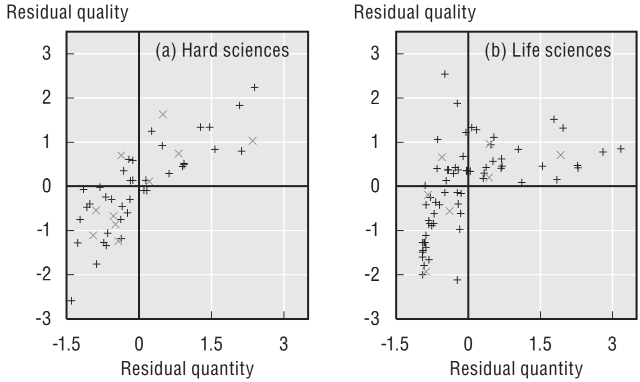
In the AERES evaluation system France is geographically divided into four parts, one part being evaluated each year. In the 2008 evaluation, a method was used which is considered more precise than that used in previous years and this facilitates comparison with the UK approach. However, since only 41 institutions were evaluated, of which 10 were traditional universities, the amount of data available for the French system is lower than for the UK equivalent. Furthermore, only a global mark is attributed to cumulated research groupings which can include several groups with heterogeneous levels. Consequently, the fine-grained analysis at the level of the research groups or departments is lost. This is clearly a weak point compared to the British system of evaluation and the AERES intends to change it in the near future. Nonetheless, in order to make the comparison with the British system, we translate the AERES grades A+, A, B, C into 4*, 3*, 2* and 1*.

It is useful to compare the French AERES system to the UK RAE for hard sciences and life sciences, although the French data refer to results on a broader institutional scale and fewer data are available. In the case of the hard sciences, we checked that the integrated UK data, obtained by merging pure and applied mathematics, statistics, physics, earth sciences, chemistry, computer science and informatics, yield results which overlap with their French counterparts. The comparison is properly made between the normalised residuals for the integrated quality and quantity measures. Quantity N is normalised to $(N - N_{av})/\sigma$, where N_{av} is the mean size and σ the corresponding standard deviation. A similar transformation normalises the quality residuals. In order to compare the life sciences, UK data for the biological sciences, pre-clinical and human biology have been integrated with agriculture, veterinary and food sciences. The resulting plots for the hard and life sciences are depicted in Figures 2(a) and 2(b) respectively. A convincing degree of overlap is evident for both plots, indicating compatibility between the British and French systems. Having established this, and having demonstrated the finer detail available in the British system, the next step is to analyse the latter at the detailed level of individual disciplines in order to investigate the model from a policy and managerial perspective. To do this, we first need to explain the British higher education research structure in more detail.

The United Kingdom's higher education research system

The UK's higher education academic research base is organised into a number of representation groups[2] (see, for example, Newman [2009] for an overview). The most famous of these is the Russell Group of research-intensive, leading universities, most of which have medical schools. It was established in 1994, two years after the Further and Higher Education Act, which conferred university status on former polytechnics, central institutions or colleges of

Figure 2. **Comparison of the French AERES system with the British RAE**



Note: Normalised quality residuals plotted against normalised quantity residuals for France's AERES and the UK's RAE for (a) the hard sciences and (b) the life sciences. The French data correspond to the symbols × and the integrated British data to +.

higher education. The Russell Group has 20 members and aims to help ensure that "resources are directed to where a critical mass of research can be undertaken to ensure the highest international standards with the greatest impact" (Russell Group, 2011). Russell Group universities are sometimes considered the UK equivalent to the US' Ivy League and its members receive approximately two-thirds of all research funding in the United Kingdom.

The 1994 Group represents 19 of the United Kingdom's top smaller research-intensive universities. It was founded in 1994 to defend the interests of its members following the establishment of the Russell Group by larger research-intensive universities earlier that same year. The Group "provides a central vehicle to help members promote their common interests in higher education, respond efficiently to key policy issues, and share best methods and practice. Each member undertakes diverse and high-quality research, while ensuring excellent levels of teaching and student experience" (1994 Group, 2011). Besides these, the remaining UK higher education institutes are represented by University Alliance, the Million+ Group and Guild HE, or are unaffiliated.

A recent bibliometric-based study has found that, league tables of university rankings notwithstanding (Salmi and Saroyan, 2007), there is little difference between the research quality of the Russell Group and the 1994 Group (Adams and Gurney, 2010). Confirmation of this observation undermines arguments that benefit continues to accrue through increase of scale. For if a continual policy of concentration of funding were indeed to lead to better quality research, one would expect that the research quality of the Russell Group of larger universities with their larger research groups, would be significantly superior to that of the 1994 Group, which typically contains smaller universities with smaller groups. The main objective here is to explain this finding and draw policy conclusions from the mathematical model underlying this explanation. First, we examine the validity of the model.

Causality and caveats

Based on the model described by Kenna and Berche (2010, 2011), we maintain that the principle causal link between quality and quantity in group research is that the latter drives the former through communication channels. While the opposite causal arrow (quality driving quantity) is, no doubt, also an important factor, we believe that this is a secondary feedback mechanism and is not the dominant driver. Besides the support that the statistical analysis lends to the mathematical model (Kenna and Berche, 2011), evidence for this assertion comes from the existence of the breakpoint N_c and the Ringelmann effect (Ringelmann, 1913). The latter is a sociological phenomenon whereby productivity increases with size, but the gain reduces with each new member added to a group. The Ringelmann effect in this context is a co-ordination transition which occurs when the group size exceeds the upper critical mass of the average number of individuals with whom one can meaningfully communicate. Thus the Ringelmann effect here is a "breaking" mechanism, countering the otherwise linear rise of quality with quantity. If the causal arrow were in the opposite direction (quality driving quantity), there would be no communication-induced Ringelmann transition and no breakpoint N_c. In this case, quality would rise with quality until the maximum possible score $s = 100$ is reached. No group at RAE achieved this theoretical maximum.

On the other hand, one may argue that one advantage of scale is that it allows the support of early-stage researchers, whose output may be mediocre, within an environment of excellence. Indeed, this is important for nurturing and developing new talent. In this respect, it is unlikely that all members of a group are excellent and this could be another explanation for the Ringelmann phenomenon. However, at RAE, the "group" is defined by the submitting university. Since the stakes are high, and the quality-related post-evaluation income depends upon the quality profile attained at RAE, it is tactical for universities not to submit mediocre research. Typically, younger researchers (PhD students and postdoctoral researchers) contribute to the research environment element of RAE rather than directly to the research outputs or esteem aspects. This means that the groups submitted are actually smaller than the total number of research-active personnel in a given department and the group sizes discussed here refer to the number of established researchers.

With these caveats in mind, we proceed to offer an explanation for the similarity of research quality between large and smaller research-intensive universities, as reported by Adams and Gurney (2010).

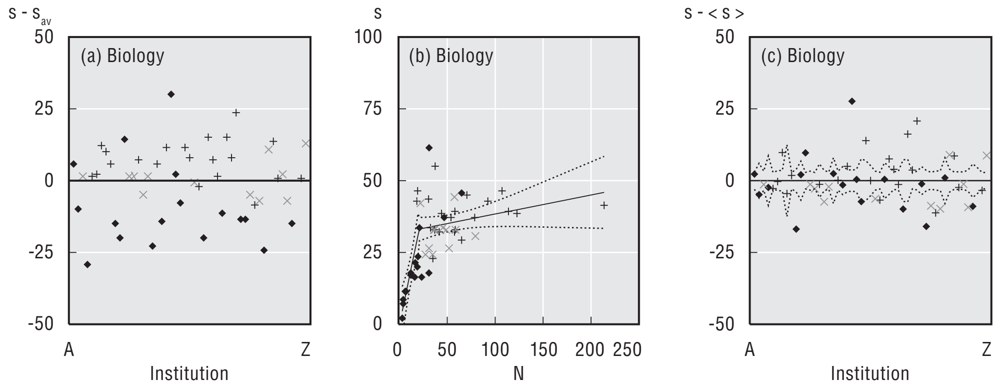
The dependency of research quality on quantity above and below the upper critical mass

This section provides an explanation for the similarity of research quality between the Russell and 1994 groups; it is useful to return to Figure 1(a), which relates to physics, as it offers two notable observations. The first is that the Russell (+) and 1994 (×) groups outperform the others (◆) in terms of their quality of research. One argument as to why this is the case is that benefit accrues through increasing scale and that increased concentration of resources leads to increased research strength. If this were the case, however, one would expect the Russell Group to outperform the 1994 Group. Indeed, the mean quantity of researchers in physics groups in a Russell Group university is $N_{av} = 63$. This is about twice the mean size for physics groups in the 1994 Group of universities, which have $N_{av} = 31$. However, the mean quality score for the Russell Group is $s_{av} = 40 \pm 5$ and that for the 1994 Group is $s_{av} = 39 \pm 8$. The proximity of the mean quality scores indicates that there is little difference, on average, between the two sets of universities, despite their differing sizes. This is consistent with the findings of Adams and Gurney (2010). This is the second observation associated with Figure 1(a) and it requires an explanation.

The explanation is found in Figure 1(b) where the same data are plotted against the size of the research groups. The solid line is the fit to the model (1) and the dashed curves represent the resulting 95% confidence intervals. The reason for the comparable qualities of the Russell Group and the 1994 Group and their superiority over other universities is clear now: although the average Russell Group research body is double the size of its equivalent in the 1994 Group, they both exceed $N_c \approx 25$ and therefore lie on the plateau where the concentration of more staff into these groups does not significantly increase overall average research quality. The physics groups of the other universities lie mainly to the left of the figure: these small and medium-sized groups have linear dependency of quality on quantity.

Similar analyses may be performed for other research areas. For example, the average size of biology research groups (Figure 3) in the Russell and 1994 groups are $N_{av} = 66$ and 44, respectively. Without the existence of the upper critical mass breakpoint, one would expect a proportionate superiority of the Russell Group over the 1994 Group in terms of quality. However, the corresponding mean quality scores are $s_{av} = 39 \pm 8$ and 32 ± 7, respectively. The reason for their proximity is that both sets of universities tend to have research groups in excess of the upper critical mass which is $N_c = 21 \pm 4$ for this discipline (Figure 3b). Similar to Figure 1 which relates to physics, the validity of the model is illustrated by the reduction of the standard deviation from 12.5 in Figure 3(a) to 8.5 in Figure 3(c).

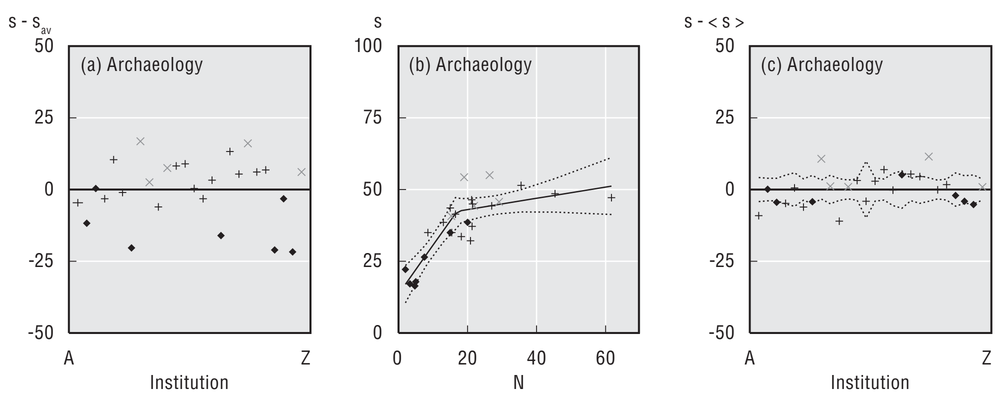
Figure 3. **Quality analysis for biology**



Note: Quality analysis for biology, analogous to Figure 1. The average size of research teams in the Russell and 1994 groups are N_{av} = 66 and 44, respectively, exceeding the upper critical mass which is N_c = 21 ± 4. The corresponding mean quality scores are similar at s_{av} = 39 ± 8 and 32 ± 7, respectively.

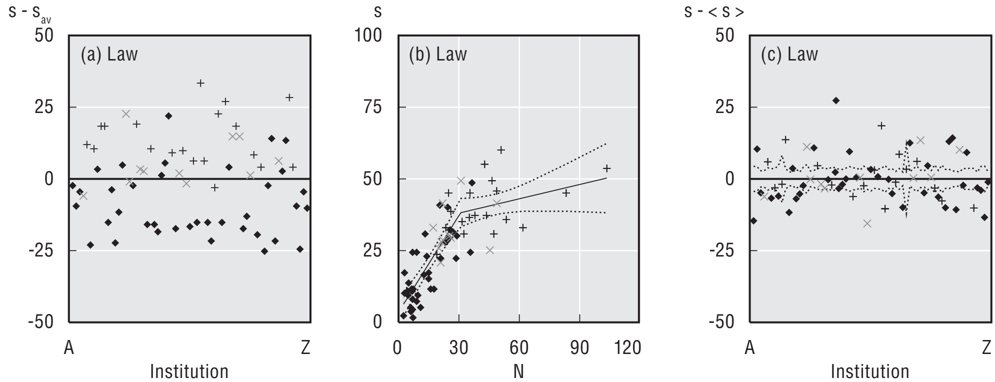
Figure 4 contains analogous plots for archaeology. The Russell and 1994 groups' archaeology teams mostly exceed the upper critical mass N_c = 17 ± 3. This explains the similarity in their average quality scores, which are s_{av} = 41 ± 7 and 48 ± 9, respectively. Figures 5-9 contain similar plots for law, education, applied mathematics and history, as well as architecture and town and country planning. In each case, the comparable levels of research quality associated with the large Russell and 1994 groups may be explained by the existence of the upper critical point and the reduction of the dependency of quality on quantity for $N > N_c$.

Figure 4. **Quality analysis for archaeology**



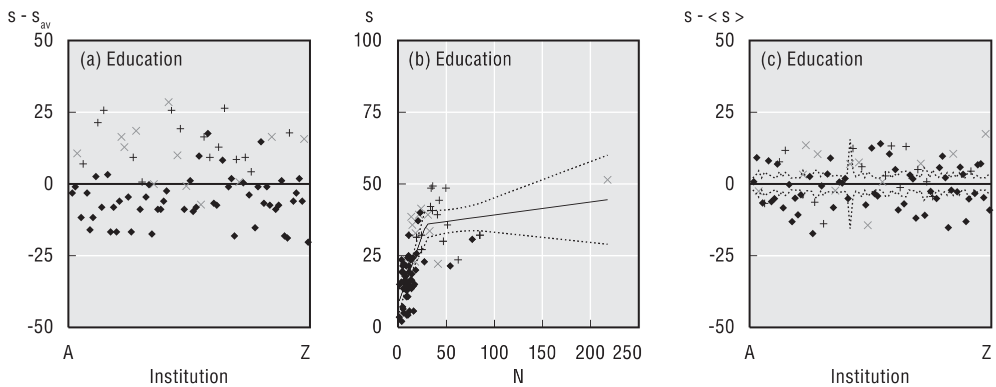
Note: Quality analysis for archaeology with upper critical mass N_c = 17 ± 3. The Russell and 1994 groups archaeology teams mostly exceed this value and their quality scores at s_{av} = 41 ± 7 and 48 ± 9 are hence comparable.

Figure 5. **Quality analysis for law**



Note: Quality analysis for law, analogous to Figure 1. The mean quality scores for the Russell and 1994 groups are s_{av} = 40 ± 10 and 32 ± 9, respectively. Their mean sizes N_{av} = 45 and 27 are at, or above, the upper critical mass N_c = 31 ± 4.

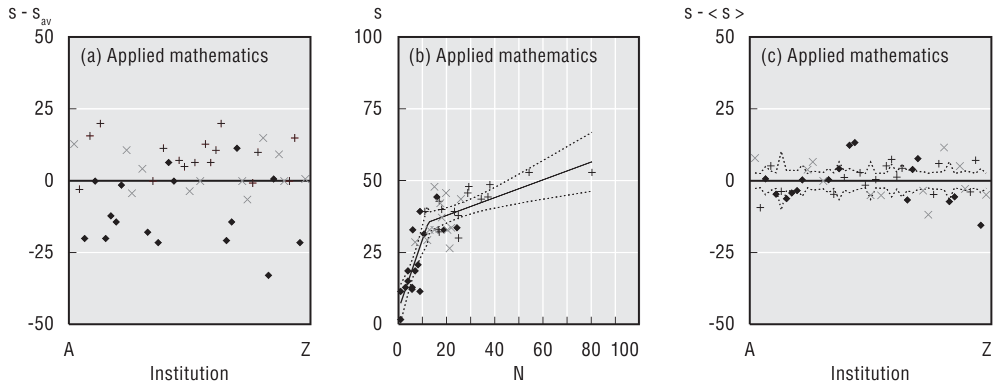
Figure 6. **Quality analysis for education**



Note: Quality analysis for education, analogous to Figure 1. The Russell and 1994 groups' research teams mostly exceed the upper critical mass N_c = 29 ± 5 for the discipline leading to comparable average quality scores are s_{av} = 37 ± 9 and 33 ± 11, respectively.

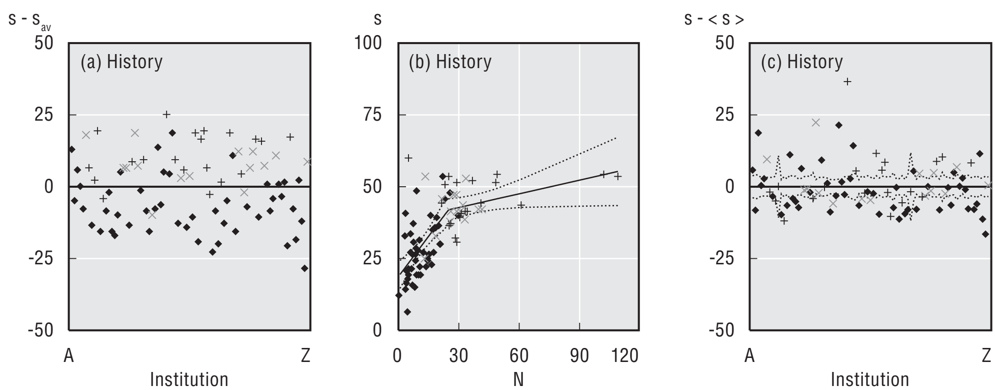
To summarise, the explanation for the superiority of the Russell Group and 1994 Group universities over the others lies in the different interaction patterns of their large groups compared to small and medium-sized groups. Individuals in large groups can interact with a maximal number of collaborators, while such interactions are limited in small and medium-sized groups. However, since one cannot sustain an indefinite number of interactions with ever-increasing numbers of colleagues, once the group grows larger than the upper critical mass, communication links are saturated and the addition of more researchers does not significantly increase research quality. Thus the

Figure 7. **Quality analysis for applied mathematics**



Note: Research quality analysis for applied mathematics. Again, although the Russell Group teams (whose mean size is N_{av} = 30) are on average twice as large as the 1994 Group teams (N_{av} = 17), their quality scores are similar, with s_{av} = 41 ± 8 and s_{av} = 36 ± 8 respectively. Again this is explained by the fact that these groups tend to exceed the upper critical mass, which is N_c = 13 ± 2 for the discipline. Other research teams are mostly small or medium sized and therefore tend to have lower quality.

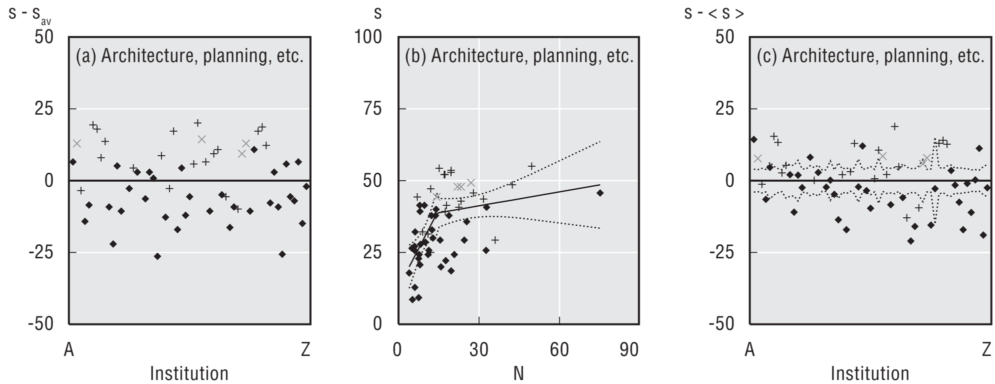
Figure 8. **Quality analysis for history**



Note: Quality analysis for history analogous to Figure 1. Again, the comparability of the average quality scores for the Russell (s_{av} = 46 ± 9) and 1994 (s_{av} = 42 ± 8) groups is attributable to them tending to have sufficiently large research groups: their average sizes N_{av} = 39 and 28 exceed the discipline's upper critical mass N_c = 25 ± 5.

Russell and 1994 groups have comparable research standards. This explains Adams and Gurney's (2010) findings, as discussed in the Introduction. A policy supporting medium-sized groups to drive them towards the upper critical mass is therefore expected to enhance the quality of research in the sector overall.

Figure 9. **Quality analysis for architecture and town and country planning**



Note: Quality analysis for architecture and town and country planning analogous to Figure 1. The average team size for the Russell and 1994 groups are N_{av} = 21 and 22, respectively, in excess of the upper critical mass N_c = 14 ± 3. At s_{av} = 44 ± 10 and s_{av} = 47 ± 3, their mean qualities are also comparable.

Conclusions

The primary aim of this paper is to explain why the research quality of certain smaller universities may be on a par with that of larger, elite establishments. Evidence that this is indeed the case supports earlier observations of parity of research impact by Adams and Gurney (2010). An explanation for this phenomenon is important for policy and management in higher education and the explanation offered herein is based on a new understanding of the notion of critical mass in research. Kenna and Berche (2010, 2011) argue that the former notion of critical mass as a minimum, threshold number of researchers required for a stepwise increment in research quality is no longer valid. Instead, research quality tends to rise continuously with group quantity but only up to a limit termed the upper critical mass. This is interpreted as the average maximum number of colleagues with whom a given individual in a research group can meaningfully interact. Once the group exceeds this size, it tends to fragment into sub-groups and research quality no longer improves significantly if the size is increased. There is also a lower critical mass, which small research groups should strive to achieve for the sake of stability. The existence and properties of the upper critical mass explain why the research quality of the Russell Group of large research-intensive universities is not significantly better than that of the 1994 Group. Both tend to contain research groups which are large, in the sense that they are above the discipline-dependent upper critical masses in terms of size.

While the research quality of the Russell and 1994 groups are on a par, they are both significantly above that of other universities in the UK's higher education sector. This is because other institutes tend to have small- or medium-sized research groups. According to this analysis, the best way to amplify the strength of the sector is to support medium-sized groups to drive them towards the upper critical mass. This is the principle policy-orientated message of this paper.

The bulk of the analysis contained herein refers to the UK higher education sector, as its Research Assessment Exercise provides the most suitable data for analysis. However, by integrating the RAE data and comparing it to the equivalent French system, it is clear that the two are compatible. For this reason, it is expected that the lessons drawn may also have practical managerial and policy implications outside of the United Kingdom.

The authors:

Dr. Ralph Kenna
Deputy Director
Applied Mathematics Research Centre
Coventry University
Priory Street
Coventry CV1 5FB
United Kingdom
E-mail: r.kenna@coventry.ac.uk

Professor Bertrand Berche
Head of Department of Physics
Statistical Physics Group
Institut Jean Lamour (Laboratoire associé au CNRS UMR 7198)
CNRS – Nancy Université – UPVM
F – 54506 Vandœuvre-lès-Nancy Cedex
France
E-mail: Bertrand.Berche@ijl.nancy-universite.fr

This research is supported by the EU programme FP7-People-2010-IRSES (Project Number 269139). In addition, RK is grateful for an invited Professorship at the University of Nancy.

Notes

1. Occam's razor is a principle attributed to the 14th century logician William of Ockham, according to which "entities should not be multiplied unnecessarily". It suggests that we should tend towards simpler theories until we can trade some simplicity for increased explanatory power.

2. We distinguish between research groups and representation groups: the former are the collectives of research staff included in the submission to the United Kingdom's RAE. These collectives are not always identical to administrative departments within universities, but they are assumed to be coherent groups of individuals for research-evaluation purposes. We refer to them in lower case (*e.g.* "physics group"). Representation groups, on the other hand, are formal collectives of universities as described here. We normally capitalise the latter (*e.g.* "Russell Group").

References

1994 Group (2010), "Enterprising Universities: Using the Research Base to Add Value to Business", *Policy Report*, *www.1994group.ac.uk/publications*, accessed 9 October 2011.

1994 Group (2011), *www.1994group.ac.uk/aboutus*, accessed 9 October 2011.

Adams, J. and Gurney K. (2010), "Funding Selectivity, Concentration and Excellence: How Good is the UK's Research?", published online by the Higher Education Policy Institute, 25 March, accessed 9 October 2011. See also Z. Corbyn (2010), "Data Disprove Case for Distributing Research Funds on Historical Basis", *Times Higher Education*, No. 1940, 25 March, p. 17.

Agence nationale de la recherche (2011), "Initiatives d'excellence IDEX", *www.agence-nationale-recherche.fr/investissementsdavenir/documents/ANR-AAP-IDEX-2010.pdf*, accessed 9 October 2011.

Evidence (Thompson Reuters) (2009), "The Use of Bibliometrics to Measure Research Quality in UK Higher Education Institutions", *Report to Universities UK*, *www.universitiesuk.ac.uk/Publications/Documents/bibliometrics.pdf*, accessed 9 October 2011.

Evidence (2010), "The Future of the UK University Research Base", *Report for Universities UK*, *www.universitiesuk.ac.uk/Publications/Documents/UUK-FutureOfResearch-LiteratureReview.pdf*, accessed 9 October 2011.

Harnad, S. (2008), "Validating Research Performance Metrics Against Peer Rankings", *Ethics in Science and Environmental Politics*, Vol. 8, pp. 103-107.

Harnad, S. (2009), "Open Access Scientometrics and the UK Research Assessment Exercise", *Scientometrics*, Vol. 79, pp. 147-156.

Harrison, M. (2009), "Does High Quality Research Require Critical Mass?", in D. Pontikakis, D. Kriakou and R. van Baval (eds.), *The Question of R&D Specialisation: Perspectives and Policy Implications*, European Commission, *JRC Technical and Scientific Reports*, pp. 57-59.

Kenna, R. and B. Berche (2010), "The Extensive Nature of Group Quality", *Europhysics Letters*, Vol. 90, Ref. 58002.

Kenna, R. and B. Berche (2011), "Critical Mass and the Dependency of Research Quality on Group Size", *Scientometrics*, Vol. 86, pp. 527-540.

Mahdi, S., P. D'Este and A. Neely (2008), "Citation Counts: Are They Good Predictors of RAE Scores? A Bibliometric Analysis of RAE 2001", *http://ssrn.com/abstract=1154053*, accessed 9 October 2011.

Million (2010), "A Postgraduate Strategy for Britain: Expanding Excellence, Innovation and Opportunity", *www.millionplus.ac.uk/research/index*, accessed 9 October 2011.

Ministère de l'Enseignement supérieur et de la Recherche (2011), "L'opération campus : Un plan exceptionnel en faveur de l'immobilier universitaire", *www.enseignementsup-recherche.gouv.fr/pid20637/l-operation-campus.html*, accessed 9 October 2011.

Newman, M. (2009), "Do you Want to Be in my Gang?", *Times Higher Education*, No. 1923, pp. 32-37.

Oppenheim, C. and M.A.C. Summers (2008), "Citation Counts and the Research Assessment Exercise, Part VI: Unit of Assessment 67 (music)", *Information Research*, Vol. 13, Paper No. 342, June.

Popper, K. (1959), *The Logic of Scientific Discovery*, Basic Books, New York, NY.

Raan, A.F.J. (van) (2005), "Fatal attraction: Conceptual and Methodological Problems in the Ranking of Universities by Bibliometric Methods", *Scientometrics*, Vol. 62, No. 1, pp. 133-143.

Raan, A.F.J. (van) (2006), "Comparison of the Hirsch-index with Standard Bibliometric Indicators and with Peer Judgement for 147 Chemistry Research Groups", *Scientometrics*, Vol. 67, Issue 3, pp. 491-502.

RAE (Research Assessment Exercise) (2008), *www.rae.ac.uk*, accessed 9 October 2011.

Ringelmann, M. (1913), "Recherches sur les moteurs animés : Travail de l'homme", *Annales de l'Institut National Agronomique*, Vol. 12, No. 1. pp. 1-40.

Russell Group (2010), "The Concentration of Research Funding in the UK: Driving Excellence and Competing Globally", *www.russellgroup.ac.uk/uploads/Concentration-of-research-funding.pdf*, accessed 9 October 2011.

Russell Group (2011), *www.russellgroup.ac.uk/Aims-and-Objectives.aspx*, accessed 9 October 2011.

Salmi, J. and A. Saroyan (2007), "League Tables as Policy Instruments: Uses and Misuses", *Higher Education Management and Policy*, Vol. 19, No. 2, pp. 31-68.

Sauvons la Recherche (2011), *http://sauvonslarecherche.fr/spip.php?article3306*, *www.sauvonsluniversite.com/spip.php?article4085*, accessed 9 October 2011.

University Alliance (2009), "Concentration and Diversity: Understanding the Relationship between Excellence, Concentration and Critical Mass in UK Research", *www.university-alliance.ac.uk/2009/12/university-alliance-report-shows-that-selectivity-not-concentration-has-driven-excellence-in-uk-research/*, accessed 9 October 2011.

From:

Higher Education Management and Policy

Access the journal at:

https://doi.org/10.1787/17269822

Please cite this article as:

Kenna, Ralph and Bertrand Berche (2011), “Critical masses for academic research groups and consequences for higher education research policy and management”, *Higher Education Management and Policy*, Vol. 23/3.

DOI: https://doi.org/10.1787/hemp-23-5kg0vswcm27g