TOMAS COATES ULRICHSEN

Knowledge Exchange Framework Metrics: A Cluster Analysis of Higher Education Institutions

A Technical Report for Research England

November 2018

About the Author

Tomas Coates Ulrichsen is a Research Associate at the Centre for Science, Technology and Innovation Policy (CSTI) at the University of Cambridge. His core research interests lie in understanding how a nation's university base, through its linkages and partnerships with industry, can drive technological innovation, economic development and support industrial transformation.

In addition to this research, Tomas undertakes advisory work for key UK government agencies responsible for funding university-industry knowledge exchange and his work has been instrumental in shaping the direction of key funding programmes in this area. Prior to joining CSTI, Tomas was an Assistant Director at Public and Corporate Economic Consultants (PACEC) where he led projects evaluating knowledge exchange and innovation funding programmes, analysing the roles and impacts of universities in the innovation system, and better understanding the knowledge exchange process and support system. He is a member of the Knowledge Exchange Framework (KEF) Steering Group and the Technical Advisory Group on KEF metrics. He was a member of the McMillan Review of Good Practice in Technology Transfer.

Contents

About the Author 2

1 Introduction 3

2 Conceptual Framework 4

3 Methods and data 6

3.1 Variables and data 8

3.1.1 Scale and focus of existing knowledge base 8

3.1.2 Scale and focus of knowledge generation 9

3.1.3 Physical assets 10

3.2 Approach 11

4 Results 13

4.1 Robustness of clusters 16

5 Local Economic Context 16

6 Discussion and conclusions 18

References 20

Appendix A Cluster membership for model with untransformed variables 22

Appendix B Cluster Characteristics 23

1 Introduction

This technical report presents a cluster analysis of English Higher Education Institutions (HEIs) to inform the development of the Knowledge Exchange Framework Metrics (hereafter KEF). This analysis was undertaken at the request of the KEF Technical Advisory Group (TAG).

At the outset is important to recognise the diversity of types of HEIs that exist in a national innovation system such as that of the UK. This diversity of institution sees different types of HEIs contributing in different ways to different socio-economic, technological, industrial and regional challenges. Importantly, structural differences between HEIs, coupled with their local economic context, shape both KE opportunities and barriers. Recognising this, the analysis in this report sought to identify groups of HEIs based on similarities in the structural characteristics that shape KE opportunities and challenges, to enable more appropriate comparisons of knowledge exchange (KE) performance.

It is very important that cluster analyses are driven by a conceptual understanding of KE. The broad approach adopted builds on discussions at the initial KEF Technical Advisory Group (TAG) meeting and assumes that KE opportunities for HEIs are underpinned by the knowledge and physical capabilities available to them. These provide a ‘capability base’ which can be thought of as quasi-fixed in the short- to medium-term, but can change over the longer-term through investments in research, teaching and related physical capital. In adopting this approach, assessments of KE performance should then focus on how well a university, given its knowledge and physical assets, is able to pursue KE opportunities and, through these, deliver socio-economic impacts.

The report is structured as follows. The next section presents the overall conceptual framework developed to guide the cluster analysis. Section 3 then outlines the methods and data used in the analysis. Section 4 presents the results. Section 5 explores the distribution of the types of HEIs identified across different types of economic contexts, building on evidence that shows that the local economic context plays an important role in shaping KE opportunities available to HEIs. Section 6 concludes.

2 Conceptual Framework

HEIs play important roles in the innovation system of a nation. However, the diversity of this type of actor is frequently overlooked, with institutions treated as similar, generating and diffusing knowledge, and developing the next generation of the labour force, particularly in simple ranking systems (Howells et al., 2008). These simple rankings typically compare metrics of 'performance' with no attempt to control for structural differences between HEIs. By contrast, this diversity must be celebrated and strengthened in order for the national innovation system to meet the many and complex knowledge needs across a broad range of industrial, technological, regional, and societal challenges. This sentiment was indeed echoed in the major review of UK science and technology policies by Lord Sainsbury (Sainsbury, 2007), who concluded that universities with different economic missions "*should carry out all three activities – research, teaching and knowledge transfer – but the way they perform them will be very different*".

One method for capturing this diversity is to identify groups of broadly similar HEIs based on the functions they perform within the innovation system (Howells et al., 2008; McCormick and Zhao, 2005). Institution-level performance comparisons can then be made within the group of similar HEIs rather than across groups. Cross-group comparisons can also be very useful, but should be limited to examine how different *types* of HEIs are contributing to the innovation system, or to enable individual institutions to explore the practices and performance of HEIs undertaking different types of functions in the system.

Early attempts to cluster HEIs go back to the 1970s with the work of King on the UK system, and the creation of the Carnegie Classification in the United States (Howells et al., 2008; King, 1970). Perhaps the best known and long-lasting is the Carnegie Classification. This was created in response to a realisation by the Carnegie Foundation for the Advancement of Teaching that there was no classification system of HEIs that differentiated institutions *along the key dimensions that were important to its work* and that this limited their ability to make appropriate recommedations on the major issues facing the sector. It sought to emphasize the diversity of the US HE system and enable institutions to compare their practices and performance with other HEIs performing 'similar' functions in the system, and contrast them to those undertaking 'different' functions (McCormick and Zhao, 2005).

In developing a classification system for UK HEIs to enable comparisons of KE performance and practice, it is important to focus on those structural dimensions that shape the nature and scale of

KE opportunities available to an HEI and the linkages that form with external partners. To give an example, it would seem to be unfair to compare the KE performance of a very large, research-intensive university undertaking world-leading research across a broad range of disciplines but with a heavy clinical medicine and engineering focus, with a small specialist arts institution. Their knowledge bases are fundamentally different which lead to fundamentally different opportunities for KE. Focusing on these types of structural dimensions that, at least in the short-term, cannot easily be changed, should help the identification of institutions with similar nature and scale of KE opportunities and focus efforts to explore how efforts at different strategic and operational levels of the HEI (leadership, KE support, academic) could help to improve KE efficiency, effectiveness and ultimately overall performance.

In thinking about the potential opportunities for KE, Molas-Gallart et al. (2002) in their early work on what was then called the 'third stream', argued that universities have sets of knowledge and physical capabilities that are developed over time as they undertake their core activities of research and teaching, and invest in physical capital. In terms of HEIs' knowledge capabilities, evidence has shown that different sectors demand knowledge from different combinations of disciplines (Cohen et al., 2002). Furthermore, Hughes and Kitson (2012) showed that while KE was prevalent across the variety of different disciplines, particular mechanisms such as commercialisation were more limited to specific areas such as life sciences and engineering.

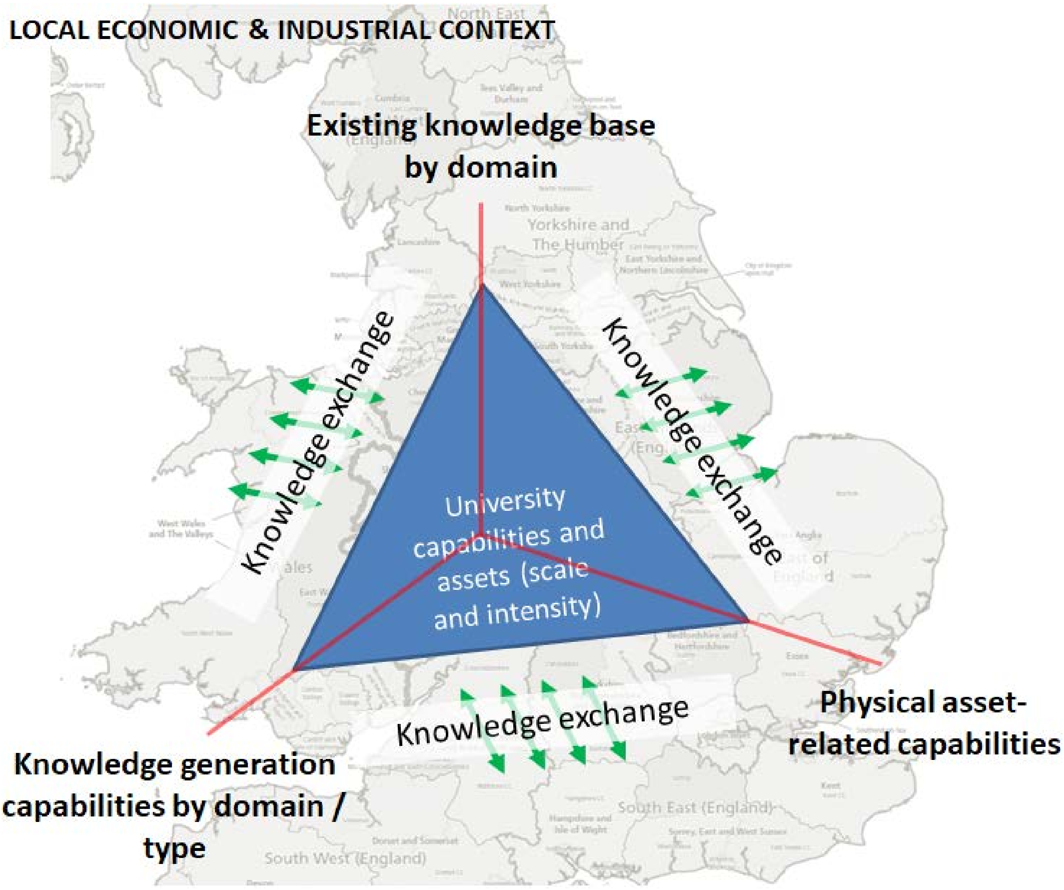
We also know that HEIs of all types – research-intensive, teaching-intensive and specialists – engage in wide varieties of KE from commercialisation, to contract and collaborative research, to consulting, provision of training, and the provision of testing and other facilities and equipment related services. In addition they perform an important 'public space' role that has the potential to bring together different actors in the innovation system and stimulate connections that may otherwise not form (Cohen et al., 2002; D'Este and Patel, 2007; Hughes and Kitson, 2012; Lester, 2005; Perkmann et al., 2013). While much of the public space role of universities is driven by the social networks within an innovation system an HEI can foster, investments in physical capital can provide an important platform on which the social capital can develop. Crucially KE engagements draw from both new knowledge generated through research as well as from the existing knowledge held within HEIs that can be deployed to address an external partner's needs. As such the knowledge capabilities of HEIs need to cover both their knowledge generation aspects as well as the existing knowledge held within the institution.

The scale of HEIs is also believed to shape the KE opportunities of HEIs (Howells et al., 2008). Larger HEIs may be able to internalise a wider range of KE support services and deliver a wider range of functions into the innovation system that smaller institutions would struggle to provide absent of partnering with other organisations (HEIs, innovation intermediaries or others).

Lastly, evidence has also shown that the local socio-economic and industrial economic context plays an important role in shaping how HEIs engage in KE (Huggins et al., 2012; Lester, 2005).

The analysis that follows thus assumes that the set of knowledge and physical capabilities developed through long-term investments in research, teaching and physical capital form a 'capability base' which shape the set of KE opportunities an HEI can pursue absent of significant changes to this base; i.e. they shape the KE potential of an HEI. These opportunities are additionally shaped by the scale of the HEI and the local economic context within which an HEI is situated. The conceptual framework is set out in Figure 1.

Figure 1 Conceptual framework



The conceptual framework is used to guide the analyses to identify groups of universities with similar structural characteristics. It distinguishes the scale and intensity of capabilities along three key dimensions: (i) existing knowledge base; (ii) knowledge generation; and (iii) physical assets.

3 Methods and data

To identify groups of similar HEIs in terms of their structural 'capability base' driving KE opportunities, I develop a method built around a statistical cluster analysis that follows an approach similar to that used in previous exercises looking to identify groups of HEIs (HEFCE, 2009; Howells et al., 2008), and in the strategic management field to identify groups of firms within an industry (Ketchen and Shook, 1996; Short et al., 2007). There are, of course, other methods for identifying groups of similar institutions, for example based on heuristics, expert allocation to groups, or self-selection. The advantage of a statistical cluster analysis is that it minimises subjectivity in the allocation of HEIs to groups, and focuses on revealed differences based on data.

It is important that the variables entering the cluster analysis are based on a conceptual understanding of those factors that drive the model – here the differences in KE opportunities between HEIs. The statistical analysis itself cannot distinguish between relevant and irrelevant variables, and inclusion of the latter could influence the results.

There are also different types of cluster analysis methods that can be broadly categorised into hierarchical and non-hierarchical, each with advantages and disadvantages (for a good discussion see e.g. Ketchen and Shook, 1996). Hierarchical methods proceed in steps, developing 'tree-like' structures that either add observations to clusters (agglomerative) or delete them from clusters (divisive). These have the advantage that the number of clusters emerges from analysis. They are also repeatable. However, they only pass through data once and an HEI cannot move cluster once assigned to a cluster. The solutions can also be unstable to dropping observations particularly where sample sizes small. Non-hierarchical methods are an iterative approach, partitioning samples into a pre-specified number of clusters. Following the specification of the initial positions of each cluster, observations are allocated to the nearest one. As each observation is added, the cluster centroids are recomputed. Multiple passes are made through the data allowing observations to change cluster, until convergence of membership is achieved. This is a key advantage of non-hierarchical cluster methods which, as a result, are less impacted by outliers. However, they suffer from some drawbacks, not least results can depend on choice of initial positions. It is often the case that, observations are able to move cluster, they tend not to move to distant clusters, making the choice of initial position very important. In addition, unlike hierarchical methods, one has to specify the number of clusters in advance, rather than let it emerge from the process.

To overcome the limitations of each type of method, scholars have developed approaches that combine elements of both hierarchical and non-hierarchical methods (Ketchen and Shook, 1996). For example, some use a hierarchical cluster analysis to determine the number of clusters and identify initial cluster positions. This information is then fed into a second stage that deploys non-hierarchical methods to determine final cluster membership.

There are a number of other important considerations when performing cluster analyses. Some of the factors influencing HEI KE opportunities are highly correlated, particularly when looking at scale effects. In feeding variables into a clusters analysis, highly correlated variables can lead to overweighting of a particular construct in the model. If, therefore, we wish different constructs to be weighted in a more balanced way, one has to deal with collinearity between variables. This can be dealt with through techniques such as principle components analysis (PCA). However, as one discards some components, we have to accept some loss of information in this process.

In addition, factors influencing KE opportunities have very different scales and variances, with some having very large scales and potentially significant 'distances' between the maximum and minimum values, with others do not. These purely scale differences can dominate the cluster results. While in some cases this can be desirable, in other cases it may not. For the latter, variables can be transformed or standardized (to a mean of zero and standard deviation of one) to account for differences in variable scales (Ketchen and Shook, 1996).

HEI activity – particularly around knowledge generation – is highly skewed, with a relatively small number of HEIs generating much of the activity. Left unattended, this can lead to challenges in discriminating HEIs with less of the particular activity. Following the practice in other cluster analyses of HEIs such as the Carnegie Classification of institutes of higher education in the United States[1], it is helpful to transform the data in order to deal with this issue. One method is to log-

[1] http://carnegieclassifications.iu.edu/, accessed on 23rd May 2018

transform the data; another used in the Carnegie Classification method is to running any analysis on the rank scores of the variables (ordered low to high) rather than on the scale.

In running cluster analyses, one also has to choose the distance measure used (e.g. Euclidean distance (derived through the use of Pythagorean formulae), Manhattan distance (based on the sum of the absolute differences between values), or other types of measures such as those based on the correlations of profiles).

3.1 Variables and data

Guided by the conceptual framework, I explored the available data along each of the three key dimensions. To be used in this exercise data had to be available annually for all HEIs in England. As a result, the primary source of data was the Higher Education Statistics Agency (HESA).

3.1.1 Scale and focus of existing knowledge base

The first dimension centres on the scale and focus of the existing knowledge base available within HEIs. This is largely held within academic and research staff, and the student population. Different types of staff and students may hold different types of knowledge which lead to different types of KE opportunities. In addition, we know from existing studies into KE that knowledge from different disciplines have different KE opportunities in different parts of the economy and society. As such, within this dimension it is important to capture differences in the composition of the staff and student populations and differences in the disciplinary portfolios of existing knowledge across HEIs. The variables selected used are summarised in Table 1.

Table 1 Variables within dimension 1: scale and focus of existing knowledge base

Category	Variables	Source
Number of academics by function	• Teaching/research • Teaching only • Research only	HESA
Portfolio of academics by discipline (proportion)	• Clinical medicine • Allied health other medical, and dentistry • Agriculture, forestry and veterinary science • Physical sciences and mathematics • Biological sciences • Engineering and materials science • Computer science • Architecture and planning • Social sciences and law • Business and management studies • Humanities, languages and education • Creative and performing arts, and design	HESA
Educational focus of HEIs	• Student FTEs at undergraduate level (full-time/part-time) • Student FTEs involved in taught postgraduate (full-time/part-time) • Student FTEs involved in research postgraduate (full-time/part-time)	HESA

3.1.2 Scale and focus of knowledge generation

The second dimension centres on the scale and focus of knowledge generation activity within HEIs. Again, it is important to capture differences between disciplines here both because KE opportunities arising from different knowledge domains differ, but also because the scale of resources required to undertake research in different disciplines can vary significantly (for example between lab-based science and engineering research, and research in the humanities). The quality of research – particularly in engineering and physical sciences has also been found to affect KE opportunities and the attraction of R&D investments (Abramovsky et al., 2007; Belderbos et al., 2014; Laursen et al., 2016; Perkmann et al., 2011; Siedschlag et al., 2013).

When firms engage with HEIs they do so for a variety of motivations, not least to access and co-develop knowledge to feed into their innovation activities. While some – typically large, technology-intensive firms – seek to co-fund relatively fundamental research, others are looking to access and develop new knowledge that is closer to application. HEIs differ not just in the discipline portfolio of research being undertaken, but also in the type of research in terms of how far it is from application in real world settings. To *proxy* for the type of research being undertaken, I assume that – crudely – different funders of will fund different types of research, with the Research Councils tending to fund more fundamental research where considerations of application are a secondary (although still important) consideration, while industry, government departments and charities will fund research based around a specific application problem.

It is also important to capture both the scale and intensity of knowledge generation activity within HEIs. The intensity helps to distinguish which HEIs are undertaking relatively more research activity after controlling for scale of institutions. However, there is also some evidence to suggest that the scale itself matters in shaping some types of KE opportunities (Perkmann et al., 2011; Ulrichsen, 2015). For example, large firms looking to develop long term strategic partnerships see a critical mass of research activity within the HEI as an important part of the value proposition to engage.

The variables selected within this dimension are summarised in Table 2.

Table 2 Variables within dimension 2: scale and focus of knowledge generation

Category	Variables	Source
Scale of knowledge generation by domain		
Scale of knowledge generation activity in different knowledge domains	• Recurrent research income (QR) • Research grants and contracts income by STEM, SSB, AH • Research quality by STEM, SSB, AH (number of academic FTEs getting 4* publications in REF2014)	HESA
Scale of knowledge generation of different types	• Research grants and contracts from different sources: • UK research councils • Charities • Government bodies / local authorities, health/hospital authorities • Industry	HESA
Scale of international linkages in research	• Research grants from overseas	HESA
Intensity of knowledge generation by domain		

Knowledge generation intensity of HEIs	• Proportion of academic FTEs submitting to REF • Proportion of students undertaking postgraduate research	HESA
Knowledge generation intensity by discipline	• Research grants and contracts income per academic by STEM, SSB, AH • Proportion of researchers generating 4* publications in REF2014 by STEM, SSB, AH	HESA, REF2014
Knowledge generation type intensity	• Research grants and contracts income from different sources (RCs, charities, gov't, industry) per academic	HESA
Research internationalisation intensity	• Research grants and contracts income from overseas per academic	HESA

3.1.3 Physical assets

The third dimension centres on the scale and intensity of investments in physical assets that have the potential to underpin KE opportunities. Some KE opportunities are based around the use of facilities and equipment to achieve particular KE objectives, such as the use of a wind tunnel to test the aerodynamic performance of a prototype vehicle, or a media company using an HEIs digital media suite[2]. It proved very challenging to identify decent proxies for the scale and intensity of physical assets available within an HEI to underpin KE opportunities that distinguished between those knowledge-related physical assets and general physical capital available such as accommodation or generic meeting rooms that could easily be provided by other, private sector providers[3]. As such, I focused on the amount and intensity of investments made by an HEI into research-related capital infrastructure. A recent evaluation of such investments found that many had spillover uses in terms of KE (PACEC, 2012). The study concluded that the "*research facilities that resulted from the funding have been increasingly made available to outside organisations, which has increased the effectiveness of knowledge exchange activities. In particular, this improved availability has strengthened the relationships between industry and universities and colleges*."

The variables selected within this dimension are summarised in Table 3.

Table 3 Variables within dimension 3: scale and intensity of physical asset investment

Category	Variables	Source
Scale of physical asset investment	• Scale of spending on research-related capital infrastructure	HESA
Intensity of physical asset investment	• Intensity of capital spending (spend per academic)	HESA

[2] HEBCI Section B, Table 2 guidance, available at https://www.hesa.ac.uk/collection/c16032/hebci_b_table_2, accessed on 25th May 2018
[3] This distinction follows the HEBCI Section B, Table 2 guidance

3.2 Approach

The overall approach used in the cluster analysis is summarized below:

Identified and categorised specialist institutions using heuristics:

- Assessed degree of concentration of academic activity across the 45 discipline groups provided by HESA using:
  - o The Herfindahl Index (which measures concentration) based on the number of full time equivalent academics in different disciplines of greater than 0.4

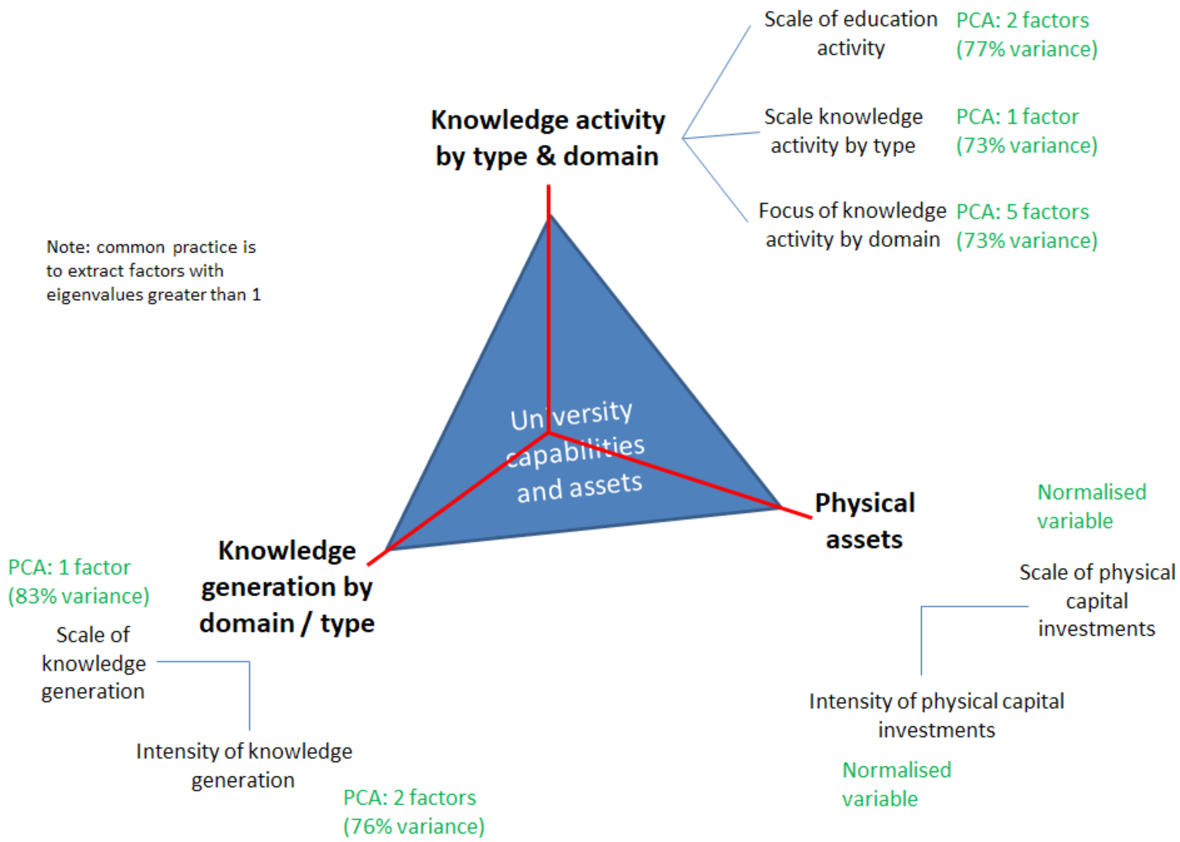
  OR

  - o A share of academics of greater than 50% in any single discipline (based on the 45-way split provided by HESA)
- Allocated specialists to one of the following categories: (i) STEM-focused, which could be further divided into those focusing on biological and veterinary sciences, engineering and technology, and agriculture; (ii) social sciences (including business), education and humanities; and (iii) creative and performing arts and design.
- The Arts University Bournemouth – was classified manually into the specialist arts group as, while it was a borderline case in terms of the above criteria, on manual inspection it appears to be a specialist institution similar to others in this group.
- The National Film and Television School (NFTS) only entered the English HE sector very recently and therefore lacked the data coverage over the period 2012 – 2016 compared to other institutions. Using 2016 data on academic staff FTEs by discipline is was clear that NFTS is a specialist institution in social sciences (including business), education and humanities. Due to this lack of data, while it was allocated to this group of HEIs, it does not enter the analysis of sector characteristics in Appendix B.

Prepare data for cluster analysis of broad-discipline HEIs

- Identified specialist HEIs (N=32) and separated from the sample
- Additionally, the University of London (institutes) was removed from the sample due to its unique characteristics in the English HE system. This resulted in 99 HEIs being clustered using the statistical cluster approach
- Developed metrics based on variables identified in section 3.1 to capture both the scale and intensity of key knowledge and physical asset dimensions
- Transformed variables using the natural logarithm to discriminate HEIs more fully at the lower ends of the distributions of variables where many HEIs are clustered. The exercise was repeated using untransformed variables; the details and resulting clusters are provided in Appendix A.
- Ran a principal component analysis (with orthogonal rotation) on variables within each dimension to deal with high correlations between variables. The resulting components were used in the model and were standardized to a mean of zero and standard deviation of one. Where only one variable exists on a particular dimension, the variable was standardized to a mean of zero and standard deviation of one. The resulting number of components and the proportion of variance captured are shown in Figure 2.

Figure 2 Results of principal component analysis based on logged variables where appropriate



Performing the cluster analysis

- A two-stage cluster analysis was performed:
  - The first stage deployed the hierarchical Wards linkage cluster method in order to both inform the selection of the number of clusters and determine the starting points for the second stage;
  - The second stage deployed the non-hierarchical kmedians cluster method (which is less sensitive to outliers) using the number of clusters and starting points from the first stage.
- Following common practice, the Euclidean distance was used as the distance measure
- The model was run using both transformed variables (preferred model and the focus of the report) and untransformed variables (see Appendix A)
- The first stage suggested three main clusters based on the cluster dendrogram plot. However, these groups were very large with diverse membership. The five cluster solution also provided decent results in the second phase, with sufficiently high values of the Calinski–Harabasz stopping rule (another method for helping to determine the number of clusters, which is particularly useful in non-hierarchical cluster methods where visual representations of the hierarchies are not possible). The clusters have broadly similar numbers of HEIs which is helpful.

Cluster stability

- Examined the stability of the clusters by randomly removing 10 HEIs (approximately 10% of the sample) and repeated the two-stage cluster analysis. I then used the Adjusted Rand Index[4] to examine the consistency of the resulting clusters. An index of one indicates a perfect match between two cluster solutions while an index of zero indicates no match.

4 Results

The results of the cluster analysis are shown in the following figures. Recall that the approach was designed to reveal systematic differences in the structural characteristics that lead to differences in KE opportunity potential. The process does not seek to make any value judgement on whether one cluster is in some way better or more valuable than another; only that they are structurally *different*. Thus to minimise the potential to impose subjective biases in the interpretation of the resulting clusters based on the labels assigned (e.g. through using categories such as 1, 2, 3; or high research-intensive/less research-intensive), I randomly assigned letters to label each cluster and ordered their presentation based on these letters.

Figure 3 Segmentation of English HE sector by cluster

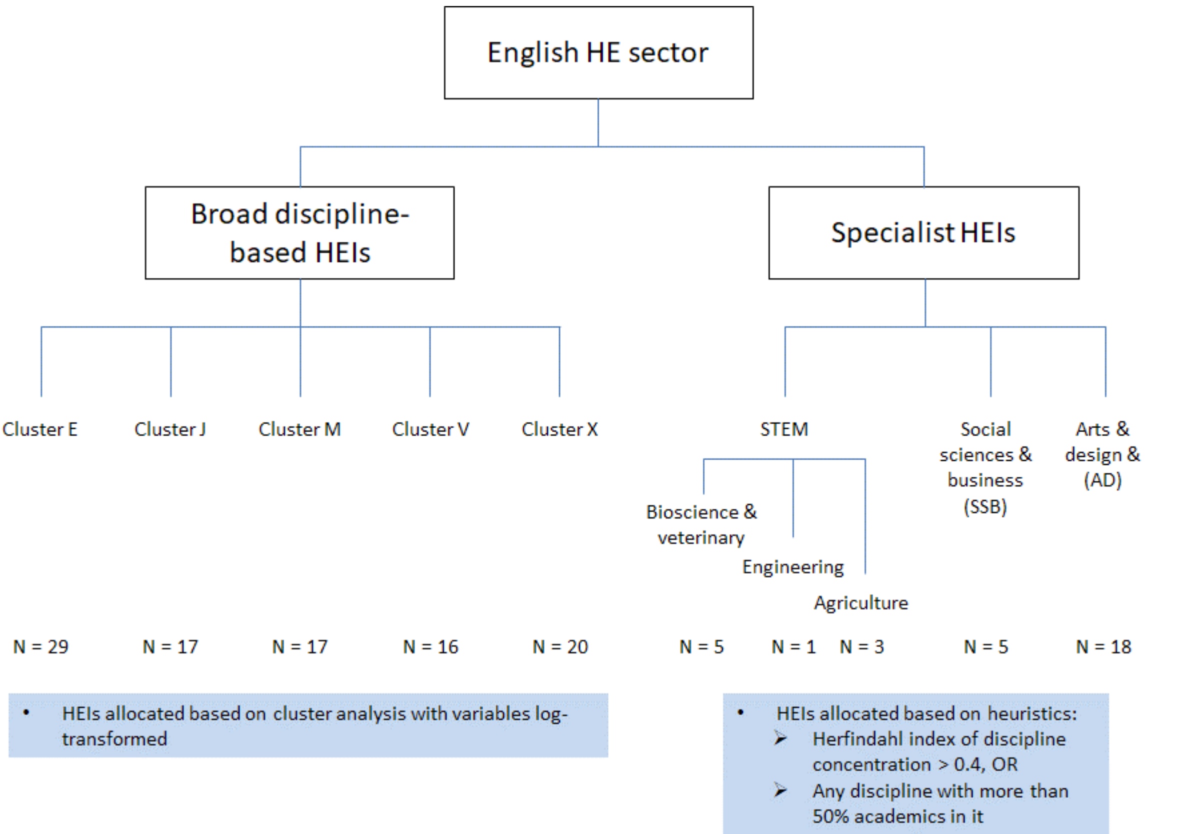


Figure 3 shows how the English HE sector segments into different clusters and the number of HEIs in each group. Figure 4 lists the membership of each cluster for the broad discipline HEIs, and Figure 5 does so for the specialist institutions.

[4] See https://www.stata.com/meeting/france17/slides/France17_Halpin.pdf for more information, accessed on 25th May 2018

Figure 4 Cluster membership (log transformed variables)

Cluster E	Cluster J	Cluster M	Cluster V	Cluster X
Anglia Ruskin	B'ham City	Bath Spa	Birmingham	Bath
Aston	Bolton	Buck's New	Bristol	Birkbeck
Bedfordshire	Canterbury	Buckingham	Cambridge	Brunel
Bournemouth	Chester	Chichester	Imperial	Durham
Bradford	Derby	Cumbria	King's College	East Anglia
Brighton	East London	Edge Hill	Leeds	Essex
Central Lancs	Gloucestershire	Falmouth	Liverpool	Exeter
City University	Leeds Beckett	Leeds Trinity	Manchester	Hull
Coventry	London Met	Liverpool Hope	Newcastle	Keele
De Montfort	South Bank	Newman	Nottingham	Kent
Goldsmiths	Northampton	St Mary Tw'ham	Oxford	Lancaster
Greenwich	Roehampton	Solent	Queen Mary	Leicester
Hertfordshire	Staffordshire	Marjon	Sheffield	LSE
Huddersfield	Sunderland	Suffolk	Southampton	Loughborough
Kingston	Teesside	West London	UCL	Reading
Lincoln	Wolverhampton	Winchester	Warwick	Royal Holloway
John Moores	Worcester	York St John		SOAS
Manchester Met				Surrey
Middlesex				Sussex
Northumbria				York
N'ham Trent				
Open				
Oxford Brookes				
Plymouth				
Portsmouth				
Salford				
Sheffield Hallam				
UWE				
Westminster				

Figure 5 Cluster membership: specialist institutions

	Specialists: STEM *	Specialists: Social sciences & business	Specialists: Arts & design
BIO/VET	ICR	UC Birmingham	Arts B'mouth
	Liver Trop Med	Bishop G'teste	Dance & Drama
	Sch of Hygiene	Heythrop	Courtauld
	Royal Vet Coll	L'don Business	Creative Arts
	St George's	National Film**	Guildhall
ENG	Cranfield		Leeds Art
AGR	Harper Adams		Liver Perf Arts
	Royal Agr Coll		Arts London
	Writtle		Norwich Arts
			Plymouth Art
			Ravensbourne
			Rose Bruford
			Royal Ac Music
			Royal Coll Art
			Royal Coll Mus
			Speech & Drama
			RNCM
			Trinity Laban

Notes:
* BIO/VET: biosciences and veterinary sciences; ENG: engineering; AGR: agriculture
** National Film and Television School was allocated to the specialist social sciences and business group based on 2016 data on academic staff

Table 4 summarises the key characteristics for each of the different broad-discipline HEI clusters along the key dimensions of the framework.

Table 4 Cluster characteristics for broad-discipline HEIs

Cluster	Characteristics
Cluster E	• Large universities with broad discipline portfolio across both STEM and non-STEM generating a mid-level amount of world-leading research across all disciplines • Significant amount of research funded by gov't bodies/hospitals; 9.5% from industry. • Large proportion of part-time undergraduate students, and small postgraduate population dominated by taught postgraduates.
Cluster J	• Mid-sized universities with limited funded research activity and generating limited world-leading research • Academic activity across STEM and non-STEM including other health, computer sciences, architecture/planning, social sciences and business, humanities, arts and design • Research activity funded largely by government bodies/hospitals; 13.7% from industry
Cluster M	• Small universities with limited funded research activity and generating limited world-leading research • Academic activity across disciplines, particularly in other health domains and non-STEM • Much of research activity funded by gov't bodies/hospitals; 14.7% from industry.
Cluster V	• Very large, very high research intensive and broad-discipline universities undertaking significant amounts of world-leading research • Research funded by range of sources incl. RCs, gov't bodies, charities and 10.2% from industry. • Discipline portfolio: significant activity in clinical medicine and STEM • Student body includes significant numbers of taught and research postgraduates.
Cluster X	• Large, high research intensive and broad-discipline universities undertaking a significant amount of world-leading research • Much of research funded by RCs and gov't bodies; 8.5% from industry • Discipline portfolio balanced across STEM and non-STEM with less or no clinical medicine activity • Large proportion of taught postgraduates in student population

See Appendix B for detailed analysis of cluster characteristics

4.1 Robustness of clusters

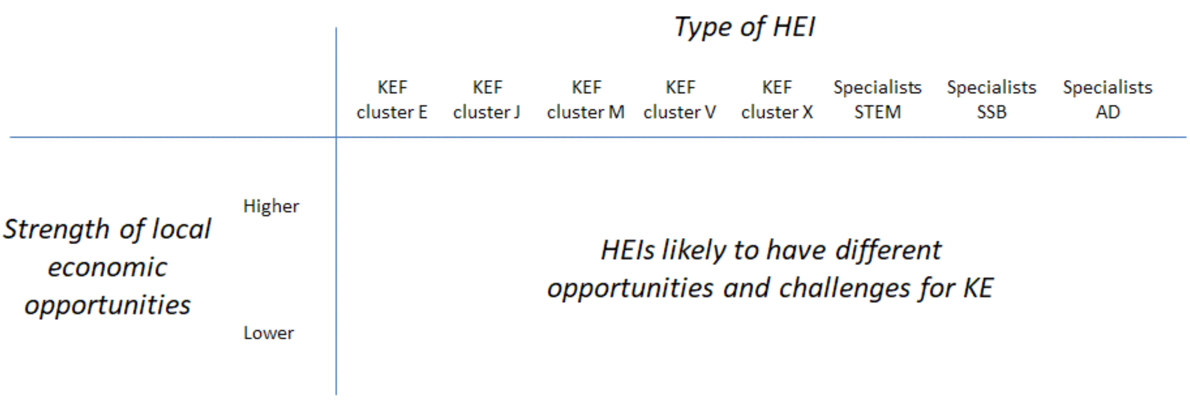
To examine the robustness of the cluster membership 10% of HEIs were randomly removed and the analysis re-run. The similarity of the resulting clusters were compared against the baseline full sample using the Adjusted Rand Index (ARI). A value of 1 indicates a perfect match while 0 indicates no overlap in membership at all. This was repeated 10 times. The results are presented in Table 5.

Table 5 Cluster stability: Adjusted Rand Index for each scenario

Scenario	ARI
1	1.0000
2	1.0000
3	0.9756
4	0.8584
5	0.9652
6	0.9756
7	0.7257
8	0.9756
9	0.9364
10	0.9423
Average	**0.93548**

5 Local Economic Context

KE opportunities are in part driven by the structure and competitiveness of the local economy within which HEI is situated and the ability of local firms to engage with, and absorb knowledge from, HEIs (Lester, 2005; Huggins et al. 2012). The KEF TAG was keen to examine how HEIs are distributed across different types of local economies.



To examine this issue, I focused on the Local Enterprise Partnerships (LEPs) as a key sub-national policy space. I grouped them using a cluster approach (similar to that used to group HEIs) based on similarities in the strength of their local economies. The strength of the LEP-level economy was based on the approach and data available from the UK Competitiveness Index (UKCI) (Huggins and Thompson, 2016) which ranks LEPS based on their strength across three components: inputs, outputs and outcomes. The two-stage cluster analysis suggested two core groups of local economy: one composed of LEPs with higher competitiveness, and one of LEPs with lower competitiveness.

The allocation of LEPs to the two clusters is shown in Figure 6 with the mean UKCI scores for each of the three components provided in Table 6.

Figure 6   Allocation of LEPs to local economy clusters

Cluster 1 : Higher competitiveness LEPs

- Buckinghamshire Thames Valley
- Cheshire and Warrington
- Coast to Capital
- Coventry and Warwickshire
- Enterprise M3
- Gloucestershire
- Greater Cambridge & Greater Peterborough
- Hertfordshire
- London
- Northamptonshire
- Oxfordshire
- Solent
- South East Midlands
- Swindon and Wiltshire
- Thames Valley Berkshire
- West of England

Cluster 2: Lower competitiveness LEPs

- Black Country
- Cornwall and Isles of Scilly
- Cumbria
- Derby, Derbyshire, Nottingham and Nottinghamshire
- Dorset
- Greater Birmingham and Solihull
- Greater Lincolnshire
- Greater Manchester
- Heart of the South West
- Humber
- Lancashire
- Leeds City Region
- Leicester and Leicestershire
- Liverpool City Region
- New Anglia
- North Eastern
- Sheffield City Region
- South East
- Stoke-on-Trent and Staffordshire
- Tees Valley
- The Marches
- Worcestershire
- York, North Yorkshire and East Riding

Table 6   Mean UKCI scores for each component by cluster

Component	Cluster 1 : Higher competitiveness LEPs	Cluster 2: Lower competitiveness LEPs
Input	112	86
Output	108	87
Outcome	102	96

Figure 7   HEI cluster membership by LEP cluster

	Cluster E	Cluster J	Cluster M	Cluster V	Cluster X	Specialists: STEM	Specialists: SSB	Specialists: AD
Higher competitiveness locations	Bedfordshire Brighton City University Coventry Goldsmiths Greenwich Hertfordshire Kingston Middlesex Open Oxford Brookes Portsmouth UWE Westminster	Chester East London Gloucestershire London Met South Bank Northampton Roehampton	Bath Spa Buck's New Buckingham Chichester St Mary Tw'ham Solent West London Winchester	Bristol Cambridge Imperial King's College Oxford Queen Mary Southampton UCL Warwick	Bath Birkbeck Brunel LSE Reading Royal Holloway SOAS Surrey Sussex	Cranfield ICR Royal Agr Coll Royal Vet Coll Sch of Hygiene St George's	Heythrop L'don Business National Film	Arts London Courtauld Creative Arts Dance & Drama Guildhall Ravensbourne Rose Bruford Royal Ac Music Royal Coll Art Royal Coll Mus Speech & Drama Trinity Laban
Lower competitiveness locations	Anglia Ruskin Aston Bournemouth Bradford Central Lancs De Montfort Huddersfield Lincoln John Moores Manchester Met Northumbria N'ham Trent Plymouth Salford Sheffield Hallam	B'ham City Bolton Canterbury Derby Leeds Beckett Staffordshire Sunderland Teesside Wolverhampton Worcester	Cumbria Edge Hill Falmouth Leeds Trinity Liverpool Hope Newman Marjon Suffolk York St John	Birmingham Leeds Liverpool Manchester Newcastle Nottingham Sheffield	Durham East Anglia Essex Exeter Hull Keele Kent Lancaster Leicester Loughborough York	Harper Adams Liver Trop Med Writtle	UC Birmingham Bishop G'teste	Arts B'mouth Liver Perf Arts Norwich Arts Plymouth Art RNCM Leeds Art

Figure 7 reveals how the HEI clusters split across types of LEP areas. It highlights that each HEI cluster has institutions based in more competitive LEPs and less competitive LEPs. It is possible that the KE opportunity potential even within a cluster may thus be different, with those in less competitive LEPs facing different opportunities and challenges to engagement locally than those in more competitive areas.

6 Discussion and conclusions

This report presents a clustering of English HEIs into groups with similar sets of knowledge and physical assets in order to support the discussions of the Knowledge Exchange Framework Technical Advisory Group around how we might facilitate fair comparisons between institutions in the KEF metrics exercise. This recognises that the diversity of HEIs in the UK national innovation system is critical, with different universities working with different types of economic and social actors, and contributing in different ways to different specific socio-economic, technological, industrial and regional challenges. Importantly, these structural differences between HEIs, coupled with their local economic context, shape KE opportunities and barriers; i.e. their KE opportunity potential.

Any attempt to develop a metrics framework to explore KE performance needs to account for this diversity of KE opportunity potential. Assessments of KE performance can then focus on how well a university, given its particular 'quasi-fixed' knowledge and physical asset base, is able to marshal these resources to pursue KE opportunities and, through these, deliver socio-economic impacts.

The cluster analysis HEIs was informed by a conceptual framework that distinguishes different sets of resources that shape the KE opportunity potential of an HEI. In particular, the framework distinguishes the scale and intensity of capabilities along three key dimensions: (i) existing knowledge base; (ii) knowledge generation; and (iii) physical assets. This guided the choice of variables which fed into a two-stage statistical cluster analysis which developed clusters of HEIs that maximise similarity within a group and differences between them.

The analysis revealed five clusters of broad-discipline HEIs exhibiting quite different characteristics along the three key dimensions. In addition, specialist institutions in STEM, social sciences and business, and in arts and design, were separately identified using heuristics, and grouped together reflecting their unique character and KE opportunity potential compared to broad discipline HEIs. Furthermore, each cluster has HEIs that are based in more competitive local economies and less competitive ones. The local context will additionally shape the KE opportunity potential.

It is critical to understand that the cluster analysis identifies groups of HEIs with broadly similar sets of knowledge and physical assets. By focusing on structural characteristics of HEIs rather than KE performance, the approach deliberately avoids make any value judgement that one group is somehow 'better' than another; rather it identifies groups that are ***structurally different*** from each other. The conceptual framework suggests that these differences will lead to differences in KE opportunity potential which should be accounted for in any KE performance measurement system.

The analysis also throws up a number of challenges for Research England in implementing a KE performance measurement system. First, while the arts and design cluster of HEIs is relatively large (with 17 HEI members, similar to the size of the broad-discipline clusters), the STEM and social science & business clusters of HEIs have very few members (with 9 and 4 members respectively). In addition, the STEM cluster includes biosciences, engineering and agriculture focused specialist institutions, each of which will have quite distinct and different sets of KE opportunities with particular sectors of the economy. Research England will need to reflect on how to fairly treat these specialist institutions alongside the much larger number of broad-discipline HEIs.

Second, Research England will need to reflect on how often to update the cluster analysis. While the approach sought to focus on quasi-fixed (i.e. medium- to long-term) structural characteristics, some HEIs are making strategic choices and investing significantly at the moment which will affect their underlying knowledge and physical asset base, which will shape their KE opportunity potential. One suggestion is that this is repeated on a three-year period, or to coincide with the production of HEIF strategies by HEIs as part of the funding process.

Lastly, the cluster analysis was based largely on data available through the Higher Education Statistics Agency (HESA) with universal coverage across all English HEIs. As new data becomes available that capture additional structural features of HEIs that shape their KE opportunity potential, the cluster model should be periodically reviewed and adapted to ensure it remains fit for purpose.

In conclusion, it is hoped that the clusters of HEIs presented in this report help to enable fairer comparisons of KE performance within the English HE system. These clusters are driven by structural differences in the scale, focus and intensity of the knowledge and physical assets of HEIs which are believed to shape their KE opportunity potential.

References

Abramovsky, L., Harrison, R., Simpson, H., 2007. University Research and the Location of Business R&D. The Economic Journal 117, C114–C141. https://doi.org/10.1111/j.1468-0297.2007.02038.x

Belderbos, R., Van Roy, V., Leten, B., Thijs, B., 2014. Academic research strengths and multinational firms' foreign R&D location decisions: evidence from R&D investments in European regions. Environment and Planning A 46, 920–942. https://doi.org/10.1068/a45536

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2002. Links and impacts: the influence of public research on industrial R&D. Management science 48, 1–23.

D'Este, P., Patel, P., 2007. University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry? Research Policy 36, 1295–1313. https://doi.org/10.1016/j.respol.2007.05.002

HEFCE, 2009. Evaluation of the effectiveness and role of HEFCE/OSI third stream funding (Issues paper No. 2009/15). HEFCE, Bristol, UK.

Howells, J., Ramlogan, R., Cheng, S.L., 2008. The role, context and typology of universities and higher education institutions in innovation systems: a UK perspective. DP4, Impacts of Higher Education Institutions on Regional Economies Initiative, http://ewds. strath. ac. uk/Portals/8/typology. doc.

Huggins, R., Johnston, A., Stride, C., 2012. Knowledge networks and universities: Locational and organisational aspects of knowledge transfer interactions. Entrepreneurship & Regional Development 24, 475–502. https://doi.org/10.1080/08985626.2011.618192

Huggins, R., Thompson, P., 2016. UK Competitiveness Index 2016.

Hughes, A., Kitson, M., 2012. Pathways to impact and the strategic role of universities: new evidence on the breadth and depth of university knowledge exchange in the UK and the factors constraining its development. Cambridge journal of economics 36, 723–750.

Ketchen, D.J., Shook, C.L., 1996. The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. Strategic Management Journal 17, 441–458.

King, J., 1970. The topology of universities. Higher Education Review; Croydon, etc. 2, 52–61.

Laursen, K., Masciarelli, F., Reichstein, T., 2016. A Matter of Location: The Role of Regional Social Capital in Overcoming the Liability of Newness in R&D Acquisition Activities. Regional Studies 50, 1537–1550. https://doi.org/10.1080/00343404.2015.1041370

Lester, R., 2005. Universities, innovation, and the competitiveness of local economies. A summary Report from the Local Innovation Systems Project: Phase I. Massachusetts Institute of Technology, Industrial Performance Center, Working Paper Series.

McCormick, A., Zhao, C., 2005. The Carnegie Classification of US institutions of higher education. Retrieved February 9, 2013.

Molas-Gallart, J., Salter, A., Patel, P., Scott, A., Duran, X., 2002. Measuring third stream activities. Final report to the Russell Group of Universities. Brighton: SPRU, University of Sussex.

PACEC, 2012. Evaluation of research capital funding to higher education institutions 2006 to 2008, Report by PACEC to the four UK higher education funding bodies and the Department for Business Innovation and Skills. HEFCE, Bristol, UK.

Perkmann, M., King, Z., Pavelin, S., 2011. Engaging excellence? Effects of faculty quality on university engagement with industry. Research Policy 40, 539–552. https://doi.org/10.1016/j.respol.2011.01.007

Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Broström, A., D'Este, P., Fini, R., Geuna, A., Grimaldi, R., Hughes, A., Krabel, S., Kitson, M., Llerena, P., Lissoni, F., Salter, A., Sobrero, M., 2013. Academic engagement and commercialisation: A review of the literature on university–industry relations. Research Policy 42, 423–442. https://doi.org/10.1016/j.respol.2012.09.007

Sainsbury, D., 2007. The Race to the Top: A Review of Government's Science and Innovation Policies. HMSO, Norwich, UK.

Short, J.C., Ketchen, D.J., Palmer, T.B., Hult, G.T.M., 2007. Firm, strategic group, and industry influences on performance. Strat. Mgmt. J. 28, 147–167. https://doi.org/10.1002/smj.574

Siedschlag, I., Smith, D., Turcu, C., Zhang, X., 2013. What determines the location choice of R&D activities by multinational firms? Research Policy 42, 1420–1430. https://doi.org/10.1016/j.respol.2013.06.003

Ulrichsen, T.C., 2015. Assessing the Economic Impacts of the Higher Education Innovation Fund: a Mixed-Method Quantitative Assessment. HEFCE, Bristol, UK.

Appendix A Cluster membership for model with untransformed variables

The cluster approach was repeated with untransformed (albeit standardized) variables. The results are shown in Figure 8. The effects of transforming the variables (using logs) on discrimination between HEIs with less activity along the dimensions is evident, with large differences in the size of cluster membership.

Figure 8 Cluster membership for broad-based HEIs using untransformed variables

Cluster D	Cluster O	Cluster S	Cluster T	Cluster Y	
• Birmingham	• Bath	• Bath Spa	• Cambridge	• Anglia Ruskin	• John Moores
• Bristol	• Birkbeck	• Buckingham	• Imperial	• Aston	• London Met
• King's College	• Brunel	• Canterbury	• Oxford	• Bedfordshire	• South Bank
• Leeds	• Durham	• Chester	• UCL	• B'ham City	• Manchester Met
• Liverpool	• East Anglia	• Chichester		• Bolton	• Middlesex
• Manchester	• Essex	• Cumbria		• Bournemouth	• Northampton
• Newcastle	• Exeter	• Edge Hill		• Bradford	• Northumbria
• Nottingham	• Hull	• Falmouth		• Brighton	• N'ham Trent
• Queen Mary	• Keele	• Gloucestershire		• Buck's New	• Open
• Sheffield	• Kent	• Leeds Trinity		• Central Lancs	• Oxford Brookes
• Southampton	• Lancaster	• Liverpool Hope		• City University	• Plymouth
• Warwick	• Leicester	• Newman		• Coventry	• Portsmouth
	• LSE	• Roehampton		• De Montfort	• Salford
	• Loughborough	• St Mary Tw'ham		• Derby	• Sheffield Hallam
	• Reading	• Marjon		• East London	• Solent
	• Royal Holloway	• Suffolk		• Goldsmiths	• Staffordshire
	• SOAS	• Winchester		• Greenwich	• Sunderland
	• Surrey	• Worcester		• Hertfordshire	• Teesside
	• Sussex	• York St John		• Huddersfield	• UWE
	• York			• Kingston	• West London
				• Leeds Beckett	• Westminster
				• Lincoln	• Wolverhampton

Comparing the membership of clusters based on the untransformed variables to that of the method using log-transformed variables (Table 7), the effects of the transformation become clear:

- Cluster V splits neatly into two clusters D and T in the untransformed method
- Cluster X maps directly to cluster O in the untransformed method
- Clusters E, J and M re-form into two clusters S and Y in the untransformed method
- The specialist HEIs are allocated based on the same set of heuristics as set out in section 3.2 and remain the same

Table 7 Overlaps of cluster membership between log-transformed and untransformed methods

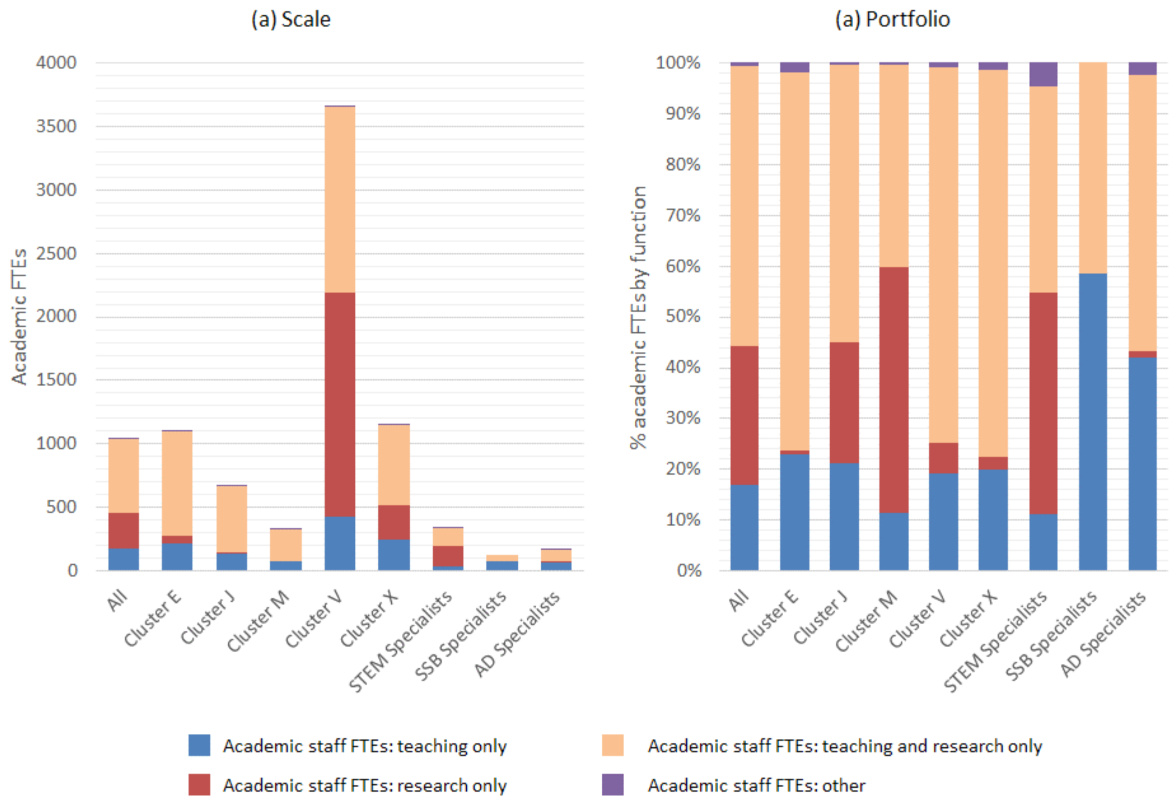
		Clusters based on untransformed variables							
		Cluster D	Cluster O	Cluster S	Cluster T	Cluster Y	STEM Specialists	SSB Specialists	AD Specialists
Clusters based on log-transformation	Cluster E					29			
	Cluster J			5		12			
	Cluster M			14		3			
	Cluster V	12			4				
	Cluster X		20						
	STEM Specialists						9		
	SSB Specialists							5	
	AD Specialists								18

Appendix B Cluster Characteristics

This appendix presents the detailed analysis of the structural characteristics for the HEI clusters presented in section 4 that result from the analysis based on the log-transformed variables.

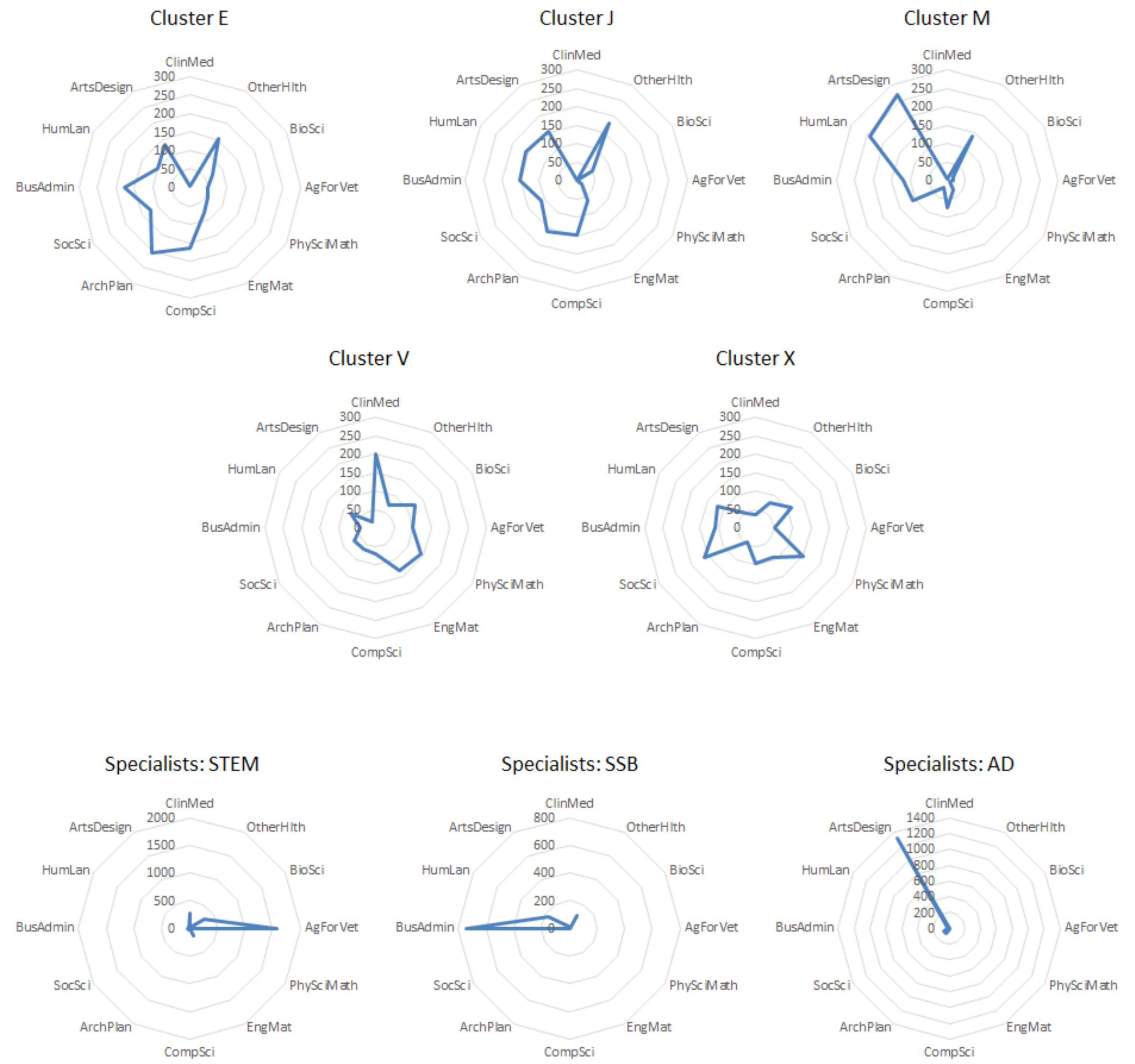
Dimension 1: Existing knowledge base

Figure 9 Academic staff (full-time equivalent, FTE) by function



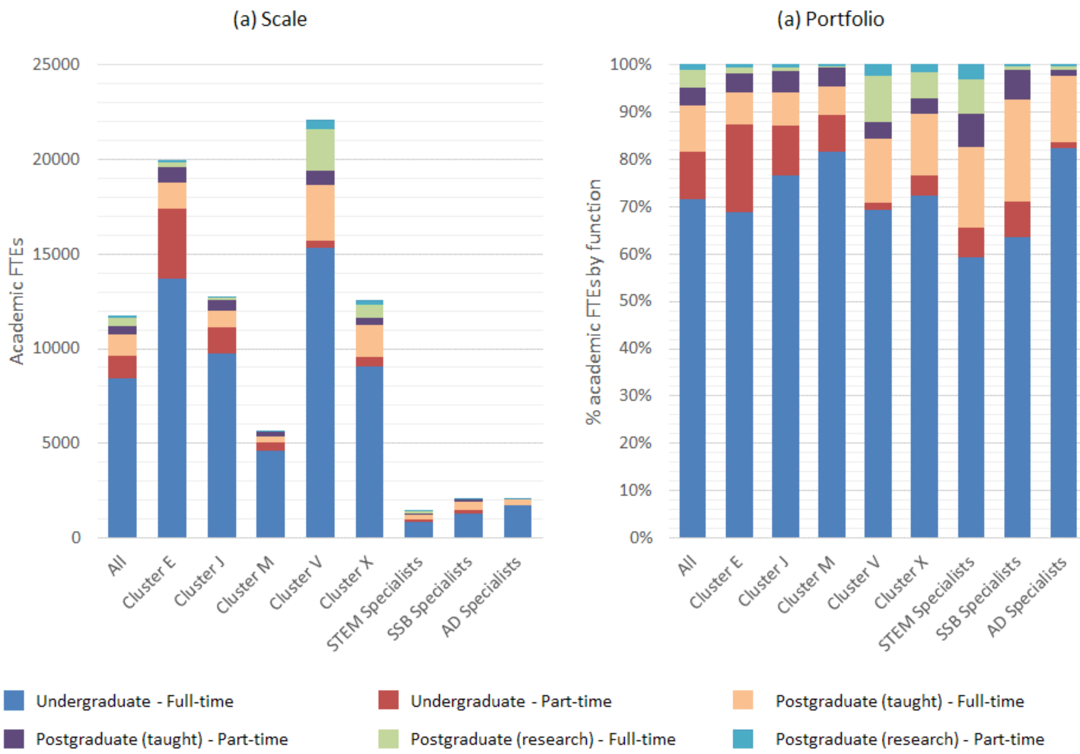
Source: Higher Education Statistics Agency (HESA)

Figure 10 Knowledge activity: disciplinary domains (average for England = 100)



Source: Higher Education Statistics Agency (HESA)

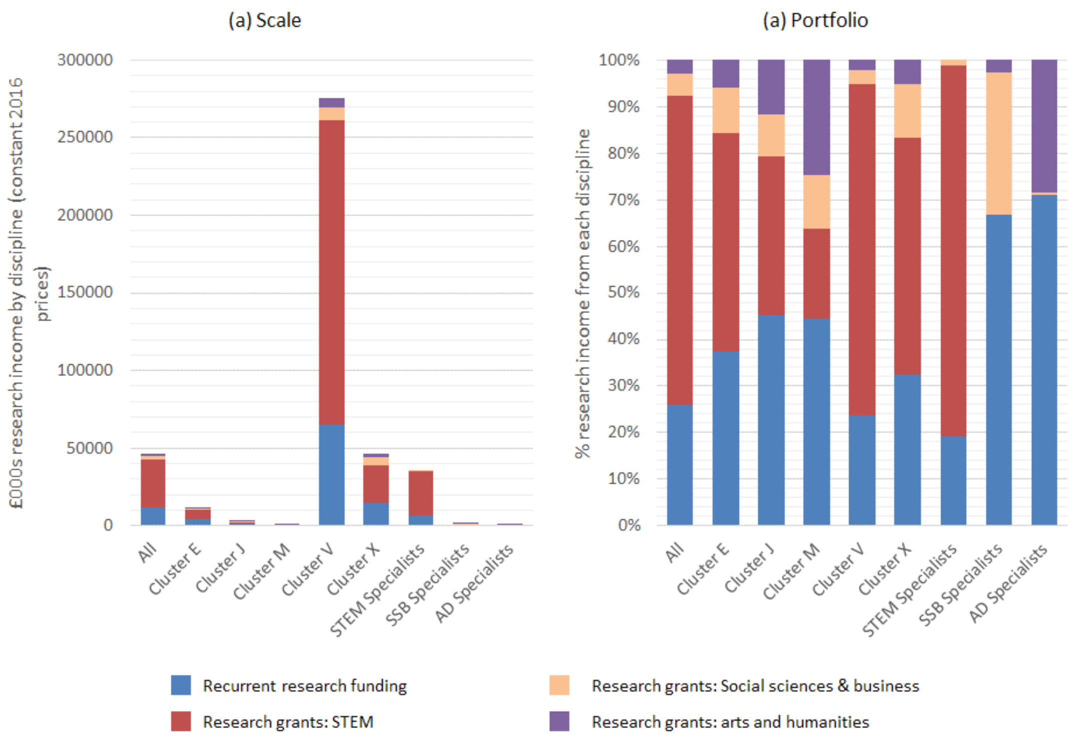
Figure 11 Education function by level



Source: Higher Education Statistics Agency (HESA)

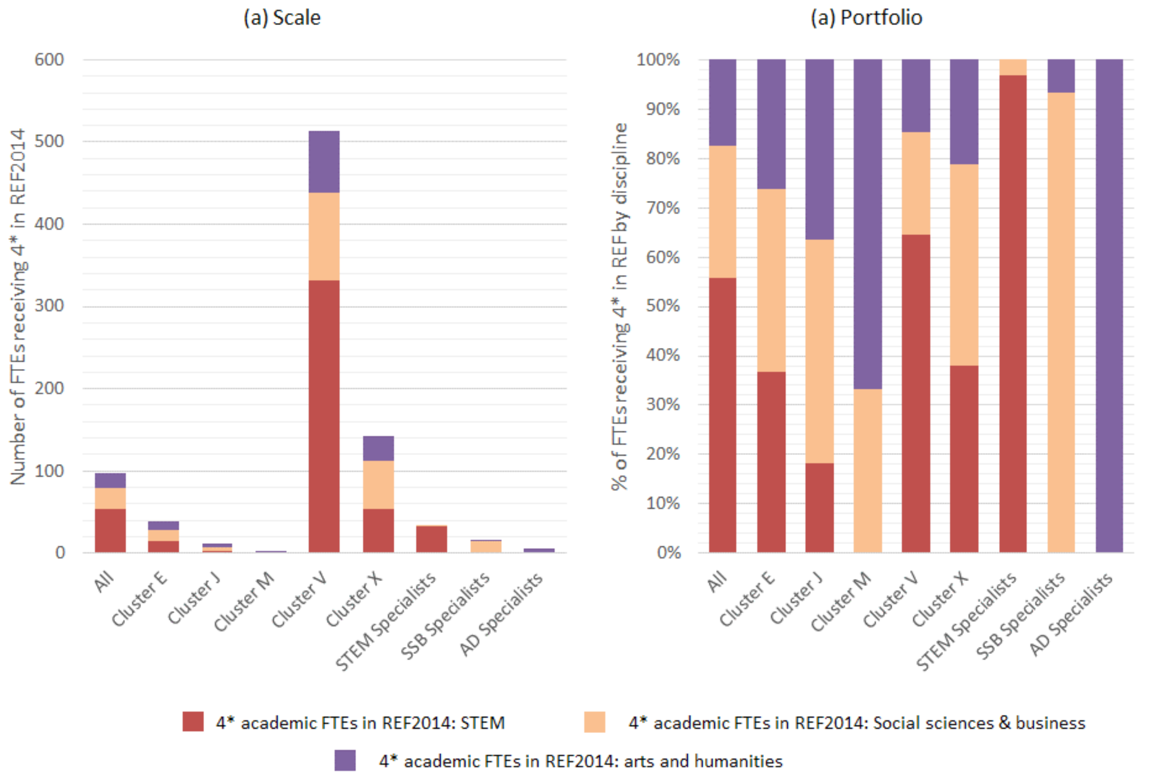
Dimension 2: Knowledge generation

Figure 12 Research income by discipline



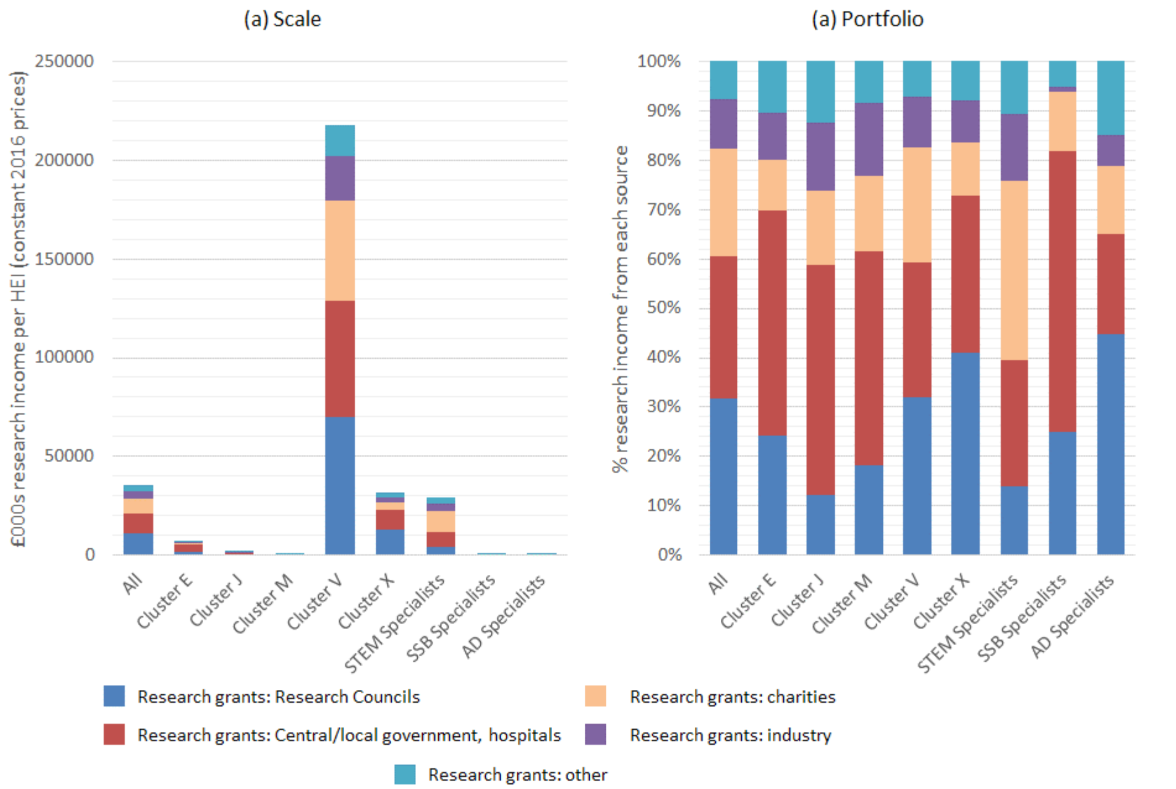
Source: Higher Education Statistics Agency (HESA)

Figure 13   4 REF academic FTEs by discipline*



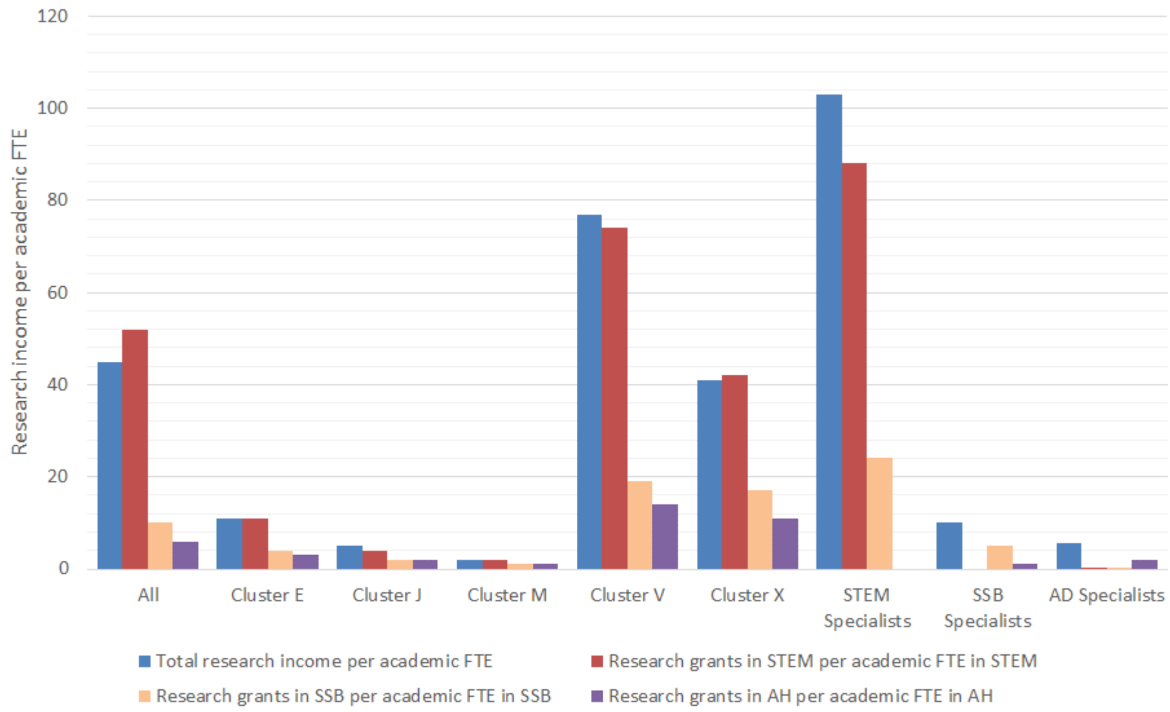
Source: Higher Education Statistics Agency (HESA)

Figure 14   Research income by partner type



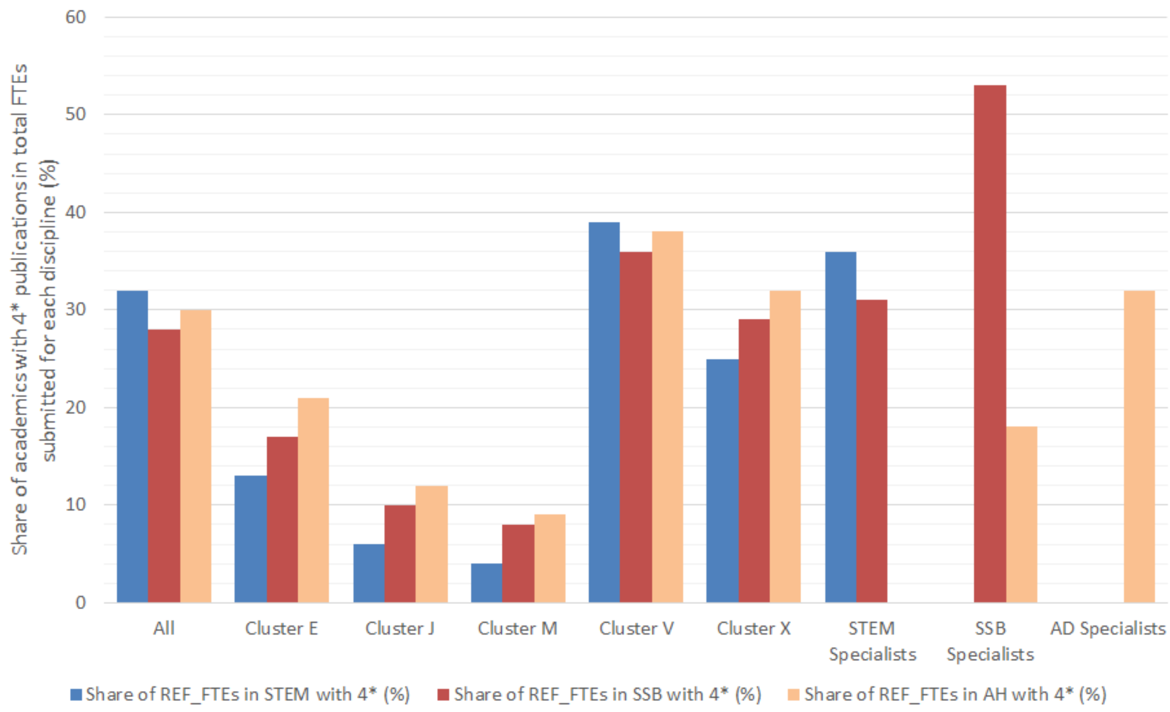
Source: Higher Education Statistics Agency (HESA)

Figure 15 Research intensity by discipline (£000s income per academic)



Source: Higher Education Statistics Agency (HESA)

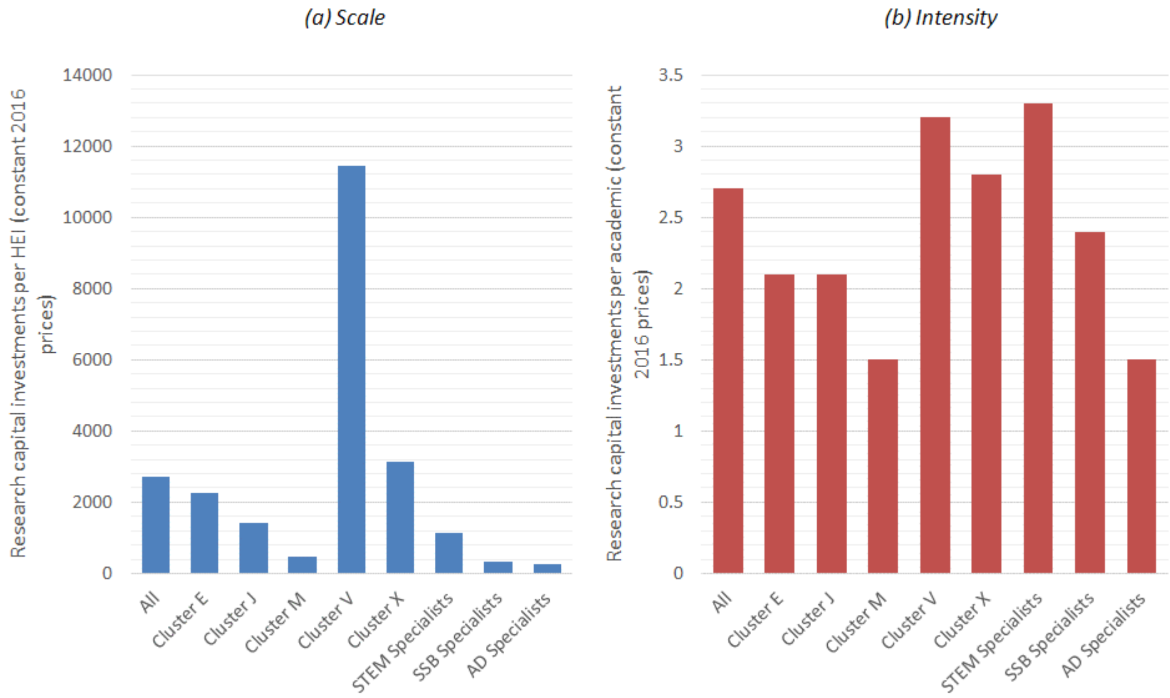
Figure 16 Intensity of academic FTEs gaining 4* in REF by discipline (share of academic FTEs submitting to REF)



Source: Higher Education Statistics Agency (HESA)

Dimension 3: Physical asset development

Figure 17 Research-related physical capital investments: scale and intensity



Source: Higher Education Statistics Agency (HESA)