TOMAS COATES ULRICHSEN

Knowledge Exchange Clusters of English Higher Education Providers: An Update

A Technical Report for Research England

May 2023

About the Author

Tomas Coates Ulrichsen is Director of the Policy Evidence Unit for University Commercialisation and Innovation (UCI) at the University of Cambridge. He set up the Unit in 2020 with an ambition to create much needed capacity in the UK to improve the data, evidence, and insights available to support the work of policymakers and university leaders in developing more effective university approaches for commercialising research and driving innovation in the economy. He is an expert adviser to governments on issues affecting knowledge exchange and research commercialisation, and over the past decade his work has been instrumental in informing UK policies and funding programmes for knowledge exchange, in particular the Higher Education Innovation Fund (HEIF). In 2018 he developed the original clusters of English universities that currently underpin the Knowledge Exchange Framework.

Prior to setting up UCI, he was a researcher at the University of Cambridge Centre for Science, Technology and Innovation Policy. Prior to that he was an Assistant Director of a leading UK economic development consultancy, Public and Corporate Economic Consultants (PACEC), where he directed and managed a range of strategic economic development and policy evaluation projects for public policy clients in the area of knowledge exchange, innovation and regional economic development. This included leading the first major evaluation of the Higher Education Innovation Fund (HEIF) in England.

Contents

1 Introduction and background

In 2018 a cluster analysis of English Higher Education Providers (HEPs) was undertaken to inform the development of the Knowledge Exchange Framework (KEF) and aid fair comparisons of knowledge exchange (KE) performance (Ulrichsen, 2018). Based on data from 2014 to 2016, it grouped non-specialist HEPs into five clusters using a statistical method that captured differences in key structural characteristics likely to shape their opportunities for KE. Specialist HEPs were assigned into three groups using a set of heuristics: specialist providers focusing on the arts and design; science, technology, engineering and mathematics; and those focusing on social sciences, business and the humanities.

Following a consultation with the sector the clusters went on to form the basis for the first iteration of the KEF, launched in March 2021, and its second iteration, KEF2, in September 2022. The clustering of HEPs in this way has allowed us to both reveal and celebrate the diversity of HEPs operating within England delivering different types of KE that unlock different types of economic and societal impacts, and examine differences in KE performance between broadly similar types of HEPs.

Five years on since the original KE cluster analysis, Research England requested an update to ensure the assignment of HEPs to a particular cluster reflects the latest data, including the results from the most recent Research Excellence Framework (REF) assessment published in 2021.

This technical report presents the updated cluster analysis and the assignment of each eligible HEP to a cluster based on the latest data available. It deliberately leaves the conceptual framework, methodology and metrics unchanged, to focus on identifying those HEPs that switch cluster based on changes to their structural characteristics likely to shape the nature and scale of potential KE opportunities available to them[1].

The 2023 analysis focuses on English HEPs eligible for Research England KE funding, and appropriate for inclusion in the exercise[2]. Should further HEPs become eligible for RE KE funding, we may need to revisit the clustering exercise and approach to ensure fair comparisons for all eligible institutions. The focus of this exercise is limited to England.

The report is structured as follows. We first reproduce the conceptual framework (section 2) and clustering method (section 3) developed for the 2018 cluster analysis. While these remain

[1] The conceptual framework, methodology, data underpinning the 2018 cluster analysis, and the resulting cluster membership were consulted on with the English HEP sector (Research England, 2019). This included the decision that the specialist cluster of social science, business and humanities HEPs was too small to lead to meaningful KE performance comparisons within this type of HEPs. As part of the consultation HEPs in this cluster were manually assigned other clusters. This affected three HEPs, with Bishop Grosseteste University and University College Birmingham being allocated to cluster M, and the London Business School allocated to cluster V.

[2] Note that the University of London (institutes and activities within its organisation, as opposed to the HEPs that are affiliated with it) is eligible for KE funding. Research England has agreed directly with the provider that its inclusion in the KE clustering exercise would not be appropriate due to the unique characteristics and structure of this provider. Further information on the structure of University of London can be found here: https://www.hesa.ac.uk/support/providers/uol-members, accessed on 4th May 2023.

unchanged, we present them here to aid the reader, ensuring the framework, methods, data and cluster analysis are all in a single document. Section 4 then presents the results of the updated 2023 cluster analysis and the cluster assignments for each of the HEPs. We look at any changes between the 2018 and 2023 exercises. Section 5 summarises the findings and some issues moving forward.

2 Conceptual underpinnings

This section largely replicates section 2 from the 2018 KE cluster analysis report (Ulrichsen, 2018).

HEPs 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). Simple rankings of HEPs typically compare metrics of 'performance' with no attempt to control for structural differences between HEPs. However, diversity within the HE system should 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 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 HEPs 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 HEPs rather than across groups. Cross-group comparisons can also be very useful, but should be limited to examine how different *types* of HEPs are contributing to the innovation system, or to enable individual institutions to explore the practices and performance of HEPs undertaking different types of functions in the system.

Early attempts to cluster HEPs 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 HEPs that differentiated institutions *along the key dimensions that were important to its work* and that this limited their ability to make appropriate recommendations 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 HEPs 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 HEPs 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 HEP 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 including with significant activities in clinical medicine and engineering, 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 HEP (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 HEPs’ knowledge capabilities, evidence has shown that different sectors demand knowledge from different combinations of disciplines (Cohen et al., 2002; Hughes et al., 2022). Furthermore, Hughes et al. (2016) showed that while KE was prevalent across the variety of different disciplines, particular mechanisms such as technology transfer through spinouts and intellectual property (IP) licensing were much more prevalent to specific areas such as life sciences and engineering.

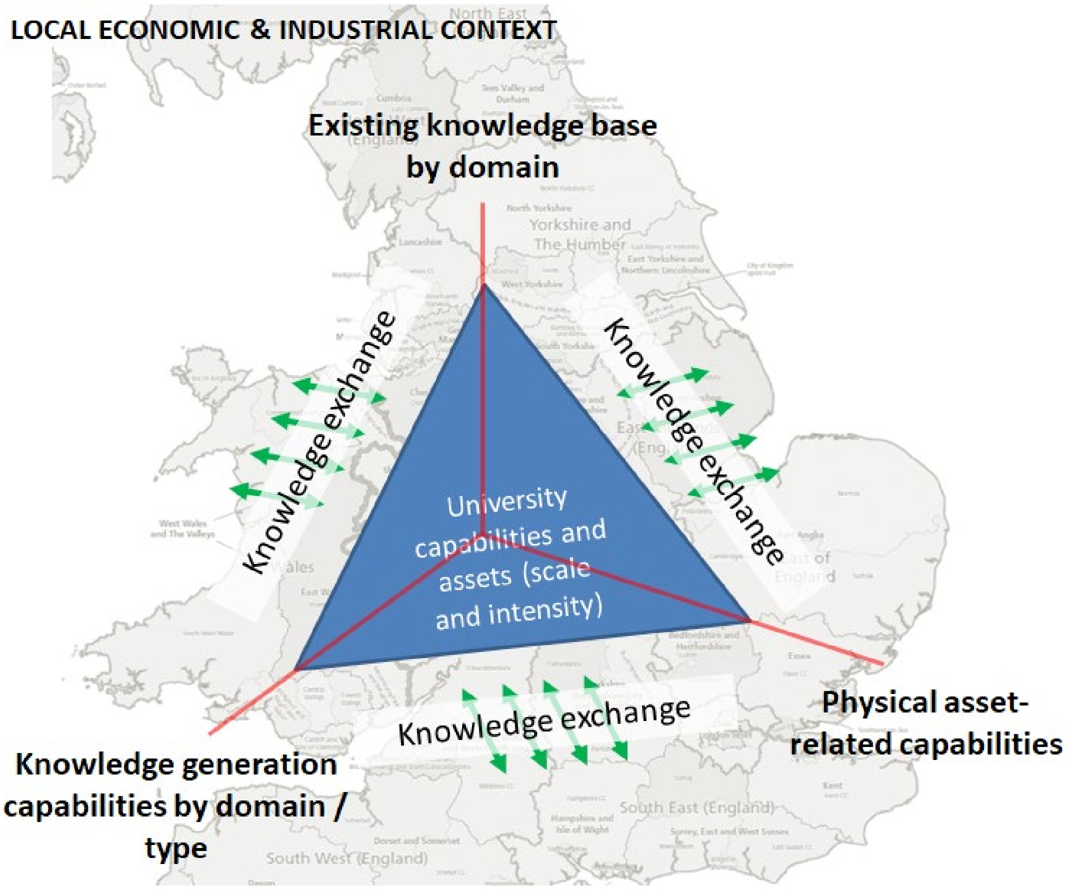
We also know that HEPs of all types – research-intensive, teaching-intensive and specialists – engage in wide varieties of KE from contract and collaborative research, to consulting, provision of training, and the provision of testing and other facilities and equipment related services. These KE activities engage partners across the public, private and third sectors, and with civil society. In addition HEPs 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 HEP can foster, investments in physical (and increasingly in virtual) capital can provide important platforms on which these social networks can develop. Crucially KE engagements draw from both new knowledge generated through research as well as from the existing knowledge held within HEPs that can be deployed to address an external partner’s needs. As such the knowledge capabilities of HEPs need to cover both their knowledge generation aspects as well as the existing knowledge held within the institution.

The scale of HEPs is also believed to shape their KE opportunities (Howells et al., 2008; Ulrichsen, 2014). Larger HEPs 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 (HEPs, 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 HEPs 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 HEP can pursue absent of significant changes to this base; i.e. they shape the KE *potential* of an HEP. These opportunities are additionally shaped by the scale of the HEP and the local economic context within which it 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 HEPs 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

The 2023 KE cluster analysis follows the same method as that developed for the 2018 exercise (Ulrichsen, 2018, section 3).

To identify groups of similar HEPs in terms of their structural 'capability base' likely to shape the nature and scale of their KE opportunities, we 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 similar HEPs (HEFCE, 2009; Howells et al., 2008; Ulrichsen, 2018), or similar firms within an industry in the strategic management field (see e.g. 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 HEPs to groups, and focuses on revealed differences based on the 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 HEPs. 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 broadly be categorised into hierarchical and non-hierarchical methods, 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 HEP cannot move cluster once assigned. The solutions can also be unstable to dropping observations particularly where sample are sizes small.

Non-hierarchical methods provide 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 based on some measure of distance. 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 cluster membership is achieved. This is a key advantage, resulting in these methods being 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, while observations are able to move cluster, they tend not to move to distant clusters. This makes 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 HEP 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 this 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 (for example to a mean of zero and standard deviation of one) to account for differences in variable scales (Ketchen and Shook, 1996).

HEP activity – particularly around knowledge generation – is also highly skewed, with a relatively small number of HEPs generating much of the activity. Left unattended, this can lead to challenges in discriminating HEPs with less of the particular activity. Following the practice in other cluster analyses of HEPs such as the Carnegie Classification of institutes of higher education in the United States[3], it is helpful to transform the data in order to deal with this issue. One method is to log-transform the data; another used in the Carnegie Classification method is to run 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. Many options exist here, for example the 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

The variables and data that underpin the 2023 cluster analysis replicate those used in the 2018 exercise. Data had to be available annually for all HEPs in England to be included in the analysis, and could be easily accessible for others to replicate the study. This resulted in the primary source of data being that available through the Higher Education Statistics Agency (HESA) as part of Jisc (hereafter referred to as 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 HEPs. This is largely held within the academic and research staff populations, 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 lead to different KE opportunities with different parts of the economy and society. It is therefore important to capture differences not just in the scale of staff and student populations, but also in the disciplinary portfolios of existing knowledge across HEPs. The variables selected are summarised in Table 1.

[3] http://carnegieclassifications.iu.edu/, accessed on 4th May 2023

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 HEPs	• 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 HEPs. 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 HEPs 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. HEPs 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 – crudely – that 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 HEPs. The intensity helps to distinguish which HEPs 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 HEP 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 REF2021)	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	• 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 REF2021 by STEM, SSB, AH	HESA, REF2021
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 HEP's digital media suite[4]. It proved very challenging to identify decent proxies for the scale and intensity of physical assets available within an HEP 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[5]. As such, we focused on the amount and intensity of investments made by an HEP 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 research capital spending (spend per academic FTE)	HESA

3.2 Approach

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

Eligible population of HEPs

- Working with Research England we identified 139 English HEPs for inclusion into the cluster analysis. This covers all HEPs eligible for their core KE funding programmes.
- Of these, five HEPs were until very recently members of the Conservatoire of Dance and Drama and now operate independently:
    - Central School of Ballet Charitable Trust Ltd
    - National Centre for Circus Arts
    - Contemporary Dance Trust Limited
    - Rambert School of Ballet and Contemporary Dance
    - Northern School of Contemporary Dance

  As a result, they lack sufficient data to be formally included in the data-driven cluster analysis. However, it is clear, both from their previous membership of a specialist arts institution, and a review of their websites, that they would be classed as specialist institutions. Given these HEPs remain eligible for KE funding, we suggest they are manually allocated to the specialist arts cluster.

[4] HEBCI Section B, Table 2 guidance, available at https://www.hesa.ac.uk/collection/c22032/hebci_b_table_2, accessed on 4th May 2023

[5] This distinction follows the HEBCI Section B, Table 2 guidance

- The London Academic of Music and Dramatic Art (LAMDA) and the Royal Academy of Dramatic Art (RADA) were previously members of the Conservatoire of Dance and Drama, and became independent in 2019. They now report data independently to HESA, with two years of data now available.
- As in 2018, HEPs in Scotland, Wales and Northern Ireland were out of scope for this study.

Identified and categorised specialist institutions using heuristics:

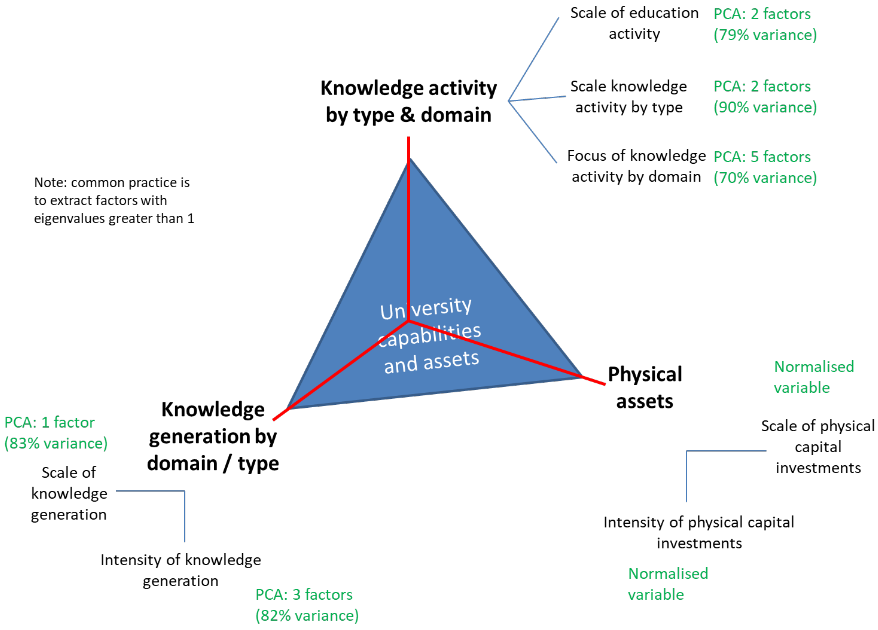
- Assessed degree of concentration of academic activity across the 45 discipline groups provided by HESA using:
  - The Herfindahl Index (which measures concentration) based on the number of full time equivalent academics in different disciplines of greater than 0.4

  OR
  - 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**: STEM-focused, which could be further divided into those focusing on biological and veterinary sciences, engineering and technology, and agriculture; (ii) **SSBH**: social sciences (including business), education and humanities; and (iii) **Arts**: creative and performing arts and design.
- The Arts University Bournemouth – was classified manually into the specialist arts group. 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 2 years of data available for LAMDA and RADA was sufficient to assign these institutions into the Arts specialist group.
- The National Film and Television School (NFTS) was classified as an Arts specialist in the 2018 exercise. This was retained for the current cluster analysis.

Prepare data for cluster analysis of broad-discipline HEPs

- Identified specialist HEPs (N=35) and separated from the sample
- The University of London (institutes) was removed from the sample due to its unique characteristics in the English HE system. This resulted in 99 HEPs 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 HEPs more fully at the lower ends of the distributions of variables where many HEPs are clustered.
- Ran a principal component analysis (with oblique rotation and kaiser normalisation) on variables within each dimension to deal with high correlations between variables. This process results in standardised components being used in the model. 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
- As with the 2018 cluster analysis, the model was run using log-transformed variables
- 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 HEPs which is helpful.

Cluster stability

- Examined the stability of the clusters by randomly removing 10 HEPs (approximately 10% of the sample) and repeated the two-stage cluster analysis. I then used the Adjusted Rand Index[6] 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. This was repeated 50 times, each time removing a different 10 HEPs from the sample.

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

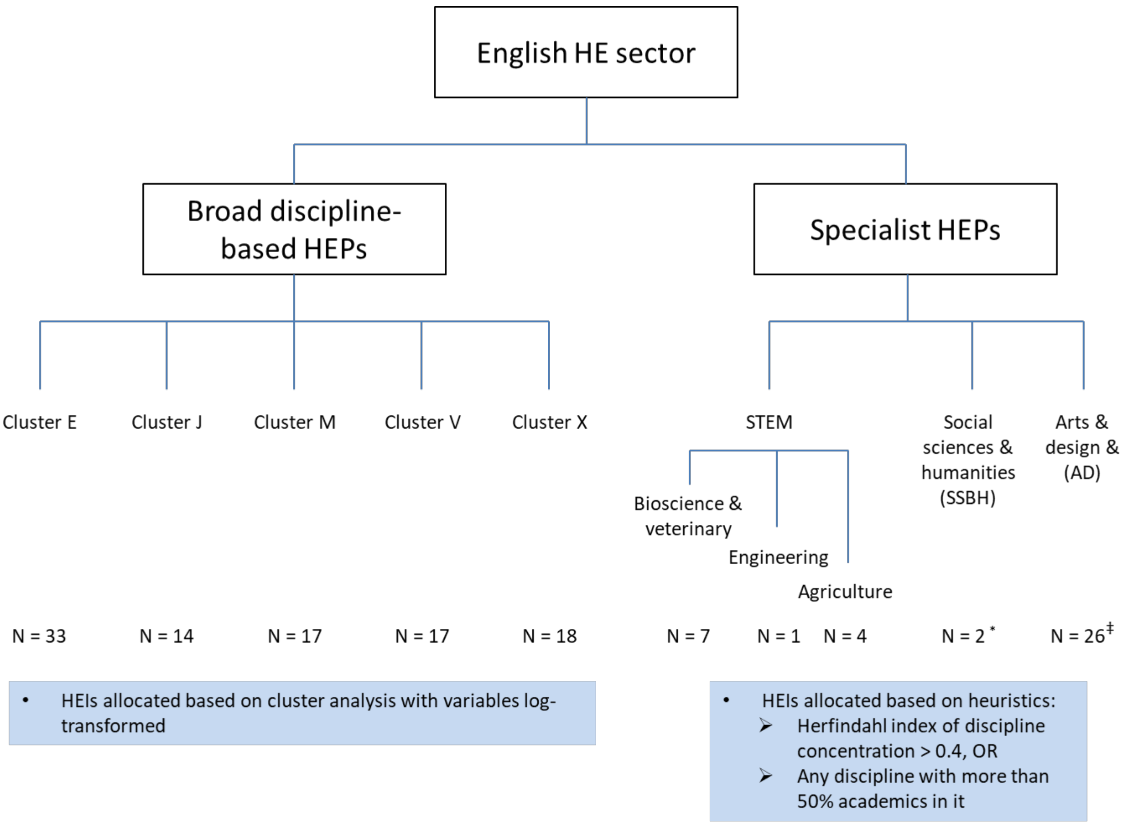
4 The updated knowledge exchange clusters 2023

The statistical cluster analysis allocated HEPs to one of five clusters (Figure 3 and Figure 4) based on similarities across a set of quasi-fixed structural characteristics likely to shape the nature and scale of potential KE opportunities available to them. Note that 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*. The cluster sizes range from 14 (cluster J) to 33 (cluster E).

As outlined in section 3.2, and replicating the 2018 exercise, we identify specialist HEPs and allocate them to one of three clusters based on heuristics (Figure 5). This results in:

- 12 HEPs in the specialist STEM cluster
- 2 HEPs in the specialist social sciences, business and humanities cluster. Given the very small size of this cluster, we suggest these HEPs retain the manual assignments to other clusters as agreed as part of the 2019 consultation exercise.
- 26 HEPs in the specialist arts and design cluster. This includes the five HEPs that were previously part of the Conservatoire of Dance and Drama. As they continue to be eligible for RE KE funding, we suggest they are manually assigned to the specialist arts cluster.

Figure 3 Segmentation of English HE sector by cluster



Notes:
* It is suggested that HEPs in this cluster are manually assigned to the clusters agreed as part of the 2019 consultation exercise
‡ This includes the five HEPs that were previously part of the Conservatoire of Dance and Drama. As they continue to be eligible for RE KE funding, it is suggested they are manually assigned to the specialist arts cluster

Figure 4 Cluster membership (log transformed variables)

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

Note: HEPs highlighted in green have changed cluster compared with the 2018 exercise. The letter in parentheses is the 2018 cluster assignment.

Figure 5 Cluster membership: specialist institutions

	Specialists: STEM *	Specialists: Social sciences & humanities	Specialists: Arts & design		
BIO/VET	AECC	Bishop G'teste *(suggest retaining manual assignment to cluster M)*	Arts B'mouth	Cen Sch of Ballet	*HEPs are eligible for KE funding but lack sufficient data. Given sufficient evidence of specialist arts focus, suggest manually assigning to this cluster*
	Brit Osteopathy	L'don Business *(suggest retaining manual assignment to cluster V)*	Courtauld	Nat Cen for Circus Arts	
	ICR		Creative Arts	Contemporary Dance	
	Liver Trop Med		National Film	Rambert School	
	Sch of Hygiene		Guildhall	Northern School of Contemporary Dance	
	Royal Vet Coll		LAMDA		
	St George's		Leeds Art		
ENG	Cranfield		Leeds Conservatoire		
AGR	Harper Adams		Liver Perf Arts		
	Hartpury Uni		Arts London		
	Royal Agr Coll		Norwich Arts		
	Writtle		Plymouth Art		
			RADA		
			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 but manually switched to the arts group following the sector consultation

Figure 4 highlights in green those HEPs that have moved cluster compared with the 2018 exercise (with their 2018 cluster assignment in parentheses). Detailed inspection of the data suggests that most of the HEPs that switched cluster were situated near the 'periphery' of their previous cluster. Changes to their structural characteristics between the two data periods (2014-16 and 2019-21), as captured by the cluster variables, result in their placement in a different cluster.

Table 4 captures the scale of movement of HEPs between clusters. Overall 12 HEPs move clusters (excluding London Business School and Bishop Grosseteste University that were manually assigned in 2018 to clusters V and M respectively due to the SSBH cluster being too small). Clusters M, V and X see few changes. The updated cluster E gains five HEPs from the 2018 cluster J. Cluster J, while losing five HEPs to cluster E, gains one HEP from each of the 2018 clusters X and E, and two from cluster M. The arts and STEM specialist clusters remain the same as those used in the first two iterations of the KEF.

Table 4 Changes to cluster membership across the HEP sample

		KE cluster 2023							
		V	X	E	J	M	STEM	ARTS	SSBH
KE cluster 2018	V	16							1*
	X	1	18		1				
	E			28	1				
	J			5	10	2			
	M				2	15			1†
	STEM						12		
	ARTS							19	
	n/a							2‡	

* London Business School (manually assigned to cluster V in 2018 – we suggest this is retained following this exercise)
† Bishop Grosseteste University (manually assigned to cluster M in 2018 – we suggest this is retained following this exercise)
‡ LAMDA and RADA were both manually added to the Arts cluster since the 2018 exercise

Robustness of clusters

The robustness of the clusters was examined by randomly removing 10% of HEPs from the sample and the cluster analysis re-run. The similarity of the resulting clusters are then compared against the baseline full sample using the Adjusted Rand Index (ARI) (Hubert and Arabie, 1985). A value of 1 indicates a perfect match while 0 indicates no overlap in membership at all. This was repeated 50 times. The average ARI was 0.875. The full ARI results for each run are presented in Appendix B.

Cluster characteristics

Table 5 summarises the key characteristics for each of the different broad-discipline HEP clusters along the key dimensions of the framework. Appendix A provides a detailed analysis of cluster characteristics.

Table 5 Cluster characteristics for broad-discipline HEPs

Cluster	Characteristics
Cluster E	• Large universities with broad discipline portfolio across both STEM and non-STEM excellent research across all disciplines. • Many academics have both a teaching and research focus or teaching only focus • Significant amount of research funded by government bodies/hospitals (cluster average 45%); 9% from industry and 12% from charities. • Large proportion of part-time undergraduate students. Smaller postgraduate population dominated by taught postgraduates.
Cluster J	• Mid-sized universities with a more teaching focus (although research is still in evidence). • Academic activity across STEM and non-STEM including other health, computer sciences, social sciences and humanities • Research activity funded largely by government bodies/hospitals (41%) and charities (20%); 9% from industry. • Smaller postgraduate population dominated by taught postgraduates.
Cluster M	• Smaller universities, often with a teaching focus. Few research-only academics • Academic activity across disciplines, particularly in other health domains, social sciences and humanities. • Research activity typically funded by non-UKRI sources, covering government bodies/hospitals (38%) and industry (27%); 14% from charities.
Cluster V	• Very large, very high research intensive and broad-discipline universities undertaking significant amounts of excellent research. • High proportion of research-only academic staff • Research funded by range of sources including UKRI (34%), other government bodies (26%), charities (24%) and industry (11%). • Significant activity in clinical medicine and STEM disciplines • 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 excellent research. • High proportion of research-only academic staff • High proportion of research funded by UKRI (45%); 29% from other government bodies; 8% from industry and 12% from charities. • Discipline portfolio balanced across STEM and non-STEM although less or no clinical medicine activity. • Large proportion of taught postgraduates in student population

5 Summary and moving forward

This technical report updates the 2018 cluster analysis of English HEPs that produced the clusters that currently underpin the KEF to aid fair comparisons of KE performance across the sector. The aim of the exercise is to ensure that the assignment of HEPs to a cluster reflects the latest data, including the results from the most recent Research Excellence Framework (REF) assessment published in 2021. The updated cluster analysis deploys the same conceptual framework, method and metrics as used in 2018. This allows us to focus on identifying HEPs that might move cluster based on changes in their structural characteristics likely to shape the nature and scale of potential KE opportunities available to the HEP.

To facilitate comparisons with previous years, the updated 2023 cluster analysis focuses on allocating non-specialist HEPs to one of five clusters using statistical methods, and specialist HEPs to one of three groups using heuristics. Overall the characteristics of the five non-specialist clusters remain largely similar to those identified in the 2018 analysis. At the HEP-level, of the 134 English institutions included in the analysis, just twelve move cluster, with clusters E and J seeing most changes. Where changes happen, a detailed inspection of the data suggested they were near the 'periphery' of their 2018 cluster. Changes to their structural characteristics – as captured by the metrics that underpin the cluster analysis – between 2014-16 and 2019-21 result in them exhibiting greater similarities with a different group of HEPs in a neighbouring cluster.

Once again, the SSBH cluster is too small to enable meaningful comparisons as part of the KEF and we suggest retaining the manual assignment of these HEPs to other clusters as agreed as part of the 2019 consultation exercise with the sector.

Looking forward, the composition of the English HEP system continues to evolve, with new HEPs occasionally entering and leaving. In addition, Research England recently announced a strategic ambition to develop the national capability in KE data, evidence and metrics. As this takes shape, with the hope for new and improved data on KE and the underlying drivers and conditions that influence opportunities, it may be sensible not just to re-run the existing cluster model, but to update both the conceptual and statistical model, and how we treat HEPs with very different levels of experience in delivering KE. This will ensure that our efforts to identify similar groups of HEPs to inform benchmarking exercises such as the Knowledge Exchange Framework are as fair as possible and based on the latest thinking, evidence and data.

References

Abramovsky, L., Harrison, R., Simpson, H., 2007. University Research and the Location of Business R&D. Econ. J. 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. Environ. Plan. 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. Manag. Sci. 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? Res. 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 High. Educ. Inst. Reg. Econ. Initiat. Httpewds Strath Ac UkPortals8typology Doc.

Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classif. 2, 193–218. https://doi.org/10.1007/BF01908075

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

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. Camb. J. Econ. 36, 723–750.

Hughes, A., Kitson, M., Salter, A., Angenendt, D., Hughes, R., 2022. The Changing State of Business-University Interactions in the UK 2005 to 2021. National Centre for Universities and Busines and Univerity of Cambridge Centre for Business Research, London, UK.

Hughes, A., Lawson, C., Kitson, M., Salter, A., 2016. The Changing State of Knowledge Exchange: UK Academic Interactions with External Organisations 2005-2015. National Centre for Universities and Business, London, UK.

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

King, J., 1970. The topology of universities. High. Educ. Rev. 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. Reg. Stud. 50, 1537–1550. https://doi.org/10.1080/00343404.2015.1041370

Lester, R., 2005. Universities, innovation, and the competitiveness of local economies. Summ. Rep. Local Innov. Syst. Proj. Phase Mass. Inst. Technol. Ind. Perform. Cent. Work. Pap. Ser.

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

Molas-Gallart, J., Salter, A., Patel, P., Scott, A., Duran, X., 2002. Measuring third stream activities. Final Rep. Russell Group Univ. Brighton SPRU Univ. 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. Res. 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. Res. Policy 42, 423–442. https://doi.org/10.1016/j.respol.2012.09.007

Research England, 2019. Knowledge Exchange Framework Consultation. Research England, Bristol, UK.

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. Strateg. Manag. 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? Res. Policy 42, 1420–1430. https://doi.org/10.1016/j.respol.2013.06.003

Ulrichsen, T.C., 2018. Knowledge Exchange Framework Metrics: A Cluster Analysis of Higher Education Institutions. Cambridge, UK.

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

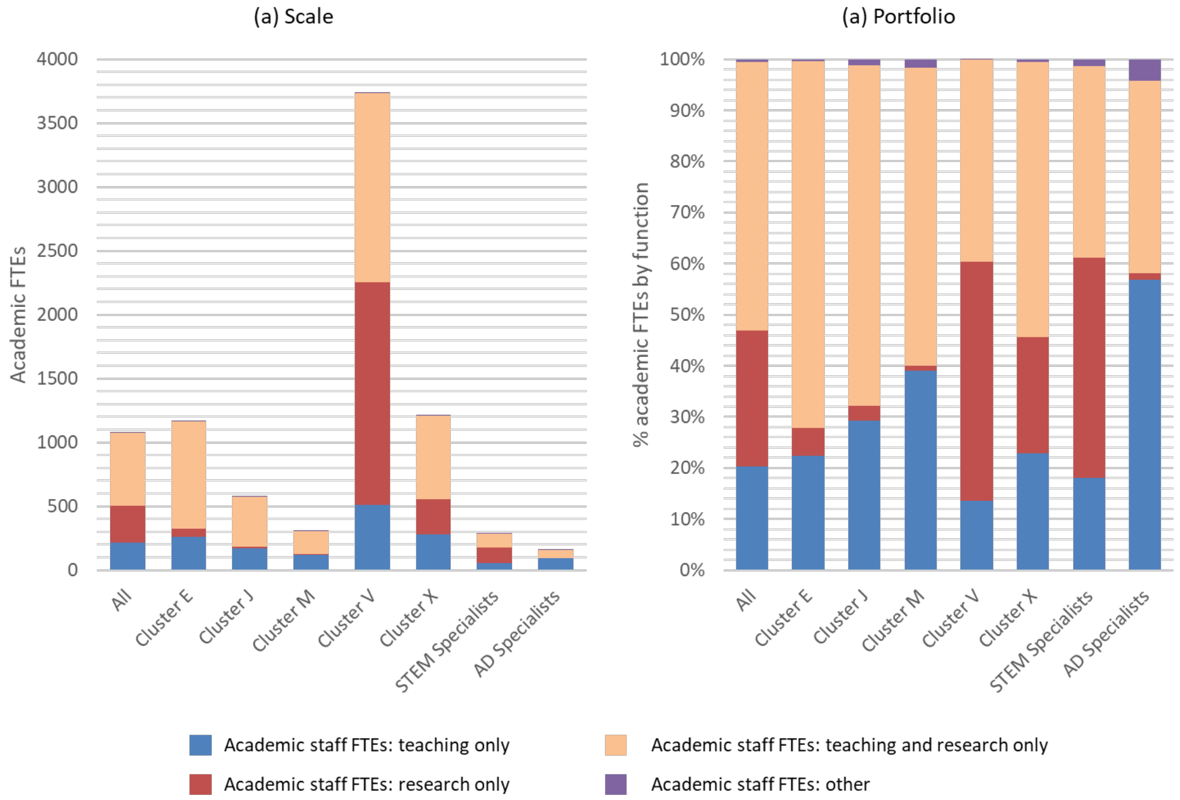
Ulrichsen, T.C., 2014. Knowledge Exchange Performance and the Impact of HEIF in the English Higher Education Sector. HEFCE, Bristol, UK.

Appendix A Cluster Characteristics

This appendix presents the detailed analysis of the structural characteristics for the HEP clusters presented in section 4.

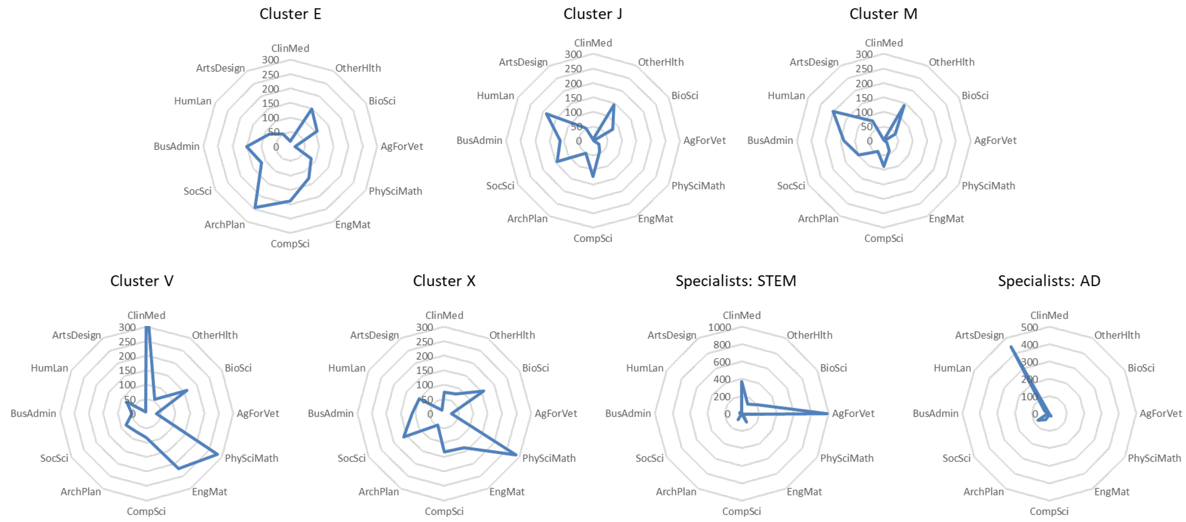
Dimension 1: Existing knowledge base

Figure 6 Academic staff (full-time equivalent, FTE) by function, averages for 2018/19 – 2020/21



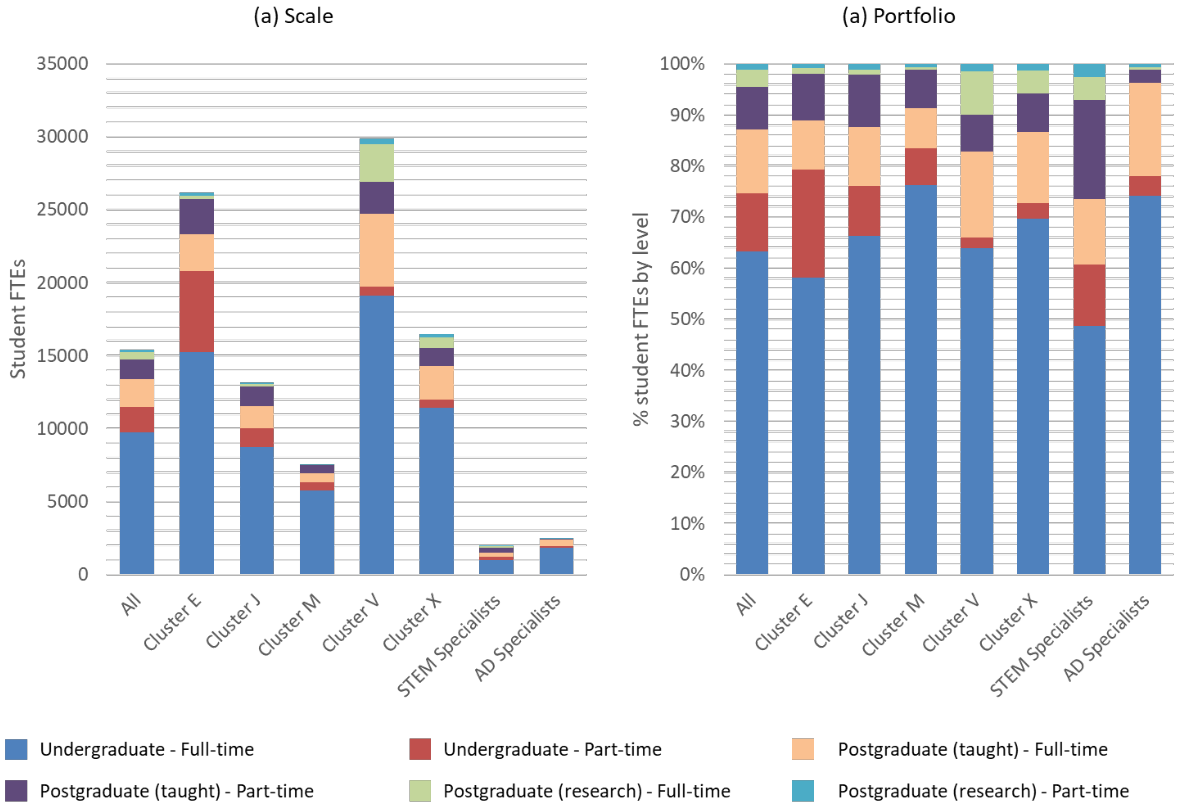
Source: Higher Education Statistics Agency (HESA)

Figure 7 Knowledge activity: disciplinary domains (average for England = 100), based on averages for 2018/19 – 2020/21



Source: Higher Education Statistics Agency (HESA)

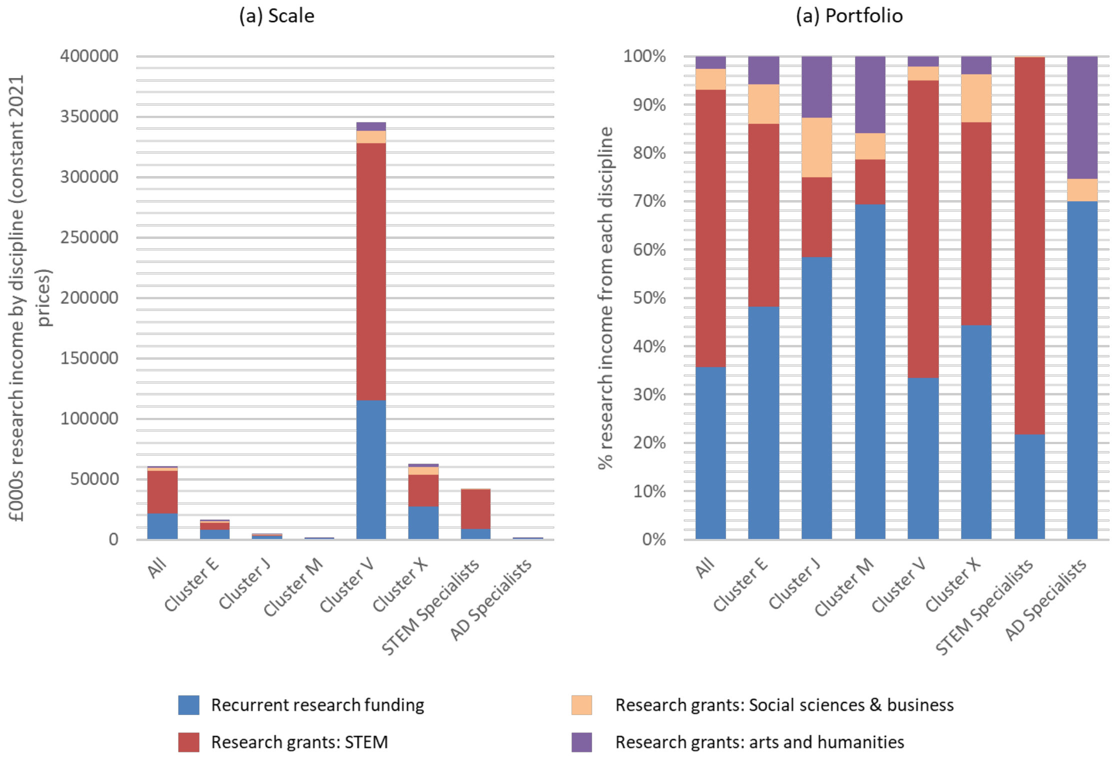
Figure 8 Education function by level, based on averages for 2018/19 – 2020/21



Source: Higher Education Statistics Agency (HESA)

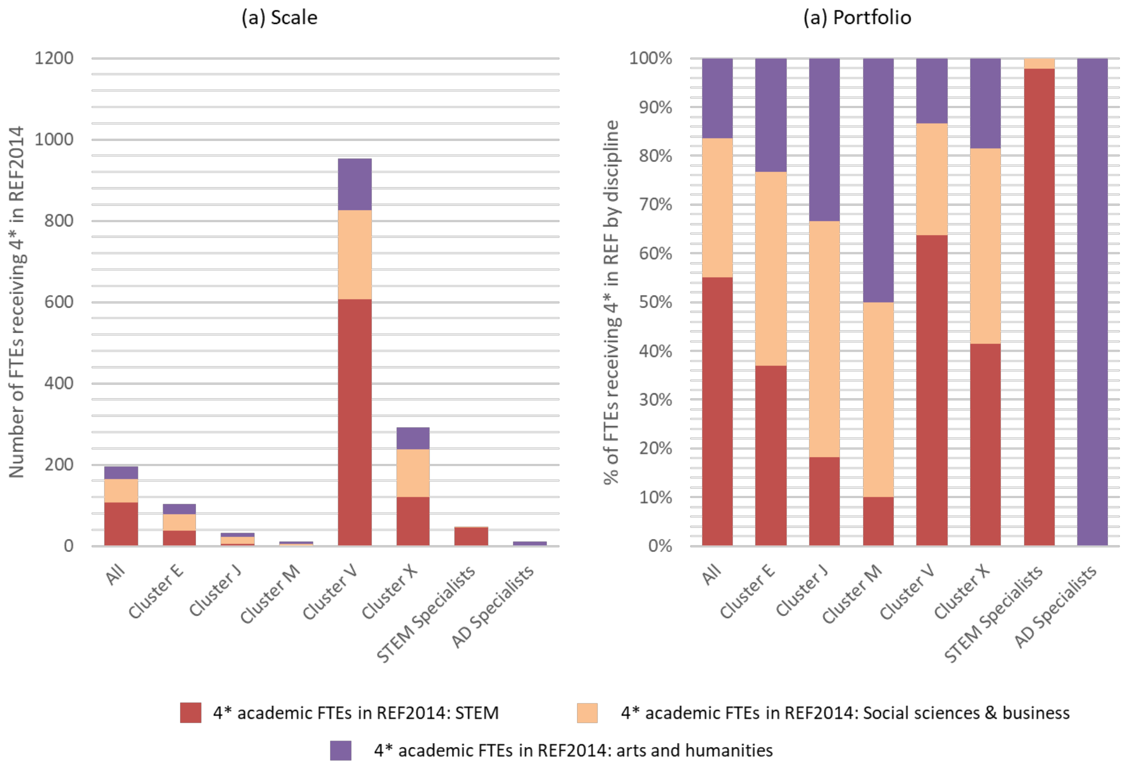
Dimension 2: Knowledge generation

Figure 9 Research income by discipline, based on averages for 2018/19 – 2020/21



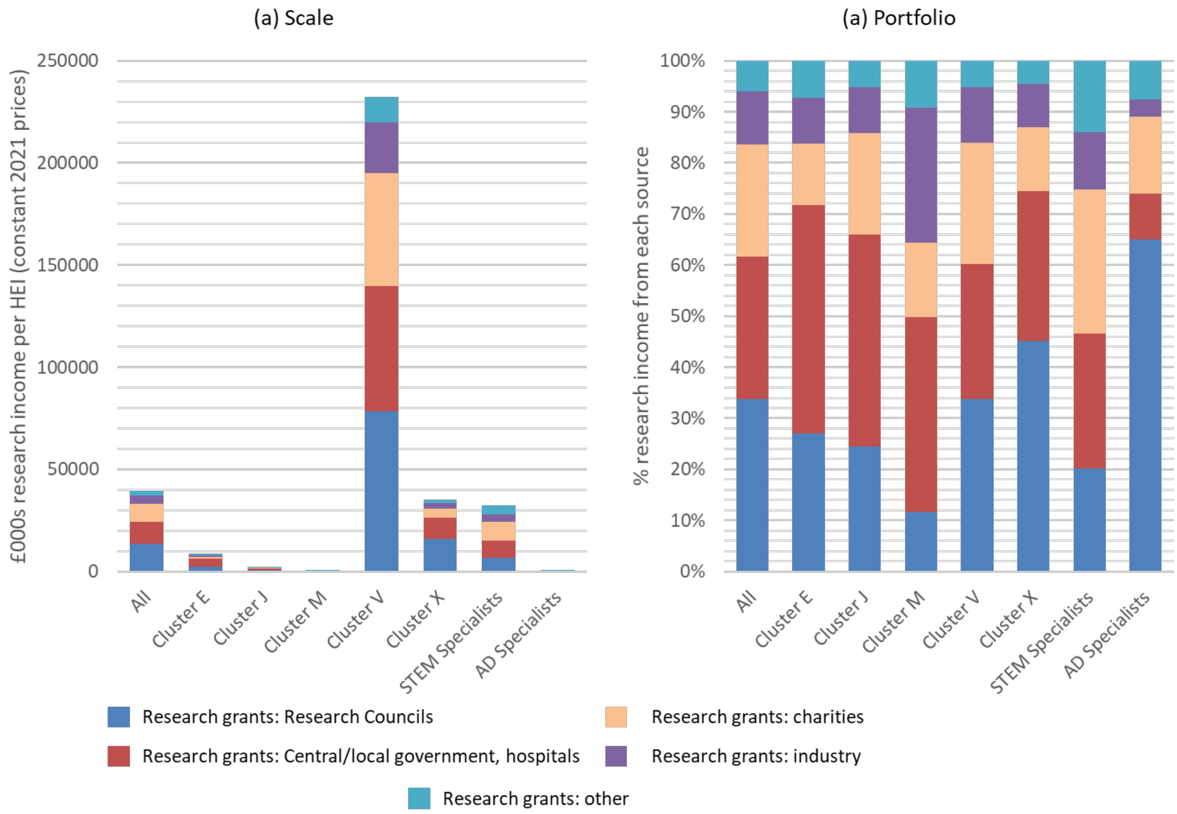
Source: Higher Education Statistics Agency (HESA)

Figure 10 4* REF academic FTEs by discipline, based on 2021 REF assessment



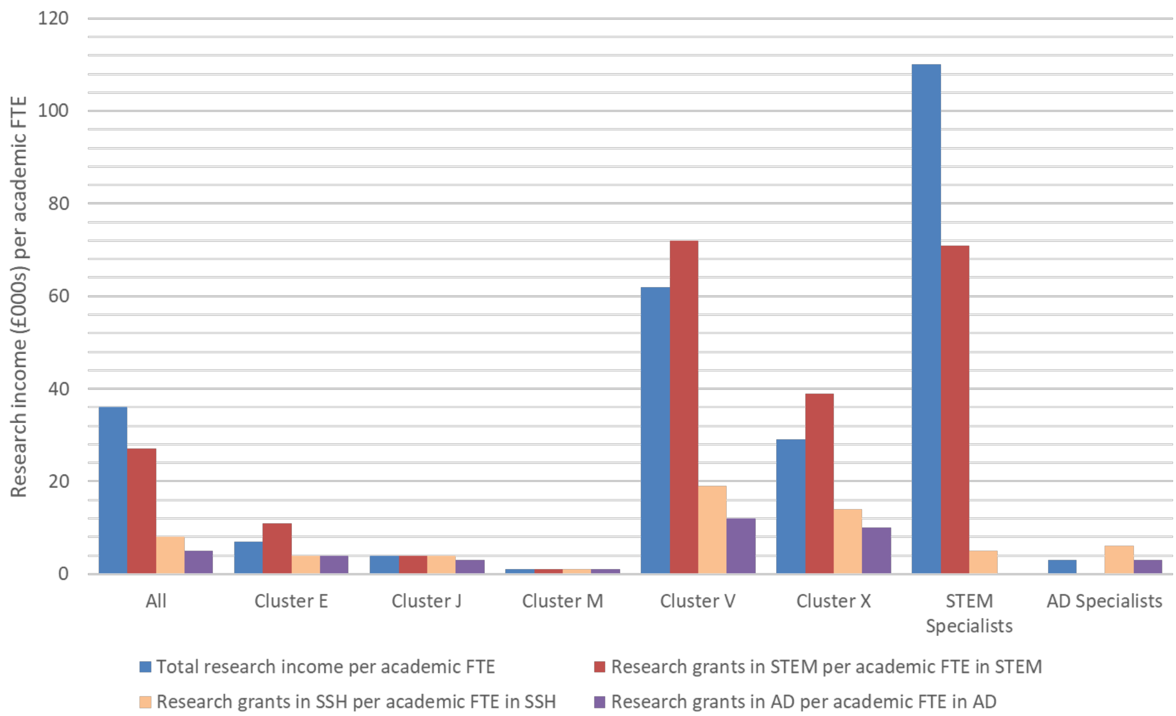
Source: Higher Education Statistics Agency (HESA) and REF2021 database (available at https://www.ref.ac.uk/)

Figure 11 Research income by partner type, based on averages for 2018/19 – 2020/21



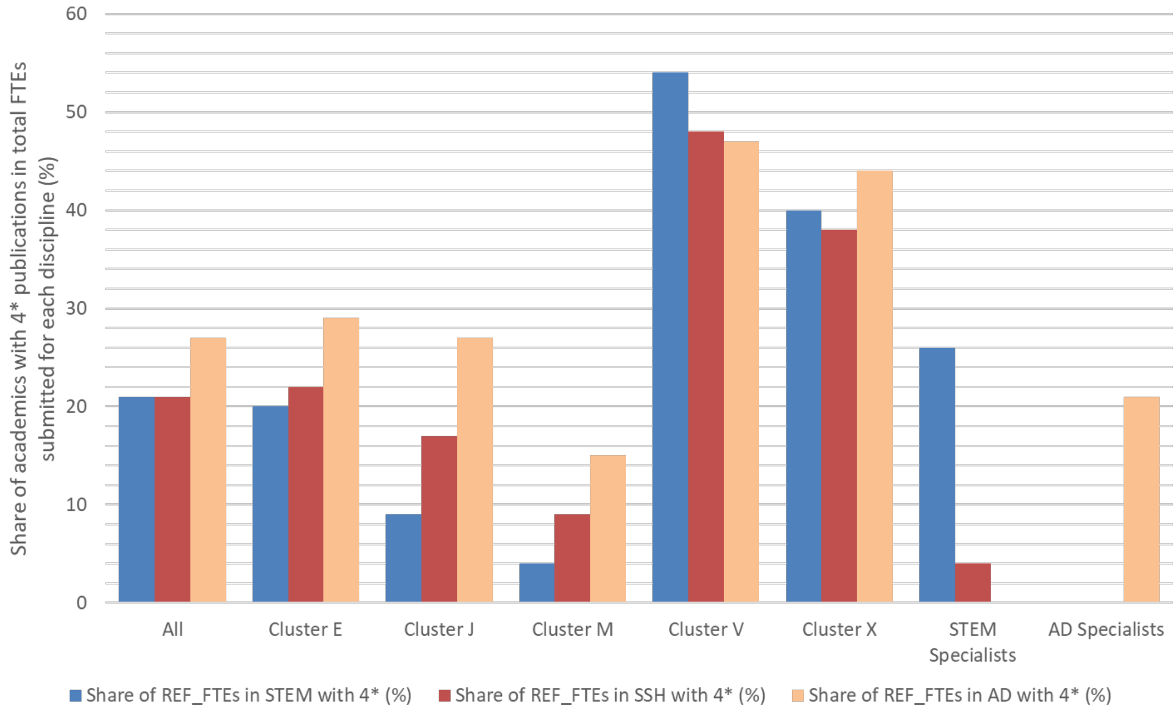
Source: Higher Education Statistics Agency (HESA)

Figure 12 Research intensity by discipline (£000s income per academic), based on averages for 2018/19 – 2020/21



Source: Higher Education Statistics Agency (HESA)

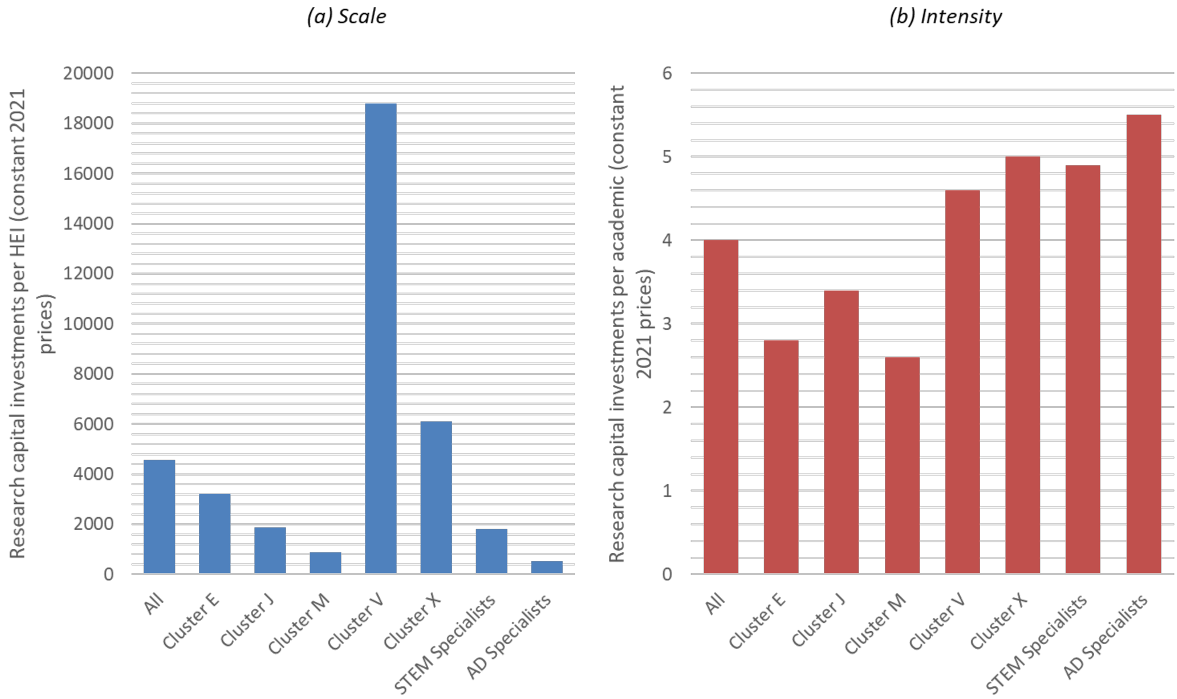
Figure 13 Intensity of academic FTEs gaining 4 in REF by discipline (share of academic FTEs submitting to REF), based on 2021 REF assessment*



Source: Higher Education Statistics Agency (HESA) and REF2021 database (available at https://www.ref.ac.uk/)

Dimension 3: Physical asset development

Figure 14 Research-related physical capital investments: scale and intensity, based on averages for 2018/19 – 2020/21



Source: Higher Education Statistics Agency (HESA)

Appendix B Robustness checks

Repetition	ARI
1	0.938
2	0.809
3	0.835
4	1.000
5	0.893
6	0.793
7	0.729
8	0.841
9	0.824
10	0.919
11	0.979
12	0.822
13	1.000
14	0.867
15	1.000
16	0.799
17	0.864
18	0.760
19	0.810
20	0.764
21	0.651
22	0.943
23	1.000
24	0.980
25	0.942
26	1.000
27	0.957
28	0.975
29	0.868
30	0.811
31	0.745
32	0.975
33	0.978
34	0.719
35	0.977
36	0.809
37	0.976
38	0.791
39	0.925
40	0.977
41	0.771
42	0.975
43	0.809
44	0.895
45	0.615
46	0.977
47	0.780
48	0.957
49	0.733
50	1.000
Average ARI	0.875