GEDII GENDER DIVERSITY IMPACT – Improving research and innovation through gender diversity

WWW.GEDII.EU

Anne Laure Humbert, Elisabeth Anna Guenther

Project Deliverable

D3.2 Measuring gender diversity in research teams: methodological foundations of the Gender Diversity Index

12th March 2018 – **Final**

Project Consortium

Universitat Oberta de Catalunya, Spain
Project Coordination
Dr. Jörg Müller
Dr. Milagros Sáinz
Dr. Rosa Borge
Dr. Julià Minguillón
Dr. Julio Meneses
Dr. Sergi Fàbregues

Hochschule Furtwangen, Germany
Prof. Dr. Ulrike Busolt
Sandra Klatt
Wiebke Kronsbein

Örebro University, Sweden
Dr. Ulf Sandström
Dr. Anne-Charlott Callerstig
Prof. Dr. Liisa Husu

Verband der Elektrotechnik Elektronik Informationstechnik e.V., Germany
Dr. Walter Börmann
Nicole Bienge
Andreas Landwehr
Annika Gereke

Oxford Brookes, UK
Dr. Anne Laure Humbert
Dr. Elisabeth Anna Guenther

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 665851.

Table of Contents

Introduction

A core objective of the GEDII project is to provide empirical evidence on the potential link between gender diversity within teams and research performance/innovation. This report builds upon a previous report (Humbert and Guenther 2017), which provided preliminary considerations and results in efforts to build the Gender Diversity Index. The Gender Diversity Index is a composite indicator that is applicable at team level and provides a summary measure of gendered processes. A composite indicator is obtained when individual indicators are compiled in a single measure on the basis of a multi-dimensional concept. For the Gender Diversity Index, this means aggregating several team level gender diversity measures into a coherent conceptual and statistical whole.

The rationale for building the Gender Diversity Index is to provide not only a descriptive account that enables teams, funders or other stakeholders to assess and monitor gender diversity, but which can also be used for further analysis. Following on from at least a decade of discussion into whether gender diversity is linked to organisational performance, the Gender Diversity Index can be used to test this relationship in the context of research in science, technology, engineering and math (STEM). A conceptual framework for gender diversity, its meaning and how it could be related to research performance and innovation were set out in D3.1 (Humbert and Guenther 2017). While D3.1 focused on theoretical considerations and their possible statistical implementation, the focus of this report is to present the statistical decisions that were taken in light of both theory and empirical material.

The development of the Gender Diversity Index was informed by the best-practice methodological guide developed jointly by the EC's JRC and the OECD (Nardo et al. 2008) and other resources provided by COIN – the Competence Centre on Composite Indicators and Scoreboards (see https://composite-indicators.jrc.ec.europa.eu/). It is important to remember that composite indicators can be powerful instruments but can also be misused. Because they provide apparently objective summary measure, they can provide valuable information. At the same time, composite indicators are distrusted because of this very reason: apparent lack of transparency and the over-simplification of complex issues are some of their main criticisms. From a gender theory perspective, composite indicators are also problematic in that they invariably rely on a binary understanding of gender through reliance on sex as a proxy. These limitations are acknowledged. Where possible, the analysis of sex-disaggregated data is embedded within a wider frame of gender statistics (UNECE 2010) that recognises unequal gendered power relations.

This methodological report describes the steps adopted in the construction of the Gender Diversity Index. First, a brief overview of the team survey is provided. This is followed by an account of what metrics were used and by the results of the multivariate analysis. Next, the aggregation steps are examined and an assessment of the robustness of the Gender

Diversity Index. The Gender Diversity Index is then described and exemplified in detail through some team cases, before some concluding remarks are provided.

Overview of Data Collection

The construction of the Gender Diversity Index is based on data collected via an online survey of research teams and their members in research institutes, universities and private for-profit companies. This section provides an overview of the survey and data collection (see D4.2 for a more detailed account).

The survey design relied on an extensive literature review that combined insights from team science, research performance and assessment and gender studies (Müller et al. 2016) as well as on a conceptual framework for the Gender Diversity Index (Humbert and Guenther 2017). The survey was complex in that it aimed at capturing team level information. As such it was important to (1) obtain information both at the individual and team level and (2) maximise the response rate within each team to provide reliable estimates. To this end, two questionnaires were designed: one for team members and another one for a team contact person, usually but not necessarily the team leader. The questionnaires were designed collaboratively within the GEDII consortium in over a dozen iterations. Moreover, extensive feedback from experts of the advisory board and a small pilot study in the UK, Spain and Germany contributed to the final version of the questionnaires. To obtain a high response rate, the consortium strived for a response time of a maximum of 15 minutes.

The members' questionnaire included questions on the respondents' characteristics, the organisation of work, team leadership (Berger et al. 2012), gender stereotypes (Miller and Borgida 2016), a short version of the Team Climate Inventory (Kivimaki and Elovainio 1999) and performance measures such as volume of publications and patents. The team contact person was also asked for general information about the team such as the number of women and men, the size of the organisation, policies within the organisation and area of work.

Once the questionnaires were finalised, they were entered in the online survey platform Unipark (https://www.unipark.com/). This is the academic branch of Questback, a for-profit online survey company, based in Germany. After the survey was completed, all data were erased from its servers. The initial geographical scope was restricted to five EU countries (Germany, Lithuania, Spain, Sweden, UK), each of which represents a different welfare and gender equality regime (Lewis 1992; Sainsbury 1996; Esping-Andersen 1990), and later completed by research teams in a further eight countries (Denmark, Finland, Italy, the Netherlands, Norway, Poland, Portugal and Switzerland). Furthermore, as STEM can encompass a wide breadth of subject areas, the project is restricted to biomedical engineering and transport research.

Research teams were identified through datamining the Web of Science and PATSTAT database to identify researchers that had published a paper or patent either in transport or biomedical engineering/biomedicine between 2001 and 2016. Based on this initial list of authors and patent holders, research teams were compiled using external information available publicly. This resulted in a list of research teams in each country and research area of interest which acted as the sampling frame for the survey. This list provided contact details for each team member, records of their publications and patents obtained from the datamining and where possible, the identified team leader. The compilation of information about research teams proved to be easier for research institutes and universities than for private for-profit companies, where there was less information about their staff on the web.

The administration of the survey followed two approaches. The first relied on being able to identify the team leader, who would then be sent an initial email, inviting them to participate in the survey. Where no information was available for a suitable contact person, this invitation was in some cases sent to another relevant contact such as the Equality and Diversity Officer in a HR Department. This was most often the case for private companies, where in addition phone calls were made to establish this initial contact. While this approach was somewhat successful in universities and public research institutions, however, it had limited success with private organisations. The second approach relied on sending out emails in bulk to all authors identified through the datamining process. This communication invited them and their teams to participate in the survey.

Once initial contact was established and the team agreed to participate in the survey, a team ID was generated. The contact person was asked to fill out the team-level questionnaire and to verify the list of team members. The teams were informed that an up-to-date list of team members would be needed to generate a bibliometric performance profile of the team, which was shared with them. The contact person could then opt to forward an email to their team members, including a link to the survey with a team ID in the URL, or alternatively for the team members to be contacted directly.

By 31 January 2018, 159 teams and 1357 individuals had completed the survey. For the computation of the Gender Diversity Index, to minimise the standard errors of estimates, only teams which met the following criteria were used: 100% response rate for teams with 4 respondents, 50% response rate for teams with 5 to 9 respondents, and 40% response rate for teams with 10 or more respondents (101 out of 159 teams).

The survey was used to provide indicators, distinguishing between gender diversity that is demographic, functional or cognitive (Table 1). Only indicators within the first two types of gender diversity were considered for the Gender Diversity Index, since these two aspects are most apt at capturing gender inequalities in research at the team level. Cognitive gender diversity will be examined for its mediating effects between demographic/functional gender diversity and performance in further analyses.

Table 1: Potential indicators to populate the Gender Diversity Index and further analyses

Demographic gender diversity	**Functional gender diversity**	**Cognitive gender diversity**
Age	Subject area	Communication
Ethnicity/minority status	Team tenure	Power and influence
Disability	Team time contribution	Team/organisational climate
Marital status	Team role	Trust
Care responsibilities	Type of contract	Leadership style
Education	Working time	Gender/diversity climate
	Experience	
	Tasks	

Source: based on Humbert and Guenther, 2017

The Gender Diversity Index relies on indicators of demographic diversity that encompass age, marital status, care responsibilities and education. The inclusion of ethnicity and disability were considered as variables for demographic gender diversity. However, due to the low number of members of an ethnic minority group and people with chronic health issues in the sample – as well as in the overall target group – these two variables could not be included. Age is a relevant indicator, not only because of the chrononormative notion of the 'ideal' scientist (Riach, Rumens, and Tyler 2014; Bourdieu 1994), but also due to the fact that women working in academia tend to be younger than men (EC 2015), suggesting that women are more likely to opt out and/or have been historically under-represented. Moreover, the Gender Diversity Index includes two indicators for family status: living together with a significant other and care responsibilities. Academia is considered a greedy institution (Wright et al. 2004; Cervia and Biancheri 2017) asking faculty to put in long working hours into a very demanding job. This creates a conflict – especially for women – to combine care responsibilities with the demands of the work place (Armenti 2004; Beddoes and Pawley 2014; Cervia and Biancheri 2017), leading to a lower fertility rate for women academics (Stanfors 2014) and women postponing childbirth up to the point that they dispense with having children at all (Metz-Göckel, Selent, and Schürmann 2010; Mason and Goulden 2004). Women academics tend to have a lower marriage rate and fewer children (Mason and Goulden 2004) and marriage, as well as the presence of young children reduces a woman's chances to obtain a tenure-track position (Wolfinger, Mason, and Goulden 2008). Finally, educational background is also introduced as an indicator for demographic diversity. Having a doctoral degree is used as an indicator for educational attainment, as a doctoral degree is often a requirement for careers in research and academia, although less so in the private sector. Within the EU28, 48% of all doctoral graduates in 2015 were women, however, women were under-represented in the area of engineering, manufacturing and construction (29%) as well as information and computer technology (21%, Eurostat, 2018 [educ_uoe_grad02]). Moreover, there is an attrition between master degree and doctoral degree, with a lower proportion of women obtaining a doctorate (Eurostat 2018 [educ_uoe_grad02]). Altogether, these different aspects of gendered diversity impact the career chances of graduates and are therefore likely to hallmark the gendered composition of research teams.

Functional diversity refers to processes related to working conditions, such as decision-making power (being in a leadership/senior role or not), long team tenure (being above or below the average team tenure) and contract (having a permanent or fixed-term contract). This builds on the body of research that shows that women are more likely to opt out of scientific careers than men (among many others: Blickenstaff, 2005; EC, 2015; Etzkowitz and Ranga, 2011; Fox and Stephan, 2001; Herman et al., 2013; Howe-Walsh and Turnbull, 2014; Mavriplis et al., 2010; Ranga and Etzkowitz, 2010) and that informal practices as well as implicit norms hinder women's progression (Aiston 2015; Faulkner 2007; Long and Fox 1995; Fox 2015; van den Brink and Benschop 2012). In addition, including the seniority of team members as an indicator for functional gender diversity considers the effects of such implicit micro-practices of exclusion. Team tenure is important as it influences the informal power position of team members. Long team tenure can be seen as a status cue and team members with long team tenure are more likely to receive deference (Joshi and Knight 2015; Bunderson 2003). Contractual agreements, furthermore, can impact the individual's integration into a team. Especially in the context of academia, fixed-term contracts appear to be increasingly common for early career scholars (Ylijoki 2010), leaving them in a more precarious position and in fragmented career (Clegg and Baumeler 2010; Knights and Clarke 2014). From this perspective, a permanent contract not only indicates more security for the individual but also allows for long term commitment to a team and greater potential to contribute to the overall team's performance. This is reflected in some national practices, such as in Germany, where legal regulations stipulate that academics obtain a permanent contract after a given period of time: early career researchers are only allowed to work for six years on fixed-term contracts.

This section has provided a brief account of the data collection process and potential indicators. Aside from selecting indicators to populate the Gender Diversity Index, it is necessary to produce suitable metrics. These metrics are described next.

Metrics of the Gender Diversity Index

For each potential indicator, two separate metrics were computed. The first one is a measure of representation, i.e. at the horizontal level. It provides information about whether women and men are equally represented. This representation is calculated in the category that is deemed more desirable or more inclusive (expressed as level 2, in contrast to level 1). For example, when it comes to equal representation within senior team roles or in permanent/tenured contracts, it is easy to identify that it is desirable to ensure equal access to more prestigious and less precarious positions (Morley 2018). Similarly, inclusivity entails that those that have care responsibilities are able to take part in research, particularly since it is difficult to combine care responsibilities with long working hours and precarious work contracts, which makes it more difficult for women to opt for motherhood when pursuing an academic career (Sang, Al-Dajani, and Özbilgin 2013; Benschop et al. 2013;

Metz-Göckel, Selent, and Schürmann 2010; Kemelgor and Etzkowitz 2001). This measure of representation depicts the situation of both women and men in science, and is calculated as follows:

$$\alpha_i = 2\times\left(1-(\mathrm{w}_{2i}^2+\mathrm{m}_{2i}^2)\right)$$

where w and m represent the proportions of women and men, respectively, at level 2 within team i. In the case where there are no women or men at level 2, then the metric returns a score of 0, indicating there can be no gender diversity as there are no opportunities for people to realise themselves. In other instances, the metric returns a score bound between 0 and 1, where the highest score is achieved where there is gender parity, meaning an equal distribution of women and men. This measure of representation, therefore, captures potential under-representation of either women or men. Moreover, it is non-linear and benefits conditions close to parity (Humbert and Guenther 2017).

The second measure is a measure of attrition, i.e. at the vertical level. It measures the extent to which women and men can equally progress within functional levels and are equally represented according to their demographic diversity. In the case of functional gender diversity, it is easy to understand that women and men should be, for example, equally represented at different levels of seniority – in other words, the pipeline should not be leaking (Dubois-Shaik and Fusulier 2015; Blickenstaff 2005). In the case of demographic diversity, attrition can be understood as the possibility for both women and men to be equally represented as parents/carers compared to people without care responsibilities or to have a cohabiting partner compared to those who are single (Mary Frank Fox 2005; Wolfinger, Mason, and Goulden 2008). Mathematically, this measure is expressed as:

$$\delta_i = \begin{cases} \dfrac{\mathrm{w}_{2i}}{\mathrm{w}_{1i}}, & \text{if } \mathrm{w}_{2i} < \mathrm{m}_{2i} \text{ and } \mathrm{w}_{2i} < \mathrm{w}_{1i} \\ \dfrac{\mathrm{m}_{2i}}{\mathrm{m}_{1i}}, & \text{if } \mathrm{m}_{2i} < \mathrm{w}_{2i} \text{ and } \mathrm{m}_{2i} < \mathrm{m}_{1i} \\ 1, & \text{otherwise} \end{cases}$$

with level 1 or 2 in team i, where w and m refer to proportions in the population of reference for women and men respectively. The metric returns a score of 0 where there are no women or men at level 2, since there can be no gender diversity where progression is not possible. The metric is bound between 0 and 1, and returns the highest score when there is no attrition between the two levels.

These metrics are applied to each potential indicator of gender diversity, and used to carry out a multivariate analysis that informs the development of the Gender Diversity Index. The next section considers this aspect in greater details.

Multivariate analysis

The aggregation of indicators to form the Gender Diversity Index is informed by a multivariate analysis that is used to construct a meaningful measurement framework (Nardo et al., 2008). The selection of indicators followed an iterative process, informed both by conceptual and statistical considerations, where different models where considered and assessed. Teams were included were they met the following criteria: full response rate for teams with four respondents; a minimum response rate of 50% for teams with between 5 and 9 respondents; a minimum response rate of 40% for teams with 10 or more respondents. This reduced the number of teams from 159 in the full sample to 101, and hereby the margin of error for the estimate of the proportion for each indicator.

Using principal component analysis provides information as to what meaningful linear combinations of variables exists along a number of different factors (Hair et al. 2006). In the case of the Gender Diversity Index, the most suitable option was obtained through a seven-factor solution with a varimax rotation. This solution accounts for 93% of the variance, well above the threshold of 70% (Jolliffe 1986), and unsurprising given the large number of factors. Furthermore, the model fit is adequate with a Kaiser-Meyer-Olkin amounting to 0.65 (Kaiser 1974). There are two related items within each factor, each with high reliability as evidenced by Cronbach α values ranging from 0.76 to 0.97 (Cronbach 1951). Factors loadings are high (all above 0.7) for all indicators (Table 2).

Table 2: Principal Component Analysis

Component	1	2	3	4	5	6	7	*MSA*	Indicator
Seniority δ	0.93							*0.68*	Attrition of the under-represented sex with/out a senior role
Seniority α	0.92							*0.70*	Representation of wo/men with a senior role (senior researcher or team leader)
Tenure δ		0.94						*0.59*	Attrition of the under-represented sex under/above average team tenure
Tenure α		0.93						*0.65*	Representation of wo/men above average team tenure
Age δ			0.93					*0.58*	Attrition of the under-represented sex under/above average age in the team
Age α			0.89					*0.65*	Representation of wo/men above average age in the team
Contract δ				0.93				*0.61*	Attrition of the under-represented sex with/out a permanent contract
Contract α				0.89				*0.66*	Representation of wo/men with a permanent contract
Care responsibilities δ					0.93			*0.56*	Attrition of the under-represented sex with/out current caring responsibilities for children under 16 or dependent adults
Care responsibilities α					0.88			*0.66*	Representation of wo/men with current caring responsibilities for children under 16 or dependent adults
Education δ						0.94		*0.60*	Attrition of the under-represented sex with/out a doctorate
Education α						0.81		*0.73*	Representation of wo/men with a doctorate
Marital status δ							0.91	*0.61*	Attrition of the under-represented sex with/out a cohabiting partner
Marital status α							0.71	*0.77*	Representation of wo/men with a cohabiting partner
Proportion of Variance	*15%*	*15%*	*14%*	*13%*	*13%*	*13%*	*11%*		
Cumulative Proportion	*15%*	*29%*	*43%*	*56%*	*70%*	*82%*	*93%*		
Cronbach's α	*0.97*	*0.95*	*0.94*	*0.90*	*0.89*	*0.88*	*0.76*		
Kaiser-Meyer-Olkin								*0.65*	

This suggests a structure for the Gender Diversity Index that consists of seven pillars, each corresponding to attrition and representation relating to different gendered process within a research team. Potential solutions with fewer factors were considered, however, none provided a satisfactory solution both conceptually and statistically. Using this framework, it is possible to then turn our attention to aggregation and weighting as the next steps needed to build the Gender Diversity Index.

Aggregation and weighting

The aggregation and weighting of a composite indicator are decisions that are not neutral, in so far as they have an impact on the overall score. In building the Gender Diversity Index, a multi-modelling approach is adopted. This means computing the scores generated by a set of different assumptions, before selecting a model based on the distribution of the scores generated.

Weighting is applied at the level of indicators (Table 3) and at the level of the pillars (Table 4). For indicators, two types of weights are considered (equal weights and PCA weights) and three types at pillar level (equal weights, PCA weights and expert weights). Although it appears that using equal weights is a neutral choice, it is in fact far from neutral since the overall score of the Gender Diversity Index can be disproportionately affected by indicators with high correlations. Deriving weights from Principal Component Analysis is useful to take into account the different contributions of indicators based on their correlations (Nardo et al. 2008). Finally, expert weights were developed together with a member of the project advisory board at the level of pillars only and consisted of three sets of weights. The first set of expert weights distributes weights equally between demographic diversity (50%) and functional diversity (50%). The second set of expert weights gives more weight to demographic diversity (60%) than functional diversity (40%): this option gives more importance to grounds of diversity that are more aligned to anti-discrimination legislation. The third set of expert weights reverses this relationship and attributes less weight to demographic diversity (40%) than to functional diversity (60%). This might be justified by arguing that functional diversity is more pertinent to research performance and innovation, since the way in which the team is organised may lead more directly to these outcomes.

Table 3: Sets of weights considered at the indicator level

			Equal weights	PCA weights
Demographic diversity	Education	δ	0.5	0.574
		α	0.5	0.426
	Care responsibilities	δ	0.5	0.528
		α	0.5	0.472
	Marital status	δ	0.5	0.622
		α	0.5	0.378
	Age	δ	0.5	0.522
		α	0.5	0.478
Functional diversity	Team tenure	δ	0.5	0.505
		α	0.5	0.495
	Seniority	δ	0.5	0.505
		α	0.5	0.495
	Contract type	δ	0.5	0.522
		α	0.5	0.478

Table 4: Sets of weights considered at pillar level

		Equal weights		Expert Set 1		Expert Set 2		Expert Set 3		PCA	
Demographic diversity	Education	0.571	0.143	0.500	0.125	0.600	0.200	0.400	0.100	0.529	0.114
	Care responsibilities		0.143		0.125		0.200		0.100		0.143
	Marital status		0.143		0.125		0.200		0.100		0.116
	Age		0.143		0.125		0.200		0.100		0.156
Functional diversity	Team tenure	0.429	0.143	0.500	0.167	0.400	0.133	0.600	0.200	0.471	0.176
	Team role		0.143		0.167		0.133		0.200		0.150
	Contract type		0.143		0.167		0.133		0.200		0.145

Aggregation is also done sequentially at the level of indicators and pillars, i.e. first indicators then pillars. Three methods were initially considered for the Gender Diversity Index: the arithmetic mean, the geometric mean and the harmonic mean. The difference between these three types of averages is in the compensation they allow, with the geometric and harmonic means being increasingly more punitive than the arithmetic mean. However, because both the geometric and harmonic means rely on a product, they were not retained because they were deemed too sensitive in the many instances where team obtained a

score of 0 in a given pillar. When different assumptions are combined, it becomes possible to compute ten versions of the Gender Diversity Index, according to the characteristics outlined in Table 5.

Table 5 Characteristics of different versions of the Gender Diversity Index

	Weights indicators	**Weights pillars**
GDI.1	Equal	Equal
GDI.2a	Equal	Expert 1
GDI.2b	Equal	Expert 2
GDI.2c	Equal	Expert 3
GDI.3	Equal	PCA
GDI.4	PCA	Equal
GDI.5a	PCA	Expert 1
GDI.5b	PCA	Expert 2
GDI.5c	PCA	Expert 3
GDI.6	PCA	PCA

The methodology adopted to compute the Gender Diversity Index is the one that minimises the average of the squared distance to the median score. The method that minimises the distance to the median score is that of GDI.6. It relies on arithmetic aggregation combined with PCA weights at both the indicator and pillar level. Results providing both scores and ranks information are provided in Table 6. Basic robustness checks confirm that with this model, all teams score within 0.05 points of the median, including 80% within 0.01 points (Figure 1). Similarly, 100% of teams rank within five ranks of the median and 75% within one rank (Figure 2).

Figure 1 Distance from median score

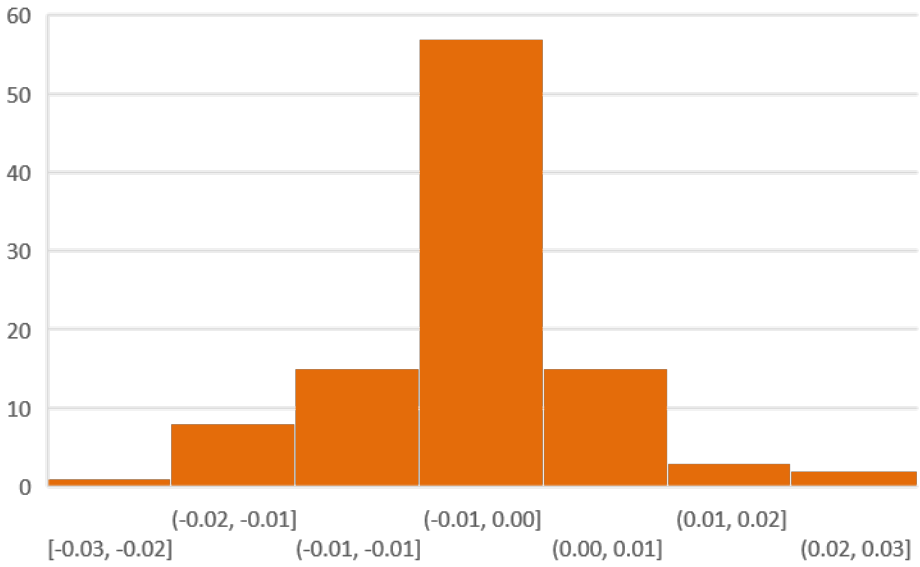

Figure 2 Distance from median rank

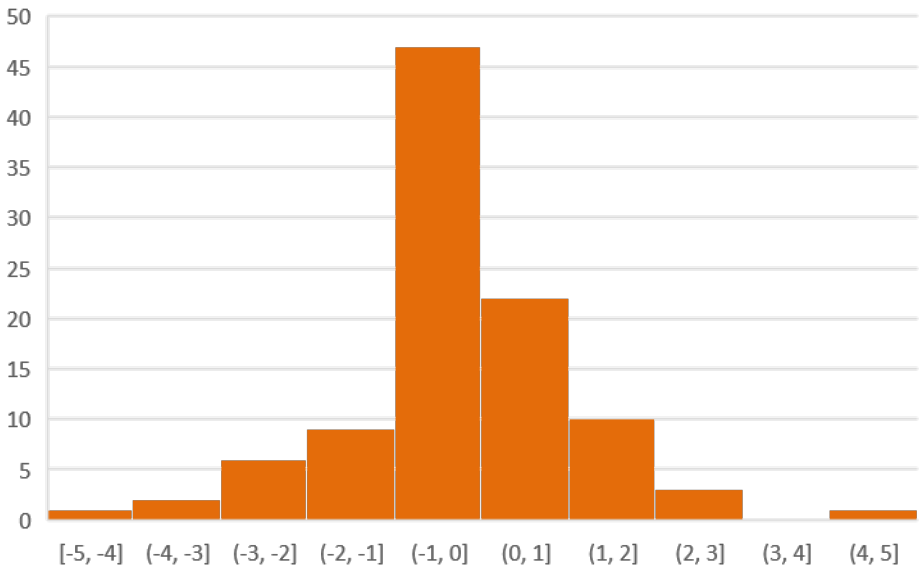

Table 6 Results of the Gender Diversity Index – Distribution of Scores and Ranks

Team ID	**Gender Diversity Index**	Minimum score	Median score	Maximum score	**Rank**	Lowest rank	Median rank	Highest rank
58	**0.94**	0.93	0.94	0.94	**1**	3	2	1
32	**0.94**	0.93	0.94	0.95	**2**	2	1	1
95	**0.93**	0.93	0.93	0.94	**3**	3	3	2
21	**0.92**	0.91	0.92	0.92	**4**	5	5	4
9	**0.91**	0.90	0.91	0.93	**5**	6	5	4
47	**0.91**	0.90	0.91	0.91	**6**	7	6	5
15	**0.91**	0.91	0.91	0.91	**7**	7	7	4
13	**0.87**	0.83	0.87	0.88	**8**	15	8	8
29	**0.86**	0.85	0.86	0.88	**9**	12	10	8
80	**0.86**	0.83	0.86	0.89	**10**	13	10	8
20	**0.85**	0.85	0.86	0.86	**11**	13	11	9
46	**0.85**	0.85	0.86	0.87	**12**	12	10	10
8	**0.85**	0.83	0.85	0.86	**13**	13	12	11
92	**0.84**	0.83	0.83	0.85	**14**	15	14	14
41	**0.82**	0.78	0.82	0.83	**15**	19	16	14
61	**0.82**	0.80	0.82	0.85	**16**	18	16	13
78	**0.81**	0.80	0.81	0.82	**17**	19	17	16
75	**0.80**	0.77	0.80	0.82	**18**	21	18	16
83	**0.78**	0.76	0.78	0.80	**19**	23	20	18
87	**0.78**	0.75	0.78	0.82	**20**	25	20	17
33	**0.77**	0.76	0.77	0.79	**21**	22	22	21
35	**0.77**	0.75	0.77	0.80	**22**	24	21	19
82	**0.76**	0.74	0.77	0.80	**23**	26	22	17
81	**0.76**	0.73	0.76	0.80	**24**	28	24	20
25	**0.75**	0.75	0.76	0.77	**25**	25	24	20
37	**0.74**	0.71	0.71	0.74	**26**	30	29	26
86	**0.74**	0.71	0.74	0.75	**27**	31	27	26

Team ID	**Gender Diversity Index**	Minimum score	Median score	Maximum score	**Rank**	Lowest rank	Median rank	Highest rank
45	**0.74**	0.71	0.73	0.75	**28**	28	28	23
79	**0.74**	0.70	0.73	0.75	**29**	32	27	26
2	**0.70**	0.67	0.69	0.70	**30**	38	32	30
97	**0.70**	0.66	0.69	0.71	**31**	39	33	31
26	**0.69**	0.69	0.71	0.73	**32**	34	30	27
74	**0.69**	0.66	0.69	0.75	**33**	39	33	23
19	**0.68**	0.68	0.68	0.69	**34**	37	33	31
40	**0.68**	0.65	0.68	0.71	**35**	38	35	30
67	**0.67**	0.65	0.67	0.69	**36**	39	37	35
31	**0.66**	0.63	0.67	0.71	**37**	43	37	33
50	**0.66**	0.63	0.66	0.70	**38**	44	39	32
98	**0.65**	0.54	0.65	0.69	**39**	50	39	33
16	**0.64**	0.62	0.63	0.65	**40**	44	41	40
22	**0.63**	0.58	0.63	0.68	**41**	44	41	35
66	**0.63**	0.63	0.65	0.69	**42**	42	41	35
4	**0.62**	0.56	0.63	0.68	**43**	47	42	36
36	**0.60**	0.59	0.61	0.62	**44**	48	44	41
65	**0.60**	0.56	0.59	0.61	**45**	50	46	45
23	**0.60**	0.54	0.60	0.65	**46**	52	45	40
10	**0.59**	0.49	0.58	0.61	**47**	59	49	46
14	**0.58**	0.53	0.58	0.60	**48**	54	51	46
43	**0.58**	0.48	0.57	0.61	**49**	60	51	45
55	**0.57**	0.56	0.58	0.59	**50**	52	48	47
70	**0.56**	0.56	0.58	0.60	**51**	52	49	46
85	**0.56**	0.56	0.57	0.60	**52**	54	53	42
90	**0.55**	0.49	0.54	0.60	**53**	66	54	43
69	**0.54**	0.49	0.52	0.55	**54**	65	59	54
51	**0.54**	0.47	0.54	0.60	**55**	62	56	48
3	**0.54**	0.54	0.54	0.55	**56**	57	55	51

Team ID	**Gender Diversity Index**	Minimum score	Median score	Maximum score	**Rank**	Lowest rank	Median rank	Highest rank
6	**0.54**	0.53	0.54	0.55	**57**	58	55	50
44	**0.53**	0.43	0.53	0.57	**58**	68	58	53
27	**0.52**	0.51	0.51	0.52	**59**	64	60	56
52	**0.52**	0.46	0.52	0.54	**60**	65	60	58
99	**0.51**	0.51	0.54	0.55	**61**	61	56	53
1	**0.51**	0.44	0.51	0.54	**62**	66	62	57
12	**0.49**	0.49	0.51	0.51	**63**	63	62	57
88	**0.48**	0.48	0.49	0.49	**64**	68	64	58
57	**0.47**	0.46	0.49	0.52	**65**	65	64	60
100	**0.47**	0.41	0.46	0.47	**66**	72	70	66
54	**0.46**	0.41	0.46	0.50	**67**	73	67	62
7	**0.46**	0.46	0.48	0.49	**68**	68	67	63
91	**0.46**	0.42	0.46	0.50	**69**	70	68	64
94	**0.44**	0.39	0.42	0.44	**70**	74	73	70
18	**0.43**	0.41	0.44	0.48	**71**	73	72	61
64	**0.43**	0.43	0.44	0.45	**72**	72	71	67
73	**0.43**	0.41	0.44	0.47	**73**	73	71	69
53	**0.39**	0.33	0.38	0.42	**74**	77	74	69
62	**0.36**	0.26	0.37	0.40	**75**	82	75	75
63	**0.35**	0.27	0.35	0.41	**76**	81	76	74
42	**0.34**	0.30	0.31	0.35	**77**	81	80	77
96	**0.33**	0.32	0.33	0.33	**78**	79	78	76
93	**0.33**	0.33	0.33	0.34	**79**	79	77	75
49	**0.32**	0.29	0.32	0.32	**80**	81	80	79
48	**0.31**	0.30	0.32	0.34	**81**	81	79	76
38	**0.29**	0.19	0.28	0.29	**82**	86	83	82
68	**0.28**	0.25	0.27	0.29	**83**	89	83	80
71	**0.27**	0.24	0.27	0.28	**84**	85	84	82
39	**0.26**	0.20	0.26	0.30	**85**	85	85	82

Team ID	**Gender Diversity Index**	Minimum score	Median score	Maximum score	**Rank**	Lowest rank	Median rank	Highest rank
77	**0.25**	0.18	0.25	0.28	**86**	88	88	86
84	**0.25**	0.19	0.25	0.29	**87**	88	87	84
60	**0.24**	0.24	0.26	0.28	**88**	88	86	83
5	**0.24**	0.18	0.24	0.27	**89**	89	89	88
34	**0.16**	0.13	0.17	0.21	**90**	90	90	90
17	**0.13**	0.09	0.13	0.13	**91**	92	91	91
11	**0.11**	0.09	0.11	0.14	**92**	92	91	91
24	**0.00**	0.00	0.00	0.00	**93**	93	93	93
28	**0.00**	0.00	0.00	0.00	**93**	93	93	93
30	**0.00**	0.00	0.00	0.00	**93**	93	93	93
56	**0.00**	0.00	0.00	0.00	**93**	93	93	93
59	**0.00**	0.00	0.00	0.00	**93**	93	93	93
72	**0.00**	0.00	0.00	0.00	**93**	93	93	93
76	**0.00**	0.00	0.00	0.00	**93**	93	93	93
89	**0.00**	0.00	0.00	0.00	**93**	93	93	93
101	**0.00**	0.00	0.00	0.00	**93**	93	93	93

Checking for robustness

In line with the COIN best practice recommendations (Nardo et al. 2008), the robustness of the Gender Diversity Index was assessed by examining the correlation structure between the indicators (Table 7), between the indicators and the pillar/index scores, and between the pillars and the Index itself. This shows that there are some very strong correlations between some indicators within the same pillars, such as those associated with senior roles, team tenure and age. However, this is likely related to the indicators picking up on the same phenomenon. Since there is no evidence of strong collinearity among indicators not grouped together, with correlations not exceeding 0.6 and no negative coefficients, the correlation structure of the indicators is deemed satisfactory.

The association of the pillars to underlying indicators (Table 8) shows that all indicators contribute to both the pillars and the overall Gender Diversity Index since all correlation coefficients of interest are above 0.5. A previous iteration of the Gender Diversity Index had considered including two further pillars – one for ethnicity and one for disability – but were subsequently omitted because they failed to meet the requirements of a composite indicator at this stage of the robustness assessment. This assessment provides evidence that all the indicators included in the Gender Diversity Index contribute to the pillars and the scores. Finally, the robustness checks show strong correlations between each of the pillars and the Gender Diversity Index (Table 9), confirming that the model used for the Gender Diversity Index is statistically coherent.

Going back down to the details

This report has described the process of development of different measures and their aggregation into a composite measure. In this section, the report examines the results of the Gender Diversity Index in greater details using examples of research teams. In the following, the meaning of the Gender Diversity Index and its potential interpretation is exemplified using some anonymised cases of the team survey.

Table 7 Correlations between indicators

	Seniority δ	Seniority α	Tenure δ	Tenure α	Care δ	Care α	Contract δ	Contract α	Education δ	Education α	Marital status δ	Marital status α	Age δ	Age α
Seniority δ	1.00													
Seniority α	**0.94**	1.00												
Tenure δ	0.33	0.34	1.00											
Tenure α	0.33	0.37	**0.90**	1.00										
Care δ	0.13	0.14	0.10	0.10	1.00									
Care α	0.33	0.39	0.27	0.32	**0.80**	1.00								
Contract δ	0.23	0.27	0.21	0.25	0.28	0.29	1.00							
Contract α	0.35	0.45	0.24	0.31	0.22	0.37	**0.82**	1.00						
Education δ	0.27	0.29	0.24	0.22	0.28	0.19	0.24	0.19	1.00					
Education α	0.41	0.48	0.35	0.41	0.29	0.41	0.32	0.37	**0.79**	1.00				
Marital status δ	0.24	0.24	0.14	0.22	0.42	0.36	0.22	0.15	0.31	0.33	1.00			
Marital status α	0.44	0.51	0.39	0.51	0.32	0.52	0.29	0.45	0.27	0.51	**0.62**	1.00		
Age δ	0.35	0.38	0.35	0.32	0.14	0.20	0.24	0.35	0.34	0.34	0.21	0.33	1.00	
Age α	0.39	0.46	0.34	0.41	0.15	0.30	0.30	0.48	0.33	0.45	0.22	0.50	**0.89**	1.00

Table 8 Correlations between Indicators and Pillars/Index

	Seniority δ	Seniority α	Tenure δ	Tenure α	Care δ	Care α	Contract δ	Contract α	Education δ	Education α	Marital status δ	Marital status α	Age δ	Age α
Seniority	**0.98**	**0.98**												
Tenure	0.34	0.36	**0.98**	**0.98**										
Care	0.23	0.27	0.19	0.21	**0.96**	**0.94**								
Contract	0.30	0.37	0.23	0.29	0.26	0.35	**0.96**	**0.95**						
Education	0.34	0.39	0.30	0.31	0.30	0.30	0.29	0.28	**0.96**	**0.92**				
Marital Status	0.33	0.36	0.25	0.35	0.42	0.45	0.26	0.27	0.33	0.42	**0.96**	**0.82**		
Age	0.38	0.43	0.36	0.37	0.15	0.25	0.28	0.42	0.35	0.40	0.22	0.42	**0.98**	**0.97**
GDI	**0.66**	**0.71**	**0.62**	**0.67**	**0.49**	**0.64**	**0.57**	**0.65**	**0.54**	**0.70**	**0.51**	**0.72**	**0.62**	**0.70**

Table 9 Correlations between Pillars and Index

	Seniority	Tenure	Care	Contract	Education	Marital Status	Age	GDI
Seniority	1.00							
Tenure	0.36	1.00						
Care	0.25	0.20	1.00					
Contract	0.34	0.27	0.31	1.00				
Education	0.37	0.31	0.31	0.30	1.00			
Marital Status	0.35	0.30	0.46	0.28	0.39	1.00		
Age	0.41	0.37	0.21	0.36	0.39	0.32	1.00	
GDI	**0.70**	**0.66**	**0.59**	**0.64**	**0.64**	**0.63**	**0.68**	1.00

Case 1

Team 58 achieves the best score (0.94) amongst the teams that participated in the study. Despite the high score, women are under-represented in this team as they represent less than third of its 13 respondents (four women and nine men). This example shows that the Gender Diversity Index does not represent a score for gender balance within the team. The high score is the result of two factors. First, the balance at level 2 in the team is better than overall. For example, women represent 40% of team members with care responsibilities and 50% of team members in a senior role. Second, there is little attrition in this team across measures. For instance, the representation of women and men is approximately the same for those without or with a doctorate.

						Women	Men
Education δ	1	0.95	0.94	Level 1	No doctorate	3	7
						30%	70%
Education α	0.89			Level 2	Doctorate	1	2
						33%	67%
Care δ	1	0.98		Level 1	No current care responsibilities	2	6
						25%	75%
Care α	0.96			Level 2	Current care responsibilities	2	3
						40%	60%
Marital status δ	1	0.98		Level 1	Not married or cohabiting	2	6
						25%	75%
Marital status α	0.96			Level 2	Married or cohabiting	2	3
						40%	60%
Age δ	1	0.98		Level 1	Below team average age	2	6
						25%	75%
Age α	0.96			Level 2	Above team average age	2	3
						40%	60%
Seniority δ	1	1		Level 1	Junior researcher/Other	3	8
						27%	73%
Seniority α	1			Level 2	Senior researcher/Team leader	1	1
						50%	50%
Tenure δ	1	0.95		Level 1	Below average team tenure	2	4
						33%	67%
Tenure α	0.89			Level 2	Above average team tenure	2	4
						33%	67%
Contract δ	0.75	0.75		Level 1	Temporary / casual	3	6
						33%	67%
Contract α	0.75			Level 2	Permanent / tenured	1	3
						25%	75%

Case 2

Team 21 achieves a score of 0.92 on the Gender Diversity Index, and where women represent 67% of respondents in the team (10 out of 15). This result is driven by high scores in most pillars. One exception is the score for attrition in relation to junior and senior roles at 0.44. This is related to women being under-represented in senior position, while they are over-represented in junior positions. This illustrates how the Gender Diversity Index results not only pick up on attrition, but also on the reversal in representation. However, these results also show the importance of not interpreting scores of the Gender Diversity Index without considering the numbers to which they related, because of the sensitivity of the measure where there are small numbers. For instance, if the team included another woman in a senior role, this would considerably increase the score achieved in the corresponding pillar.

						Women	Men
Education δ	1	1	0.92	Level 1	No doctorate	7	2
						78%	22%
Education α	1			Level 2	Doctorate	3	3
						50%	50%
Care δ	0.78	0.8		Level 1	No current care responsibilities	4	3
						57%	43%
Care α	0.89			Level 2	Current care responsibilities	4	2
						67%	33%
Marital status δ	1	0.98		Level 1	Not married or cohabiting	4	1
						80%	20%
Marital status α	0.96			Level 2	Married or cohabiting	6	4
						60%	40%
Age δ	1	0.97		Level 1	Below team average age	5	2
						71%	29%
Age α	0.94			Level 2	Above team average age	5	3
						62%	38%
Seniority δ	0.44	0.66		Level 1	Junior researcher/Other	9	3
						75%	25%
Seniority α	0.89			Level 2	Senior researcher/Team leader	1	2
						33%	67%
Tenure δ	1	0.98		Level 1	Below average team tenure	7	3
						70%	30%
Tenure α	0.96			Level 2	Above average team tenure	3	2
						60%	40%
Contract δ	1	1		Level 1	Temporary / casual	8	3
						73%	27%
Contract α	1			Level 2	Permanent / tenured	2	2
						50%	50%

Case 3

In team 38, six team members (four women and two men) participated in the study. The team achieves a low score (0.29) in the Gender Diversity Index. The results of the pillars are very polarized, with two high scores for care responsibilities and age but all other pillars showing a lack of gender diversity. An interesting feature of this team is that it is not only women-dominated, but also generally dominated by women at level 2 particularly when it comes to functional diversity. In the team, the two tenured and senior members that responded are women. This contrasts with the lower levels were representation is more equal. This shows that the Gender Diversity Index focuses on inequalities regardless of whether they affect women or men, and instead provides a measure of relative representation from a gender perspective (rather than a women's empowerment perspective).

						Women	Men
Education δ	0	0	0.29	Level 1	No doctorate	2	2
						50%	50%
Education α	0			Level 2	Doctorate	2	0
						100%	0%
Care δ	1	1		Level 1	No current care responsibilities	2	1
						67%	33%
Care α	1			Level 2	Current care responsibilities	1	1
						50%	50%
Marital status δ	0	0		Level 1	Not married or cohabiting	0	2
						0%	100%
Marital status α	0			Level 2	Married or cohabiting	4	0
						100%	0%
Age δ	1	0.95		Level 1	Below team average age	2	1
						67%	33%
Age α	0.89			Level 2	Above team average age	2	1
						67%	33%
Seniority δ	0	0		Level 1	Junior researcher/Other	2	2
						50%	50%
Seniority α	0			Level 2	Senior researcher/Team leader	2	0
						100%	0%
Tenure δ	0	0		Level 1	Below average team tenure	2	2
						50%	50%
Tenure α	0			Level 2	Above average team tenure	2	0
						100%	0%
Contract δ	0	0		Level 1	Temporary / casual	2	2
						50%	50%
Contract α	0			Level 2	Permanent / tenured	2	0
						100%	0%

Conclusion

The Gender Diversity Index is a composite indicator that aims to capture gendered processes within research teams. A composite indicator aggregates several measures into one figure, making it easier to compare complex situations. This report has presented the methodological steps used in developing the Gender Diversity Index. Following a brief overview of the survey used to populate this composite measure, the results of the multivariate analysis and multi-model procedures used to determine how to aggregate indicators were provided. Basic robustness checks were also presented, demonstrating that the Gender Diversity Index's methodology is appropriate.

The Gender Diversity Index conceptualises gender equality as parity between women and men, and on equal ability to contribute to science and innovation. This reflects the need to consider the field specificities across different subject areas. For example, even though women are, in general, the under-represented gender group in science and technology, there are scientific subjects in which women are over-represented. In 2015, for instance, only 29% of all engineering, manufacturing or construction doctoral graduates in the EU28 were women, while 59% of doctoral degrees in agriculture, forestry, fisheries and veterinary sciences were awarded to women (Eurostat 2018 [educ_uoe_grad02]). Accordingly, scores reflect these inequalities regardless of whether they affect women or men.

The Gender Diversity Index focuses on two aspects: (1) the horizontal representation of women and men in the category that is either most desirable (e.g. senior role) or inclusive (e.g. care responsibilities) and (2) the vertical attrition of women and men between categories. A high score, reflecting high levels of gender diversity, is achieved when a team achieves a high score in representation combined with low levels of attrition in achieving this representation across different aspects of diversity. The Gender Diversity Index does not measure gender representation in other categories (e.g. in non-senior roles), as it gives precedence to aspirational categories. For example, this approach is based on the ideal that women role models encourage retention of women in field where they are usually under-represented (Herrmann et al. 2016; Drury, Siy, and Cheryan 2011). This also goes against the grain of approaches that seek to simply populate the pipeline with more members of the under-represented group and simply wait for equality to be reached naturally. As a result, the Gender Diversity Index does not pick up on whether the pipeline is gender diverse. However, by ensuring that there is gender diversity in the more desirable and inclusive categories, it is expected that there would be alignment in other categories, as failing that, it would ultimately lead to a decrease of gender diversity in the team over time. This is visible in the data, such as in Case 4, where a team achieves a very high score that reflects the high level of gender diversity across all aspects of diversity, despite only few early career women working in the team. This emphasises the need to examine the score of the Gender Diversity Index together with the data at team level.

The Gender Diversity Index is not without limitations. The main one is related to conducting measurement at the team level, where invariably numbers can be very small. This calls for great care in the interpretation of the measures, as it leads to some imprecision with estimates and as it is difficult to talk in percentage terms of sex distribution with sometimes little more than a handful of cases. The measures of the Gender Diversity Index, particularly when drilling down to the pillar level can be very crude and should not be interpreted independently of the nominal data of the teams. Pillars will be particularly sensitive to changes in the data with smaller numbers. However, the overall scores should provide a valuable starting point to understand gender diversity processes and what areas to address first in promoting gender diversity in research teams.

The Gender Diversity Index provides a first measure – at team level – of gender diversity that goes beyond parity and extends to some key processes within the team. The results can be used on their own to provide a guide as to how gender diverse a research team is: indeed, within the scope of the GEDII project, a web-based self-assessment tool will be developed and made available in 2018. The results can also be used within further analysis, notably to assess the empirical link between gender diversity in research teams and the level research performance.

References

Aiston, Sarah Jane. 2015. “Whose Academy? Gender and Higher Education.” *Investigating Higher Education: A Critical Review of Research Contributions. London: SHRE/Routledge*, 80–97.

Armenti, Carmen. 2004. “Women Faculty Seeking Tenure and Parenthood: Lessons from Previous Generations.” *Cambridge Journal of Education* 34 (1): 65–83. doi:10.1080/0305764042000183133.

Beddoes, Kacey, and Alice L. Pawley. 2014. “‘Different People Have Different Priorities’: Work–family Balance, Gender, and the Discourse of Choice.” *Studies in Higher Education* 39 (9): 1573–85. doi:10.1080/03075079.2013.801432.

Benschop, Yvonne, Marieke van den Brink, Hans Doorewaard, and Joke Leenders. 2013. “Discourses of Ambition, Gender and Part-Time Work.” *Human Relations* 66 (5): 699–723. doi:10.1177/0018726712466574.

Berger, Rita, Marina Romeo, Joan Guardia, Montserrat Yepes, and Miguel Angel Soria. 2012. “Psychometric Properties of the Spanish Human System Audit Short-Scale of Transformational Leadership.” *The Spanish Journal of Psychology* 15 (1): 367–76. doi:10.5209/rev_SJOP.2012.v15.n1.37343.

Blickenstaff, Jacob Clark. 2005. “Women and Science Careers: Leaky Pipeline or Gender Filter?” *Gender and Education* 17 (4). Taylor & Francis Group: 369–86. doi:10.1080/09540250500145072.

Bourdieu, Pierre. 1994. *Homo Academicus*. Standford University Press.

Bunderson, J. Stuart. 2003. “Recognizing and Utilizing Expertise in Work Groups: A Status Characteristics Perspective.” *Administrative Science Quarterly* 48 (4). SAGE Publications: 557–91. doi:10.2307/3556637.

Cervia, Silvia, and Rita Biancheri. 2017. “Women in Science: The Persistence of Traditional Gender Roles. A Case Study on Work–life Interface.” *European Educational Research Journal* 16 (2–3): 215–29. doi:10.1177/1474904116654701.

Clegg, Stewart, and Carmen Baumeler. 2010. “Essai: From Iron Cages to Liquid Modernity in Organization Analysis.” *Organization Studies* 31 (12): 1713–33. doi:10.1177/0170840610387240.

Cronbach, Lee J. 1951. “Coefficient Alpha and the Internal Structure of Tests.” *Psychometrika* 16 (3). Springer: 297–334.

Drury, Benjamin J., John Oliver Siy, and Sapna Cheryan. 2011. “When Do Female Role Models Benefit Women? The Importance of Differentiating Recruitment from Retention in STEM.” *Psychological Inquiry* 22 (4): 265–69. doi:10.1080/1047840X.2011.620935.

Dubois-Shaik, Farah, and Bernard Fusulier. 2015. “Academic Careers and Gender Inequality: Leaky Pipeline and Interrelated Phenomena in Seven European Countries.”

EC, European Commission. 2015. “She Figures.” doi:10.2777/744106.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge: Polity.

Etzkowitz, Henry, and Marina Ranga. 2011. “Gender Dynamics in Science and Technology: From The ‘leaky Pipeline’ to The ‘vanish Box’.” *Brussels Economic Review* 54 (3): 131–47.

Eurostat. 2018. “Graduates by Education Level, Programme Orientation, Sex and Field of Education [educ_uoe_grad02].”

Faulkner, Wendy. 2007. “`Nuts and Bolts and People’: Gender-Troubled Engineering Identities.” *Social Studies of Science* 37 (3): 331–56. doi:10.1177/0306312706072175.

Fox, M. F., and P. E. Stephan. 2001. “Careers of Young Scientists: Preferences, Prospects and Realities by Gender and Field.” *Social Studies of Science* 31 (1): 109–22. doi:10.1177/030631201031001006.

Fox, Mary Frank. 2005. “Gender, Family Characteristics, and Publication Productivity among Scientists.” *Social Studies of Science* 35 (1). Sage PublicationsSage CA: Thousand Oaks, CA: 131–50. doi:10.1177/0306312705046630.

———. 2015. “Gender and Clarity of Evaluation among Academic Scientists in Research Universities.” *Science, Technology, & Human Values* 40 (4): 487–515. doi:10.1177/0162243914564074.

Hair, Joseph F, William C Black, Barry J Babin, Rolph E Anderson, and Ronald L Tatham. 2006. *Multivariate Data Analysis 6th Edition*. New Jersey: Pearson Education.

Herman, Clem, Suzan Lewis, and Anne Laure Humbert. 2013. “Women Scientists and Engineers in European Companies: Putting Motherhood under the Microscope.” *Gender, Work and Organization* 20 (5): 467–78. doi:10.1111/j.1468-0432.2012.00596.x.

Herrmann, S D, R M Adelman, J E Bodford, O Graudejus, M A Okun, and V S Y Kwan. 2016. "The Effects of a Female Role Model on Academic Performance and Persistence of Women in STEM Courses." *Basic and Applied Social Psychology* 38 (5). Taylor & Francis: 258–68. doi:10.1080/01973533.2016.1209757.

Howe-Walsh, Liza, and Sarah Turnbull. 2014. "Barriers to Women Leaders in Academia: Tales from Science and Technology." *Studies in Higher Education* 41 (3): 415–28. doi:10.1080/03075079.2014.929102.

Humbert, Anne Laure, and Elisabeth Anna Guenther. 2017. "D3.1 The Gender Diversity Index, Preliminary Considerations and Results." Project Deliverable. doi:10.17862/cranfield.rd.5110978.v1.

Jolliffe, Ian T. 1986. "Principal Component Analysis and Factor Analysis." In *Principal Component Analysis*, 115–28. Springer.

Joshi, Aparna, and Andrew P. Knight. 2015. "Who Defers to Whom and Why? Dual Pathways Linking Demographic Differences and Dyadic Deference to Team Effectiveness." *Academy of Management Journal* 58 (1): 59–84. doi:10.5465/amj.2013.0718.

Kaiser, Henry F. 1974. "An Index of Factorial Simplicity." *Psychometrika* 39 (1): 31–36. doi:10.1007/BF02291575.

Kemelgor, Carol, and Henry Etzkowitz. 2001. "Overcoming Isolation: Women's Dilemmas in American Academic Science." *Minerva* 39 (2): 153–74. doi:10.1023/A:1010344929577.

Kivimaki, M, and M Elovainio. 1999. "A Short Version of the Team Climate Inventory: Development and Psychometric Properties." *Journal of Occupational and Organizational Psychology* 72: 241–46. doi:10.1348/096317999166644.

Knights, David, and Caroline A. Clarke. 2014. "It's a Bittersweet Symphony, This Life: Fragile Academic Selves and Insecure Identities at Work." *Organization Studies* 35 (3): 335–57. doi:10.1177/0170840613508396.

Lewis, Jane. 1992. "Gender and the Development of Welfare Regimes." *Journal of European Social Policy* 2 (3). Sage Publications: 159–73.

Long, J. Scott, and Mary Frank Fox. 1995. "Scientific Careers: Universalism and Particularism." *Annual Review of Sociology*. doi:10.1146/annurev.soc.21.1.45.

Mason, Mary Ann, and Marc Goulden. 2004. "Marriage and Baby Blues: Redefining Gender Equity in the Academy." *Annals of the American Academy of Political and Social Science* 596 (November): 86–103. doi:10.1177/0002716204268744.

Mavriplis, Catherine, Rachelle Heller, Cheryl Beil, Kim Dam, Natalya Yassinskaya, Megan Shaw, and Charlene Sorensen. 2010. “Mind the Gap : Women in STEM Career Breaks” 5 (1).

Metz-Göckel, Sigrid, Petra Selent, and Ramona Schürmann. 2010. “Integration Und Selektion. Dem Dropout von Wissenschaftlerinnen Auf Der Spur.” *Beiträge Zur Hochschulforschung* 32 (1): 8–35.

Miller, Andrea L., and Eugene Borgida. 2016. “The Separate Spheres Model of Gendered Inequality.” *PLoS ONE* 11 (1): 1–34. doi:10.1371/journal.pone.0147315.

Morley, Louise. 2018. “Gender in the Neo-Liberal Research Economy: An Enervating and Exclusionary Entanglement?” In *Gender Studies and the New Academic Governance*, 15–40. Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-19853-4_2.

Müller, Jörg, Sandra Klatt, Anne-Charlott Callerstig, and Ulf Sandström. 2016. “D1 . 1 Conceptual Framework.” Project Deliverable. www.gedii.eu.

Nardo, Michela, Michaela Saisana, Andrea Saltelli, Stefano Tarantola, Anders Hoffman, and Enrico Giovannini. 2008. *Handbook on Constructing Composite Indicators. Methodology and User Guide.* Paris: OECD Publishing. doi:10.1787/533411815016.

Ranga, Marina, and Henry Etzkowitz. 2010. “Athena in the World of Techne: The Gender Dimension of Technology, Innovation and Entrepreneurship.” *Journal of Technology Management and Innovation* 5 (1): 1–12. doi:10.4067/S0718-27242010000100001.

Riach, Kathleen, Nicholas Rumens, and M. Tyler. 2014. “Un/doing Chrononormativity: Negotiating Ageing, Gender and Sexuality in Organizational Life.” *Organization Studies* 35 (11): 1677–98. doi:10.1177/0170840614550731.

Sainsbury, Diane. 1996. *Gender, Equality and Welfare States*. Cambridge University Press.

Sang, Katherine, Haya Al-Dajani, and Mustafa F. Özbilgin. 2013. “Frayed Careers of Migrant Female Professors in British Academia: An Intersectional Perspective.” *Gender, Work & Organization* 20 (2): 158–71. doi:10.1111/gwao.12014.

Stanfors, Maria. 2014. “Fertility and the Fast-Track: Continued Childbearing among Professionals in Sweden, 1991-2009.” *Demographic Research* 31 (1): 421–58. doi:10.4054/DemRes.2014.31.15.

UNECE, United Nations Economic Commission for Europe. 2010. “Developing Gender Statistics: A Practical Tool.”

http://www.unece.org/fileadmin/DAM/stats/publications/Developing_Gender_Statistics.pdf.

van den Brink, Marieke, and Yvonne Benschop. 2012. “Gender Practices in the Construction of Academic Excellence: Sheep with Five Legs.” *Organization* 19 (4): 507–24. doi:10.1177/1350508411414293.

Wolfinger, Nicholas H., Mary Ann Mason, and Marc Goulden. 2008. “Problems in the Pipeline: Gender, Marriage, and Fertility in the Ivory Tower.” *The Journal of Higher Education* 79 (4). Taylor & Francis: 388–405. doi:10.1353/jhe.0.0015.

Wright, Mary. C., Carla. B. Howery, Nandini Assar, Kathleen McKinney, Edward L. Kain, Becky Glass, Laura Kramer, and Maxine Atkinson. 2004. “Greedy Institutions: The Importance of Institutional Context for Teaching in Higher Education.” *Teaching Sociology* 32 (2): 144–59. doi:10.1177/0092055X0403200201.

Ylijoki, Oili-Helena O.-H. 2010. “Future Orientations in Episodic Labour: Short-Term Academics as a Case in Point.” *Time & Society* 19 (3): 365–86. doi:10.1177/0961463X10356220.

Annex: Items used in the questionnaire

The following questionnaire items are used for the construction of the Gender Diversity Index.

Sex

Sex is defined by the participant's self-identification to either sex-group. If one person does not identify with either woman or men, they are offered a third undefined option.

I am (gender)	
1	A woman
2	A man
3	Other

The following question was asked only in cases when someone ticked "Other"

On the previous question on "gender" you marked "Other". If you prefer not to identify with either woman or man, some sections of this questionnaire can't be used, since one aim of this study is to construct a gender diversity index which operates with a binary gender variable, despite other identities in-between. However, we encourage you to respond to all items in any case, contributing to the overall picture of your team.

Identification man or woman

1	I would tend to identify as a woman
2	I would tend to identify as a man
3	Neither

Education

The Gender Diversity Index compares team members who hold a doctorate to those who do not. The basis for this is the following question:

Which is your highest level of education?

- Secondary education
- Bachelor
- Master or equivalent postgraduate qualification
- Doctorate or higher
- Other, please specify:

Care Responsibilities

The Gender Diversity Index relates the gendered composition of those with current care responsibilities, both for children under 16 and dependent adults, with those who do not have current care responsibilities. The following question provided the basis for this construct:

Do you have/ have had care responsibilities for children under 16 years of age or for dependent adults? Select all that apply		
	Yes	No
Yes, for children under 16 years		
Yes, in the past for children under 16 years		
Yes, for dependent adults		
Yes, in the past for dependent adults		
No		

Marital Status

Do you live with a partner (marriage, cohabitation, civil partnership, etc)? ○ Yes ○ No

Age

Age was calculated using the birth yea.

I was born in (year):

Team Tenure

The Gender Diversity Index links the gender balance of the team members with a team tenure above average to those below team average. The following item was used to calculate team average and identify those above and below team average.

In which year did you join your team?

Seniority

Seniority is defined by team role. Those, who either have a leadership position or are senior members (i.e. team leader and senior researcher) are put in relation to the other team members.

Which of the following best describes your primary, current role in the team?
○ MA Student ○ PhD Student ○ Research or lab assistant / technician ○ Postdoc / Junior researcher ○ Senior researcher ○ Team leader ○ Other, please specify:

Contract

What type of contract do you have?
○ Temporary / casual ○ Permanent / tenured