

Dr Anne Laure Humbert, Dr Elisabeth Anna Guenther

Project Deliverable

## D3.1 The Gender Diversity Index, preliminary considerations and results

31<sup>st</sup> March 2017 – **FINAL**



The work contained in this document is distributed under the Creative Commons Attribution 4.0 License (CC BY). [View License Deed.](#)

## Project Consortium



Universitat Oberta de Catalunya, Spain

Project Coordination

Dr. Jörg Müller  
Dr. Milagros Sáinz  
Dr. Àgata Lapedriza  
Dr. Rosa Borge  
Dr. Julià Minguillón



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



Cranfield University, 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.

## Contents

1	Introduction.....	4
2	Developing the framework.....	6
2.1	The meaning of gender diversity .....	6
2.2	Relating gender diversity to team performance .....	8
2.3	Team processes and performance in STEM.....	10
3	Selection of indicators.....	12
4	Building the metrics .....	14
4.1	Measuring gender diversity.....	14
4.2	Gender diversity metrics: measuring representation.....	15
4.3	Relative gender diversity metric .....	17
4.4	Examples – examining gender differences in team roles .....	17
4.5	Further considerations.....	20
5	Multivariate analysis .....	21
6	Preliminary conclusions and further steps for development .....	25
	References.....	26

# 1 Introduction

A core objective of the GEDII project is to assess the benefits of gender diversity for science and for society. Much research has been done that attempts to link gender diversity to organisational performance in the context of business and management, with mixed outcomes. Reviews find inconsistent results (Armstrong et al., 2010; Miles & Erhardt, 2014; Müller, Klatt, Callerstig, & Sandström, 2016). One of the first scholars to look at the relationship between diversity management and performance was Kanter (1984) who noted that lowering discrimination in the workforce through effective diversity management was related to higher levels of innovation and performance. Armstrong (2010) provides a review of the potential effects to be gained in organisations by having more diversity in the workforce for both the organisation itself but also for its employees. Yet, at the empirical level, there is no actual consensus as to whether these claims can be upheld. This said, there is nonetheless starting to be an ever increasing number of studies able to demonstrate a positive link, although other studies find negative or no effects (Armstrong et al., 2010). Dezsö and Ross (2012) for example confirm the positive link between diversity and performance, showing empirically that as the proportion of women increases in leadership, financial performance improves but in a much stronger fashion where the organisational strategy has high innovation intensity – making it a very relevant finding to apply to the area of research.

However, even though the link between diversity and performance has been the subject of much debate in organisation studies, the literature in the context of research is more limited and inconclusive. The purpose of this project is therefore to transfer and test these arguments to the context of research, and more particularly within the scope of this project to biomedical engineering and transport research, and to test them empirically. The analysis is conducted in five EU countries (Germany, Lithuania, Spain, Sweden, UK), each of which represents a different welfare and gender equality regime (Esping-Andersen, 1990; Lewis, 1992; Sainsbury, 1996). To examine the relationship between gender diversity in teams and research performance, this project develops a composite indicator: the Gender Diversity Index. Its aim is to summarise the complexity of gender diversity within team processes in one single – easily interpretable and communicable – number. The Gender Diversity Index does not only represent an analytical tool, but can also be used as a practical monitoring tool for teams to assess themselves against the benchmark of other research teams.

A composite indicator is obtained when individual indicators are compiled in a single measure on the basis of a multi-dimensional concept. The Gender Diversity Index is a statistical measure that

aggregates several gender diversity measures at the team level into a statistically and conceptually coherent whole. The construction of a composite indicator needs to rely on two characteristics. First, it should include sensible development considerations from a statistical and conceptual viewpoint. Second, its limitations should be acknowledged in that it can be used to draw conclusions, but only while remembering the context in which they originate. Composite indicators are strong advocacy tools. To be useful, they need to achieve simplicity, but avoid triviality, in measuring a multi-dimensional concept (Saisana, 2016). Indeed, composite indicators make it possible to measure and summarise a complex topic into one single number. This makes them extremely powerful on the one hand by telling you ‘everything’, although at the same time they say ‘nothing’ unless they are unpacked and used in further analyses. The Joint Research Centre’s (JRC) Composite Indicators Research Group (European Commission, 2016) – also known as COIN – describes the pros of composite indicators as summarising complex multi-dimensional issues; providing a summary of the overall picture; and attracting public interest. However, this is mitigated by a list of cons which include the potential for sending misleading messages; generating simplistic conclusions; and being subject to uncertainty for example in regards to the indicators chosen or methodology adopted.

The audience of composite indicators is described as a polarised one. At one extreme are ‘enthusiastic supporters’ that use composite indicators to communicate around an issue and as a tool to enact change. At the other extreme, there are ‘sceptical economists and official statisticians’ who express concern around the validity of such measures because of the underlying methodological choices that were made (Saisana, 2016). In the context of gender, another group can be added. Many gender scholars and experts might be critical of the ability of a measure to capture the complexity of gender through quantification, particularly within the constraints of a binary system that ignores the fluidity and contextual nature of gender. Moreover, when looking at existing research and research assessment such as bibliometric studies, gender diversity is usually reduced to an individual’s sex and perhaps the sex-composition of the co-authors. The aim of the Gender Diversity Index is to provide a more sophisticated measure of gender diversity, that although based on a binary sex-category will nevertheless be sensitive to gender power relations, exemplified by vertical segregation in science which is among the most persistent expressions of gender inequality (European Commission, 2015). It is precisely because of the controversial nature of composite indicators and the difficulty to measure gendered processes that it is crucial to adhere to rigorous international quality standards. The state of the art methodology in this area is the so-called ‘decalogue’ developed jointly by the OECD and JRC (Nardo et al., 2008) and presented in Table 1.

**Table 1: The ‘decatalogue’ of composite indicator development:  
recommended methodological steps**

<b>Step</b>	<b>Topic</b>
<b>1</b>	Developing a theoretical/conceptual framework
<b>2</b>	Establishing data selection criteria
<b>3</b>	Processing data as required
<b>4</b>	Conducting multivariate analysis
<b>5</b>	Normalising indicators
<b>6</b>	Weighting and aggregation indicators
<b>7</b>	Performing uncertainty and sensitivity analyses
<b>8</b>	Relating the composite indicator to other indicators
<b>9</b>	Decomposing the results at sub-domain and indicator level
<b>10</b>	Producing data visualisation and disseminating results

In the following sections, the methodological steps considered in constructing the Gender Diversity Index are set out. This is followed by an analysis of the main features of this tool and how it can be used further to assess the effects of gender diversity in teams on research performance.

## **2 Developing the framework**

Research on gender diversity should invariably start by the question: what is meant by gender diversity? The answer is further complicated by the fact that gender diversity can be understood either as a whole (looking at gender processes within other grounds of diversity) or seeing gender as one category of difference (hence focussing mainly on diversity aspects and gender as one expression of diversity that could be examined). The aim of the Gender Diversity Index as a measurement tool is to transcend the reduction of gender to biological sex only in most quantitative research, and provide a more gender-sensitive measure of gendered processes in research. Gender diversity and team performance are seemingly straightforward concepts. Yet, to measure the effects of gender diversity on team performance, it is important to first define what we understand by these terms and how they are related.

### **2.1 The meaning of gender diversity**

The term ‘gender diversity’ is informed by two research streams: gender research and diversity research, both in teams and in organisations. We use this combination of terms to emphasise that we aim to address the variety of gender realities within team processes. Gender is part of everyday social interactions and, hence, a crucial aspect of social relations within working contexts. This shows, for

instance, in the way gender co-constitutes the symbolic order of an organisation in which practices, such as addressing people, reproduces gender (Gherardi, 1994, 2014). Gender is an inherent aspect of organisations, tasks and occupations (Acker, 1992; Britton, 2000) that affects working groups. To measure this effect, it is important to go beyond a simple head-count and instead implement a more sophisticated approach, which echoes gender processes.

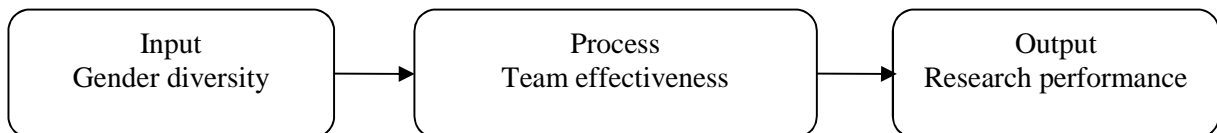
Diversity is a complex concept with different layers. It refers to ways team members vary along different faultlines (Harrison & Klein, 2007; Mayo, Kakarika, Mainemelis, & Deuschel, 2016). Diversity is a group characteristic (van Knippenberg, De Dreu, & Homan, 2004), as it is necessary to have more than one person to establish diversity. Within organisations, people differ along demographic, cognitive and functional characteristics. Demographic diversity categories are, for instance, gender, age and race/ethnicity (Joshi & Roh, 2009) as well as more life course orientated aspects such as tenure, education or marital status (Harrison & Klein, 2007). Cognitive aspects can range from the perception of time (Mohammed & Nadkarni, 2011) to how team processes, such as decision-making, division of labour or power distribution within a working team are perceived (Kilduff, Angelmar, & Mehra, 2000). Functional characteristics comprise different occupations (DiBenigno & Kellogg, 2014), different roles and aspects of seniority. Altogether, diversity can be described as “the distribution of differences among the members of a unit with respect to a common attribute, X, such as tenure, ethnicity, conscientiousness, task attitude, or pay” (Harrison & Klein, 2007, p. 1200).

Aside from the different types of categories on which members can vary, there are also several theoretical approaches. Harrison and Klein (2007) identify three major diversity aspects that affect working groups’ performance: separation, variety and disparity. The first refers to social categorisation, which focuses on subgroups that are related to larger social groups such as gender. This creates separation, which has an impact on team cohesion. The second consists of different forms of information processing, which creates a variety. Working groups could potentially benefit from a higher variety, for instance, in educational backgrounds, which could increase creativity. Third, diversity in the sense of disparate access to resources, power or status impacts working teams. Diversity as disparity therefore refers to the existent inequality within working groups. These different layers of diversity have the potential to influence team effectiveness and performance. Moreover, these three layers interact with each other, prompting Mayo et al. (2016) to suggest taking them all into consideration.

## 2.2 Relating gender diversity to team performance

The logic of research on team performance continues to largely follow the initial basic perspective of ‘input-process-output’ originally outlined by McGrath (1984) (Figure 1). This proposition stipulates that input characteristics (in our case gender diversity), affect team effectiveness (processes and dynamics) and which in turn relate to team outputs themselves (research performance).

**Figure 1: Basic model of input-process-output**



Much of the work on organisational team effectiveness remains atheoretical, despite evidence that the field has reached maturity through some of the work, according to a review conducted by Edmondson and McManus (2007). The most prevalent debates examine the relationship between diversity in teams and various measures of performance. Two opposing perspectives are put forward, and which explain the link between the two in relation to team processes. First, information processing perspectives (Mannix & Neale, 2005) argue that the relationship is positive because diversity in teams bring a wider array of perspectives on which to draw and therefore increase team effectiveness. Second, social categorisation perspectives (Tajfel, 1981; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987) instead argue that in-groups favourings disrupts team work and lowers team effectiveness. These two opposing perspectives have been integrated into a wider framework that suggests that both could affect performance through their effects on team processes and dynamics (van Knippenberg et al., 2004).

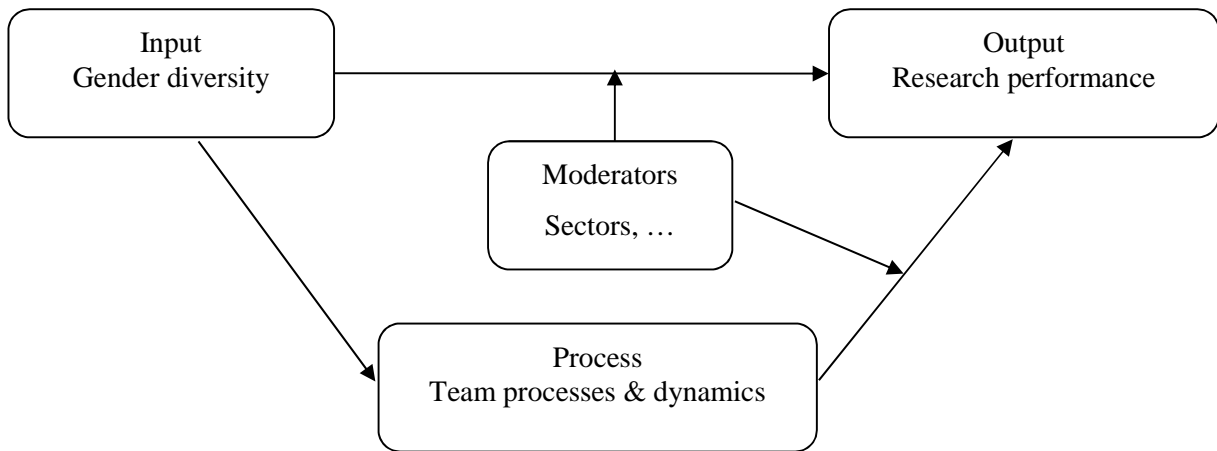
Whether gender diversity in the composition of teams has been associated with better performance has also been hotly debated. Evidence suggests that demographics can affect team dynamics in two key ways: where higher levels of expertise are attributed to those of higher status (status characteristics theory) but also where there is a greater social affinity (social identity theory) (Anderson, Willer, Kilduff, & Brown, 2012; Joshi & Knight, 2015). According to status characteristics theory, one of the reasons gender matters in the composition of research teams is that it acts as a powerful signifier of what tasks are attributed to whom, based on establishing an implicit hierarchy of expertise (Barton & Bunderson, 2014; DiTomaso, Post, Smith, Farris, & Cordero, 2007; Joshi & Knight, 2015). From the perspective of social identity theory, demographic similarity plays an important role in shaping team dynamics through socio-cognitive processes and the creation of in and out groups on the basis of shared membership (Tajfel, 1981; Turner et al., 1987). This leads to valuing different contributions unequally within a team, and favouring those of the in-group more than others.



Affinity is not defined along personal grounds, but instead along social grounds and shared identity features (Hogg, 2001). Gender itself is a particularly salient marker of difference (van den Brink & Benschop, 2012) that constitutes such social groups. One example that illustrates both status theory and social identity theory is deference. Deference is of particular importance where the work is highly specialised and skills are not shared across all members of the team – a setting which might be quite common in the context of research and innovation. Deference can be defined as yielding to others' opinions, beliefs, insights, perspectives or decisions within the process of team work. Lower levels of deference are afforded to women within teams, because their competence and status are perceived as lower (Bunderson, 2003; DiTomaso et al., 2007; Joshi, 2014). Within men-dominated contexts, levels of deference to women are higher if there is a more equal gender representation within a team, because the mechanism of favouring those of the in-group more than others will be weakened (Joshi & Knight, 2015). Similarly, differences are less pronounced with teams that have a narrower hierarchical tier structure (Anderson & Brown, 2010; Bunderson & Reagans, 2011), probably owing to lower disparities across and within levels.

Results of empirical studies exploring the link between gender and effectiveness in teams has produced mixed results, prompting a closer look at the role played by contextual effects, such as culture (Joshi & Roh, 2009). Culture is defined by Maznevski and Chudona (2000, p. 474) as a “set of deep-level values associated with societal effectiveness, shared by an identifiable group of people”. Little attention to date has been given to gendered cultures (Joshi, 2014) and to the different levels at which gender can shape teams (e.g. team diversity, organisational culture, industry/sector, country/societal). However, the extent to which gender can be viewed as culture is debatable, although one might view socialisation processes and resulting gender roles as typifying a feminine and masculine ‘culture’ or ‘cultures’ (Ely & Thomas, 2001). A greater range of cultures multiplies the perspectives available (Adler, 1997; Hofstede, 1980) which can increase potential effectiveness and hence performance. On the contrary, different cultures can hinder intra-team cooperation and decision-making, harming performance in the process, and emphasising that point that greater performance can only be realised where team integration is effective (Maznevski & Chudoba, 2000). Aside from culture, other contextual effects can be considered such as that of the organisation (Joshi, 2014) or that of diversity beliefs (Hentschel, Shemla, Wegge, & Kearney, 2013). Increasingly, research studies have examined more complex models and looked at direct and indirect effects (mediation models) as well as exploring the potential effects of moderators on research performance. It is therefore important to develop models that better account for other contextual factors which may affect team processes and dynamics (Hajro, Gibson, & Pudelko, 2015; Marcus & Le, 2013; Schneid, Isidor, Li, & Kabst, 2015; van Knippenberg et al., 2004), as summarised graphically in Figure 2.

**Figure 2: Contextual effects and the relationship between gender diversity and team performance**



In analysing the effect of gender diversity on team performance, we account for these different layers of diversity while maintaining gender as a cross-cutting issue. For the purpose of this analysis, we build upon existing conceptualisations of diversity (Harrison & Klein, 2007; Mayo et al., 2016; van Dijk, van Engen, & van Knippenberg, 2012; Zhang, 2016) and consider three aspects of gender diversity: (1) demographic: e.g. age or ethnicity; (2) functional: e.g. education, subject area or team tenure; and (3) cognitive: e.g. personality, attitudes or values.

### **2.3 Team processes and performance in STEM**

Little research on how gender diversity impacts team processes in STEM has been conducted. Moreover, most studies assess research performance at the individual level. In this section, we therefore briefly elaborate on how individuals' research assessments are gendered, before developing on how these individualised measures are intertwined with team effectiveness on a broader level.

Different measures have been introduced to assess research outputs beyond a simple count of the number of publications since this method fails to distinguish between different types of output (e.g. peer reviewed articles, books or essays for a practitioners' journal) or the quality of each paper. For example, the physicist Hirsch (2005) introduced the h-index which has become a widely applied measure. The h-index puts the numbers of papers in relation to citations, which has the effect of disregarding non-cited papers but also undervaluing highly cited papers (Egghe, 2006). This measure only includes scientific publications that are cited by peers and consequently other forms of output are not included.

Measures of research output show that women tend to publish less than men (Cole & Zuckerman, 1984; Leahey, 2006). One possible reason is that women specialise less than men (Leahey, 2006). Specialisation in this context means to repeatedly engage in the same research topic(s) and publish on it for an extensive time period or the entire career. Another reason is teaching: women faculty tend to engage more in service-orientated, student-centred teaching which reduces research productivity (Eagan & Garvey, 2015; Hurtado, Eagan, Pryor, Whang, & Tran, 2011). It is not only gender that affects research outputs, but also intersectional factors, with evidence showing that outputs are more numerous among the category of white men faculty members (Eagan & Garvey, 2015). The gender bias does not only show in the quantitative aspect of the h-index (number of publications), but also regarding the indicator for quality of research (number of citations). Articles with women as lead authors tend to get less cited than if the author was a man (Larivière, Ni, Gingras, Cronin, & Sugimoto, 2013; Maliniak, Powers, & Walter, 2013). Men tend to cite other men more often than women, and men more often self-cite their work more than women do (Maliniak et al., 2013). In addition, papers deriving from national collaborations are less cited than such with international author teams which women are less frequently involved in (Larivière et al., 2013).

Overall, women's achievement is often undervalued. Women are less likely to obtain a professorship, even when they have as good a track record as their peers that are men (Treviño, Gomez-Mejia, Balkin, & Mixon, 2015; van den Brink & Benschop, 2012). Evidence also suggests that the criteria used to assess research performance are biased. The analysis of the reviews on applications for postdoctoral fellowships at the Swedish Medical Research Council unveiled nepotism and sexism (Wennerås & Wold, 1997). The authors could demonstrate that peer reviewers tend to overestimate the achievements of men applicants and underrate those of women applicants. Moreover, applicants were reviewed more favourable if they had personal ties, for instance through their supervisor, with one of the reviewers. This matters as network activities are also gendered (van den Brink & Benschop, 2014), with women less likely to benefit from them.

Aside from different gendered processes that impact an individual's research assessment, the immediate working group has also an impact although literature on this is scarce and the results are to some extent inconsistent. Research groups can be considered "as the most important work floor unit in research" (van Raan, 2006, p. 492), particularly so in the context of STEM. The size and composition of a team can have an influence on performance measures, since for example there is less citation traffic within smaller teams (van Raan, 2006). The educational diversity of team members – or interdisciplinarity – also has an impact and results in higher citations but less publications (Leahey, Beckman, & Stanko, 2017). However, little is known about the impact of gender diversity on how the context of the research team relates to the number of research outputs and their quality.

Looking at the role of gender diversity in team processes is also particularly important since research draws on increasingly specialised skills, requiring ever greater collaboration across different specialisms (Stokols, Hall, Taylor, Moser, & Harvey, 2008). It is however not sufficient to look at the number of women and men within research teams, and instead gender processes need to be considered. Based on the body of research on gender and science, this project focuses on some of the key gendered processes within research team that may relate to performance, including both the representation and attrition at – and across – different areas such as team roles or age. In the next section, the methodological steps adopted to build a composite measure of gender diversity – the Gender Diversity Index - within research teams is set out.

### **3 Selection of indicators**

The Gender Diversity Index is unlike many other composite indicators in that it does not rely on secondary data. Instead, data to populate the indicators are obtained through a survey. This is an advantage in that it means that it is possible to draw on data that were collected for the specific purpose of measuring gender diversity in teams and therefore increases the validity of the Gender Diversity Index. There are five requirements that need to be met for indicators to populate the Gender Diversity Index: (1) clearly defined and operationalised; (2) objective and unidirectional; (3) reliable and valid; (4) harmonised and comparable across research teams; (5) easily available. A list of potential indicators was drawn from the main themes identified in the conceptual framework at the outset of the project (Müller et al., 2016). The measure adopted for the Gender Diversity Index builds on sex, as we rely on sex disaggregated data. However, the analysis and interpretation of the Gender Diversity Index is based on an analysis of gendered processes within research teams. In this section, we expand upon the three aspects of gender diversity considered within team processes, demographic, functional and cognitive aspects.

The first type of gender diversity within team processes that is considered is demographic diversity. This includes factors such as ethnicity, marital status or age, and which are located outside of the working place and closely attached to the individual. The relevance of demographic gender diversity can be illustrated through age, as age is part of the masculine norm of an ‘ideal’ scientist is the chrononormative notion (Riach, Rumens, & Tyler, 2014) of a scientific career. There is strong pressure on achieving specific career steps with a certain age, otherwise stereotypes about competencies emerge (Bourdieu, 1994). Moreover, the most important steps within a scientific career are to be made in the late 20s and 30s, which often corresponds to the time when individuals may decide to have children or not. Consequently, this is can be a very demanding life phase which in general affects women differently than men (Kelan, 2014). This is reflected in the data, which shows that women within academia tend to be younger than men (European Commission, 2015).

The second type of gender diversity in team processes that we examine is functional diversity. This includes issues such as differences in hierarchical levels, seniority and roles, team tenure or education. These characteristics are located and distributed within an organisational setting and derive their meaning from these settings. An example of the relevance of functional gender diversity is that of team roles, and particularly within the remit of the so-called leaky pipeline. At each step along an academic career, women are more likely to opt or drop out. For example, in the EU-28 in 2013, the proportion of men within STEM increased from 69% at the level of undergraduates to 87% at the level of full professors and chairs (European Commission, 2015). Women are therefore under-represented at higher levels within teams. It is important to consider hierarchical relations within research teams since they can produce negative effects on performance through impeding the sharing of information (Bunderson, van der Vegt, Cantimur, & Rink, 2016) or creating competition and conflict (Bendersky & Hays, 2012) between members. Where women are disproportionately under-represented in the lower status group, it might restrict the extent to which they feel they share the same goals and equal ability to devise a common agenda, leaving their contribution both silenced and overlooked (Tost, Gino, & Larrick, 2012).

The third aspect of gender diversity considered is cognitive diversity. Much research on cognitive diversity stems from the perceived need to measure diversity beyond team heterogeneity in demographic or functional terms (van Knippenberg et al., 2004). Mello and Rentsch (2015) conducted a review of literature on cognitive diversity and pointed out that there was little overlap between the different concepts of cognitive diversity in the literature. Based on their review, they argue that there are four layers of cognitive diversity. The first consists of traits such as personality, thinking-style or problem solving. These characteristics are regarded as innate and consistent during an adult's life. The second layer, developmental cognitive diversity, includes perceived team values or work values and can be measured by indicators such as team identification or cohesion. The third layer of cognitive diversity focuses on acquired characteristics, such as educational and functional background, as well as expertise. Finally, the fourth layer includes exposed cognitive diversity, which is the most context specific and refers to training conditions and or instructions on decision-making. In our analysis, we focus more particularly on aspects on developmental and acquired cognitive diversity, as those are likely to be affected by work relations within a research team. Mello and Rentsch's (2015) framework also suggests that there are some overlaps between cognitive and functional diversity, as seen in the case of educational background or subject area. However, for the purpose of developing the Gender Diversity Index, in line with other research (Harrison & Klein, 2007; Mayo et al., 2016; van Dijk et al., 2012; Zhang, 2016), this is considered as functional diversity.

Demographic diversity tends to be theorised as having a negative effect on teams functioning while the opposite is true for functional diversity (Mello & Rentsch, 2015). However, empirically there is little – or at least mixed – support for these hypothesised relationships. Cognitive diversity, and its role as either a moderator or mediator, is increasingly taken into consideration. Since many aspects of cognitive diversity are gendered (Müller et al., 2016), we propose that the different layers of gender diversity within team processes are used at different points of the analysis. Demographic and functional gender diversity in team processes are elements that will be considered for inclusion in the Gender Diversity Index. Cognitive gender diversity indicators, however, can be used instead as a mediators and/or moderators in analysing the relationship between gender diversity in teams and research performance. A summary of potential indicators to populate the Gender Diversity Index, informed by the conceptual framework developed within the remit of the GEDII project (Müller et al., 2016) and to be used in further analysis is provided in Table 2.

**Table 2: Potential indicators to populate the Gender Diversity Index and further analyses**

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

In considering different indicators, we prioritise those that are easy accessible, since the GEDII project foresees the development of a self-assessment tool, through which users can determine how gender diverse their team is.

## **4 Building the metrics**

Several different metric types have been considered and evaluated in the development of the Gender Diversity Index. The metrics adopted are introduced below and the rationale for these choices is set out. Illustrative cases are then provided together with further considerations.

### **4.1 Measuring gender diversity**

Gender can be broadly regarded as an “institutionalized system of social practices within society that constitute people as two significantly different categories, men and women, and organize relations of

inequality on the basis of this difference” (Ridgeway, 2007, p. 312). These relations of inequality entail two intertwined components that create unequal structures and processes in research: first the unequal distribution and access to resources and decision-making and second the different value that is culturally attributed to femininity and masculinity (Fraser, 2003). To fully measure gender equality, different articulations of gender processes within research teams, and their meaningful operationalisation in mathematical terms (Harrison & Klein, 2007) need to be developed. This also needs to be done without losing sight of the necessity to rely on data that is easy to compile, if the Gender Diversity Index is to be used not only as a research measure but also as a self-assessment tool.

One of the first methodological considerations to decide upon is how gender equality is conceptualised. Among existing gender-related composite indicators, different approaches have been adopted. For example, the Gender Gap Index developed by the World Economic Forum (Hausmann, Tyson, & Zahidi, 2006; Leopold, Ratcheva, & Zahidi, 2016) considers only gaps that are detrimental to women. They consider there is perfect equality even when women exceed men such as for example in relation to the number of third-level graduates in a European context. On the contrary, the Gender Equality Index developed by the European Institute for Gender Equality (de Bonfils et al., 2013; Humbert, Ivaškaitė-Tamošiūnė, Oetke, & Paats, 2015) states that any kind of gender gap is detrimental, regardless of whether it is women or men that are disadvantaged. The Gender Diversity Index elects to align itself with this latter approach and treat both groups equally.

## **4.2 Gender diversity metrics: measuring representation**

Several different metrics can be considered to measure gender diversity. This section reviews the possibilities and outlines their characteristics based on commonly used metrics in the literature (Harrison & Klein, 2007; Mulcahy & Linehan, 2014; Williams & Meân, 2004). Two different approaches exist: the trait approach and the diversity approach (Williams & Meân, 2004). The trait approach focuses on the proportion of women or men within the group, while the diversity approach instead takes the view that it is not the actual sex-composition of teams that matters but instead differences among women and men in relation to several different characteristics. It is on this latter approach that the proposed metrics are based.

Each of the metric considered yields a number that is bound between 0 and 1 (or is rescaled accordingly), which gives a measure of the distance of equality between women and men for each variable. A score of 1 is obtained when there are exactly 50% of women and 50% of men in any given indicator. The three metrics which were considered for the construction of the Gender Diversity Index are provided below.

$$\gamma_{ijt} = 1 - \left| 1 - \frac{x_{ijt}^w}{0.5} \right| \quad (\text{metric 1})$$

with team  $i$ , in institution  $j$ , at time  $t$ , where  $w$  refers to individual observations  $x$  for women.

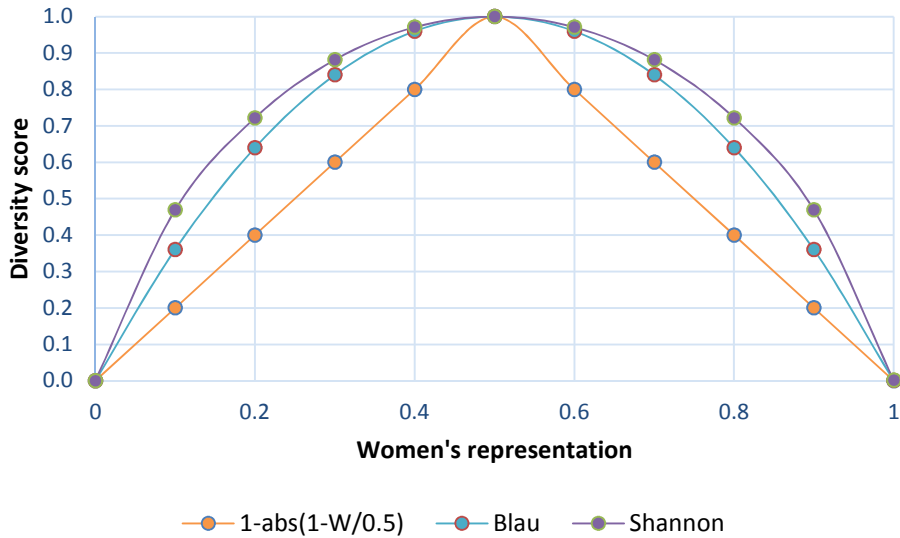
$$Blau_{ijt} = 2 \times \left( 1 - \sum_{s=1}^n P_{sijt}^2 \right) \quad (\text{metric 2})$$

$$Shannon_{ijt} = \frac{-\sum_{s=1}^n P_{sijt} \ln P_{sijt}}{0.693147180559945} \quad (\text{metric 3})$$

where in both cases  $s = w, m$  and  $n = 2$  and with team  $i$ , in institution  $j$ , at time  $t$ .

A comparison of these three metrics is provided in Figure 3. Because the Shannon Index is based on logarithms and the Blau Index a quadratic function, they are by construction more sensitive to smaller values and may thus be more appropriate to capture critical mass effects. A drawback of using a logarithm-based metric is that it is not defined where  $P_i$  is equal to 0. Even though this is easily fixed by imposing a value of 0 in those instances, it may be preferable instead to adopt the Blau Index given that it only marginally differs from the Shannon Index and simplifies the computation for end-users. We therefore propose to adopt the Blau index within the Gender Diversity Index as a measure of gender representation, thereafter referred to as the ‘ $\alpha$  metric’.

**Figure 3: Comparison of diversity metrics to capture representation**



All three metrics provide information about the extent to which teams are gender diverse, but can only do so on a single level. This means that they do not capture any potential glass ceiling effects. For example, these metrics can be used to provide a measure of the representation of women among



senior researchers and team leaders. However, none of these metrics provide information on relative terms. This means that the gender composition of research teams at the lower levels is not taken into consideration. In other words, when women and men are equally represented at higher levels, all metrics produce a score of 1 regardless of whether there is gender inequality at a lower level. To add information on potential glass ceiling effects within teams, the next section focuses on relative measures of gender diversity.

### 4.3 Relative gender diversity metric

Relative gender diversity metrics capture gender representation at different levels, as opposed to the metrics that only measure gender representation on one level. They provide information on attrition or, to use the popular metaphor, on the extent to which the pipeline is leaking. A measure of attrition is provided by She Figures (European Commission, 2015) and consists of the ratio of the percentage of women at the higher level to the percentage of women at all levels. A similar Glass Ceiling Index can be calculated for men. Where women are less represented at higher levels, the Glass Ceiling Index produces a number that is below 1. If there is no attrition, then it will produce a score of 1. There is no upper bound and values can exceed 1 where there are more women at higher level than in other levels.

For the Gender Diversity Index, an alternative metric derived from the Glass Ceiling Index is proposed. First, it is designed to ensure that scores are bound between 0 and 1. Second, it is linked to the need for the metric to provide information on attrition regardless of whether inequalities are on the side of women or men as opposed to two separate measures for women and men. Indeed, rather than focus on either women and men, the proposed metric considers attrition for the under-represented group and constraints are introduced to ensure that it is bound between values of 0 and 1. It is expressed as:

$$\delta_{ijt} = \begin{cases} \frac{x_{2ijt}^w}{x_{1ijt}^w} & \text{if } x_{2ijt}^w < x_{2ijt}^m \text{ and } x_{2ijt}^w < x_{1ijt}^w \\ \frac{x_{2ijt}^m}{x_{1ijt}^m} & \text{if } x_{2ijt}^m < x_{2ijt}^w \text{ and } x_{2ijt}^m < x_{1ijt}^m \\ 1 & \text{otherwise} \end{cases} \quad (\text{metric 4})$$

with level 1 or 2, team  $i$ , in institution  $j$  at time  $t$  where  $w$  and  $m$  refer to individual observations  $x$  and proportions in the population of reference for women and men respectively.

### 4.4 Examples – examining gender differences in team roles

We now turn to some examples to illustrate the metrics selected for the building of the Gender Diversity Index and their properties in different situations. In these examples, the role within the team

is coded as 1 for lower positions and 2 for higher positions (i.e. either senior researcher or team leader). This is based on understanding being in a higher position as ‘more desirable’ and therefore a positive outcome. Conceptually, it is based on understanding women’s lack of proportional representation in higher positions as problematic (European Commission, 2015).

The first example (Case 1) represents a situation where women are strongly over-represented at the lower level, but where there is an increase in men’s representation at the higher level to the extent that it is near parity.

***Case 1: decreasing inequalities***

	<b>Women</b>	<b>Men</b>
<b>Level 1</b>	90%	10%
<b>Level 2</b>	60%	40%

$$\alpha_1 = 2 \times (1 - (0.9^2 + 0.1^2)) = 0.36$$

$$\alpha_2 = 2 \times (1 - (0.6^2 + 0.4^2)) = 0.96$$

$$\delta = 1$$

In Case 1, the representation metric at level 1 ( $\alpha_1$ ) shows that there are considerable inequalities in representation with a score of 0.36. The representation metric at level 2 ( $\alpha_2$ ), is close to 1 with a score of 0.96, since there is near equality at that level. Because there has been an improvement in gender representation from level 1 to level 2, the attrition metric ( $\delta$ ) provides a score of 1. It is important to note that the attrition metric provided cannot discern how much progress has been made. It produces a score of 1 regardless of whether there is a marginal or a strong improvement. For this reason, we suggest that the attrition metric should not be used without the corresponding measures of representation.

***Case 2: decreasing inequalities with reversal in inequalities***

	<b>Women</b>	<b>Men</b>
<b>Level 1</b>	90%	10%
<b>Level 2</b>	40%	60%

$$\alpha_1 = 2 \times (1 - (0.9^2 + 0.1^2)) = 0.36$$

$$\alpha_2 = 2 \times (1 - (0.6^2 + 0.4^2)) = 0.96$$

$$\delta = \frac{0.4}{0.9} = 0.44$$

In Case 2, the representation metrics ( $\alpha_1$  and  $\alpha_2$ ) are the same as in Case 1. This is because inequalities in representation are the same, although there has been a reversal in the side of the inequality. Men are under-represented at level 1 but over-represented at level 2. This is captured by the attrition metric ( $\delta$ ) which drops to 0.44.

***Case 3: increasing inequalities***

	<b>Women</b>	<b>Men</b>
<b>Level 1</b>	60%	40%
<b>Level 2</b>	90%	10%

$$\alpha_1 = 2 \times (1 - (0.6^2 + 0.4^2)) = 0.96$$

$$\alpha_2 = 2 \times (1 - (0.9^2 + 0.1^2)) = 0.36$$

$$\delta = \frac{0.1}{0.4} = 0.25$$

In Case 3, the representation metrics ( $\alpha_1$  and  $\alpha_2$ ) are the same as in Cases 1 and 2 but in reverse order. This represents a case where there are increasing inequalities between the two levels. This is picked up by the attrition metric ( $\delta$ ) which provides a score of 0.25 since inequalities are increasing. It is worth noting that the nominal value of the attrition metric ( $\delta$ ) in this case is lower than in Case 2 where there was less inequality but where the attrition metric ( $\delta$ ) picks up on the reversal of the side of the inequality.

***Case 4: increasing inequalities and reversal***

	<b>Women</b>	<b>Men</b>
<b>Level 1</b>	40%	60%
<b>Level 2</b>	90%	10%

$$\alpha_1 = 2 \times (1 - (0.4^2 + 0.6^2)) = 0.96$$

$$\alpha_2 = 2 \times (1 - (0.9^2 + 0.1^2)) = 0.36$$

$$\delta = \frac{0.1}{0.6} = 0.17$$

In Case 4, the representation metrics ( $\alpha_1$  and  $\alpha_2$ ) have the same values as in Case 3 because representation is the same at levels 1 and 2 but is reversed at level 2. Consequently, the attrition metric ( $\delta$ ) drops further to 0.17, which is nominally smaller than in Case 3, because there is both decreasing inequalities and a reversal in inequalities.

#### **4.5 Further considerations**

The example above looked at hierarchical levels and roles within research teams. Since these can be measured as categorical variables, it is straightforward to construct levels. The situation is different when it comes to numerical variables, such as for instance team tenure which is measured in months. This can be solved by taking a reference point to construct categories. A simple reference point can be the average, which allows for a comparison of the proportion of women and men that are above average team tenure compared to those below that average. The main drawback of this approach is that it must be decided whether to adopt a sample average, or alternatively an average drawn from the subject-area or national context.

The attrition metric ( $\delta$ ) used relies on identifying a direction for the variable. Because the representation metrics ( $\alpha_1$  and  $\alpha_2$ ) are expressed in relative terms, there is no need to specify direction. The direction of an indicator is assessed by evaluating whether an increase or decrease in its value is associated with an increase in the value of the concept measured. In the example of team roles, it is possible to determine that being in a higher position is ‘better’ than being in a lower position because women and men should have equal opportunities for advancement. This means that vertical segregation is seen as detrimental since it does not allow for equal chances to progress to higher levels. This is in line with policy discourse, including for example She Figures (European Commission, 2015) and the famous ‘scissors’ representation of women’s attrition at higher levels within academia. However, in the case of team tenure, direction is more difficult to assess. It is therefore necessary to consider whether there is an implicit order in this instance. If framed within a logic of retention, it is possible in this instance to conclude that longer team tenure is associated with greater know-how and experience, which in turn can positively affect research performance (Joshi, 2014; Kochan et al., 2003). Staying within a team longer may also be conceptualised as a marker of well-being within and commitment to the team. At the same time, team tenure is also bound to team roles: PhD students have a fixed team tenure lasting usually three-five years; upon completion of their dissertation many have to move on to other organisations and teams. In the following section, we illustrate how the metrics proposed can be operationalised into a composite indicator on the basis of the conceptual structure of gender diversity discussed above.

## 5 Multivariate analysis

Indicators are not only selected on the basis of a literature review but also through statistical analysis such as exploratory data analysis, principal component analysis etc. (Nardo et al., 2008). The construction of the Gender Diversity Index requires an examination of the latent structure that underpins it, through the application of Factor Analysis (Hair, Black, Babin, Anderson, & Tatham, 2006). This property of composite indicators provide a methodologically robust measure which addresses concerns in the literature about the correlations that exist between different types of diversity (van Dijk et al., 2012). Usual practice requires that the factor solution retained satisfies the following criteria: eigen-values should be above or near 1 (Kaiser, 1960), loadings should ideally exceed 0.5 with no cross-loadings above that value with other factors (Hair et al., 2006), the validity of each factor should be checked through the use of a Cronbach test (Cronbach, 1951) with values reaching at least 0.6 as the lowest acceptable threshold for newly developed scales (Nunnally & Bernstein, 1994).

To better understand what the Gender Diversity Index might look like, a pilot was conducted using the data gathered via the case studies which formed part of the GEDII project (Work Package 2). We use these data to illustrate the approach only. Given the small sample size the numbers presented need to be interpreted with care. A summary of the data is provided in Table 3 with indicators constructed using the following variables:

- Sex (1: woman; 2: man);
- Number of months in the team;
- Role within the team (1: not team leader or senior researcher; 2: team leader or senior researcher);
- Qualifications (1: no doctorate; 2: doctorate)
- Age (1: below 35; 2: 35 and above)

These data were then used to compute a dashboard of potential indicators using the metrics presented in the previous section. These results are presented in Table 4.

**Table 3: Descriptive statistics for pilot teams**

Team	% women		Average team tenure (months)	Lower than average team tenure	Higher than average team tenure	Non-team leaders or senior researchers	Team leaders or senior researchers	No doctorate	Doctorate	Under 35	Over 35
1	40%	Women	7	40%	40%	25%	50%	43%	33%	44%	0%
		Men	11	60%	60%	75%	50%	57%	67%	56%	100%
2	70%	Women	17	71%	67%	67%	75%	80%	60%	80%	50%
		Men	23	29%	33%	33%	25%	20%	40%	20%	50%
3	56%	Women	64	40%	75%	75%	40%	50%	67%	40%	75%
		Men	45	60%	25%	25%	60%	50%	33%	60%	25%
4	29%	Women	37	40%	0%	25%	33%	33%	25%	0%	33%
		Men	78	60%	100%	75%	67%	67%	75%	100%	67%
5	38%	Women	29	40%	33%	33%	50%	0%	50%	50%	25%
		Men	27	60%	67%	67%	50%	100%	50%	50%	75%
6	63%	Women	109	60%	67%	40%	100%	60%	67%	80%	33%
		Men	41	40%	33%	60%	0%	40%	33%	20%	67%

**Table 4: Dashboard of potential indicators to populate the Gender Diversity Index**

Team	Sex	Team tenure			Team roles			Qualifications			Age		
	$\alpha$	$\delta$	$\alpha_1$	$\alpha_2$	$\delta$	$\alpha_1$	$\alpha_2$	$\delta$	$\alpha_1$	$\alpha_2$	$\delta$	$\alpha_1$	$\alpha_2$
1	0.96	1	0.96	0.96	1	0.75	1	0.78	0.98	0.89	0	0.99	0
2	0.84	1	0.82	0.89	0.75	0.89	0.75	1	0.64	0.96	1	0.64	1
3	0.99	0.42	0.96	0.75	0.53	0.75	0.96	0.67	1	0.89	0.42	0.96	0.75
4	0.82	0	0.96	0	1	0.75	0.89	0.75	0.89	0.75	1	0	0.89
5	0.94	0.83	0.96	0.89	1	0.89	1	1	0	1	0.5	1	0.75
6	0.93	0.83	0.96	0.89	0	0.96	0	0.83	0.96	0.89	0.42	0.64	0.89

The dashboard of potential indicators is useful in identifying areas where gender diversity has been largely achieved and those where improvements are required. In Table 4, results are colour coded from red to green (worst to best position) to provide a visual representation of where each team stands. The results are colour-coded in percentile rather than through bins. This means that the shade attributed to each indicator is relative to the overall distribution of other diversity metrics across teams and across indicators, instead of numbers being pooled into bins which are then attributed a colour. This works well where data are available for other teams, however, the self-assessment tool should consider opting for a bin colour-coding based on the results obtained from the results of the analysis in the project. This approach better unpicks the in-built variation in relation to the data in existing teams rather than that of the wider range of all possible values the metrics can take.

While a dashboard of indicators can be a useful tool, the aim of a composite indicator such as the Gender Diversity Index is to aggregate different indicators, in line with a conceptually and statistically sound structure. Methodologically, it is important to ensure that the structure used to aggregate different indicators conforms to the underlying correlation structure of the indicators used. This correlation structure can be assessed through looking at the correlation matrix, computing reliability factors and performing principal components analysis (Nardo et al., 2008). Preliminary results are presented below.

The correlation matrix (Table 5) shows there are several high positive correlations within the matrix. However, it also shows that there is tendency for related representation metrics ( $\alpha_1$  and  $\alpha_2$ ) to be strongly negatively correlated. For example, the correlation between the  $\alpha_1$  and  $\alpha_2$  for team roles is -0.73, for qualifications -0.67 and for age -0.48. This suggests that it may be necessary to drop one of the measures from the composite indicator. There are also negative correlations between lower level Blau indices ( $\alpha_1$ ) and corresponding measures of attrition ( $\delta$ ): for roles -0.57 or for age -0.73. Consequently, retaining the Blau indices for the higher levels ( $\alpha_2$ ) is more desirable. This also makes sense conceptually, since the emphasis of policy is to ensure that women and men are equally represented particularly at higher levels such as for example at professorial level.

**Table 5: Correlation matrix**

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Sex	1.00												
2 Tenure attrition	0.22	1.00											
3 Tenure lower Blau	0.51	-0.40	1.00										
4 Tenure higher Blau	0.55	0.92	-0.21	1.00									
5 Roles attrition	-0.28	-0.14	-0.04	-0.29	1.00								
6 Roles lower Blau	-0.17	0.53	-0.30	0.46	-0.57	1.00							
7 Roles higher Blau	0.08	-0.19	0.02	-0.17	0.88	-0.73	1.00						
8 Qualifications attrition	-0.43	0.60	-0.58	0.39	0.16	0.70	-0.07	1.00					
9 Qualifications lower Blau	0.16	-0.27	0.13	-0.21	-0.40	-0.40	-0.29	-0.79	1.00				
10 Qualifications higher Blau	0.29	0.79	-0.37	0.84	-0.03	0.55	0.05	0.72	-0.67	1.00			
11 Age attrition	-0.83	-0.43	-0.56	-0.59	0.12	0.14	-0.02	0.31	-0.18	-0.22	1.00		
12 Age lower Blau	0.80	0.68	0.08	0.87	-0.05	0.10	0.20	0.15	-0.26	0.77	-0.73	1.00	
13 Age higher Blau	-0.48	-0.33	-0.39	-0.32	-0.37	0.49	-0.38	0.29	-0.22	0.00	0.82	-0.48	1.00

A multivariate analysis was conducted and shows that the latent correlation structure is appropriate for the construction of a composite indicator. It identifies three domains (three components with eigen-values above 1), each with high levels of reliability (Table 6). Due to the very small sample size (six teams), the results need to be interpreted with caution and it was not possible to compute measures of fit such as KMO or BTS. Considering the high loadings, it appears that the fit is adequate. Even though the metrics capture four areas of gender diversity, the Factor Analysis merges qualifications and team tenure. This suggests that the two are highly correlated, but it remains to be seen whether that will persist when the sample size is increased. This consists of a pilot only, and the analysis will be replicated using the data gathered through the survey questionnaire.

**Table 6: Rotated component matrix and reliability coefficients**

	Component		
	1	2	3
<b>Alpha</b>	<b>0.813</b>		
Qualifications $\alpha_2$	.954		
Team tenure $\delta$	.903	-.316	
Team tenure $\alpha_2$	.842	-.450	
Qualifications $\delta$	.820	.479	
<b>Alpha</b>		<b>0.898</b>	
Age $\delta$		.956	
Age $\alpha_2$		.902	-.354
<b>Alpha</b>			<b>0.939</b>
Roles $\delta$			.988
Roles $\alpha_2$			.943

Extraction Method: Principal Component Analysis.  
 Rotation Method: Varimax with Kaiser Normalization.  
 Loadings below |0.3| are omitted



The results of the multivariate analysis provide the basis on which to aggregate indicators. Various choices to do so are available in relation to weights and aggregation methods. For this preliminary analysis, aggregation is done using the simplest methods available: equal weights are applied to all indicators and the aggregation is done using the arithmetic average. This produces the results provided in Table 7. Because of the very small sample size of the pilot study, these results are only indicative and should only be used to illustrate potential methodological steps when the survey data become available.

**Table 7: Computing the Gender Diversity Index – preliminary results**

Team	Team tenure		Qualifications		Age		Team roles		Tenure and qualifications	Age	Roles	GDI	Rank
	$\delta$	$\alpha_2$	$\delta$	$\alpha_2$	$\delta$	$\alpha_2$	$\delta$	$\alpha_2$					
1	1.00	0.96	0.78	0.89	0.00	0.00	1.00	1.00	0.91	0.00	1.00	0.64	5
2	1.00	0.89	1.00	0.96	1.00	1.00	0.75	0.75	0.96	1.00	0.75	0.90	1
3	0.42	0.75	0.67	0.89	0.42	0.75	0.53	0.96	0.68	0.59	0.75	0.67	4
4	0.00	0.00	0.75	0.75	1.00	0.89	1.00	0.89	0.38	0.95	0.95	0.76	3
5	0.83	0.89	1.00	1.00	0.50	0.75	1.00	1.00	0.93	0.63	1.00	0.85	2
6	0.83	0.89	0.83	0.89	0.42	0.89	0.00	0.00	0.86	0.66	0.00	0.51	6

This section has provided an outline of what the future Gender Diversity Index might look like and the methodological steps that will be adopted. In the next section, we summarise the work accomplished to date and provide an outline roadmap for the work ahead in developing the Gender Diversity Index.

## 6 Preliminary conclusions and further steps for development

The data obtained through the case studies allowed for a preliminary analysis of possible ways of constructing the Gender Diversity Index. Using the data obtained through the survey, these steps will be retraced and the results used to make final decisions in relation to the methodology employed. The future steps are likely to include the following:

1. Discuss the compensability allowed between different indicators and the ensuing aggregation method employed.
2. Discuss the weighting method employed, and whether the correlation structure among indicators should be considered.
3. Conduct an intermediate coherence check.
4. Consider the possibility of modelling different assumptions through sensitivity analysis.
5. Analyse the Gender Diversity Index in relation to other indicators within the survey and in relation to various measures of research performance.

The Gender Diversity Index is a measure that can be used to empirically test whether greater gender diversity in teams, particularly in the context of STEM, is related to increased research performance. The aim is to go beyond a simple head count and to capture gendered processes within a research team. Through this measure, we hope to generate more reliable insights in what way gender diversity impacts research performance. Different performance measures will be considered, including publications outputs and patents. Furthermore, it is not only the relationship between gender diversity in teams and research performance that will be assessed, but also potential mediators and moderators of that relationship to respond to calls in the literature to better take into account contextual effects on the research performance. Besides the theoretical development arising from this work, we expect that there will also be some practical implications. The Gender Diversity Index will be made available to the public through an online self-assessment tool. This tool will allow research teams to obtain a score which they can then use to monitor progress over time in regard to gender equality and benchmark themselves against other teams.

## References

- Acker, J. (1992). From sex roles to gendered institutions. *Contemporary Sociology*, *21*(5), 565–569. <http://doi.org/10.2307/2075528>
- Adler, E. (1997). Seizing the Middle Ground: Constructivism in World Politics. *European Journal of International Relations*, *3*(3), 319–363. <http://doi.org/10.1177/1354066197003003003>
- Anderson, C., & Brown, C. E. (2010). The functions and dysfunctions of hierarchy. *Research in Organizational Behavior*, *30*, 55–89. <http://doi.org/10.1016/j.riob.2010.08.002>
- Anderson, C., Willer, R., Kilduff, G. J., & Brown, C. E. (2012). The origins of deference: When do people prefer lower status? *Journal of Personality and Social Psychology*, *102*(5), 1077–1088. <http://doi.org/10.1037/a0027409>
- Armstrong, C., Flood, P. C., Guthrie, J. P., Liu, W., MacCurtain, S., & Mkamwa, T. (2010). The impact of diversity and equality management on firm performance: Beyond high performance work systems. *Human Resource Management*, *49*(6), 977–998. <http://doi.org/10.1002/hrm.20391>
- Barton, M. A., & Bunderson, J. S. (2014). Assessing member expertise in groups. *Organizational Psychology Review*, *4*(3), 228–257. <http://doi.org/10.1177/2041386613508975>
- Bendersky, C., & Hays, N. A. (2012). Status Conflict in Groups. *Organization Science*, *23*(2), 323–340. <http://doi.org/10.1287/orsc.1110.0734>
- Bourdieu, P. (1994). *Homo Academicus*. Stanford University Press.
- Britton, D. M. (2000). The Epistemology of the Gendered Organization. *Gender & Society*, *14*(3),

- 418–434. <http://doi.org/10.1177/089124300014003004>
- Bunderson, J. S. (2003). Recognizing and Utilizing Expertise in Work Groups: A Status Characteristics Perspective. *Administrative Science Quarterly*, 48(4), 557–591. <http://doi.org/10.2307/3556637>
- Bunderson, J. S., & Reagans, R. E. (2011). Power, Status, and Learning in Organizations. *Organization Science*, 22(5), 1182–1194. <http://doi.org/10.1287/orsc.1100.0590>
- Bunderson, J. S., van der Vegt, G. S., Cantimur, Y., & Rink, F. (2016). Different Views of Hierarchy and Why They Matter: Hierarchy as Inequality or as Cascading Influence. *Academy of Management Journal*, 59(4), 1265–1289. <http://doi.org/10.5465/amj.2014.0601>
- Cole, J., & Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. In M. W. Steinkamp & M. L. Maehr (Eds.), *Advances in Motivation and Achievement* (pp. 217–258). Greenwich & London: JAI Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <http://doi.org/10.1007/BF02310555>
- de Bonfils, L., Humbert, A. L., Ivaškaitė-Tamošiūnė, V., Manca, A. R., Nobrega, L., Reingardė, J., & Lestón, I. R. (2013). Gender Equality Index Report. Retrieved November 21, 2016, from <http://eige.europa.eu/content/document/gender-equality-index-report>
- Dezső, C. L., & Ross, D. G. (2012). Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal*, 33(9), 1072–1089. <http://doi.org/https://doi.org/10.1002/smj>
- DiBenigno, J., & Kellogg, K. C. (2014). Beyond Occupational Differences. *Administrative Science Quarterly*, 59(3), 375–408. <http://doi.org/10.1177/0001839214538262>
- DiTomaso, N., Post, C., Smith, D. R., Farris, G. F., & Cordero, R. (2007). Effects of Structural Position on Allocation and Evaluation Decisions for Scientists and Engineers in Industrial R&D. *Administrative Science Quarterly*, 52(2), 175–207. <http://doi.org/10.2189/asqu.52.2.175>
- Eagan, K., & Garvey, J. C. (2015). Stressing Out: Connecting Race, Gender, and Stress with Faculty Productivity. *The Journal of Higher Education*, 86(6), 923–954. <http://doi.org/10.1353/jhe.2015.0034>
- Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, 32(4), 1246–1264. <http://doi.org/10.5465/AMR.2007.26586086>
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152. <http://doi.org/10.1007/s11192-006-0144-7>
- Ely, R. J., & Thomas, D. A. (2001). Cultural Diversity at Work: The Effects of Diversity Perspectives on Work Group Processes and Outcomes. *Administrative Science Quarterly*, 46(2), 229. <http://doi.org/10.2307/2667087>
- Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge: Polity.

- European Commission. (2015). *She Figures*. <http://doi.org/10.2777/744106>
- European Commission. (2016). What is a composite indicator? Retrieved March 13, 2017, from <https://composite-indicators.jrc.ec.europa.eu/?q=content/what-composite-indicator>
- Fraser, N. (2003). Rethinking recognition: overcoming displacement and reification in cultural politics. In B. Hobson (Ed.), *Recognition struggles and social movements: Contested identities, agency and power* (pp. 21–32). Cambridge: Cambridge University Press.
- Gherardi, S. (1994). The Gender We Think, The Gender We Do in Our Everyday Organizational Lives. *Human Relations*, 47(6), 591–610. <http://doi.org/10.1177/001872679404700602>
- Gherardi, S. (2014). Organizations as Symbolic Gendered Orders. In S. Kumra, R. Simpson, & R. J. Burke (Eds.), *The Oxford Handbook of Gender in Organizations*. (pp. 76–94). Oxford: Oxford University Press. <http://doi.org/10.1093/oxfordhb/9780199658213.013.026>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis 6th Edition*. New Jersey: Pearson Education.
- Hajro, A., Gibson, C., & Pudelko, M. (2015). Knowledge Exchange Processes in Multicultural Teams: Linking Organizational Diversity Climates to Teams' Effectiveness. *Academy of Management Journal*, amj.2014.0442. <http://doi.org/10.5465/amj.2014.0442>
- Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, 32(4), 1199–1228. <http://doi.org/10.5465/AMR.2007.26586096>
- Hausmann, R., Tyson, L. D., & Zahidi, S. (2006). *The Global Gender Gap Report 2006*. Retrieved from [http://www3.weforum.org/docs/WEF\\_GenderGap\\_Report\\_2006.pdf](http://www3.weforum.org/docs/WEF_GenderGap_Report_2006.pdf)
- Hentschel, T., Shemla, M., Wegge, J., & Kearney, E. (2013). Perceived Diversity and Team Functioning. *Small Group Research*, 44(1), 33–61. <http://doi.org/10.1177/1046496412470725>
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. <http://doi.org/10.1073/pnas.0507655102>
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Thousand Oaks California: Sage Publications.
- Hogg, M. A. (2001). A social identity theory of leadership. *Personality and Social Psychology Review*, 5(3), 184–200. [http://doi.org/10.1207/S15327957PSPR0503\\_1](http://doi.org/10.1207/S15327957PSPR0503_1)
- Humbert, A. L., Ivaškaitė-Tamošiūnė, V., Oetke, N., & Paats, M. (2015). Gender Equality Index 2015 - Measuring gender equality in the European Union 2005-2012. Retrieved November 21, 2016, from <http://eige.europa.eu/sites/default/files/documents/mh0215616enn.pdf>
- Hurtado, S., Eagan, K., Pryor, J. H., Whang, H., & Tran, S. (2011). *Undergraduate Teaching Faculty: The 2010 – 2011 HERI Faculty Survey*.
- Joshi, A. (2014). By Whom and When Is Women's Expertise Recognized? The Interactive Effects of Gender and Education in Science and Engineering Teams. *Administrative Science Quarterly*,

- 59(2), 202–239. <http://doi.org/10.1177/0001839214528331>
- Joshi, A., & Knight, A. P. (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. <http://doi.org/10.5465/amj.2013.0718>
- Joshi, A., & Roh, H. (2009). The Role Of Context In Work Team Diversity Research: A Meta-Analytic Review. *Academy of Management Journal*, 52(3), 599–627. <http://doi.org/10.5465/AMJ.2009.41331491>
- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <http://doi.org/10.1177/001316446002000116>
- Kanter, R. M. (1984). *Change masters*. New York: Simon and Schuster.
- Kelan, E. K. (2014). From Biological Clocks to Unspeakable Inequalities: The Intersectional Positioning of Young Professionals. *British Journal of Management*, 25(4), 790–804. <http://doi.org/10.1111/1467-8551.12062>
- Kilduff, M., Angelmar, R., & Mehra, A. (2000). Top Management-Team Diversity and Firm Performance: Examining the Role of Cognitions. *Organization Science*, 11(1), 21–34. <http://doi.org/10.1287/orsc.11.1.21.12569>
- Kochan, T., Bezrukova, K., Ely, R., Jackson, S., Joshi, A., Jehn, K., ... Thomas, D. (2003). The effects of diversity on business performance: Report of the diversity research network. *Human Resource Management*, 42(1), 3–21. <http://doi.org/10.1002/hrm.10061>
- Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. R. (2013). Bibliometrics: Global gender disparities in science. *Nature*, 504(7479), 211–213. <http://doi.org/10.1038/504211a>
- Leahey, E. (2006). Gender differences in productivity: Research specialization as a missing link. *Gender & Society*, 20(6), 754–780. <http://doi.org/10.1177/0891243206293030>
- Leahey, E., Beckman, C. M., & Stanko, T. L. (2017). Prominent but less productive. *Administrative Science Quarterly*, 62(1), 105–139. <http://doi.org/10.1177/0001839216665364>
- Leopold, T. A., Ratcheva, V., & Zahidi, S. (2016). *The Global Gender Gap Report 2016*. Retrieved from [http://www3.weforum.org/docs/GGGR16/WEF\\_Global\\_Gender\\_Gap\\_Report\\_2016.pdf](http://www3.weforum.org/docs/GGGR16/WEF_Global_Gender_Gap_Report_2016.pdf)
- Lewis, J. (1992). Gender and the development of welfare regimes. *Journal of European Social Policy*, 2(3), 159–173. <http://doi.org/10.1177/095892879200200301>
- Maliniak, D., Powers, R., & Walter, B. F. (2013). The Gender Citation Gap in International Relations. *International Organization*, 67(4), 889–922. <http://doi.org/10.1017/S0020818313000209>
- Mannix, E., & Neale, M. A. (2005). What differences make a difference? *Psychological Science in the Public Interest*, 6(2), 31–55. <http://doi.org/10.1111/j.1529-1006.2005.00022.x>
- Marcus, J., & Le, H. (2013). Interactive effects of levels of individualism-collectivism on cooperation: A meta-analysis. *Journal of Organizational Behavior*, 34(6), 813–834. <http://doi.org/10.1002/job.1875>
- Mayo, M., Kakarika, M., Mainemelis, C., & Deuschel, N. T. (2016). A metatheoretical framework of

- diversity in teams. *Human Relations*, 1872671667924. <http://doi.org/10.1177/0018726716679246>
- Maznevski, M. L., & Chudoba, K. M. (2000). Bridging space over time: Global virtual team dynamics and effectiveness. *Organization Science*, 11(5), 473–492. <http://doi.org/10.1287/orsc.11.5.473.15200>
- McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Prentice-Hall Englewood Cliffs, NJ.
- Mello, A. L., & Rentsch, J. R. (2015). Cognitive Diversity in Teams: A Multidisciplinary Review. *Small Group Research*, 46(6), 623–658. <http://doi.org/10.1177/1046496415602558>
- Miles, P., & Erhardt, N. (2014). Top management diversity is just good business: Using CSR measures to explore the diversity-firm performance link. *Journal of Business and Economics*, 5(4), 449–461.
- Mohammed, S., & Nadkarni, S. (2011). Temporal diversity and team performance: The moderating role of team temporal leadership. *Academy of Management Journal*, 54(3), 489–508. <http://doi.org/10.5465/AMJ.2011.61967991>
- Mulcahy, M., & Linehan, C. (2014). Females and precarious board positions: Further evidence of the glass cliff. *British Journal of Management*, 25(3), 425–438. <http://doi.org/10.1111/1467-8551.12046>
- Müller, J., Klatt, S., Callerstig, A.-C., & Sandström, U. (2016). *D1.1 Conceptual Framework* (Project Deliverable).
- Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008). *Handbook on constructing composite indicators - Methodology and User Guide*. OECD publishing. Retrieved from <https://www.oecd.org/std/42495745.pdf>
- Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. In *Psychometric Theory* (Vol. 3, pp. 248–292). New York: McGraw-Hill.
- Riach, K., Rumens, N., & Tyler, M. (2014). Un/doing chrononormativity: Negotiating ageing, gender and sexuality in organizational life. *Organization Studies*, 35(11), 1677–1698. <http://doi.org/10.1177/0170840614550731>
- Ridgeway, C. L. (2007). Gender as a Group Process: Implications for the Persistence of Inequality. In S. J. Correll (Ed.), *Social Psychology of Gender* (pp. 311–333). Elsevier. [http://doi.org/10.1016/S0882-6145\(07\)24012-3](http://doi.org/10.1016/S0882-6145(07)24012-3)
- Sainsbury, D. (1996). *Gender, equality and welfare states*. Cambridge University Press.
- Saisana, M. (2016). Indices and Scoreboards: Ten Steps Guide. In *14th JRC Annual Training on Composite Indicators & Scoreboards, 26-28 September*. Ispra.
- Schneid, M., Isidor, R., Li, C., & Kabst, R. (2015). The influence of cultural context on the relationship between gender diversity and team performance: a meta-analysis. *The International Journal of Human Resource Management*, 26(6), 733–756.

- <http://doi.org/10.1080/09585192.2014.957712>
- Stokols, D., Hall, K. L., Taylor, B. K., Moser, R. P., & Harvey, R. (2008). The science of team science. *American Journal of Preventive Medicine*, 35(2), S77–S89. <http://doi.org/10.1016/j.amepre.2008.05.002>
- Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. CUP Archive.
- Tost, L. P., Gino, F., & Larrick, R. P. (2012). Power, competitiveness, and advice taking: Why the powerful don't listen. *Organizational Behavior and Human Decision Processes*, 117(1), 53–65. <http://doi.org/10.1016/j.obhdp.2011.10.001>
- Treviño, L. J., Gomez-Mejia, L. R., Balkin, D. B., & Mixon, F. G. (2015). Meritocracies or Masculinities? The Differential Allocation of Named Professorships by Gender in the Academy. *Journal of Management*, XX(X), 14920631559921. <http://doi.org/10.1177/0149206315599216>
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Basil Blackwell.
- van den Brink, M., & Benschop, Y. (2012). Gender practices in the construction of academic excellence: Sheep with five legs. *Organization*, 19(4), 507–524. <http://doi.org/10.1177/1350508411414293>
- van den Brink, M., & Benschop, Y. (2014). Gender in academic networking: The role of gatekeepers in professorial recruitment. *Journal of Management Studies*, 51(3), 460–492.
- van Dijk, H., van Engen, M. L., & van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes*, 119(1), 38–53. <http://doi.org/10.1016/j.obhdp.2012.06.003>
- van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work Group Diversity and Group Performance: An Integrative Model and Research Agenda. *Journal of Applied Psychology*, 89(6), 1008–1022. <http://doi.org/10.1037/0021-9010.89.6.1008>
- van Raan, A. F. J. (2006). Comparison of the Hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. *Scientometrics*, 67(3), 491–502. <http://doi.org/10.1556/Scient.67.2006.3.10>
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387(6631), 341–343. <http://doi.org/10.1038/387341a0>
- Williams, H. M., & Meân, L. J. (2004). Measuring gender composition in work groups: A comparison of existing methods. *Organizational Research Methods*, 7(4), 456–474. <http://doi.org/10.1177/1094428104269175>
- Zhang, Y. (2016). Functional diversity and group creativity: The role of group longevity. *Journal of Applied Behavioral Science*, 52(1), 97–123. <http://doi.org/10.1177/0021886315591364>