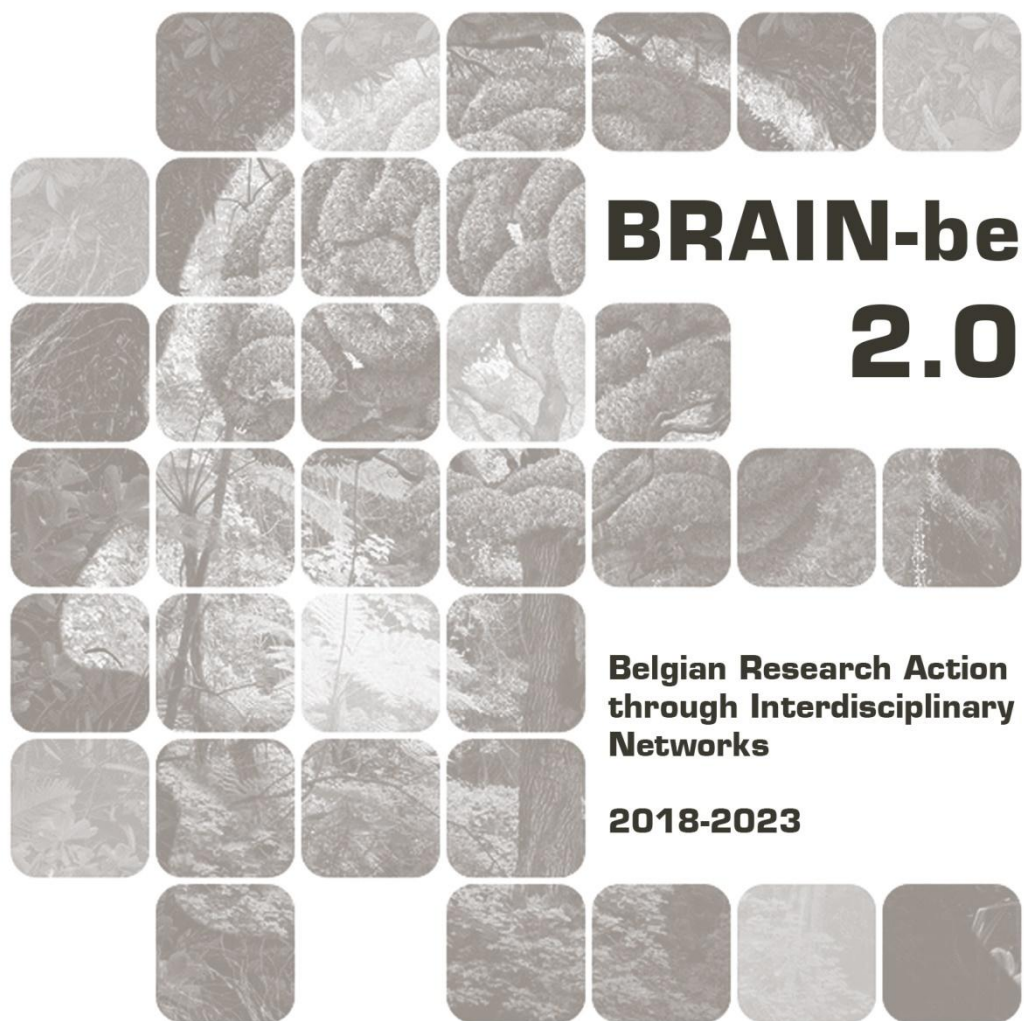


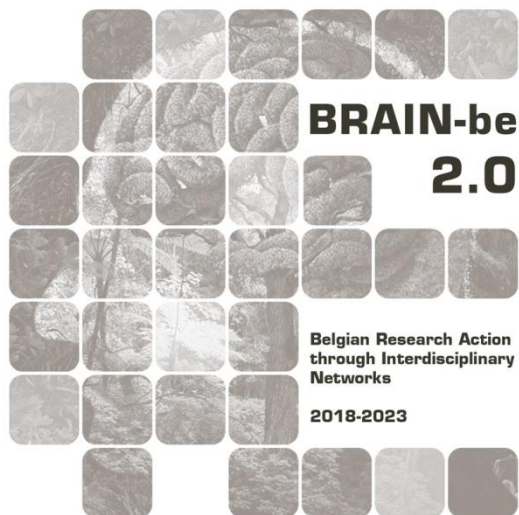


FedDiverse

Ethnic Diversity in Federal Public Services

Bert Leysen (VUB) – Eva Deros (Universiteit Gent) – Donatienne Desmette (UC Louvain)
– Pieter-Paul Verhaeghe (VUB)

Pillar 3: Federal societal challenges



NETWORK PROJECT

FedDiverse

Ethnic Diversity in Federal Public Services

Contract B2/223/P3/FedDiverse

FINAL REPORT

PROMOTORS:	VERHAEGHE PIETER-PAUL	Vrije Universiteit Brussel
	DEROUS EVA	Universiteit Gent
	DESMETTE DONATIENNE	UC Louvain

AUTHORS:	LEYSEN BERT	Vrije Universiteit Brussel
	DEROUS EVA	Universiteit Gent
	DESMETTE DONATIENNE	UC Louvain
	VERHAEGHE PIETER-PAUL	Vrije Universiteit Brussel



Published in 2025 by the Belgian Science Policy Office

WTCIII

Simon Bolivarlaan 30 bus 7

Boulevard Simon Bolivar 30 bte 7

B-1000 Brussels

Belgium

Tel: +32 (0)2 238 34 11

<http://www.belspo.be>

<http://www.belspo.be/brain-be>

Contact person: Emmanuèle Bourgeois

Tel: +32 (0)2 238 34 94

Neither the Belgian Science Policy Office nor any person acting on behalf of the Belgian Science Policy Office is responsible for the use which might be made of the following information. The authors are responsible for the content.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without indicating the reference:

Leysen, B., Derous, E., Desmette, D. Verhaeghe, P.P. FedDiverse. Federal Diversity in Federal Public Services. Final Report. Brussels: Belgian Science Policy Office 2025 – 117 p. (BRAIN-be 2.0 - Belgian Research Action through Interdisciplinary Networks)

TABLE OF CONTENTS

LIST OF TABLES	6
TABLE OF FIGURES	6
SUMMARY	8
CONTEXT	8
OBJECTIVES.....	8
CONCLUSIONS	8
KEYWORDS	8
1. INTRODUCTION	9
2. STATE OF THE ART AND OBJECTIVES	10
<i>OBJECTIVE 1: TO ANALYSE AND BENCHMARK THE CURRENT ETHNIC COMPOSITION WITHIN FEDERAL PUBLIC SERVICES.....</i>	<i>10</i>
<i>OBJECTIVE 2: TO ANALYSE ETHNIC DIFFERENCES IN MOTIVATIONS FOR, AND PERCEPTIONS OF PUBLIC SECTOR EMPLOYMENT (INFLOW).</i>	<i>11</i>
<i>OBJECTIVE 3: TO ANALYSE ACTUAL AND PERCEIVED FAIRNESS OF SELECTION PROCEDURES (INFLOW).....</i>	<i>12</i>
<i>OBJECTIVE 4: TO ANALYSE CAREER SUCCESS AND ADVANCEMENTS (THROUGHFLOW).....</i>	<i>13</i>
3. METHODOLOGY	14
3.1 LITERATURE REVIEW AND VARIABLE ASSESSMENT (WP 1.1).....	14
3.2 INTERSECTIONAL ANALYSIS OF SOCIO-ECONOMIC MONITORING AND BENCHMARKING (WP 1.2).....	15
3.3 ETHNIC DIFFERENCES IN PUBLIC SERVICE MOTIVATIONS (WP 2.1)	17
3.4 ETHNIC DIFFERENCES IN THE PERCEPTIONS OF SIGNALS IN PUBLIC SERVICE JOB ADVERTISEMENTS (WP 2.2).....	19
3.5 ACTUAL FAIRNESS OF SELECTION PROCEDURES (INFLOW) (WP 3.1)	21
3.6 PERCEIVED FAIRNESS OF SELECTION PROCEDURES (INFLOW) (WP 3.2)	22
Time 1: Immediate Post-Test Reactions	22
Time 2: Post-Feedback Reactions on Selection Procedure	24
3.7 FOCUS GROUP (WP3.1 + WP 3.2)	24
3.8 ANALYSIS OF OBJECTIVE CAREER SUCCESS AND ADVANCEMENT (THROUGHFLOW) (WP 4.1).....	28
3.9 ANALYSIS OF SUBJECTIVE CAREER SUCCESS AND ADVANCEMENT (THROUGHFLOW) (WP 4.2)	28
3.10 FOCUS GROUP (WP4.1 + WP 4.2)	31
4. SCIENTIFIC RESULTS AND RECOMMENDATIONS	32
4.1 LITERATURE REVIEW AND VARIABLE ASSESSMENT	32
4.1.1 Three approaches to measuring ethnicity	32
4.1.2 Advantages and Disadvantages	35
4.1.3 Recommendations for the Belgian Federal Public Services	40
4.2 INTERSECTIONAL ANALYSIS OF SOCIO-ECONOMIC MONITORING AND BENCHMARKING	41
4.2.1 Ethnic Origin Composition	41
4.2.2 Generation Composition	47
4.2.3 Ethnic Diversity	49
4.3 ETHNIC DIFFERENCES IN PUBLIC SERVICE MOTIVATIONS	52
4.3.1 Introduction and literature findings	52
4.3.2 Ethnic differences in public service motivations	55
4.4 ETHNIC DIFFERENCES IN THE PERCEPTIONS OF SIGNALS IN JOB ADVERTISEMENTS FOR THE PUBLIC SECTOR	59
4.4.1 Introduction and literature findings	59
4.4.2 Ethnic differences in signal perception	61
4.5 ACTUAL FAIRNESS OF SELECTION PROCEDURES (INFLOW)	65
4.5.1. Literature findings	65
4.5.2. Adverse impact calculations	67
4.6 PERCEIVED FAIRNESS OF SELECTION PROCEDURES (INFLOW)	70
4.6.1 Literature findings	70
4.6.2 Perceived fairness of selection procedures	71
4.6.3 Conclusions quantitative analyses: Perceived and actual fairness of selection procedures	80

4.6.4 Focus groups	80
4.7 ANALYSIS OF CAREER SUCCESS AND ADVANCEMENT (THROUGHFLOW).....	84
4.7.1 Literature findings	84
4.7.2 Objective career success: KPIs	87
4.7.3 Subjective career success and advancement (throughflow)	89
CAREER RESOURCES.....	92
4.7.5 Focus group	95
4.7.6 Conclusion	97
 5. DISSEMINATION AND VALORISATION	 98
 6. PUBLICATIONS	 98
 7. ACKNOWLEDGEMENTS	 98
 8. REFERENCES	 99

LIST OF TABLES

Table 1 Overview of country categorisation of requested data	16
Table 2 Overview of federal selection tests and the competences measured at the different organisational levels	26
Table 3 Dimensional overview of (dis)advantages identified in the three approaches of ethnicity	35
Table 4 Spearman's Rho correlations: PSM items x IND-COLL/ FLX-MOM scales on the Belgian sub-dataset (ethnic groups with n> 10 observations (13))	56
Table 5 Cross-Classified Multilevel Logistic Regression Models on Public Service Motivation	58
Table 6 Linear Mixed-Effects Models Predicting 1) organisational attractiveness, 2) job attractiveness, 3) job pursuit intentions	63
Table 7 Simple Effects of signals on 1) organisational attractiveness, 2) job attractiveness, 3) job pursuit intentions, by origin region	64
Table 8 Selection Rates and Adverse Impact-ratios according to the Four-Fifth Rule for federal selection tests	69
Table 9 Descriptives and Correlations of Time 1 variables	73
Table 10 Mean and Standard Deviations of Applicant Reactions towards Tests at Time 1 (N = 5,256)	75
Table 11 Descriptives and Correlations of Time 2 variables	79
Table 12 Results from focus groups, discussing findings of WP3	83
Table 13 Descriptives and zero-order correlations of study variables on the total sample (N = 1953)	90
Table 14 Descriptives and zero-order correlations of study variables on the ethnic minority sample (n = ???)	91
Table 15 Regression of SCS on Career Resources and Demographics (n = 1,953)	92
Table 16 Regression of SCS on Career Barriers and Demographics (n = 2,286)	93
Table 17 Regression of SCS on Organisational DEI climate and Demographics (n = 2,286)	94
Table 18 Regression of SCS on Ethnicity-related Individual Factors and Demographics (n = 247)	95

TABLE OF FIGURES

Figure 2 Proportions of non-Belgian origin groups per sector (2021)	42
Figure 1 Difference in proportions of origin groups in federal public sector and private sector (2021)	42
Figure 3 Gender balance breakdown of the federal public sector (2021)	43
Figure 5 Federal public sector gender balance per origin group (2021)	44
Figure 4 Gender balance breakdown per selected origin group of the federal public sector (2021)	44
Figure 6 Federal public sector gender split per age cohort for selected origin groups (2021)	45
Figure 7 Education level by origin group for the federal public sector (2017)	46
Figure 8 Breakdown of type of employment contract per origin group for the federal public sector (2021)	47
Figure 9 Changes in generational proportions across the three sector groups between 2011 and 2021	48
Figure 10 Proportions of individuals employed as civil servants in the federal public sector between 2011 and 2021 split by gender	49
Figure 11 Evolution of EDI for the subgroups of the federal public sector (2017 - 2021)	50
Figure 12 Federal public sector EDI split by gender (2021)	50
Figure 13 Overview of public sector diversity (2021)	51
Figure 14 Country positions on Minkov and Kaasa's cultural dimension scales	54
Figure 15 Absolute distributions for the questions "How important is a job that allows someone to help other people?" (left) and "How important is a job that is useful to society?" (right) (1 = very important, 2 = important, ... 5 = not important at all)	55
Figure 16 Ethnic Minority – Majority Differences in Applicant Reactions (SJT)	76
Figure 17 Ethnic Minority – Majority Differences in Applicant Reactions (File test)	76
Figure 18 Ethnic Minority – Majority Differences in Applicant Reactions (in-basket test)	76
Figure 19 Ethnic Minority – Majority Differences in Applicant Reactions (abstract reasoning)	76

Figure 20 Ethnic Minority – Majority Differences in Applicant Reactions (verbal reasoning)	76
Figure 21 Ethnic Minority – Majority Differences in Applicant Reactions (numerical reasoning)	76
Figure 22 Ethnic Minority – Majority Differences in Applicant Reactions (administrative test)	76
Figure 23 Actual and perceived fairness of test administered by FOD BOSA	80
Figure 24 Number of promotions, promotion rate and speed for Belgian vs. non-Belgian origin employees	88
Figure 25 Perceived importance of diversity interventions for career planning for ethnic minority workers (EMI) and ethnic majority workers (EMA)	97

SUMMARY

Context

Ethnic minorities are still underrepresented at higher echelons of many organisations, including the public sector. This status quo is problematic for several reasons. First, it is not only in the law but also ethical to provide fair job chances to any applicant. Unfair procedures might discourage individuals to apply for jobs, trigger complaints, initiate legal cases, and negatively affect organisations' image as well as employer branding. Second, unfairness upon organisational access or career advancement (whether actual and/or perceived) also negatively affects individuals' wellbeing (low self-esteem and commitment; depression; dropout and turnover from employment). Third, a homogeneous workforce undermines the good functioning of organisations (like organisational innovativeness), including meeting the needs of the increasingly diverse customers and stakeholders. Fourth, governments ought to set an example to others by reflecting its society and its citizens.

Objectives

The FedDiverse project identifies aims to analyse the ethnic diversity and to identify factors that promote and inhibit ethnic diversity in the federal public services during the initial screening and later career stages by investigating targeted recruitment (*applicants' public service motivations and perceptions of job advertisements*), fairness of tests (*adverse impact*) and job incumbents' career success (in both *objective* and *subjective* terms) to monitor career chances for ethnic minorities vis-à-vis majorities at the federal public services.

Conclusions

First, exploratory analyses confirm that minorities are underrepresented in the federal sector compared to the private sector, with large, organization-specific, and gendered discrepancies. Next, socio-cultural value structures associated with more conservative societies (*the origin of several major ethnic groups in Belgium*) lead to markedly higher loadings on Public Service Motivation (PSM). This finding indicates that ethnic minorities overall do not lack motivation to serve the public good. Similarly, ethnic differences in the perception of job advertisement signals are minimal with only a few minority groups preferring intrinsic job attribute signals over prosocial ones, and multicultural framings of diversity climate over colourblind ones.

Focusing on the application phase, ethnic minority applicants demonstrated significantly lower selection rates than majority applicants across all selection tests (except numeric reasoning). Crucially, this disparity exists despite both minority and majority applicants perceiving the tests as equally fair.

Post-hire, ethnic minorities experience fewer promotions overall but achieve them at a faster rate than majorities. While ethnic background is not related to subjective career success, both objective and subjective career outcomes are positively influenced by organisational factors such as perceived career support and mentoring, highlighting actionable areas for policy intervention.

Keywords

Inflow, throughflow, ethnic origin, fairness, career success, diversity, recruitment, selection, public service motivation, job advertisements, adverse impact, promotion

1. INTRODUCTION

Diversity and inclusion in public administration are complex contemporary challenges (Ding & Riccucci, 2023; Martínez Guzmán et al., 2023; McCandless et al., 2022). While diversity concerns a myriad of elements and intersecting identities, one of these which remain underrepresented in the public administration are ethnic minorities (Groeneveld & Meier 2022; Socio-Economic Monitoring 2020). Explaining this underrepresentation, however, has proven complex, with potential barriers at each stage of the attraction, selection, and attrition framework (Schneider, 1987).

Attracting skilled candidates is considered a crucial first step in arriving at a quality work force (Barber, 1998; Ployhart & Kim, 2013). Candidates from minority groups may not apply to public sector jobs to the same degree, limiting the inflow of applicants. This can be due to different motivations, or values that don't match with what a position in the labour market should entail. Not only the sector, but how the job is described may hinder attraction, such as particular job advertisement signals. A common avenue of research probes the selection-phase, with a wide range of literature testing labour market discrimination (Lippens et al., 2023; Neumark, 2016) although it is unclear to what extent this is a factor in the public sector (Leysen et al., 2025). Other inequalities too may pose barriers, such as particular standardised tests and procedures which are especially prevalent in the public sector. Finally, once hired, retention can be a pitfall for organisations in their quest for a diversified working force. Bias, both actual as well as perceived, may hinder ethnic minority candidates further in their career path.

This project examined several of these potential barriers captured by the following objectives:

- Objective 1: We reviewed the strategies to define and measure ethnic origin and analysed and benchmark the current ethnic composition within federal public services
- Objective 2: We examined ethnic differences in motivations for and perceptions of public sector employment (*inflow*)
- Objective 3: We examined the actual and perceived fairness of selection procedures (*inflow*)
- Objective 4: We examined the career success and advancement (*throughflow*)

For each FedDiverse objective, the main data, measures, and methods used to reach our findings are described in the methodology section. The presentation of the scientific results, conclusions and related recommendations constitutes the main part of the report.

2. STATE OF THE ART AND OBJECTIVES

Ethnic minorities are still underrepresented at higher echelons of many organisations, including the public sector (Groeneveld & Meier, 2022; UNIA & FOD WASO, 2022). This co-funded FedDiverse project aimed to identify factors that promote and inhibit ethnic diversity in the federal public services during the initial screening stages as well as later career stages. It consisted of four objectives, covering the current ethnic composition (1), the inflow (2 & 3) and throughflow (4) of ethnic minorities. These objectives are briefly described below. Afterwards, their methodology and main results are discussed in section 3 and 4 respectively.

Objective 1: To analyse and benchmark the current ethnic composition within federal public services.

Ethnicity or ethnic origin is central throughout this project. It is therefore crucial to reflect on how to conceptualise, define, and *measure* this central concept. **Objective 1.1** was to review and assess the strategies to define and measure ethnic origin, with a focus on assessing the diversity within federal public services. This entails a myriad of ways, ranging from self-identification to relying on administrative data capturing proxies such as nationality, country of birth, parentage, and combinations of these factors (Jacobs et al., 2009; Van Caeneghem, 2019). Measuring ethnic origin by way of *survey questions* contains several pitfalls. Simplistic, singular questions often fail to capture dynamic and multidimensional concepts of which ethnic origin is one; researchers or policy makers conducting surveys may have different understandings of ethnic origin compared to the individuals completing these surveys; and last, there is the question of the effectiveness of these measures (Burton et al., 2010). Research by Woolverton and Marks (2023), for instance, illustrates how the congruency between open-ended questions and forced choice checkbox questions was significantly higher among white adolescents compared to adolescents of colour (38.8% as opposed to 22.5%).

Official *administrative data* have their problems too. The country of birth of the individual that is ‘measured’ or in combination with that of their parents is an often opted-for proxy for ethnic origin. Questions regarding temporality can be raised particularly when categorising migrant groups: How many generations should ethnicity ‘stick’ before these individuals are categorised as belonging to the destination country? Even the measurement of generations in this context can have an impact on how people are categorised. Research by Chimienti et al. (2021) illustrates how different ways of measuring generation can drastically impact group proportions. Applied to the Swiss LIVES-FORS cohort survey, the authors found a 19-percentage point difference in the proportion of second-generation respondents in their sample depending on the operationalisation. Administrative data by and large preclude any form of agency in determining one’s ethnic origin. As such, self-identification, an often-touted requirement for ethnic origin measurement (for instance, see the *EU High Level Group on Non-discrimination, Equality and Diversity’s* guidance note in 2021) is missing entirely. This last aspect is identified by literature to be a crucial one in the formation of ethnic groups (Jenkins, 2008). Further complexities are the conceptual distinction of ‘ethnic’ and ‘national’ minorities, as well as the question of which ethnic (ethno-national) categories one must distinguish (e.g. specific groups or broader categories). For instance, China recognises over 50 ethnic groups within its borders, yet these Chinese nationals would be recorded as “Chinese” in most administrative data systems. Each country has its specific demographic composition and challenges. Consequently, conceptual frameworks should be tailored to their specific context.

In addition to clarifying conceptual outlines, mapping the status quo is of equal importance. **Objective 1.2** was to analyse the ethnic composition of the current stock of employees in the federal

public sector and benchmark it with the employee stock of other public entities and the private sector. Understanding the status of ethnic diversity in the federal public services starts with understanding the ethnic composition of the current stock of employees, its evolution, and how it compares to the employee stock of other public entities and the private sector. Past research in the form of the Socio-Economic Monitoring (2017) has performed similar analyses until 2014 (2020 in online attachments) resulting in aggregate tables of ethnic composition, subdivided by gender, contract type, or job/educational level. However, more detailed and up-to-date intersectional analyses, including benchmarking, are as of yet lacking and could inform policymakers as to the underrepresentation of specific groups and intersections. A crucial factor pertains to the specificity of the groups involved and subdomains of the public services. By design, the Socio-Economic Monitoring uses broad categories such as “EU-14”, “Maghrebi”, or “Sub-Sahara Africa” which subsume specific countries with a particular importance to Belgium due to its colonial and migration history (e.g., Congolese). Disentangling these groups from their broader categories was a key objective for this benchmarking.

Objective 2: To analyse ethnic differences in motivations for, and perceptions of public sector employment (*inflow*).

Numerous factors influence job seekers' decision to apply for a specific job. Drawing on self-determination theory, scholars typically differentiate between intrinsic and extrinsic work *motivations*. In the case of the public sector, research has emphasized the importance of *public service motivation* (PSM), entailing prosocial motivations to contribute to the common good through public organisation employment (Breugh et al., 2018; Perry & Wise, 1990). PSM may be correlated with the perceived attractiveness of the public sector as an employer (Asseburg & Homberg, 2020; Vandenabeele & Van de Walle, 2008).

Objective 2.1 examined ethnic differences in PSM. Despite previous research showing gender and age-related differences in public service motivation (DeHart-Davis et al., 2006; Pandey & Stazyk, 2008), there was but limited research on how ethnic origin may affect PSM. This is remarkable given that PSM appear to be affected by cultural and religious belongings. The level and nature of PSM differ, e.g. between the cultural communities in Switzerland (Ritz & Brewer, 2013), between countries with their historical-culturally grown administrative traditions (Asseburg & Homberg, 2020; Vandenabeele & Van de Walle, 2008) and according to religious practices (Hassan & Ahmad, 2021; Perry, 2000) (Perry 2000; Hassan & Ahmad 2021). Our analysis was structured around the hypothesis that varying levels of Public Service Motivation (PSM) across ethnic groups drove applicant self-selection, which, in turn, contributed to the underrepresentation of certain groups in public sector employment. The rationale behind this is that the historical-cultural traditions of a migrant's country of origin, in which they are socialized, play a role in shaping values and subsequently their PSM in their destination country. By examining the contextual origin of non-Western migrants, we contributed, to some extent, to the scientific literature by investigating PSM in non-Western countries, closing an often-mentioned scientific gap (Asseburg & Homberg, 2020).

Objective 2.2 involved analysing how ethnic minorities react to signals in job advertisements for the public sector to get a better employer branding. While PSM may shape work motivations affecting job searching, job seekers are subsequently confronted with *job advertisements* – a key medium by which potential applicants are attracted to openings (Griffeth et al., 2014). Key is the ‘person-job’ and ‘person-organisation’ fit between applicants and the public sector (Cable & Edwards, 2004; B. E. Wright & Pandey, 2011). Job ads allow applicants to evaluate potential fit by interpreting signals about the organisation and the position (Celani & Singh, 2011). Three types of signals are

examined. First, intrinsic and extrinsic job attributes are stronger predictors for intentions to apply for the public sector than prosocial job attributes (Asseburg et al., 2020; Keppeler & Papenfuß, 2020; Weske et al., 2020). Job seekers with high PSM react, however, differently to prosocial attributes (Asseburg et al., 2020). Most studies do not focus on ethnic differences in these perceptions (for an exception: Linos 2018, about the police). Second, signals about the level of formalization in recruitment - typical elements of public sector recruitment and selection (Chen & Rainey, 2014) - have negative consequences on intentions to apply for the public sector (Sievert et al., 2020). Again, little is known about the disparate impact for ethnic minorities. Finally, diversity climate perceptions play a role too (Avery et al., 2013). Ethnic minorities might hypothesize that they do not fit into an organisation because of 'belonging uncertainty' (Linos, 2018; Walton & Cohen, 2007). Explicit information about diversity management has been shown to support organisational attractiveness (Olsen & Martins, 2016). Examining and clarifying how ethnic origin affects PSM, and how job seekers - in particular ethnic minorities - react to different signals in job advertisements for the public sector served as the basis for concrete recommendations regarding adjustments to employer branding in advertisements and attracting a wider range of talent.

Objective 3: To analyse actual and perceived fairness of selection procedures (*inflow*).

During hiring it is common practice to administer psychological tests/assessments to evaluate applicants' talents and predict future job performances. Good test practice includes that both psychometric and social criteria are met (Derous, 2020; Evers et al., 2010). Objective 3 focuses on social criteria.

Objective 3.1 investigated actual fairness of commonly used tests and assessments at the federal public services as well as potential actions to remedy unfairness (if being observed). Psychological tests and assessments are typically developed by and for ethnic majority populations (Outtz, 2010). Therefore, these tests/assessments may be guided by biases, leading to adverse impact (AI), being a differential and unjustified dropout of underrepresented groups (like ethnic minorities) from selection procedures. First, a systematic literature review provided the latest state-of-the-art on the AI of the tests/assessments employed for the federal public services and actions to mitigate AI. This was followed by adverse impact analyses on tests/assessments used for the federal public services. Finally, intersectional effects of ethnicity with gender, age and job level were investigated. These have not been considered systematically (Derous & Pepermans, 2019) but provide a more fine-grained and realistic perspective on ethnic subgroup differences and follows international recommendations (Employer Information Report; US EEOC 2019; Else-Quest & Hyde, 2016).

Objective 3.2 focused on ethnic minority/majority applicants' perceived fairness of commonly used tests and assessments. Perceived fairness concerns justice beliefs on procedural aspects (i.e., interpersonal treatment, etc.) and distributive aspects (i.e., test outcomes) (McCarthy et al., 2017). Perceived test discrimination, then, is a specific facet that refers to applicants' perception of an unequal treatment in terms of procedural and/or distributive justice that substantially differs between ethnic minorities and the majority (Anderson, 2011). We examined (external) applicants' perceptions of procedural and distributive fairness and whether fairness perceptions differ between ethnic groups. Although the project's focus is on ethnicity, other factors, such as age, gender and job level, were considered as well from an intersectional angle. Findings regarding actual/perceived test fairness and potential actions to mitigate (perceptions of) unfairness were discussed within a focus group of relevant stakeholders.

Objective 4: To analyse career success and advancements (*throughflow*).

The barriers that ethnic minorities face in employment are, however, not limited to the recruitment and selection phase. Career success and advancement of ethnic minority employees (organisational throughflow) have been considered to an even lesser extent than actual/perceived fairness of hiring procedures (inflow). Career success is defined as the ‘positive psychological, work-related outcome or achievement one accumulates as a result of work experience’ (Seibert et al., 1999, p. 417) and encompasses ‘objective’ and ‘subjective’ career success (Abele & Spurk, 2009). Objective career success is typically operationalized by occupational titles, promotion and income growth (Apers et al., 2019). Subjective career success refers to self-referenced experiences of success, operationalized by career satisfaction measures. Addressing calls for diverse and inclusive workforces, ethnic minorities’ career advancement is studied here (Leong, 2014).

Objective 4.1 monitors objective career success by examining ethnic minorities vs. majorities’ promotion chances. First and similar to 3.1, adverse impact analyses were conducted on tests/assessments among job incumbents (i.e., internal candidates) for promotion goals. Most studies overlooked the potentially differential selection rates for ethnic minority/majority candidates for promotions (Anderson, 2011). Additionally, we investigated career progression within federal services by not only considering ethnic diversity ratios (KPI or Key Performance Indicator on the number of ethnic majorities/minorities overall and/or at different job levels or units) but by also developing new KPIs to evaluate both rate and speed with which ethnic minorities vs. majorities progress in their career (making formal promotions).

Objective 4.2 analysed subjective career success by considering perceived personal/organisational barriers/facilitators. To this end, ethnic minority and majority job incumbents were surveyed on their subjective career success, including estimated importance and perceived achievement of relevant aspects on the facets ‘learning and development’, ‘work-life balance’, ‘positive impact’, ‘positive work relationships’, ‘financial security and wealth’ (Briscoe et al., 2021), perceived job-demands-resources (like workload and autonomy, Bakker & Demerouti, 2017), and diversity climate (i.e., intention to support diversity; Hiemstra et al., 2013; Villotti et al., 2019). Second, focus groups were held with ethnic minority and majority job incumbents to discuss survey findings and identify potential positive actions to support career advancements of minority employees, like role socialisation and mentoring programs (Alanis & Ryan, 2022) and the role of allies (Gardner & Alanis, 2020; Gardner & Ryan, 2020).

3. METHODOLOGY

3.1 Literature review and variable assessment (WP 1.1)

Objective 1.1 is to review the strategies to define and measure ethnic origin, with a focus on assessing the diversity within federal public services. Project leader of WP1.1 was Prof. Verhaeghe, and the scientific researcher was Bert Leysen. This work package was approved by the Ethics Committee for the Human Sciences Vrije Universiteit Brussel with reference no. ECHW_433.

Design. First, a scoping review was conducted. Scoping reviews are preliminary assessments of the existing literature, typically conducted in advance of a more systematised review (Grant & Booth, 2009; Petticrew & Roberts, 2006). Scoping reviews aid researchers in finetuning search parameters such as keywords, exclusion criteria, and optimise search approaches. We noticed a substantial difference between Web of Science (WoS) and Scopus, and, considering them complementary rather than substitutive, opted to use both for the review. The review was further complemented with Google Scholar. The scoping review resulted in a conceptual review protocol that outlined the search parameters, key terms, scope, and potential risks and countermeasures. This was finalised after consulting the other academic partners, and a search matrix was constructed. This search matrix was combined in all possible ways to come up with one search string that could be used for each database. For instance, the search string would include both “ethnic* + group + measur*” as well as “ethnic* + measur*” with key terms in all possible orders. Conducting separate searches for each combination would result in an even larger number of duplicates, so we opted for the single search query approach. We chose to limit our search to literature published no earlier than the year 2000 due to the extensive volume of available publications. Considering the objectives of the project, next a conceptual review/synthesis method was decided upon for the actual review. This is a type of literature review that aims to provide the researcher with an understanding of the literature in a given field, including key ideas and concepts (e.g., ethnic origin), models and debates, and primarily focus on synthesizing areas of conceptual knowledge.

Analysis. After running the complete search matrix in the databases, we ended up with 9,629 references. Since multiple databases were used, a substantial number were expected to be duplicates. We used Rayyan, an advanced systematic literature review tool, to prune the dataset. The reference lists were extracted from the search databases and uploaded to Rayyan. First, clear duplicates were removed. All sources that matched 95% or below, were manually checked for duplicates. In total 2,633 sources were deleted this way, netting 6,996 remaining sources.

Next, the search results were evaluated based on relevance in three rounds. First, we reviewed the titles of the 6,996 sources. Due to the search parameters, a substantial number of hits result from the field of biology that deals with classification in nature. While we could have foreseen this, initial testing during which we limited the search parameters to a set number of research fields had mixed results. For instance, if papers were not tagged properly to belong to the field of social psychology, they would not have been included in the source list. This necessitated a larger ‘cleanup’ after extracting the sources.

The topic of ethnicity and related concepts has garnered such attention, that it is virtually impossible to map the complete field that uses it as a variable. For instance, Ratuva’s compilation on the concept itself, “The Palgrave Handbook of Ethnicity” (2019), spans over 2,000 pages, offering a substantial resource on the subject while simultaneously illustrating the expansive and ongoing nature of the debate on the concept. Since the primary interest here is how ethnicity can be measured,

defined, and conceptualised, the next step evaluates the abstracts of the remaining reference to curate a list for full text reading. This final body of literature of 314 sources was subsequently processed and the findings structurally summarized.

Finally, this range of academic literature was supplemented by non-peer-reviewed ‘grey’ literature such as websites and codebooks of statistical agencies, government websites, and documentation by research institutes resulting in a total list of 345 documents.

3.2 Intersectional analysis of Socio-Economic Monitoring and benchmarking (WP 1.2)

Objective 1.2 is to analyse the ethnic composition of the current stock of employees in the federal public sector according to the findings in WP 1.1 and benchmark it with the employee stock of other public entities and the private sector. Intersectional analyses and benchmarking rely on official administrative data (cf. *registering ethnicity approach* – see section 4.1.1). We therefore spend some attention to the variables in the below sections. Project leader of WP 1.2 was Prof. Verhaeghe, and the scientific researcher was Bert Leysen. This work package was approved by the Ethics Committee for the Human Sciences Vrije Universiteit Brussel with reference no. ECHW_433.

Data. The data was made accessible by the Crossroads Bank for Social Security, combining sources from several administrative databases. After a lengthy data request and approval process, the scientific researcher gained *in-situ* access to the requested population registers for the years 2011-2021. Aside from self-explanatory variables, such as “age”, others deserve a note of explanation. First, “employment contract type” has four possible categories: workers, employees, civil servants, and others (*arbeiders, bedienden, ambtenaren, anderen*). Next, “gender” is captured at the administrative level. This categorisation is consequently limited to either *men* or *women*. “Education level”, is based on the ISCED framework and can range from ISCED 0 (early childhood education) to level 8 (doctoral degree or equivalent tertiary education level). This variable was not available for every year requested, so analyses reflect this limitation.¹ To identify the sector of employment, the variable “CODNAC” is used, reflecting the industry classification according to the NACE-BEL 2008 coding system. To further differentiate between public sector employees, the variable “pouorg” is used, denoting the “establishing power of employer in public sector”. Codes ranging from 11 through 19 are grouped under the federal public sector, while all other non-missing values in this variable are classified as the other public sector. These classifications encompass all employees, not limited to those in statutory positions.

As to the topic at hand, two variables requiring further explanation as well as acknowledgement. To examine “origin” and “migration background”, we made use of two algorithms graciously provided by the FPS Employment, Labour and Social Dialogue (FOD WASO/ SPF ETCS) and have been used similarly in the Socio-economic Monitoring. The variable “origin” enables a more precise analysis by mapping the second generation and individuals who have become Belgian based on the nationality at birth of either the individual or their parents. Furthermore, the variable “migration background” allows us to distinguish between more recent and older migrants, migrants from various generations, and individuals who have acquired Belgian nationality. This variable and its underlying algorithm are also managed by the aforementioned FPS, which has kindly approved its usage for this project. No further modifications were made to these algorithms.

¹ Despite receiving approval for data up until 2021, the “study domain” variable was only available for 2011 in the in-situ data. Therefore, we opted to exclude these data from analyses in favor of more up-to-date ones.

The variable “origin” is a combination of the variables “current nationality,” “first nationality,” and “first nationality of the parents.” This allocation is based on:

1. The first nationality of the father, followed by the first nationality of the mother: if both parents have different nationalities at birth (Belgian or foreign), the foreign nationality is chosen. If both parents have different foreign nationalities at birth, the father's nationality is selected.
2. The nationality of the individual at birth.
3. The current nationality of the individual.

Although the variables are analogous to those employed by the FPS they originate from, the results differ in their categorisation and level of aggregation. Table 1 illustrates the categorisation employed in this report. The findings of WP 1.1 convinced us of the need – when feasibly – to differentiate between origin groups at a more detailed level rather than just broader regions. Since WP 1.2 deals with register data, the number of cases was deemed large enough to opt for this approach. However, while country-level data is ideal as it would be the most nuanced information available, GDPR considerations limited this approach. Categories that are too fine-grained run the risk of small-cell identification (Williams & Husk, 2013) so we found a compromise to isolate several of the main historically relevant immigrant origin countries, while opting for broader regions for others. The below table lists the final categorisation employed in this WP.

Table 1 Overview of country categorisation of requested data

Netherlands
Romania
Poland
Morocco
Bulgaria
Italy
Spain
France
Russia
Turkey
Portugal
China
India
Democratic Republic of the Congo
Other EU 14 (excluding Netherlands, Spain, France, Portugal)
Other EU 13 (excluding Romania, Bulgaria)
Other EU candidates (excluding Turkey)
Other Europeans (including U.K.)
Other Maghreb (excluding Morocco)
Near/Middle East (Iran, Israel, Jordan, Iraq, Syria, Lebanon, Palestine, Saudi Arabia, Yemen, Oman, United Arab Emirates, Qatar, Bahrain, Kuwait, Egypt, Pakistan, and Afghanistan)
Sub-Saharan Africa (excluding DR Congo)
North America (including Mexico)
Central America (including the Caribbean)
South America (everything south of Panama)
East and Southeast Asia (Japan, South Korea, Taiwan, Vietnam, Laos, Cambodia, Thailand and Myanmar, Malaysia, Indonesia, the Philippines, and the Andaman and Nicobar Islands)
Oceania (including Australia, excluding Indonesia)
Other Asia
Unspecified

The “origin” variable does not strictly capture ethnicity or ethnic origin. Currently, Belgium does not officially collect statistics on ethnic origin, least of all at the individual level and in a comprehensive

manner. The variable “migration background” is a combination of the previous variables along with “country of birth,” “date of registration in the National Register,” and “date of nationality acquisition” (categorised as: more than 5 years ago or 5 years ago or less). Simply summarized, from the third generation, individuals are seen as having “Belgian origin”. The different generations are calculated as follows:

1. Generation three: Belgian nationals born as Belgians, to parents that have the Belgian nationality and are born as Belgians.
2. Generation two: Belgian nationals born as Belgians, to at least one non-Belgian parent or if the origin of the parent(s) is unknown.
3. Generation one (naturalized): Naturalized not born as Belgians.
4. Generation one (not naturalized): Non-Belgian nationals.

Analyses. Measuring *ethnic composition* with these data is relatively straightforward due to their nature. Inferential statistical methods are typically used for estimations about a population based on smaller sample data. In other words, inferential statistical methods are used to generalize findings from a smaller sample to the population. Because this study uses population data, however, there is no probabilistic sample drawn from a population. The exact population parameters are already known, rendering inferential methods unnecessary (cf., discussion by Gibbs and colleagues (2015, 2017) and Figueiredo Filho and colleagues (2013)).

Whereas *ethnic composition* typically refers to the proportion of a population belonging to a particular ethnic origin (e.g., a non-European origin or specific ethnic group), *ethnic diversity* also takes into account the number of different groups. An intuitive way to express ethnic diversity is by means of the Hirschman-Herfindahl Index (HHI) (Hirschman, 1964). The HHI is calculated as follows:

$$HHI = \sum_{i=1}^N (ES_i)^2$$

where N is the number of ethnic groups within a given entity, ES_i is the ethnic share or proportion of group i. A higher HHI typically means a higher concentration of a particular ethnic group, and a lower HHI means the opposite. The maximum value this index can take is 1, designating an ethnic monopoly. The minimum value is $1/N$. Following this logic, more diversity means a lower number. Because this is counterintuitive – especially when interpreting figures – we create an Ethnic Diversity Index (EDI) by taking the complement of the HHI:

$$EDI = 1 - \sum_{i=1}^N (ES_i)^2$$

In contrast to HHI, a higher EDI signifies more ethnic diversity. The EDI is subsequently calculated for the different sectors and intersections, elucidating ethnic diversity in a variety of different settings.

3.3 Ethnic differences in Public Service Motivations (WP 2.1)

Objective 2.1 examines ethnic differences in PSM. Project leader of WP2.1 was Prof. Verhaeghe, and the scientific researcher was Bert Leysen. This work package was approved by the Ethics Committee for the Human Sciences Vrije Universiteit Brussel with reference no. ECHW_433.

Data. Analysing ethnic differences in public service motivations (PSM) and their association with historical-cultural background, necessitates the use of several datasets. First, the most recent wave (2015) of the Work Orientations module of the International Social Survey Programme (ISSP) was

used. The ISSP is an ongoing international collaboration that conducts frequent surveys on key social science topics. Data were collected between 2015 and 2017 by probability sampling and cover 37 countries. Data collection was done via telephone and face-to-face interviews as well as via self-administered questionnaires resulting in 51,668 individual cases. The Work Orientations module includes information on attitudes and motivations related to work life, working conditions, organisation, and more (ISSP Research Group, 2017).

The systematic literature review revealed two ISSP questions that have been used in previous research to capture individual-level PSM loading (Houston, 2014; S. Kim, 2017; Vandenabeele & Van de Walle, 2008; Westover & Taylor, 2010):

- Q2f “How important is a job that allows someone to help other people?”
- Q2g “How important is a job that is useful to society?”

The answers to these two questions are heavily skewed towards “important” and “very important”, so these variables were recoded into binary variables either taking the value “very important” or “not very important” in line with research by Houston (2014).

Next to demographic variables such as gender and age, this wave contains important background variables that are used to construct ethnic origin. In line with the operationalisations from WP1.1, we use the respondent’s mother’s and father’s country of birth (M_BORN, F_BORN). These were used to construct a new variable (‘parentage’) capturing an individual’s parental composition (one, two, or no foreign-born parents). In addition, a second variable (‘ethnic origin’) was constructed based on the country of origin of the parents. If both parents are born in the survey country, the respondent’s origin is labelled as such. If either parent was born in a country different from the survey country, the non-survey country was assigned to the respondent’s origin. If both parents were born outside the survey country, the father’s origin was used.

Next, we use a dataset quantifying cultural values to model the historical-cultural traditions of an immigrant’s country of origin. While initially focusing on the World Value Survey (WVS) dataset, the literature review for this work package made us evaluate other frameworks as well. Empirical research has illustrated similarities between the key dimensions of these divergent frameworks. After a thorough comparison of different typologies, we decided to work with for the two-dimensional model as proposed by Minkov and Kaasa (2022). This model is rooted in the influential framework of Hofstede, but expands on it in several important ways. First, they employ the two dimensions that are validated between by comparing cultural frameworks across different samples. Rather than focusing on the six traditional dimensions by Hofstede, the authors reduce the number of dimensions to two, in following with previous empirical cross-cultural research findings. By coopting data from the WVS, Minkov and Kaasa expand the sample to 102 countries and assign scores per country on these dimensions. Next, they address the potential issue of ‘subjective culture’. In other words, large-scale surveys examining national cultural values are typically captured by individual self-reporting. This self-reporting does not necessarily correspond with what the researchers call ‘objective culture’ or measures of societal practices at the level of nation-states, such as gender equality, political freedom, homicide rates, family structures, innovation rates, educational efforts and achievements and more. In other words, Minkov and Kaasa (2021) question to what extent these individual datapoints were reflected in country level characteristics (or vice-versa) in essence bridging two levels of analysis. As the current work package aims to use these scores at the group level, such considerations are important. They empirically demonstrate a high correlation between these two dimensions and a range of external, country-level variables indicators (such as GDP, Gini coefficient, etc.), allowing for more parsimonious statistical

modelling (i.e., *we can exclude these additional macro-level variables from the model*) and making it a suitable system for the present study.

The model of Minkov and Kaasa (2021) centres around two factors with two dimensions each. The *individualism-collectivism* dimensions (IDV-COLL) addresses the degree to which people are integrated into groups, ranging from individualist (IDV) to collectivist (COLL) societies with a stronger sense of cohesion and loyalty. The second *flexibility-monumentalism* dimensions (FLX-MON), refines similar concepts such as long-term orientation (Hofstede, 2011). Cultures scoring high on flexibility (FLX) are more open to change and are considered more adaptable. Those scoring high on monumentalism (MON) on the other hand, value stability and tradition. As indicated by Fog's conceptualization of the F1 and F2 dimensions, these dimensions – or their equivalent - are identified as recurrent in most cross-cultural research (Fog, 2021, 2023).

Based on factor analysis of wave 6 (2010-2014) and wave 7 (2017-2022) of the WVS – congruent with the ISSP dataset - Minkov and Kaasa calculate a positive/negative numerical indicator per dimension per country depending on their scoring on the two dimensions IDV-COLL and FLX-MON. These indicator numbers are linked to the countries of origin of the individual respondents in the ISSP data as well as with the countries of origin of the individual's father and mother, serving as an indicator for the quantification of cultural value structures.

Analyses. Since the observations measuring PSM are clustered at the country level thus violating core regression assumptions, and since the key objective is to model whether culture affects PSM loading of individuals, multilevel modelling is an appropriate method. Level 1 units of analysis are the individuals with PSM outcomes and control variables such as gender. Level 2 units are nation-states as well as ethnic groups with of primary interest the scores on the two dimensions of cultural values IDV-COLL and FLX-MON in addition to a range of control variables. Considering the outcome variables capturing PSM dimensions are recoded into indicator variables, multilevel logistic regression analyses are used to estimate differences in PSM. However, since there is no “pure” hierarchical nesting of ethnic groups in countries (i.e., *an ethnic group exists not solely in one country*) this assumption is violated, prompting the need to use cross-classified multilevel modelling (Heck & Thomas, 2020; Hox & Roberts, 2011). The linear mixed effects models are fitted with the “lme4” package in R.

3.4 Ethnic differences in the perceptions of signals in public service job advertisements (WP 2.2)

Objective 2.2 examines how ethnic minorities react to signals in job advertisements for the public sector. Recent research has illustrated how wording in job applications can have differentiating effects on different groups (Abraham et al., 2024). Project leader of WP2.1 was Prof. Verhaeghe, and the scientific researcher was Bert Leysen. This work package was approved by the Ethics Committee for the Human Sciences Vrije Universiteit Brussel with reference no. ECHW_433.

Design. This work package utilizes multifactorial survey experiments, also known as vignette studies. Vignette studies are a combination of traditional surveys and experimental methods (Mutz, 2011). According to Sniderman and Grob (1996) they are designed to address some of the primary shortcomings of both: experiments typically have high internal validity although there are challenges with regards to external validity. Surveys inverse these points, with typically high external validity but lower internal validity due to multicollinearity (Atzmüller & Steiner, 2010). Multifactorial survey experiments consist of a series of vignettes that are shown to a respondent. Vignettes are short but realistic descriptions of situations, people, or objects that systematically includes a combination of signals or ‘factors’ the researcher is interested in. After being confronted with this vignette, the respondent is asked to what extent they agree with a series of statements. Respondents typically

process multiple vignettes, thus allowing researchers to simultaneously test several factors as well as vary in contextual elements (Auspurg & Hinz, 2014).

Here, the vignettes are used to examine ethnic differences in the perceptions of signals in job advertisements for the federal public sector. The factorial survey experiment features a 3x2x2 design, meaning there are three ‘factors’ or signals that are varied between the different vignettes, with the first factor featuring three levels, and factor two and three each two levels. The first factor signals different job attributes which can be intrinsic, extrinsic, or prosocial. The second factor deals with standardised recruitment procedures. Such procedures are either mentioned or not in the job ad. Finally, the third factor contains signals on the diversity climate of the organisation. Here, a diversity statement either signals a multicultural environment, or a colourblind one. In total, there are twelve (3x2x2) potential combinations of signals. Based on the literature review examining research on these signals, sentences were constructed to serve as signals in the job ad text (Chen & Rainey, 2014; Holmes et al., 2021; Lee & Choi, 2016; Leslie et al., 2020; McKay et al., 2007, 2008; Ritz & Waldner, 2011; Sievert et al., 2022; Van de Walle et al., 2015; Vogel et al., 2024).

Data and sample. The outcome variables of interest pertain *organisational attractiveness*, *job attractiveness*, and *job pursuit intentions*. In other words, by varying the signals across the different vignettes, the experiment tries to establish whether any of the signals have an effect on how attracted the respondent is to the organisation after reading the job ad, how attractive the job seems, and whether or not they would apply for this job.

The conducted literature review informs the operationalisation of the outcome variables. The outcomes are measured by means of three items, resulting in a three-by-three question and answer matrix for each vignette (three items for three outcomes). Organisational attractiveness feature items from Highhouse et al., 2003; Olsen & Martins, 2016; and Vandenabeele & Jager, 2020, job attractiveness items from Koçak et al., 2022 and Van Hooft et al., 2006, and job pursuit intentions from Aiman-Smith et al., 2001; Highhouse et al., 2003; and Sievert et al., 2022.

In addition to the three-by-three matrix, a series of socio-demographic questions were asked at the end of the experiment as not to prime respondents’ perceptions of the vignettes in any way. These feature several basic items such as gender and age but also questions on the respondent’s ethno-national origin and that of their parents. Finally, we also include several items to measure public service motivation as a potential moderator for being attracted to public sector jobs, as research has shown how motivations for certain careers can differ between ethnic and racial groups (Benditt, 2015; Schuck, 2021). Since the signals are not visually indicated in the vignettes to maintain realism as much as possible, we included several attention checks throughout the survey to ensure respondents were reading sentences carefully. Failing to pass these checks resulted in immediate termination of the experiment.

The vignettes are based on a random selection of existing job advertisements for the Belgian federal government. Two series of job advertisements were selected based on function level since positions for higher educated candidates may not appeal to those with different educational credentials and vice versa. We split the experiment in a series of vignettes for those with a higher education degree, and those without. The full description of the position, regulations, and related information of the job advertisement often resulted in several pages of text. Since vignettes are typically shorter, we condensed the job advertisements to contain only the core information. This resulted in a structured format containing a brief paragraph describing the job and organisation, followed by the main tasks for this position in bullet point format, as well as a closing sentence with a call to apply.

The signals were first pretested in a separate experiment among last-year students at the VUB to ensure the respondents could correctly identify them. In case less than 80% of the respondents could identify the signals, the statement was removed from the list of usable statements for the experiment. In total, eight statements were removed, of which two particularly problematic (24% and 39% positive identification). The finalised vignettes were tested in a probability sample of the Belgian population ($n = 3461$)². After excluding incomplete surveys and cases that failed quality control questions or had unrealistic completion times, the final sample consisted of 756 participants. Since each participant evaluated and rated a random subset of five vignettes³ each, a total of 3780 observations were collected.

Analysis. We conducted three separate multilevel linear regressions, one for each of the three outcome variables. In each model, the dependent variable was the composite score from averaging the three questions for that outcome. The models included the predictor variables as fixed effects and a random intercept to account for the nested structure of the data. Main effects are modelled, followed by interactions between the signals and the different origin clusters as well as interactions with public service motivation scores. In line with Hayes and Little (2018), simple effects analyses are used to probe the interactions found. Values for missing control variables (9 observations for sector preference) are simulated by means of Multiple Imputation via the R package “mice”. The linear mixed effects models are fitted with the “lme4” package, and estimated marginal means for the comparison between origin groups are calculated with the “emmeans” package.

3.5 Actual fairness of selection procedures (inflow) (WP 3.1)

Objective 3.1 investigates actual fairness of commonly used tests and assessments at the federal public services as well as potential actions to mitigate unfairness (if being observed). It consists of three steps. Project leader of WP 3.1 was Prof. Deros and the scientific researcher was dr. Aylin Koçak. This work package was approved by the Ethics Committee of the Faculty of Psychology and Educational Science at Ghent University with reference no. 2023-32.

Design and sample. First and in close collaboration with representatives from FOD/SPF BOSA, an overview was made of the commonly used selection tests/assessments to assess applicants who apply for a job at the federal public services. Table 2 shows the different tests administered by BOSA. Second, once a decision was made on type of selection tests/assessment, a literature review was conducted to evaluate the state-of-the-art on adverse impact of the type of tests/assessments used for the federal public services. Third and finally, the adverse impact of these tests/assessments was measured. To this end, secondary data (period 2020-2022) were used, focusing on several commonly administered federal public service tests. The tests examined include the: Situational Judgment Test (SJT) ($N = 16.855$), File test (*dossierbehandeling*) ($N = 18.160$), In-basket test (*postbakoeffening*) ($N = 933$), Abstract reasoning ($N = 10.841$), Numerical reasoning ($N = 125^*$), Verbal reasoning tests ($N = 1.816$), and Administrative skills and planning ($N = 4.097$).

² Due to disciplinary variations across the project, the notation for sample size changes between sections. In Work Packages 1 and 2, N denotes the population and n denotes the sample. In Work Packages 3 and 4, N represents the complete sample size, while n is used for any subgroup or subsample.

³ Randomization of preset combinations of vignettes using D-efficient fractions (Atzmüller & Steiner, 2010; Auspurg et al., 2014).

Data and analysis. For each test, key demographic data were extracted: ethnic background, gender, and age, alongside the selection outcome (*pass/fail*). Participants' ethnic origin was measured using 11 categories based on the European Standard Classification of Cultural and Ethnic Groups, including: (a) Belgian; (b) European (outside Belgium, EU27); (c) European (outside Belgium, non-EU27); (d) North Africa, Middle East and Central Asia; (e) Sub-Saharan Africa; (f) South and Southeast Asia; (g) East Asia; (h) Latin American; (i) Caribbean; (j) North American and Australian; and (k) other. For the purpose of this analysis, these were dichotomized into two broad groups: Belgian origin (a) versus non-Belgian/ethnic minority origin (b through k). Adverse impact was assessed both statistically and practically. Statistical tests include the z-test, Chi-squared test, and Fisher's exact probability test to evaluate significant differences in selection rates. Adverse impact was also determined using the "80% rule," which stipulates that the selection rate for the minority group must be at least 80% of the selection rate for the majority group. This adverse impact analysis was performed for ethnic minority versus majority applicants overall and subsequently repeated within these two groups to examine internal disparities based on gender (men vs. women) and age (applicants aged below or above the age of 50y).

3.6 Perceived fairness of selection procedures (inflow) (WP 3.2)

Objective 3.2 investigates ethnic minority/majority applicants' perceived fairness of commonly used tests and assessments at the federal public services (FPS). To reach this goal a three-step procedure was followed. Project leader of WP3.2 was Prof. Deros and scientific researcher was dr. Aylin Koçak. This work package was approved by the Ethics Committee of the Faculty of Psychology and Educational Science at Ghent University with reference no. 2023-32.

Design and data. First, a literature review was conducted regarding the state-of-the-art of fairness of tests administered by BOSA and whose actual fairness was too being evaluated. Second, a survey was conducted among ethnic minority/majority candidates on experienced procedural justice after they participated in a selection hurdle (**Time 1**). Applicants were also invited to participate in a voluntary way after they received feedback on test performance(s) in order to measure perceived distributive justice (**Time 2**).

Third, results of **objectives 3.1 and 3.2** were combined for a more in-depth evaluation of overall test fairness/discrimination and potential actions to increase (perceptions of) fairness in a focus group of stakeholders. We evaluated each test on its objective/subjective fairness, which is typically not done and innovative. Specifically, tests being evaluated on actual/perceived fairness, may belong to one of the following groups (Anderson 2011): (1) no actual/no perceived fairness, (2) no actual/perceived fairness, (3) actual/no perceived fairness, and (4) both actual and perceived fairness. Hence, how a risk analysis of tests regarding (actual/perceived) fairness can be designed and lead future strategic decisions in organisations.

To reach Objective 3.2, primary survey data for examining *perceived fairness* of the selection procedures were gathered between February 2024 and June 2024. A two-wave (T1 and T2) design was used, gathering data on demographic factors (ethnic background, gender, age) and capturing various applicant reactions. Findings were analysed per test: Situational Judgment Test, File test, In-basket test, Abstract reasoning test, Numerical reasoning test, Verbal reasoning test, Administrative test (at Time 1, immediate post-test) and about the selection procedure as a whole (at Time 2, post-feedback).

Time 1: Immediate Post-Test Reactions

The first measurement point (T1) concerns an in-situ survey administered immediately after applicants completed a PC-based test on the *Werkenvoor.be* platform. This occurred either after a single test or the last of two consecutive tests. The T1 sample included 5,941 test takers of which 5,256 completed the survey, with a mean age of 34.33 years ($SD = 9.76$). The individuals in the sample were predominantly of Belgian origin (76.60%), female (64.8%), and largely composed of external candidates (69%).

Multivariate Analysis of Variance (MANOVA) was used to examine the influence of demographic variables, as well as other individual difference variables (such as self-efficacy, reaction sensitivity), on a set of seven key outcome variables: *procedural justice*, *general fairness*, *perceived discrimination*, *test anxiety*, *test motivation*, *test ease*, and *test attitude at Time 1 (T1)*. All items were measured on a five-point Likert-type scale (with 1 = *strongly disagree*; 5 = *strongly agree*). First, **procedural justice** was measured with 9 items adapted from Bauer et al. (2001). An example item is: 'This test allows me to show my capabilities'. Next, **general test fairness** was measured with 1 item, adapted from Gilliland (1993) and Thorsteinson and Ryan (1997), namely 'Whether I passed or not, I thought this was a fair test'. **Perceived test discrimination** was measured in terms of ethnicity⁴, age and gender, with one item each, adapted from Ghazaleh Abu (2012), e.g., 'I felt disadvantaged in this test because of my [ethnic origin]', with 1 = strongly disagree until 5 = strongly agree. Finally, the following one-item measures were also administered: **test anxiety** ('I felt comfortable taking this test [R]'), adapted from Taylor & Deane (2002), **test motivation** ('I was very motivated to do well on this test'), adapted from Arvey et al. (1990), **attitude toward the test** ('I enjoy participating in this kind of test'), adapted from Tonidandel et al. (2002) and **test ease** ('I found this test difficult [R]'), adapted from Arvey et al. (1990).

We also measured to what extent participants *perceived* that they performed well ('I think I performed well') and to what extent they expected they had passed the test ('I think I have passed the test'), each both a single item with 1 = *strongly disagree* until 5 = *strongly agree*. **Sensitivity to being rejected** in a selection context was measured with a one-item scale adapted from Ghazaleh Abu (2012), i.e., 'I mind not to be selected for the job during a selection process'). Participants' **self-efficacy** in relation to test-taking was measured with a single item adapted from Truxillo et al. (2001), i.e., 'I think I do well in this kind of test'. Further, demographics were measured too. Participants' **chronological age** was measured with the open-ended question 'What is your age (in years)?'. We asked participants 'What is your gender' and provided the following categories: 'Man', 'Woman', 'I'd rather not say', and 'other'. We measured **participants' ethnic origin**, using the following categories, based on the European Standard Classification of Cultural and Ethnic Groups: (a) Belgian; (b) European (outside Belgium) and member of the EU27; (c) European (outside Belgium) and NOT a member of the EU27; (d) North Africa, Middle East and Central Asia; (e) Sub-Saharan Africa; (f) South and Southeast Asia; (g) East Asia; (h) Latin American; (i) Caribbean; (j) North American and Australian; (k) other. Examples of each category were provided between brackets for easy recognition. To answer the research question of this work package, we grouped the abovementioned categories in two broad categories, namely *Belgian origin: (a) versus non-Belgian origin, or ethnic minority origin: (b) until (k)*. Finally, we registered **actual test performance**, i.e., whether participants passed the individual test or not (i.e., based on the minimum criteria of that specific selection) in an objective way, namely through the testing platform used by *Werkenvoor.be*.

⁴ Because the main focus of this project is on ethnicity, we report and discuss the results for ethnicity but also consider interactions with gender and age (for sake of ecological validity).

Time 2: Post-Feedback Reactions on Selection Procedure

The second measurement point (Time 2, T2) consisted of a follow-up survey (via Qualtrics) for participants who had provided consent for this at T1. This survey was deployed *after* applicants had received their formal feedback on (part of) the selection procedure at Time 1, gauging their perception on how they were evaluated, more in general (and not towards a specific test). The T2 sample included 376 participants who completed the survey with a mean age of 36.82 years ($SD = 10.17$). This smaller group showed slightly different demographics, being 78.7% Belgian origin, 64.9% female, and 66.8% external candidates.

Moderated mediation analyses were used to examine the relationships between variables. Specifically, we tested the indirect relationship between *perceived performance* and *procedural/distributive justice* as potentially mediated by attribution styles, with the ethnic background, gender and age of the applicant serving as a potential moderating variable. Like Time 1 analyses, non-Belgian groups (categories b-k, cf. *supra*) were aggregated into a single ethnic minority category to facilitate comparison.

At **Time 2 (T2)**, we asked participants to indicate their **satisfaction with their (perceived) performances** on the tests with a single item, i.e., ‘How satisfied you with your test performance’ (i.e., ‘perceived performance’, 1 = not at all; 5 = very much) as adapted from Oostrom and De Soete (2016). Secondly, **attribution styles** (McAuley et al., 1992; Russell, 1982) were measured with 9 items. Three items measured participants’ locus of causality, three measured stability and three measured personal control as regards their test performance/result. For each item, participants indicated on a 9-point Likert-type bipolar scale what they attributed their result to the most. For example, ‘the test result is caused by... with 1 = an aspect of the situation until 5 = an aspect of myself, for an item measuring locus of control (external vs. internal). Third, **applicant reactions on procedural justice** were measured again at Time 2, but this time in relation to the **test procedure/process**, instead of just 1 test. For instance, *procedural justice* was measured with 9 items adapted from Bauer et al. (2001), similar to the Survey T1, but with a slightly adapted formulation (e.g., ‘This test procedure allowed me to show my capabilities’). In a similar vein, **distributive justice** was administered with a single item adapted from Gilliland (1993) and Thorsteinson and Ryan (1997), i.e., ‘I found the outcome of this test procedure fair’, with 1 = strongly disagree until 5 = strongly agree. Next, **perceived discrimination** during the test procedure was measured in terms of ethnicity, age and gender, with one item each, adapted from Ghazaleh Abu (2012), e.g., ‘I felt disadvantaged in this test procedure because of my [ethnic origin/age/gender]. Finally, the following one-item measure was also administered: Attitude toward the test procedure (‘Up until now, I found participating in this test procedure a positive experience’ in this kind of test’), adapted from Macan et al. (1994). Finally, the same **demographics** as at Time 1 were measured at Time 2 (cf. *supra*).

3.7 Focus group (WP3.1 + WP 3.2)

In order to contextualize the expected findings, a focus group was organised in July 2024 to examine applicants’ perceptions of the selection procedures and study results. Participants ($N = 8$) were recruited via earlier survey participation in the project and the sponsors’ network within the federal government. In addition to the leading researcher and an assistant observing the discussion, the group had the following stakeholder composition: two job applicants (ethnic minority and majority), one job incumbent (ethnic minority), one selection professional, a job coach and three diversity policymakers (all ethnic majority). Participants could participate in either Dutch or French. First, focus group members

introduced themselves and their role as participants. Second, the leading researcher presented the project's findings. Third, participants discussed the results and some guiding questions (one example being: 'Is there anything you would change about the selection instruments?'). Two independent researchers analysed the data using thematic analysis. That is, the focus group was content-coded independently by two trained industrial and organisational psychologists ($M_{age} = 28.5y$ $SD_{age} = 0.71y$). Theme structures were compared, and differences were elaborately discussed among raters until agreement was reached (McDonald et al., 2019).

Table 2 Overview of federal selection tests and the competences measured at the different organisational levels

Test	Summary of test for each organisational level	Measured constructs/competences for each organisational level
Situational Judgment Test		
Level A	Candidates are presented with a situation (text) relevant to working within the federal government with additional context and out of different possible reactions, they need to indicate the extent to which these reactions are appropriate	Key competences: teamwork, service-orientation, reliability, self-development, achieving objectives
Level B		
Level C		
Level D	Candidates are presented with a situation (video) relevant to working within the federal government with additional context and out of different possible reactions, they need to indicate the extent to which these reactions appropriate	
File Test		
Level A	Candidates (a) process 4 files and (b) subsequently answer questions related to the files.	Information integration and decision making
Level B		Information analysis and problem solving
Level C		
Level D	///	///
In-basket Test		
Level A	Candidates process a mailbox containing messages (short memos, letters, full reports, etc.) within a given deadline and indicate where priorities lie and which follow-up should be planned.	Information integration and decision making
Level B		Information analysis and problem solving
Level C		Information analysis and problem solving
Level D	Candidates process a mailbox containing short messages and answers related questions.	Handling information and structuring work
Abstract Reasoning		
Level A	Candidates are presented with a series of figures that follow underlying rules and indicate which of the new figures is next in line.	Abstract reasoning (i.e., deriving rules from a set of abstract data and apply them to new information)
Level B	Candidates compare three rows of two abstract figures (e.g., squares or circles) to determine which transformation is applied from left to right, which is indicated by third figure (e.g., a dice) in between.	
Level C	Candidates are presented with a series of figures that follow underlying rules and indicate which of the new figures is next in line.	
Level D	Candidates are presented with 4 figures and need to indicate (part 1) which of 4 new figures has characteristics in common with the first set of figures or (part 2) which of 4 new figures can complement the row of figures.	

(continued)

Numerical Reasoning		
Level A	Candidates use facts and numbers in statistical tables in order to answer questions.	Numerical reasoning (i.e., the accuracy and speed of making connections between pieces of numerical information) and mathematical competences.
Level B	Candidates use facts and numbers in statistical tables in order to answer questions.	
Level C	Candidates are (part 1) presented with incorrect multiplications (consisting of letters) and indicate the number of multiplications that are incorrect and are (part 2) presented with a table with an incorrect number and indicate what the correct number should be.	Reviewing numerical information.
Level D	Candidate are presented with a table that contains information on inventory and answer questions regarding the table that require fast and accurate calculations.	Handling numerical information.
Verbal Reasoning		
Level A	Candidates are presented with a series of texts and answer statements in terms of true/false/not inferable based on the information in the texts.	Verbal reasoning (i.e., the accuracy and speed of making connections between pieces of verbal information).
Level B		
Level C	Candidates are presented with a series of short texts and answer a multiple-choice question based on the texts.	
Level D		
Administrative Test		
Level A	N/A	N/A
Level B	Candidates (part 1) organise data and answer related questions and (part 2) review two columns of data to indicate the number of errors in the columns.	Ability to review and organise administrative data.

3.8 Analysis of objective career success and advancement (throughflow) (WP 4.1)

Objective 4.1 monitors objective career success by investigating ethnic minorities vs. majorities' promotion chances from different and new angles (diversity KPIs). Project leader of WP4.1 was Prof. Derous and the scientific researcher were dr. Aylin Koçak and dr. Philippe Sterkens. This work package was approved by the Ethics Committee of the Faculty of Psychology and Educational Science at Ghent University with reference no. 2023-32.

Design and data. The investigation into **objective career success** during the *throughflow* stage (WP 4.1) focuses on comparing promotion opportunities for ethnic minority employees against the majority group within the federal government. Ethnic minority and majorities' career progression can be monitored by means of KPIs (Key Performance Indicators; part of HR analytics). However, specific KPIs for doing so, are still limited. We aimed developing three KPIs to measure ethnic minorities' career progression within the federal government. Three key performance indicators (*KPIs*) of career progression that were developed and monitored are: **(a) the Number of formal upward promotions** (i.e., *promotions to higher organisational or job levels*); **(b) the Promotion rate** (*the ratio of promotions to organisational tenure, expressed in months*); and **(c) the Promotion speed** (*the number of months until the next promotion*). This innovative objective was realised through secondary data analysis. Specifically, the analysis used secondary administrative data from Persopoint, encompassing 43,172 observations across 29 federal organisations, collected from the past 40 years.⁵ This data included 10,857 distinct individuals, with a mean age of 38.03 years ($SD = 12.67$). The sample consists of 53% women and 94% of the Belgian ethnic majority.

Analysis. Two statistical approaches were used. First, Poisson regressions were conducted for the count variables: the *number of promotions* and *promotion speed* (i.e., counted in months between promotions). Next, a univariate Analysis of Variance (ANOVA) was used to analyse the promotion rate. All statistical models systematically controlled for the confounding effects of age, gender, and the specific organisation to ensure robust comparisons between ethnic groups.

3.9 Analysis of subjective career success and advancement (throughflow) (WP 4.2)

Objective 4.2 is to evaluate ethnic minorities' subjective career success. Literature suggests that effects of perceived test fairness might even be larger for internal than for external candidates given the impact on internal candidates' job performance and job satisfaction. Surprisingly, however, till today only a limited number of studies considered internal candidates' ethnic background and fairness perceptions of promotion decisions. Therefore, a *concise literature review* and a *survey* were conducted among ethnic minority and majority job incumbents who work at the federal government to evaluate their subjective career success and relevant contextual factors (like diversity climate). Project leader of WP4.2 was Prof. Derous and the scientific researcher was dr. Aylin Koçak. This work package was approved by the Ethics Committee of the Faculty of Psychology and Educational Science at Ghent University with reference no. 2023-32.

⁵ Data on employees' careers was collected through the Persopoint career administration system which is only recently being used by several federal organisations, each with their own (recent) onboarding date. While this administration system offers a thorough and complete registration of one's career steps, this information is limited to the onboarding date of the organisation within the system. This career data is merged with earlier career data that was registered for each organisation separately. However, the way in which this earlier data was registered differs between organisations. Hence, the merging and cleaning of this career data was a very time consuming and complex task. Future initiatives might benefit from a standardised approach in which the career data are collected for different federal organisations.

The analysis of *subjective career success* during the throughflow stage (**WP 4.2**) investigates employee perceptions - both among ethnic minority and majority groups. We examine factors across four domains that potentially influence an individual's subjective career outcome: (a) *career resources*, (b) *career barriers*, (c) *organisational Diversity, Equity, and Inclusion (DEI) climate*, and (d) *ethnicity-related individual characteristics* (i.e., acculturation, meta-stereotyped beliefs, perceived career discrimination, experienced ethnic allyship, and core self-efficacy.). The first three factors are subject to organisational influence and control, while the fourth reflects personal attributes. In addition to assessing subjective success, we also explored the relationship between these four sets of factors and the employees' objective career success, as measured by the number of promotions received.

Design and data. First, **subjective career success** was measured with 19 items of the DAIA-SCC (Briscoe et al., 2021) on a five-point Likert-type scale (i.e., 'Thinking about my career success, I am ... with this career aspect; with 1 = *not at all satisfied* to 5 = *very satisfied*). We measured the following six career aspects: *Learning & Development* (e.g., 'doing work that gives the opportunity to learn'), *Work-Life Balance* (e.g., 'having time for non-work interests'), *Positive Impact* (e.g., 'helping others'), *Positive Work Relationships* (e.g., 'experiencing positive relations with others'), *Financial Security* (e.g., 'having financial security'), and *Financial Success* (e.g., 'achieving wealth'). Cronbach's alpha was .90.

Second, **career resources** encompass four key career resource dimensions, namely (1) human capital factors (i.e., occupational expertise, human capital development), (2) environmental career resources (i.e., organisational career support, social career support, mentoring), (3) motivational career resources (i.e., career clarity/goals), and (4) career management behaviours (i.e., career preparation). All items were measured on a five-point Likert-type scale (1 = *strongly disagree* until 5 = *strongly agree*). *Human capital* was measured with six items referring to occupational expertise (3 items adapted from Hirschi et al., 2018) and job-related human capital development (3 items adapted from Kiazad et al., 2020). Cronbach's alpha was .80. *Environmental career resources* were measured with 11 items on organisational career support (3 items adapted from Hirschi et al., 2018), social career support (4 items adapted from Hirshi et al., 2018), and perceived mentoring (4 items, adapted from Rechlin, 2016). Cronbach's alpha was .90. Motivational career resources were measured with three items referring to clarity of career goals (Hirshi et al., 2018). Cronbach's alpha was .91. *Career management behaviours* (like following workshops) were measured with three items adapted from Hirshi et al. (2018). Cronbach's alpha was .71.

Third, **ethnic/age/gender career barriers** were measured with a three-item scale each, adapted from Luzzo and McWhirter (2001) (e.g., 'In my future career I will probably be treated differently because of my ethnic background'), rated on a five-point Likert-type scale with 1 = *strongly disagree* until 5 = *strongly agree*. Cronbach's alphas were .93 (ethnic career barriers), .87 (age career barriers) and .93 (gender career barriers).

Fourth, **organisational diversity, equity and inclusion (DEI) climate** was measured with 15 items, referring to employees' perceptions of (a) organisational diversity, (b) organisational inclusion, (c) organisational equity of practices, and (d) supervisors' support for diversity and inclusion. All items were measured on a five-point Likert-type scale (1 = *strongly disagree* until 5 = *strongly agree*). *Organisational diversity* was measured with four items adapted from Hofhuis et al. (2016), e.g., 'In my organisation there is room to work according to one's own culture'. Cronbach's alpha was .83. We also measured *organisational inclusion* with five items adapted from Nelissen et al. (2017), e.g., 'In my organisation we are attentive to the needs of people of foreign origin'. Cronbach's alpha was .86. Further, perceived *equity of organisational practices* was measured with 4 items adapted from Nishii (2013)'s inclusion dimension 'Foundation of equitable employment practices', e.g., 'The

employment/HR practices of my organisation are fairly implemented'. Cronbach's alpha was .90. *Supervisor diversity and inclusion* was measured through 3 items adapted from Ward et al. (2022), e.g., 'I feel that my immediate supervisor does a good job of managing people with diverse backgrounds'. Cronbach's alpha was .89. Taken together, overall reliability was .90.

Further, **ethnicity-related individual factors** included items on (a) acculturation, (b) meta-stereotyped beliefs, (c) perceived career discrimination, (d) experienced ethnic allyship, and (e) core self-efficacy. Unless otherwise stated, all variables were again measured on a five-point Likert-type scale (1 = *strongly disagree* until 5 = *strongly agree*). Acculturation was measured with 16 items as adapted from Beirens and Fontaine (2011), e.g., 'I listen to music of my ethnic group.' The acculturation measure was only administered among ethnic minority participants and was included in the model as two outcomes, namely cultural adaptation (Cronbach's alpha was .83) and cultural maintenance (Cronbach's alpha was .87). Cronbach's alpha of all 16 items was .78. *Meta-stereotyped beliefs* related to one's ethnic origin were measured with 4 items adapted from Ryan et al. (2015), e.g., 'When I interact with people of Belgian origin, I feel that they interpret my behaviours in terms of the fact that I have a foreign origin'. Cronbach's alpha was .76. *Perceived career discrimination* was measured with 1 item adapted from Ghazaleh Abu (2012), i.e., 'During my career, I have felt disadvantaged in career opportunities because of my ethnic background'. *Experienced allyship* was measured from the perspective of the participant with an ethnic minority background with two items based on Collier-Spruel and Ryan (2022) and Kutlaca and Radke (2023). Items referred to either allyship coming from people with a different (i.e., 'I have people of Belgian origin who act as allies for me') or the same ethnic origin (i.e., 'I have people of my origin who act as allies for me'). (No coefficient alpha calculated given two items). *Core self-efficacy* was measured with 12 items adapted from Judge et al. (2003), e.g., 'I am confident I get the success I deserve in life'. Cronbach's alpha was .82. Finally, *ethnic identification*, or the extent to which one identified with their origin, was measured based on Phinney and Ong (2007) with eight items, e.g., 'I identify with people of my own ethnic group'. Cronbach's alpha was .905.

Finally, and at the end of the survey, participants were asked about possible **interventions to increase inclusion and decrease discrimination in careers based on ethnicity**. First, participants were presented two types of practices, and we asked for both of them whether they believed that these are important in decreasing career disadvantages based on ethnicity: i.e., 'inclusive practices/initiatives organised by the organisation' versus 'Inclusive practices/initiatives organised by individual employees', with 1 = *not at all important* to 5 = *very important*. Second, participants were presented with 9 possible interventions that are reported in the literature on diverse career planning (Leimon et al., 2011) and were asked for each one how important they are in helping those who are disadvantaged in their career based on their ethnicity, e.g., 'networking initiatives', with 1 = *not at all important* to 5 = *very important*. Answers to these questions were used in the focus group on WP4 (see further) and are useful for policy recommendations.

Sample. A total of 3,680 federal employees participated in the survey (October 2023- February 2024), although they did not complete every part of it (see further). Participants had a mean age of 44.17 years ($SD = 11.01$) and an average organisational tenure of 15.52 years. The individuals in the sample were predominantly from Belgian origin (88,90%) consisted of 58.4% women and 88.60% Belgian employees. About half of participants (52.7%) were Dutch-speaking, and the other half were French-speaking (47.3%). In terms of gender, 58.4% identified as female, 40.7% as male, 0.4% as 'other', and 0.5% chose not to disclose their identified gender. Participants had an average organisational tenure of 15.52 years within the federal government and 5.35 years outside of the federal government.

They were employed in a total of 44 different federal organisations⁶, mostly from: SPF Finances/FOD Financiën (34.3%); Office national de l'emploi/Rijksdienst voor Arbeidsvoorziening (8.7%); SPF Emploi, Travail et Concertation sociale/FOD Werkgelegenheid, Arbeid en Sociaal Overleg (6%); Institut national d'assurances maladie-invalidité/Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (5.4%); Institut national d'assurances sociales pour travailleurs indépendants/Rijksinstituut voor de Sociale Verzekeringen der Zelfstandigen (4.5%); SPF Intérieur/FOD Binnenlandse Zaken (3.9%); SPF Santé publique, Sécurité de la Chaîne alimentaire et Environnement/FOD Volksgezondheid, Veiligheid van de Voedselketen en Leefmilieu (3.7%); SPF Justice/FOD Justitie (3%). About half of the sample (54.1%) worked at level A (master's degree required) within the federal government, 30.5% at level B (bachelors' degree required), 12.8% at level C (high school degree required) and 2.5% at level D (no required degree).

Analyses. A series of regression analyses (GLM or Generalize Linear Modelling) were conducted to investigate whether subjective career success (interval-level variables) depended on (a) career resources ($n = 1,953$), (b) career barriers ($n = 2,286$), (c) organisational DEI ($n = 1,953$), and (d) ethnicity-related individual characteristics (like job level, level of acculturation, perceived allyship, perceived discrimination, ethnic identification) ($n = 247$). Similarly, a series of Poisson regression analyses were conducted to regress objective career success (count variables) on these factors ($n = 19$).

3.10 Focus group (WP4.1 + WP 4.2)

In order to contextualize the expected findings, a focus group was organised in July 2024 to examine applicants' perceptions of the selection procedures. Participants ($n = 6$) were recruited via earlier survey participation in the project and the sponsors' network within the federal government. In addition to the leading researcher and an assistant observing the discussion, the group had the following stakeholder composition: one job incumbent (ethnic minority), one selection professional, a job coach and three diversity policymakers (all ethnic majority). Participants could participate in either Dutch or French. Two independent researchers analysed the data using thematic analysis. That is, the focus group was content-coded independently by two trained industrial and organisational psychologists ($M_{age} = 28.5y$ $SD_{age} = 0.71y$). First, focus group members introduced themselves and their role as participants. Second, the leading researcher presented the project's findings. Third, participants discussed the results and some guiding questions (one example being: 'What are your expectations regarding the career progression responses of ethnic minority employees?'). Theme structures were compared, and differences were elaborately discussed among raters until agreement was reached (McDonald et al., 2019).

⁶ Considering effects for different FOD/SPFs was not the goal of this study. As agreed upon, we present the results for the total federal employees' sample, only.

4. SCIENTIFIC RESULTS AND RECOMMENDATIONS

4.1 Literature review and variable assessment

Ethnic origin can be ‘measured’ in several different ways. This study aims to structure disparate **measurement methods**, evaluate their respective **advantages** and **limitations**, and formulate practical recommendations for future research and practice. Before clarifying the different measurement methods, our first question is whether these different ways of measuring matter in the outcome they generate. Studies in the field of Public Health reveal substantial differences between different conceptualizations. Yao et al. (Yao, Bullen, et al., 2022; Yao, Meissel, et al., 2022) distinguish two common methods to classify polyethnic individuals into single categories by means of a survey. On the one hand, there is administrative-prioritisation which uses a predetermined hierarchy. Self-prioritisation, on the other hand, allows individuals to indicate a main ethnic group they identify with. These different ways of measuring result in over 50% difference in demographic proportions of the same group. Similar research found over 60% discrepancy, in particular among individuals with multiple ethnic identification (Yao et al., 2021). Lachowsky et al. (2020) too see different sample sizes by investigating public health surveys and calculating samples either by administrative-prioritisation, single/combination (similar to administrative-prioritisation but allows for multiple ethnicity combinations), and total response (for similar research, see also Boven et al., 2020; Shiao, 2019). Differences in outcomes can also be observed. Research in New Zealand examined the effect of five ethnic classification methods on adolescent mental health outcomes and found significant differences within ethnic groups both in specific health outcomes and effect sizes of these outcomes (Yao, Bullen, et al., 2022), potentially affecting policymakers’ perspective and decisions on public health policies. The way ethnic origin is measured thus has a significant impact on the resulting proportions or associated outcomes.

One manner of classifying these ways of measuring is via the distinction between assigned and self-declared ethnic categories depending on whether the individual has the option to select their ethnicity or not (Karakas & Özbilgin, 2019). Similar juxtapositions can be found in the work of Chandra and Wilkinson (2008) and their distinction between ethnic structure (dealing with the distribution in a population) and ethnic practice (the activation of ethnic categories in different contexts). Additionally, research by Schneider and Heath (2020) identified this conceptual disparity and assigned the labels of origin on the one hand, and identity on the other, alluding to the above mentioned dimensions of ancestry and descent, and culture respectively.

Focusing on the different advantages and disadvantages, these various ways of measuring ethnicity can be categorised into three approaches. While these approaches are not necessarily explicitly identified as such in existing literature, they aptly encompass several aspects, perspectives, and foci that characterize respective research and address gaps. They also complement each other. Consequently, there is no best way of measuring ethnicity. The context – both in terms of research questions and topics, and practical considerations – ought to inform the method of measuring.

4.1.1 Three approaches to measuring ethnicity

Within the first approach of **registering ethnicity**, data are situated primarily methods that rely on **proxies** such as ancestry, citizenship, and nationality obtained on the basis of (a combination of) **administrative data**. To that extent, it may be more accurate to label this as the measurement of ethno-national origin rather than the broader term ethnicity. Conceptually, this approach fixes ethnicity externally to the individual and projects it back in time as it often takes into account the birth nationality

of the individual, the nationality of the parents and grandparents, birth countries, and migration background. The algorithm applied to this determination is, to some extent, a political decision, such as whether a third-generation individual is still considered non-native and thus part of an ethnic minority.

In practical terms, categories are frequently based on (historical) **nations** or geographic regions, although this depends on how administrative data is recorded initially. For instance, some countries also record 'race' or 'ethnic group' which can consequently be used as such in research using this data. Typically, countries of origin are enumerated, and when relevant, further organised into larger, contextually significant units. States generally record both citizens and non-citizens within the framework of such nation-states, with a primary focus on the country of birth rather than narrower regions or affiliations with specific ethnic/linguistic groups. Consequently, categories based on countries are the most prevalent. In cases where groups of particular origins are notably small due to the migration history of a country, these groups may be amalgamated.

These broader categories are not necessarily stable throughout time. Especially with regards to Europe, increasingly regional categories are used, such as 'EU-member states'. This instability makes statistical comparison and analysis more cumbersome. Nation-states, although not everlasting, have proven to be more resilient than regional alliances (in addition to being more detailed) and are thus preferred over broader, regional groups.

Registry data is not always available, nor is it necessarily the most appropriate way of capturing ethnicity in all situations. A second way of measuring ethnicity is by asking direct questions such as "*To which ethnic group do you belong?*" or "*What is your ethnicity?*". The medium of these questions varies from large-scale surveys (e.g., population surveys or censuses) to in-person questions for targeted surveys or interviews. Central in this method is the operationalisation of ethnic origin via ethnic identification by the respondent or a third party and is commonly employed in English-speaking countries. We label this approach **identifying ethnic origin**.

Of importance is how answers to such questions are registered. **Open questions** allow respondents to answer with the concepts or words they prefer, regardless of whether these words are commonly seen as an ethnic group or not. Aspinall (2012) has been an advocate of this method, arguing how otherwise "census ethnicity data collection is reductionist" (p. 361). In addition, the wide range of answers given, although potentially lacking consistency, may reflect the superdiverse nature of contemporary societies more accurately than 'essentialist' categorisations.

A second option are **closed-question options**. The respondent chooses from a list of options or writes in their answer in case this list does not suit their preference. These two ways can be combined in one survey as well. Moreover, the survey can force individuals and households to limit their answers to just one category or allow multiple selections, resembling what is called "mutually exclusive methods and non-mutually exclusive methods" often employed in healthcare practice (Callister et al., 2007; Yao, Bullen, et al., 2022; Yao et al., 2021).

The individual registering the answers to such questions can differ as well. In many cases surveys are completed by the subject (i.e., *the person that is being questioned*), some data may be recorded by a third party such as an external surveyor. For instance, in research on U.S. healthcare systems, surveys on health including a race/ethnicity question are often completed by hospital staff. Also, a census can be (and often is) on a household level. This means that potentially just one member of the household fills out the survey for all members, potentially misidentifying some members.

A third group of measuring devices is inspired by (social)-psychological research (e.g., the *Social Identity Theory of Intergroup Behaviour* by Tajfel and Turner (2004 (1986)) or the *Ego Identity Model* by

Erik Erikson (1968)). This strand of research aims to measure and examine several specific dimensions of ethnicity, such as the strength of ethnic behaviour, motivations, belonging, day-to-day contexts, and more and is commonly captured by the term *ethnic/racial identity* (ERI). Compared to the previous measurement devices, the focus here is primarily on **ethnic identity** dimensions rather than ethnic origin groups or categories.

Several scales and measurement tools were developed to measure specific dimensions of ERI (Yip et al., 2019). Some focus more on ethnicity, such as the *Multigroup Ethnic Identity Measure* (MEIM) and its revised version (MEIM-R), the *Ethnocultural Identity Behaviour Index* (EIBI), the *Ethnic Identity Scale* (EIS), and the *Cross Ethnic-Racial Identity Scale-Adult* (CERIS-A). While there are more scales that only mention ‘race’ as a point of study – for instance, the *Cross Racial Identity Scale* (CRIS, see Vandiver et al., 2012) which formed the foundations for CERIS-A – we limit the scope to those dealing with ethnicity. Central here is the measurement of individual and group dimensions such as importance and salience of one’s ethnic identity, self-categorising and labelling, and commitment and attachment.

Developed by Phinney (1992) and updated by Phinney & Ong (2007) to address specific measurement issues, *MEIM* and *MEIM-R* have been used and tested widely (e.g., Brown et al., 2014; Maehler et al., 2019; Yoon, 2011). The original version (*MEIM*) featured twelve items representing two factors: *exploration* and *commitment* but was tested via a different statistical technique to understand the relationship between the items (i.e., exploratory factor analysis instead of confirmatory factor analysis for the revised version). The revised version (*MEIM-R*) forewent several behavioural items (e.g., ‘*being active in an ethnic organisation*’ and ‘*participating in cultural practices*’), adjusted the tense of the wording in some questions in addition to modifying the battery of test items to reflect the two subscales of exploration and commitment better. Behavioural items were removed as they are also considered an aspect of acculturation and thus separate from ethnic identity (Phinney, 1992; Phinney & Ong, 2007).

Other, less frequently used scales also appear in the literature. *EIBI*, developed by Yamada and colleagues (1998), considers items dealing with cultural activities, social interaction, and language opportunities but has seen little use (e.g., see Rotter et al., 2016). The same can be said for the *EIS* developed by Umaña-Taylor and colleagues (2004). This scale assesses three domains of identity formation (exploration, resolution, and affirmation) through a 46-item questionnaire (e.g., Douglass & Umaña-Taylor, 2015; Worrell & Gardner-Kitt, 2006; Yoon, 2011). A more recent scale, *CERIS-A*, was developed by Worrell and colleagues (2019) and measures seven “ethnic-racial identity attitudes” (assimilation, miseducation, self-hatred, anti-dominant, ethnocentricity, multiculturalist inclusive, and ethnic-racial salience) via 41 items. Internal consistency (i.e., whether different questions measuring one attitude consistently elicit similar responses or not) was retested by the same authors in 2021. They also confirmed the survey’s construct validity, allowing other authors to make further use of the scale as well (e.g., Day-Vines et al., 2022).

4.1.2 Advantages and Disadvantages

Table 3 Dimensional overview of (dis)advantages identified in the three approaches of ethnicity

Dimensions	Approaches		
	<i>I. registering ethnicity</i>	<i>II. identifying ethnic origin*</i>	<i>III. assessing ethnic identity*</i>
	1. <i>Stability/ Fluidity</i>	+ stable categorisations	- fluid categorisations + changes can be studied
	2. <i>Objectivity/ Subjectivity</i>	+ more objective	- fluid identifications + changes can be studied
	3. <i>Practical Considerations</i>	- less objective	- less objective + subjectivity can be studied
	4. <i>Agency</i>	+ time- and cost effective + valid repeat measurements	- surveying more time and cost-intensive - answer categories may not be comparable - repeat measurements difficult
	5. <i>Context and Bias</i>	- no agency for those being measured - community input not possible	- surveying more time and cost-intensive - answer categories may not be comparable - repeat measurements difficult
	6. <i>Details and Nuance</i>	+ less contextual influence - only registered nation-state categories - potential selection bias due to assumed comprehensiveness	+ agency to self-categorise + community input possible
		+ additional information can be linked on individual level - no subnational details	+ agency to self-identify + community input possible
		+ subnational categorisation possible - potential bias due to priming or contextual influence	+ subnational and other forms of identification possible - potential bias due to priming or contextual influence
		+ required information can be included + subnational details can be included	+ required information can be asked + subnational details can be included + behavioural, motivational, and day-to-day situations can be studied

**In approach II. and III. it is assumed that open-ended questions are possible and included.*

Table 3 summarizes the main advantages and disadvantages of the three approaches. These can be summarized over six different dimensions:

Dimension One: Stability versus Fluidity

The benefit of the registering ethnicity approach is the stable nature of its measurement of ethnic origin. Repeated yearly measurements can be done in near-identical ways, resulting in data that can be directly compared over time to identify specific trends or changes. International comparisons are also possible as long as the same algorithms are followed and data gathering is equivalent (e.g., groups based on recognised nation-states). In addition, registry data can be used for more targeted research. Detailed demographic data can be linked on an individual level (such as gender, age, educational attainment, place of residence, sector of employment, etc.) to construct intersectional analyses that may benefit researchers and policymakers.

The two other approaches are more challenging in this respect. Research has shown that respondents do not reliably mark the same category consistently and that group affiliation is fluid, changing easily over time. For instance, Rademakers and van Hoorn (2021) found longitudinal evidence of ethnic switching in individual-level panel data in the U.S., Indonesia, and India. Booth and colleagues (2022) also found changes in self-classification among students in their research on ethnic-racial identity development among adolescents. 22% of the 732 followed students changed their ethnic-racial classification at least once during their research (Booth et al., 2022). Other studies yield comparable findings. In examining Canadian data Liebler and Hou (2020) concluded that some individuals change responses to ‘race questions’ over five years, with ‘White’ groups being the most stable, and multiple race response groups the least. Answer stability was linked to social statuses and experiences (e.g., mixed ethnic heritage, immigration status) rather than economic (e.g., income level) and personal characteristics (i.e., education, age). Research by Posner (2004) set in the U.S. has illustrated utilizing longitudinal data that an individual’s self-identified race changes over time, in response to changes in, amongst other elements, social position.

Dimension Two: Objective/indirect versus subjective/direct measurement

Registering ethnicity approaches can be considered more ‘objective’ or indirect (Karakas & Özbilgin, 2019). While the term objective carries a normative connotation, a crucial element is that the individual being categorised (the subject) in general cannot influence the measurement to the same extent as with other approaches and is consequently an indirect way of measurement. In addition, measurable characteristics are used as proxies for ethnicity, often nationality, citizenship, country of birth, or even names. In other words, individuals are labelled in a comparable way as opposed to the other approaches that can be more open to interpretation.

If the categorising is done by a third party, the individual doing the categorising (e.g., via surveys) can introduce subjectivity as well. For instance, Ahmed and colleagues show how respondents were classified as Roma based on “negative social characteristics” such as poverty or low education rather than on specific ethnic markers (Ahmed et al., 2007). This difference between observed and self-reported ethnic and racial categorisation is illustrated by research conducted by Saperstein (2006) as well. The author found substantive differences in the General Social Survey in the U.S., showing how different subjectivities can result in different findings. Other context-related factors can be minimised by using the registering ethnicity approach. For instance, experiencing high levels of discrimination may affect ethnic identification, consequently influencing measurements. Respondents may be less likely to

identify with a majority group compared to respondents experiencing less or no discrimination (Jasinskaja-Lahti et al., 2012).

Although the *registering ethnicity* approach does not ‘suffer’ from these factors to the same extent as other approaches, there is still a specific context at work. The nation-state is often the primary vehicle for collecting these data, establishing the allowed categories (often in terms of other nation-states), and other technical considerations, such as frequency of data collection, and accessibility.

Dimension Three: Practical considerations

For policymakers especially, practical considerations can be crucial. In case registers are available - as they are in most Western European countries including Belgium (Kukutai et al., 2015) - it is both time- and cost-effective to enumerate groups based on origin data of the individual and/or family members. This precludes the need for costly surveys with potentially complicated sampling strategies to reach representative data. While such challenges can be met to some extent, there are nevertheless arguments to be made for the relative ease with which these registers can be consulted.

However, registry data has downsides as well. The *identifying ethnic origin* approach may be better suited in some scenarios. For instance, if there are large groups that are labeled by their country of origin, subnational differences can mask in-group variation that may very well contribute to the topic under consideration. In this case, complementary usage of registry data to identify the subpopulation (e.g., people with Moroccan parents) to administer a survey to is recommended.

Dimension Four: Agency

The lack of agency is the principal disadvantage of the *registering ethnicity* approach compared to the *identifying ethnic origin*- and *assessing ethnic identity* approaches. Since many if not most ethnic minorities are in an inferior power position vis-a-vis the respective majority populations, it is paramount to empower individuals and groups by allowing space for using the identity categories they prefer. Register data precludes the input of the individuals to identify as they see fit, potentially silencing specific minority groups that could take issue with the labels provided by the official registration systems, commonly based on recognised nation-state borders.

Historically, population counts have been a historical area of struggle for self-determination and self-identification. Preventing input also “keep(s) out politics and turn(s) identity designation into an administrative and statistical exercise” (De Zwart, 2012, p. 311), leaving little room for identify politics. Open questions and the possibilities of mixed options here are recommended if this is an area of concern.

Agency is also relevant when the possible categories or labels are established. It can be challenging to be familiar with all the (historical) connotations different labels carry, or with important ethnic subdivisions that nation-state-level groupings simply cannot capture or willfully gloss over. In constructing categories input by members of the communities can be an important asset in allowing agency and capturing detail and nuance. However, if ethnicity is ‘real’ and needs investigation in the first place, it is conceivable that the ‘community’ follows ethnic lines. In other words, it is important to have some knowledge of the ethnic groups below the level of national categories in order not to reify and maintain existing power imbalances within minority groups, reflecting, among other things, political or historical cleavages.

Finally, agency also matters with regards to the collection and use of personal data, such as ethnicity or origin data. Approach two and three (*identifying ethnic origin* and *assessing ethnic identity*) allow for informed consent to be directly obtained from the individual. As such, individuals not in favor

of data collection can simply opt out or not participate with the survey. For the use of administrative data, this is for obvious reasons not possible as much of this data is registered in order to participate in society (e.g., address, marital status, work history for tax purposes and more).

Dimension Five: Context and Bias

Context is an important factor to consider when measuring ethnic origin. Overall, *registering ethnicity* approaches tend to be less susceptible to contextual variations albeit with one caveat. Due to methodological nationalism, population registers typically employ categories based on recognised nation-states. The (imagined) boundaries of ethnic groups and categories of self-identification are, however, not necessarily equivalent to those proclaimed by nation-states. In other words, nationality or country of origin may not be a suitable proxy for ethnic origin, potentially obscuring considerable ethnic cleavages with groups actively engaged in conflict. Policies that fail to take these aspects into account may inadvertently cause harm to certain groups or reinforce existing sub-national or supra-national power relations.

In addition, since the *registering ethnicity* approach is assumed to comprehensively register the population, it counterintuitively may be prone to oversight and selection bias. A register implies that it contains individuals who are *registered*. Individuals who are *not* registered are typically not considered. These individuals may disproportionately belong to specific vulnerable ethnic groups (e.g., belonging to specific refugee groups). Policies targeting social challenges could end up using data that disproportionately misses the most vulnerable groups in need of these policies thus missing their mark.

Administrative data, however, is cumbersome to update or adjust according to new understandings of the demographic landscape (De Zwart, 2012). In contrast, following an *identifying ethnic origin* approach, updating the categories in a survey can be rather simple. One can simply adjust the labels used, the format, the order of questions, and more for the next survey. Yet this approach too can be biased by means of *priming*. Priming can occur through the preset categories provided in a survey. For instance, if a question on ethnic origin offers only regional or country names, respondents may feel forced to choose from those options, even if they would have preferred a different self-identification. This can lead to a bias where the data wrongly suggests people identify primarily with nation-states

The *identifying ethnic origin* approach also includes self-reporting which too can introduce bias. Dembosky et al. (2019) found Black individuals were most likely to not report, and Croll & Gerteis (2019) noted similar issues with open-ended questions. Furthermore, Woolverton and Marks (2023) showed white adolescents' answers were more consistent between different question formats than those of adolescents of colour.

External factors too may bias origin categorisations. Classical constructivist works on ethnicity (e.g., Gans, 1979; Lie, 2008; Waters, 1990) argue how ethnicity is not always salient in all contexts but gets 'activated' during certain events or at specific times. Polyethnic individuals – often with parents of different ethnicities - can also oscillate between two or more options, depending on the circumstances. Consequently, conducting a survey risks capturing whichever option is salient at the time of the survey.

Moreover, international research illustrates how structural and historical contexts may potentially influence how people categorise themselves today. Research in the U.S. has examined the relationship between historical anti-miscegenation laws and contemporary racial classifications within Black-White mixed households. O'Connell et al. (2022) found that historical laws still influence racial self-categorisation, skewing comparisons of certain groups between states. Both their research and that of Pickett et al. (2019)) show that racial classification varies spatially within the same country. This

aligns with findings in Brazil from Chor et al. (2019), where a higher proportion of African ancestry in a city increased the likelihood of self-declaring as “Black.” Research by Veerman and Platt (2021) illustrates how in more diverse classrooms children with a migrant background choose a greater number of ethnic identities compared to those in less diverse classrooms. Johfre et al. (2021) found how even (not) taking genetic ancestry tests can impact self-categorisation and self-identification. Their research found that people identifying as multiracial were more likely to take such tests and that people taking such tests were more likely to report multiple ancestral origins in addition to reporting multiracial self-identification.

Dimension Six: Detail and Nuance

In a recent editorial Harewood & Rosenfield (2021) asked the reader how much sense it makes to use categories such as “Asian”, a continent comprising over four billion people. What we in Europe see as “Chinese” for instance, is not necessarily an ‘ethnic origin’ as such and reflects methodological nationalist perspectives on places where more nuances may be required. China in itself has a large ethnic diversity, with around 56 ethnic-cultural groups registered (Y. Zhou, 2019). Similarly, consider the case of India. First, the sheer number of its inhabitants complicates comparisons with other nation-states (1.408 billion in 2021 compared to 11.59 million Belgian people according to the World Bank). Next, ethnic conflicts within state borders take place today, such as in Manipur between Kukis and Meiteis (Singha, 2017). While these cleavages do not necessarily always translate into divisions in overseas ethnic groups, awareness of these may aid prevention of potential issues. Last, even regardless of such complications in the ‘homeland’, several challenges require a more nuanced understanding. For instance, regional differences in literacy in India are vast, ranging from 66.4% in Andhra Pradesh to 96.2% in Kerala, with gender literacy gaps respectively 13.9 and 2.2 percentage points in 2022 (Swargiary & Roy, 2022). As such, when considering topics from educational challenges to employment barriers, the situation may differ substantially between two families that are categorised under the same moniker (i.e., “Indian”), illustrating the potential pitfalls of a methodological nationalist perspective on the world and on a society’s composition.

Research has illustrated the “heterogeneous nature” of populations (Aspinall, 2002) and the importance of considering these sub-national levels of categorisation. Elling and Harris (2021), in a study set in Iran, investigated differences in ethnic categorisation captured via closed and open-ended questions from the 2016 Iran Social Survey. They found discrepancies between these groups with open-ended questions allowing for more detailed responses, changing the proportions of particular groups when compared to common, broader categorisations. Xu and Zhang (2022), operationalizing ethnicity via name and mother tongue as listed on historical census data of the early twentieth-century U.S., found high levels of occupational segregation within categories based on national groups. Among the Russian-born immigrants (“Russians”), Jews were more concentrated in trade, ethnic Germans in agriculture, and Polish immigrants in the manufacturing sector, potentially affecting socio-economic conditions in myriad ways.

However, *registering ethnicity* approaches are not necessarily devoid of any detail and nuance. A wealth of socio-economic variables can be linked on an individual level: employment sector, educational attainment, field of education, living address, household composition and family members, and more. In our increasingly diverse society, it becomes paramount to take into account this “diversification of diversity” or superdiversity (Vertovec, 2007) rather than just focusing on ethnic or national background. Whereas the format of traditional census questionnaires struggles with capturing

this superdiversity (Aspinall, 2012), extensive use of different registries that allow such linkages can open new avenues of research.

4.1.3 Recommendations for the Belgian Federal Public Services

The approaches discussed have their advantages and disadvantages. The idea of "more data is better" needs to be balanced by practical considerations. With an eye on the agency of individuals and mapping disenfranchised groups, self-identification within an identifying ethnic origin approach, as recommended by the EU High Level Group on Non-discrimination, Equality and Diversity is preferred. However, the massive recurrent cost and time investment, sampling difficulties, in addition to the expected time- and context variability of answers on a population scale can make this a challenging approach that is more suited to complement existing methods.

Belgium as a bureaucratically developed nation-state has a wide range of databases that can be linked via organisations such as KSZ/BCSS. Factors such as employment status, employment sector, nationality, and country of birth of parents and individuals, but even educational attainment and more can be captured at the level of the individual, allowing the measurement of not only diversity but to some extent superdiversity by means of the *registering ethnicity* approach albeit with a concern for GDPR guidelines. Organisations that aim to keep track of changes in ethnic composition benefit since registers will not change as easily as an individual's self-identification on a repeated survey may. In this case, administrative data is a suitable, default approach to frequently monitor ethnic diversity in the public administration.

Yet the federal public administration is also contained enough to go beyond the mere use of administrative data (*registering ethnicity* approach). For instance, access and sampling are less of an issue, since all employees are registered. This makes surveys a valid and practical solution for specific research questions and topics. An *identifying ethnic origin* approach can be proposed in the form of a (yearly) well-constructed survey. Including open write-in options and the option to check multiple categories, this method allows more nuance and details (such as religious affiliation or even language use) to be registered in addition to the commonly used proxies such as nationality at birth of the individual and the parents. This format may also provide space for much needed sub-national detail.

Specific research themes dealing with ethnicity, related motivations, behavioural aspects as well as identity formation, often proposed as an important potential explanatory mechanism may also benefit from one of the several scales proposed (*assessing ethnic identity* approach). This would also be in line with both academic and other recommendations. In line with the Socio-Economic monitoring, it is also recommended to 'measure' or take into account third-generation individuals. The algorithm employed in WP 1.2 of this report can be used in case the registering ethnicity approach is the sole option. If not, information about the grandparents can be added to surveys.

More complex is what 'groups' we should identify to begin with. Register data, embedded in a nation-state framework, suffers from methodological nationalism and consequently only identifies countries. Therefore, the recommendation is to allow more detail that can be captured via surveys. With explicit informed consent, this also prevents any additional privacy concerns regarding access to existing administrative data for extracting more information.

4.2 Intersectional analysis of Socio-Economic Monitoring and benchmarking

This section examines ethnic composition and diversity between three groups, denoted as ‘sectors’ here. Primarily, the difference is between the public sector and private sector employees. Within the group of public sector, the division is made between those of whom the establishing power falls under the federal jurisdiction or not. As such, *other public sector* contains entities from the regional level to the provinces and local municipalities, whereas the *federal public sector* includes advisory councils and special services (e.g., *Federal Advisory Council on the Employment of Foreign Nationals*), the parliament and parliamentary institutions, public social security institutions (e.g., *National Employment Office (RVA/ONEM)*, *National Pensions Office (RVP/SFPD)*, *the National Social Security Office (RSZ/NSSO)*), institutions of public utility and scientific institutions (e.g., *Belgian Royal Library, Sciensano, Federal Planning Bureau*), state-owned enterprises (e.g., *NMBS/SNCB, Bpost*), public services and administrations (e.g., the different *Federal Public Services (FOD/SPF)*), the federal police, and the armed forces. While the allotted space does not allow visual representations of all intersections, the main findings are included in this section, primarily focusing on ethnic origin.

4.2.1 Ethnic Origin Composition

Figure 1 depicts the proportion of each non-Belgian group within each sector for 2021. Almost all origin groups are underrepresented in the federal public sector compared to both the private sector and other public sector entities. Figure 2 compares the relative proportions of these numbers between the private sector and the federal public sector. The percentages show the difference between the proportion of a group in the private sector compared to the group’s proportion in the federal public sector. For instance, if a group has a proportion of 1% in the federal public sector, and 5% in the private sector, the figure shows this group as 400% more represented in the private sector. All but one origin group is more represented in the private sector, with some showing particularly high differences. The Bulgarian and Romanian groups are 631.4% and 519.9% more represented in the private sector compared to the federal sector. In contrast, Western European groups are less overrepresented, with figures ranging from 13.6% for individuals of Italian origin to 80.9% for those of Dutch origin. The only exception is the Congolese origin group, which is 13.9% less represented in the private sector. Meanwhile, other Sub-Saharan African groups are 80.7% overrepresented in this sector. In line with the findings and recommendations of WP 1.1, this is a concrete instance in which broader classifications would miss important nuance. Although the Congolese origin group is commonly subsumed within the broader category of Sub-Saharan African origin, the representation of the Congolese group is clearly different as opposed to the other Sub-Saharan African groups.

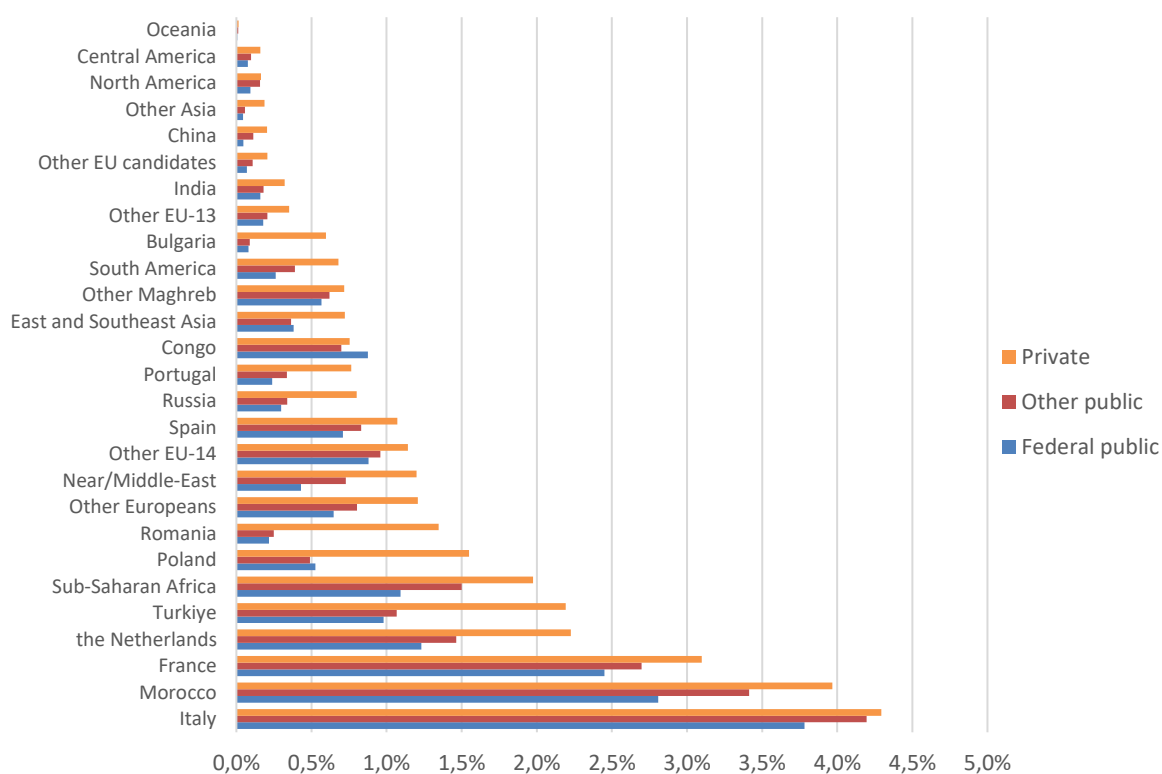
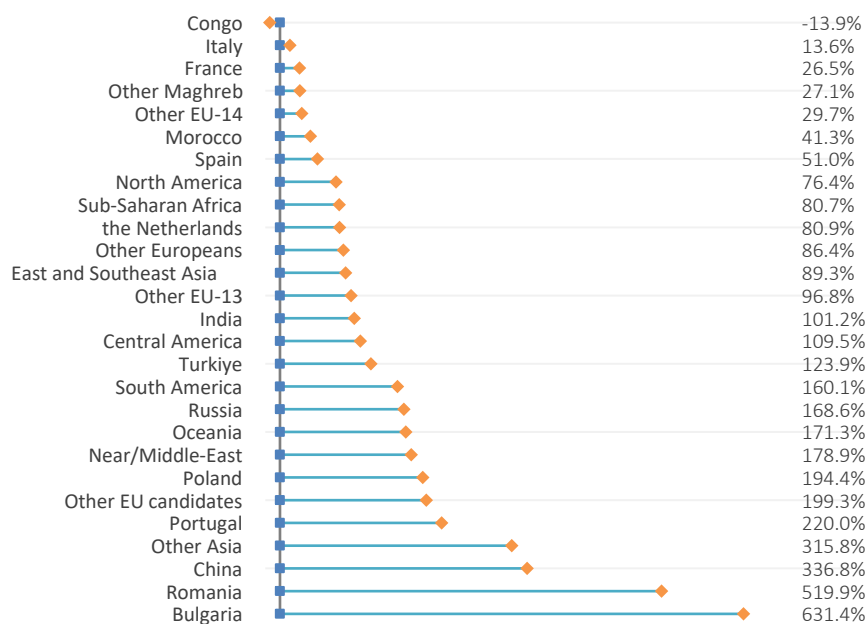


Figure 1 Proportions of non-Belgian origin groups per sector (2021)



federal public sector & private sector (2021)

Figure 2 Difference in proportions of origin groups in federal public sector and private sector (2021)

In the federal public sector overall, men are overrepresented in both the Belgian origin and non-Belgian origin groups, accounting for 63.2% and 63.7% in 2021, respectively. These figures have only slightly decreased since 2011, by 3.3 and 0.9 percentage points, respectively. This picture, while useful, hides substantial internal differences. Figure 3 illustrates how some employers, such as the *armed forces* or the *police* are dominated by men with twice (68% vs 32%) or even more than four times their number (87.3% vs 12.7%), whereas in *public services and administrations* and *public social security institutions* women comprise the majority of employees (53.1% and 66.8% respectively).

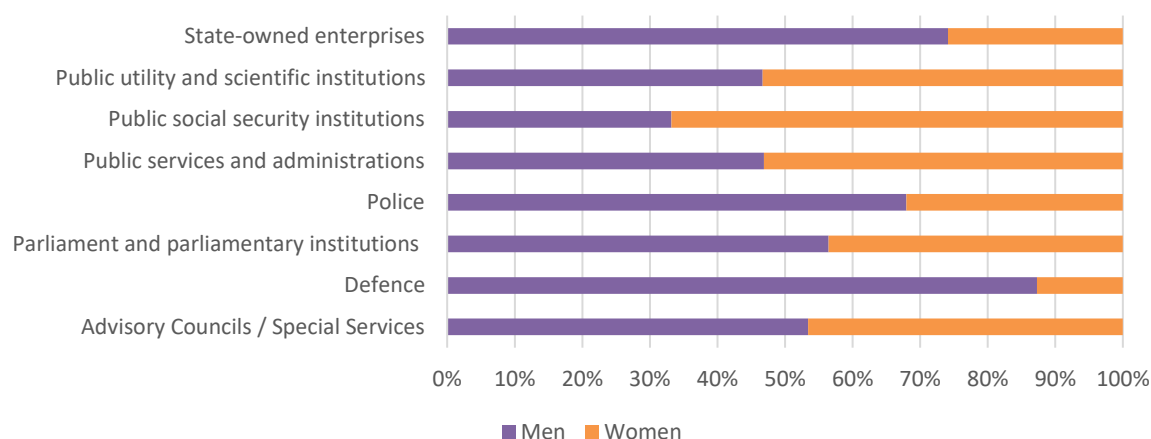


Figure 3 Gender balance breakdown of the federal public sector (2021)

Figure 4 illustrates the gender balance within the overall federal public administration per origin group. Groups of Near/Middle East origin, Congolese, Sub-Saharan African, as well as Moroccan and other Maghreb origin show the strongest imbalances. For this first group, women consist of only 23% of the origin group employed in the federal public sector. Figure 5 breaks down these proportions further by examining the gender balance within several origin groups for the different institutions within the federal public sector. Note that parliamentary institutions and advisory councils are excluded from this figure, as some gender and origin groups had no employees represented in these categories. For *public social security institutions*, the Moroccan and Other Maghreb origin groups show a much higher proportion of women than men employed. *Public utility and scientific institutions* are fairly balanced, except for those of Near or Middle Eastern origin, where men dominate (70.4%). *State-owned enterprises* are also predominantly occupied by men, especially within the Near/Middle Eastern origin group. *Federal services and administration*, despite an overall tendency toward employing more women, see more men than women employed among the Congolese, other Sub-Saharan African, and Near/Middle Eastern origin groups. The Moroccan and Other Maghreb groups, however, align with the overall trend favouring female employment. Finally, the armed forces and police services reflect the broader global pattern, with substantially more men employed than women.

To examine the intersection of age and gender, we focus on the four origin groups with the greatest gender imbalance in the federal public sector in 2021. Figure 6 shows that among individuals of Congolese and Moroccan origin, the imbalance is less pronounced in the younger age cohorts. However, starting from around ages 30 and 35, the gap widens. For those with origins in the Near and Middle East and Sub-Saharan African origin, the gender gap is already evident in the youngest age groups and remains consistent across later age cohorts.

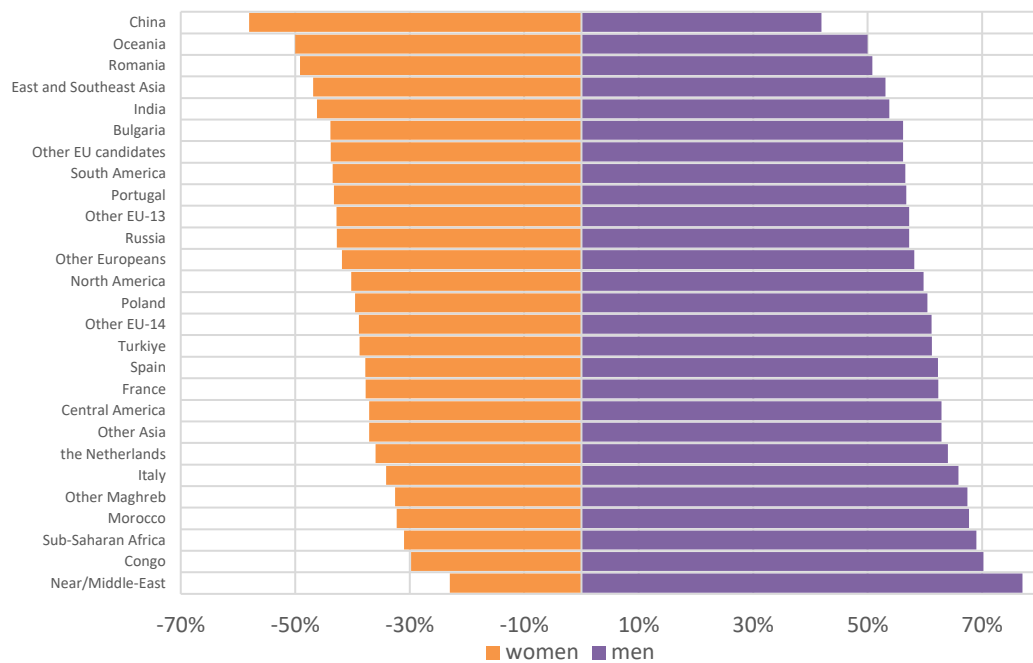


Figure 4 Federal public sector gender balance per origin group (2021)

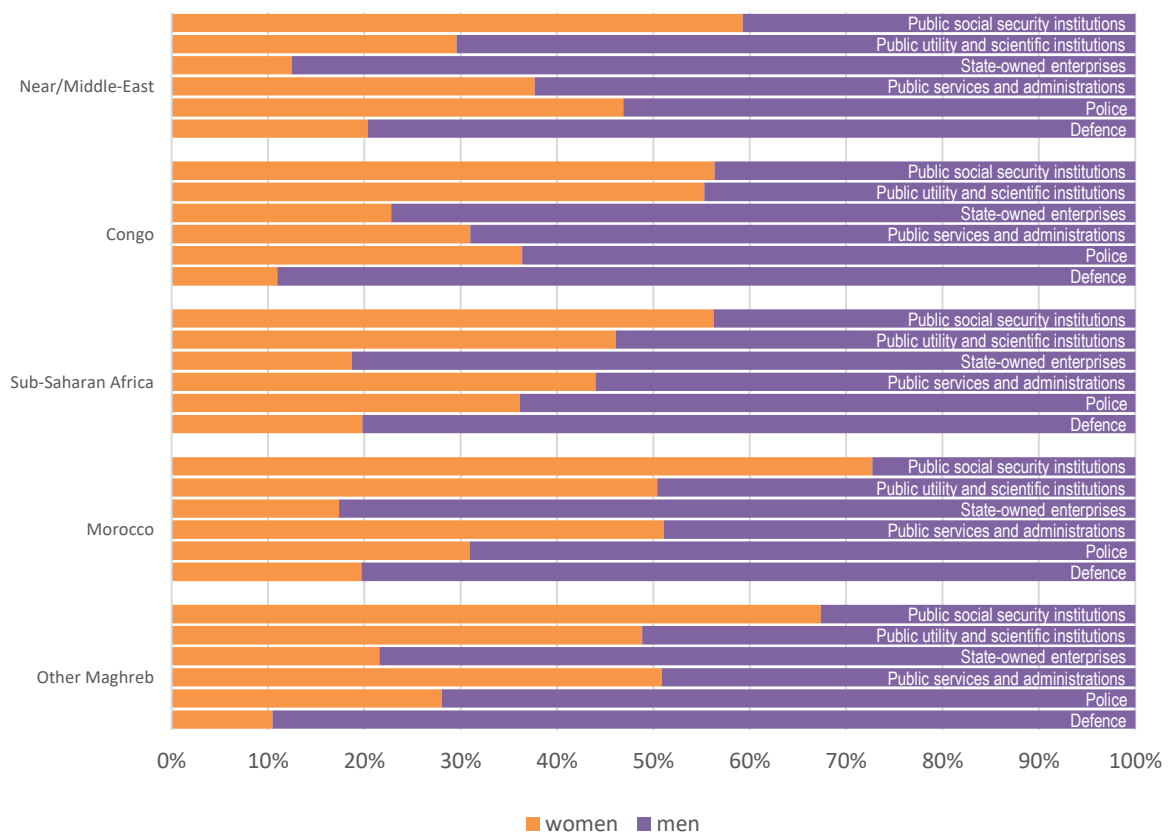


Figure 5 Gender balance breakdown per selected origin group of the federal public sector (2021)

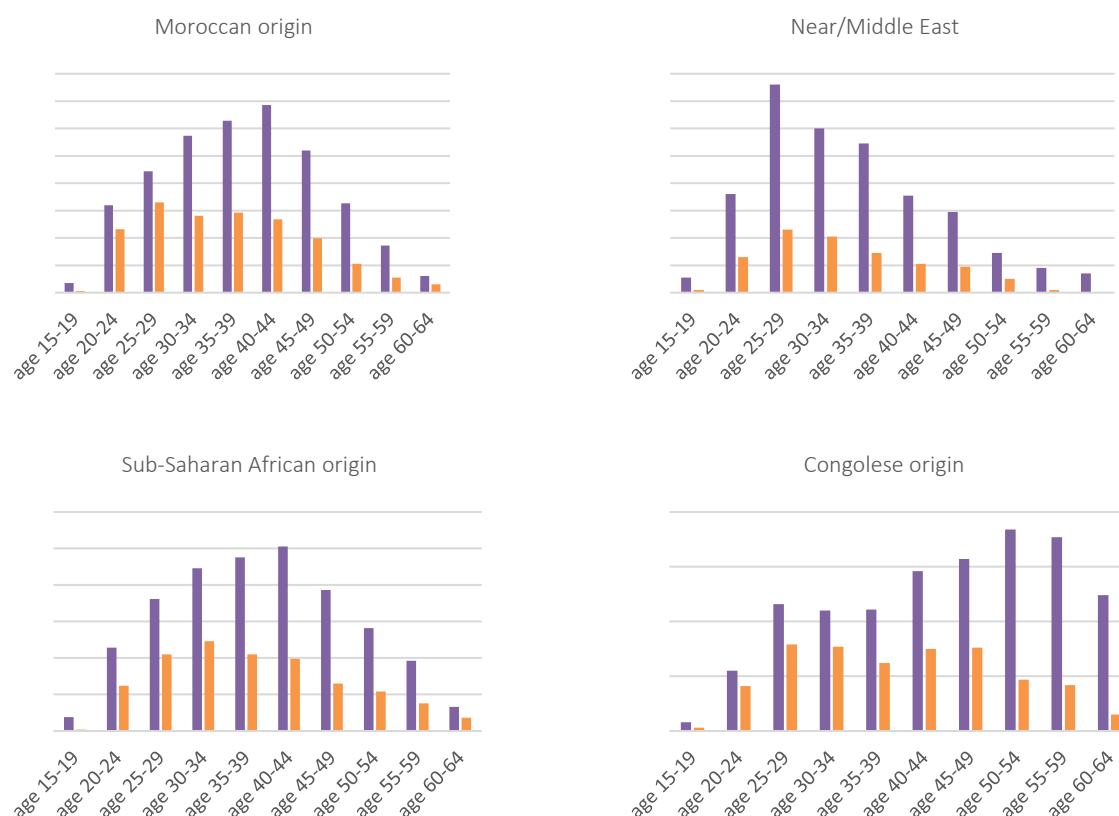


Figure 6 Federal public sector gender split per age cohort for selected origin groups (2021)

Since the various ISCED levels make it difficult to visually distinguish differences as there are nine categories, a simplified comparison is provided in Figure 7 for 2017⁷. ISCED level is clustered as *lower educated* (ISCED 0 – 2), *medium educated* (ISCED 3-4) or *higher educated* (ISCED 5-8). The origin groups with the highest proportion lower educated employees in the federal public sector originate from Portugal (22.9%), Turkey (21.1%), and several EU countries. The origin areas with the highest proportion of higher educated workers are China (64.5%), Oceania (57.1%), followed by Congo (47.9%) and East and Southeast Asia (46.3%). Finally, a note on missing values. While no dataset is perfect, in particular with education level, some ethnic groups have substantially more missing educational data. For instance, for the Bulgarian and Romanian origin group 31.5% and 33.7% missing values respectively.

⁷ This was the most recent year for which this information was available at the time of the data request.

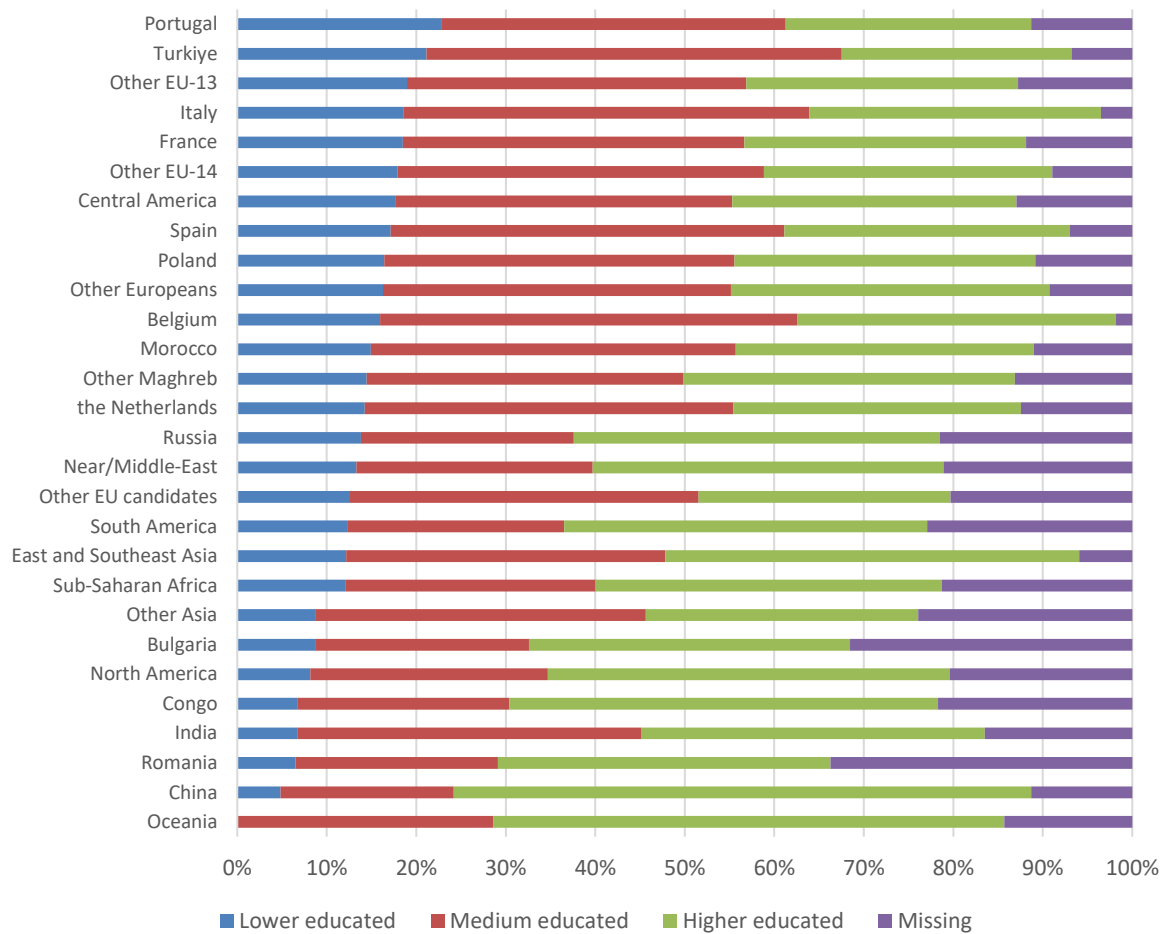


Figure 7 Education level by origin group for the federal public sector (2017)

Next, we consider the type of employment contract of individuals in the federal public sector. More specifically, we consider whether individuals are employed as workers, employees, or as civil servants. Figure 8 shows how the proportion of civil servants is largest in the Belgian group (74.9%), whereas the average among non-Belgian origin groups is 49.3%. Both civil servants as well as workers have declined since 2011 for both Belgian and non-Belgian origin groups. The Bulgarian group consists both of the highest proportion of workers (tied with the Central American origin group at 9.3%) as well as the lowest share in civil servants (27.2%).

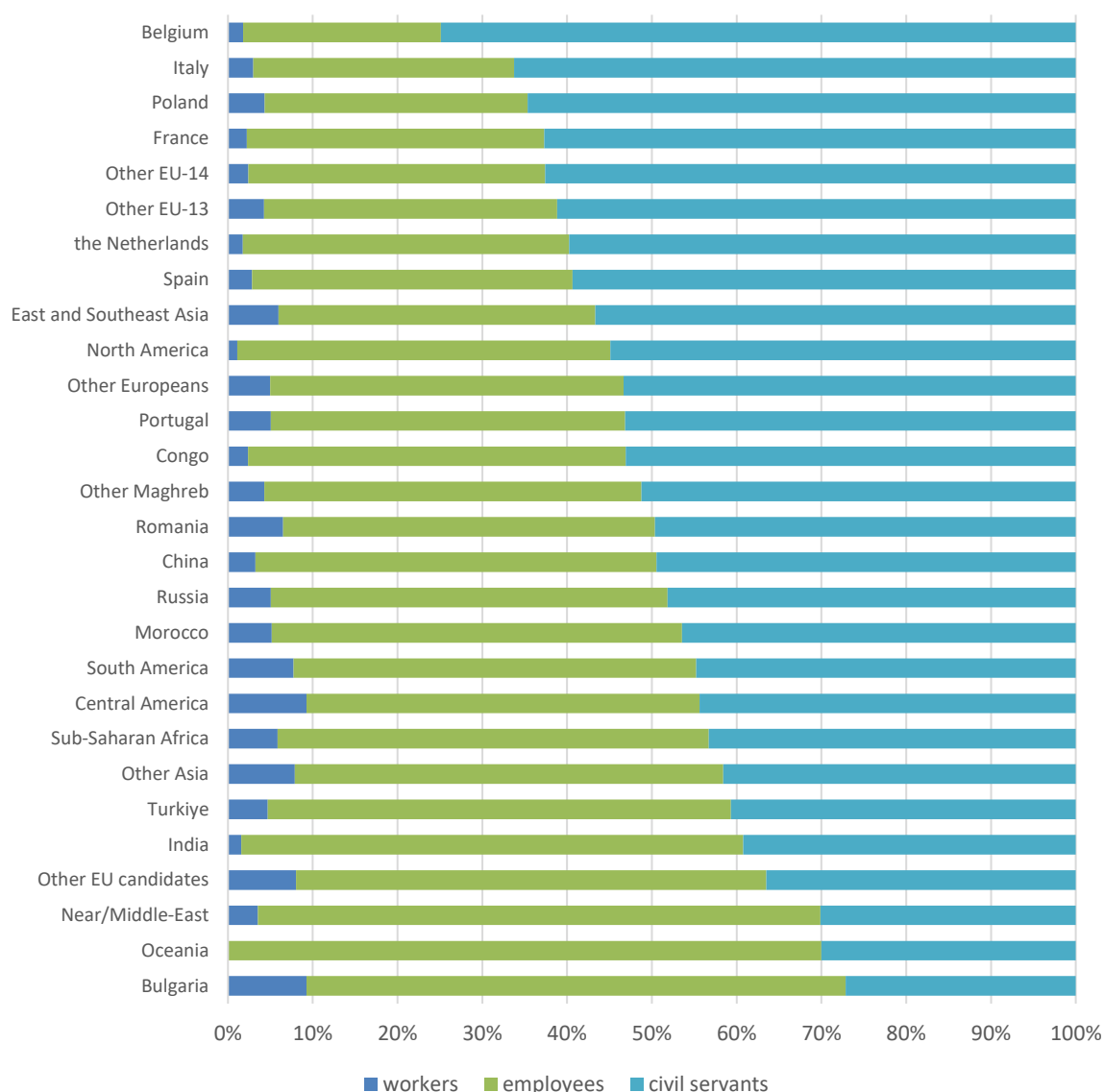


Figure 8 Breakdown of type of employment contract per origin group for the federal public sector (2021)

4.2.2 Generation Composition

Focusing on ethnic origin – while elucidating – can be complemented by examining the composition of generations across several variables. Generation here is used in the context of migration. We examine individuals with a first- (G1), second- (G2), and third generation (G3) migration background. In other words, first generation individuals are those that moved to Belgium, whereas second generation individuals are their offspring. From third generation onwards, individuals with a migration background are considered as “Belgian” in the existing variable construction, but this choice is to some extent arbitrary. To this last category belong all individuals that were born from Belgian parents (i.e. Belgian nationals born as Belgians). Among the first generation-cohort, the distinction is made between those that are naturalized as Belgian (G1_nat), and those that are not naturalized (G1_nn).

Figure 9 shows the overall changes in generational proportions across the three sectors between 2011 and 2021. The proportions of both second- and third-generation groups decrease in all sectors, with the third-generation group experiencing a more pronounced decline in the private sector. Conversely, the proportion of the first-generation group increases across all three sectors, ranging from a 1.4% and 1.5% increase for the non-naturalized G1 group in the two public sector categories, to a 3.6% increase in the private sector.

Federal public sector	2011	2021	Δ
First generation - not naturalized (G1_nn)	1.4%	2.9%	1.5%
First generation - naturalized (G1_nat)	2.3%	4.5%	2.3%
Second generation (G2)	18.3%	16.1%	-2.2%
Third generation (G3)	78.1%	76.5%	-1.6%

Other public sector	2011	2021	Δ
First generation - not naturalized (G1_nn)	4.1%	5.5%	1.4%
First generation - naturalized (G1_nat)	3.6%	5.4%	1.8%
Second generation (G2)	17.2%	15.1%	-2.1%
Third generation (G3)	75.1%	73.9%	-1.1%

Private sector	2011	2021	Δ
First generation - not naturalized (G1_nn)	9.4%	12.9%	3.6%
First generation - naturalized (G1_nat)	5.4%	7.9%	2.5%
Second generation (G2)	16.9%	15.0%	-1.9%
Third generation (G3)	68.4%	64.2%	-4.1%

Figure 9 Changes in generational proportions across the three sector groups between 2011 and 2021

Figure 10 examines changes in the proportions of individuals employed as civil servants in the federal public sector between 2011 and 2021 split by gender. While the overall trend indicated a (slight) decrease across various origin groups (including Belgians – see 4.2.1), the proportion of first-generation naturalized civil servants increases over time. This rise can partly be attributed to the overall growth in the number of employees within this group, regardless of contract type. In contrast, the non-naturalized first-generation group similarly experiences an overall increase in proportion employed, but this growth is not reflected in their representation among civil servants. The gender split reveals that both increases and decreases are less strong for women compared to men. For third generation women, the proportion over time remains more or less stable.



Figure 10 Proportions of individuals employed as civil servants in the federal public sector between 2011 and 2021 split by gender

4.2.3 Ethnic Diversity

Whereas ethnic *composition* typically refers to the proportion of a population belonging to a particular ethnic origin (e.g., a non-European origin or specific ethnic group), ethnic *diversity* also takes into account the number of different groups involved within an entity. Ethnic diversity here is expressed by means of the Ethnic Diversity Index (EDI). This EDI ranges between 0 and 1, with 0 designating no ethnic diversity (i.e., *all members have the same ethnic origin*), and 1 perfect diversity.

Figure 11 illustrates the Ethnic Diversity Indices for the subgroups of the federal public sector, and contrasts these with those of 2017 – the earliest year for which we had complete data for this calculation. The figure suggests overall ethnic diversity has improved in each of the federal institutions examined between 2017 and 2021. The smallest increase can be observed in the federal *parliamentary institutions* (EDI 0.337 up from 0.334 in 2017) and the largest increase in *state-owned enterprises* (from 0.387 to 0.446). Overall, in 2021, federal institutions of *public utility and scientific institutions* were the most diverse, with an EDI of 0.503, while federal parliamentary institutions were the least diverse (0.337).

In addition, we can add the intersection with gender to consider ethnic diversity within gender groups. EDI split between administrative gender groups (Figure 12) shows how in 2021, ethnic diversity was higher among female employees in all but two institutions: *federal state-owned enterprises* (7.5% lower), and *public services and administrations* (2.2% lower). This gender difference is particularly pronounced in the employees of the *parliament and parliamentary institutions*. Here, women are 46.7% more ethnically diverse when compared to their male colleagues.

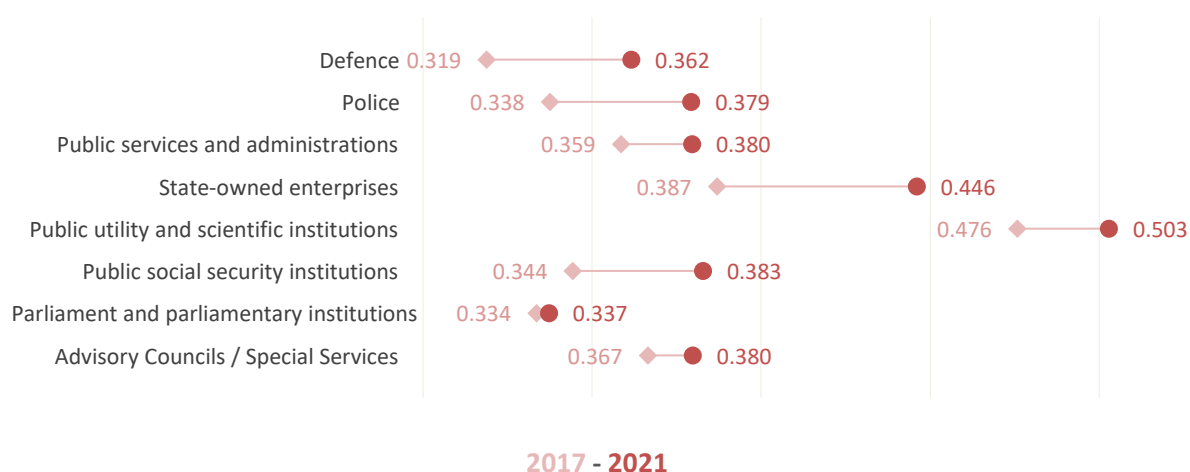


Figure 11 Evolution of EDI for the subgroups of the federal public sector (2017 - 2021)

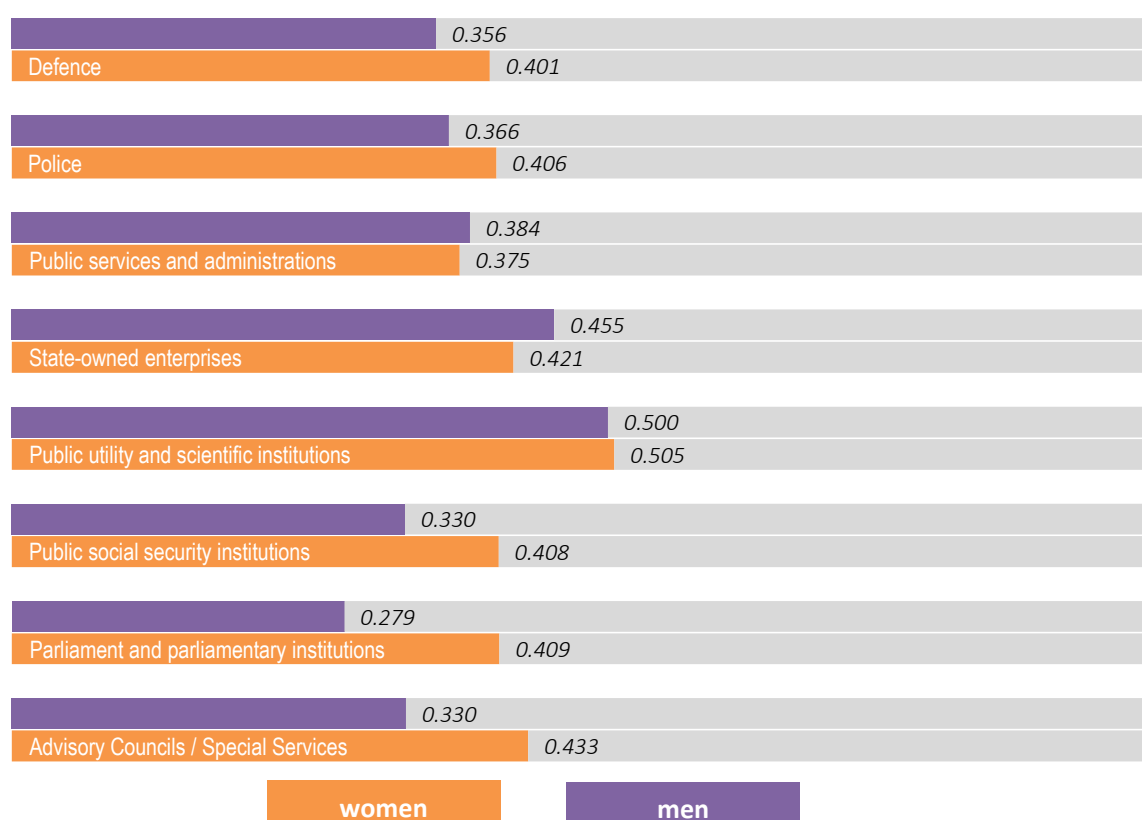


Figure 12 Federal public sector EDI split by gender (2021)

While overall ethnic diversity in the federal public sector has increased, how does it compare to other non-federal public administrations and services?⁸ One key consideration is the large variety of institutions and organisations that make up the public sector.⁹ Spanning both international institutions and local administrations, ethnic diversity varies widely. International institutions score especially high, with 2021 EDIs of 0.895 for *educational* occupations, and 0.918 for *embassies and representational* occupations. Due to the international nature of the functions encompassed by these categories, a higher degree of ethnic diversity is to be expected. At the other end of the spectrum are the *Flemish parliament and parliamentary institutions* with an EDI of 0.157, *provincial development companies* (i.e., organisations in charge of implementing the socio-economic policy of the provincial governments) scoring 0.181, followed by several Flemish organisations and institutions.

Several findings stand out. Parliamentary institutions across the board seem to score low(est) on ethnic diversity, both on a federal level as well as across the regions. In other words, these institutions employ predominantly individuals of Belgian origin. Next, employees in the Walloon Region as well as the French-speaking Community public services are, on average, more ethnically diverse than their Flemish/Dutch counterparts. Finally, while *parliamentary institutions* and the *provincial development companies* are relatively small groups in terms of the number employees employed, this is not the case for all categories that score low on ethnic diversity. The third lowest ranking entry, *Flemish Administration*, and in particular the *Flemish Educational* group, are large employers, suggesting the difference in ethnic diversity is not merely explained by employer or institutional size in the market.

Finally, the Brussels Region public sector is the most diverse, closely followed by ‘special cases,’ which include various groups within international organisations (0.759 and 0.708, respectively). The lowest levels of ethnic diversity are observed in the Flemish Region and Community (0.251). Comparatively, the federal public sector scores average to low (0.407), with only the provincial and Flemish levels being less ethnically diverse.

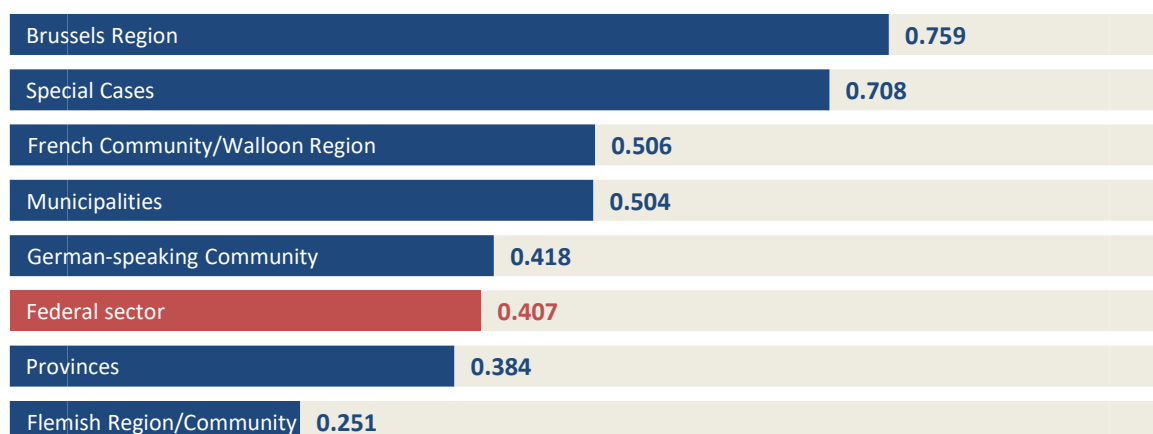


Figure 13 Overview of public sector diversity (2021)

⁸ The Flemish Region and Community are combined in the source data, so these are listed here as such.

⁹ Only occupational groups with at least 100 employees are considered here.

4.3 Ethnic differences in public service motivations

4.3.1 Introduction and literature findings

Coined by Rainey (1982), *public service motivation* (PSM) was popularized in the 1990s via research by Perry and Wise (1990). Initially, PSM was understood as “*an individual's predisposition to respond to motives grounded primarily or uniquely in public institutions and organisations*” (p.368). More recently, it generally indicates the motivation of people to want to contribute to society. In parsing recent literature on PSM, Vandenabeele et al. (2018) untangles two common aspects of definitions, that is, the willingness to contribute to public processes, and the potential to set aside one's own interests in doing so. Although not limited to public sector employment, research has shown that PSM is positively correlated with public sector recruitment (Asseburg et al., 2018; Piatak et al., 2020; Woo & Kim, 2024), and thus forms a potential avenue to examine underrepresentation of certain groups.

PSM research has primarily focused on correlates of PSM either in the form of antecedents such as gender, or outcomes such as organisational performance. As to what causes PSM, Vandenabeele, Ritz, and Neumann (2018) point to explanations centring around *socialisation processes* at different levels of interaction, *self-determination theory* (R. M. Ryan & Deci, 2017) as to the satisfaction of basic psychological needs, as well as *social-identity theory* (Tajfel & Turner, 2004) in how some aspects of people's identities can support PSM, focusing the attention to *values* and *value structures* as a potential underlying factor.

Empirical research gives some substance to this claim as well. Ritz and Brewer (2013) for instance, find different PSM levels at the subnational level in Switzerland. The authors illustrate how Swiss German public employees show significantly higher levels of overall PSM compared to their Swiss French counterparts, pointing in the direction of cultural-administrative traditions as way of explanation. A non-Western study by Hassan and Ahmad (2021) investigates the relationship between PSM and an Islamic work ethic by sampling 419 employees in the Kurdistan Region in Iraq. Their analyses reveal a significant relationship in both public and private organisations, stressing the importance of cultural and religious factors.

In addition to subnational group level differences in PSM, the literature also points to cross-country distinctions. Asseburg and Homberg (2020) find significant differences in effect sizes between groups of countries, arguing for more research into the associations between PSM and different traditions. In a meta-analysis examining the effects of age and gender on PSM, Parola et al. (2019) too concluded that culture is in fact a key mediating variable. They bring to call an increased focus on this aspect for future PSM research, suggesting how “consistency of PSM correlations across cultures cannot, and given the evidence, should not, be taken as a given” (p.1402). Country differences were also found by Vandenabeele and Van de Walle (2008), Westover and Taylor (2010) as well as Houston (2014). Yet despite these conclusions suggesting how the extent of PSM tends to vary both between as well as within countries, potentially mediated by culture, there is a notable absence of research focusing on ethnic minorities – a target group that at least nominally combines these two elements.

Operationalizing cultural values is, however, not straightforward. Based on the influential work by Kluckhohn, Rokeach and Schwartz on values, Sagiv and Roccas (2017) summarize values as “cognitive representations of basic motivations” or needs in the form of “abstract, desirable goals”. Compared to norms or attitudes, values are deemed relatively stable from adolescence onwards (Cieciuch et al., 2016), transcending specific situations, actions, or contexts. As such, they can be linked to a much wider range of behaviour – from specific ones such as donating to specific charities, career choices, to more general everyday behaviour such as keeping promises. Values are considered to be

important for individuals since they serve as criteria or standards for a wide range of everyday “evaluations” (Rokeach, 1973), thus serving as a foundation for both motivations and action.

Stability and commonalities in values, however, do not equate to uniformity across cultures. Although there are similarities between cultures of what are deemed the most and least important values, the overall structure or framework of personal values can differ substantially (Roccas & Sagiv, 2010). Cieciuch (2017) refers in this context to phylogenetic and ontogenetic perspectives on values. Whereas the phylogenetic perspective on values departs from a more general, biologically grounded approach of values as basic needs, the ontogenetic perspective considers an interaction between basic motivations and (latent) culture as the key process for individual value construction. For this reason, Cieciuch also refers to the former type of value as “biologically determined needs” and to the latter as “value preferences”. In sum, culture, as a crucial element in the socialisation of individuals, moderates the construction of personal value systems (“value preferences”).

Although agency and individual characteristics are important factors in value systems, there is nevertheless empirical substance to the argument for a clustering at the nation-state level. Akaliyski et al. (2021) empirically test the validity of the nation-state as cultural entities via a range of multi-national survey projects such as the World Values Survey (WVS), the European Values Study (EVS), the European Social Survey (ESS), the International Social Survey Program (ISSP), and the Global Barometers Project (GBP). Instead of clearly marked containers, they consider culture as a gravitational field with the nation at the centre of concentric circles and find how nation-state clusters explain a significant share of the variation found in cultural values – more so than other aggregations such as those based on religion, ethnic group, or language. As to which value system to employ for this work package, we refer to the methodology section in 3.3. A range of datasets are available, such as the World Values Survey (Inglehart, 2004), research by Hofstede (2011), but also the cultural dimensions based on Schwartz’ values framework (Schwartz, 2012; Schwartz et al., 2012). Based on these, cross-cultural research has validated dimensions that appear common between these surveys and are consistent between countries. We make use of a two-factor dimensional system by Minkov and Kaasa, centring around two factors with two dimensions each: the *individualism-collectivism* dimensions (IDV-COLL) and the *flexibility-monumentalism* dimensions (FLX-MOM). Figure 14 positions the cultures that are included in the original dataset on both scales, with several highlighted.

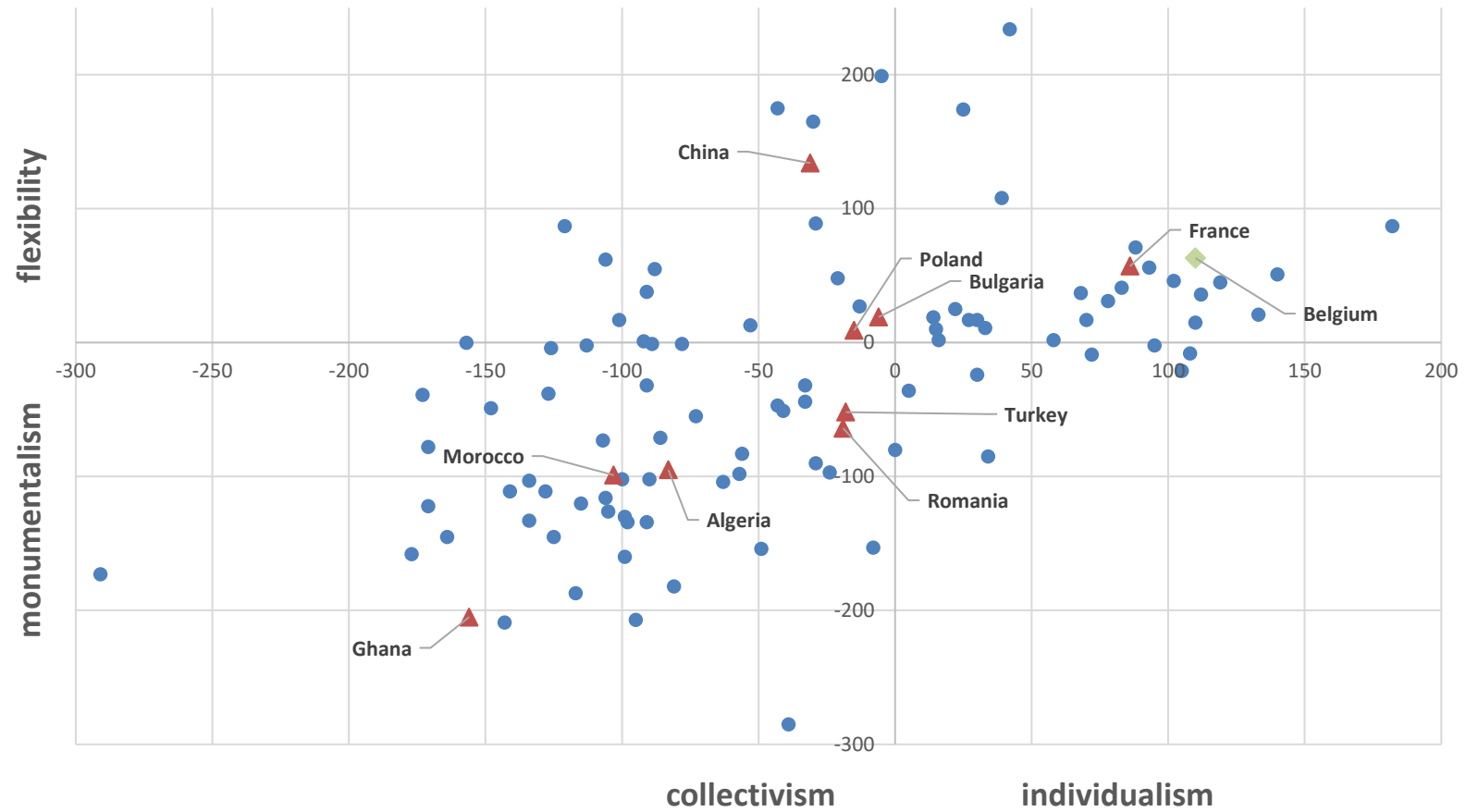


Figure 14 Country positions on Minkov and Kaasa's cultural dimension scales

4.3.2 Ethnic differences in public service motivations

First, we explore the bivariate relationship between PSM scores and ethnic groups. As a first step, we look at three groups: individuals with native parents respective to the country of measurement, with mixed parents (one parent is native respective to the country of measurement, the other is not), and two non-native parents. The mean PSM scores (operationalized by the questions “How important is a job that allows someone to help other people?”; “How important is a job that is useful to society?” 1 = very important, 5 not important at all) are highest for those with mixed parents: 2.04 and 2.01 on a 1 to 5 Likert scale (SD 0.894 and SD 0.901). Next is the native group, with mean PSM scores for the two questions respectively of 2.02 and 1.98 (SD 0.872 and 0.868). Finally, those with two non-native parents score have the lowest mean: 1.95 and 1.91 (SD 0.862 and SD 0.853).

Next, we examine the correlation between the PSM scores and the Minkov cultural dimensions. Before this examination, however, we test whether the two scales of the Minkov dimensions ‘overlap’ statistically.¹⁰ A multicollinearity test of the Minkov scales shows a correlation of 0.565, staying well below levels that would prove cause for concern (0.8).

As discussed in the methodology section, from hereon we use a dichotomized version of the PSM scores (“very important” vs. the other replies – see Figure 15 for a visual representation of its skewedness towards important). Although having less statistical power, we use Spearman's Rho over Pearson's r , prioritizing validity over statistical power. Whereas the Pearson correlation calculates whether the data fit a linear form well and requires both variables to be quantitative and approximately normally distributed, the Spearman correlation does not have any strict requirements regarding the shape of the data. Instead, it “ranks” the values of both the independent and dependent variables separately, and models the consistency of the direction between both (Gravetter & Wallnau, 2012).

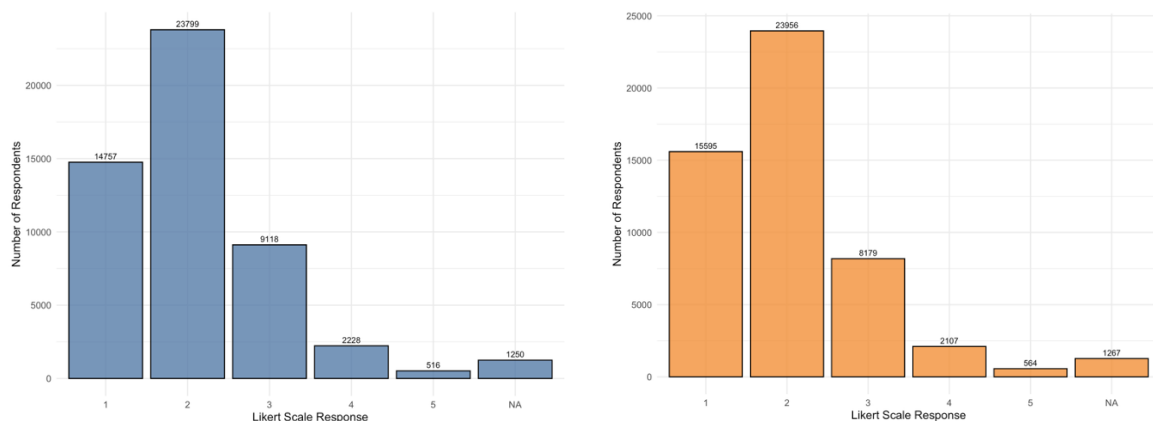


Figure 15 Absolute distributions for the questions “How important is a job that allows someone to help other people?” (left) and “How important is a job that is useful to society?” (right) (1 = very important, 2 = important, ... 5 = not important at all)

We calculated these correlation coefficients first on the Belgian observations (Table 4 Spearman's Rho correlations: PSM items x IND-COLL/ FLX-MOM scales on the Belgian sub-dataset (ethnic groups with $n > 10$ observations (13)) Table 4). For the question focusing on helping others, we find no significant correlation with both the IND-COLL scale, nor the FLX-MOM scale ($r_s -0.429$, $p = 0.188$; $r_s -0.388$, $p = 0.31$). For the question probing how important a job's usefulness to society is, there

¹⁰ As these scales have been empirically validated in prior research, this is likely not a cause for concern.

is a significant correlation with both scales (IND-COLL: $r_s = -0.782$, $p = 0.007$; $r_s = -0.691$, $p = 0.023$). This suggests that as cultures become more individualistic and flexible, having a job that is useful to society becomes *less* important.

Table 4 Spearman's Rho correlations: PSM items x IND-COLL/ FLX-MOM scales on the Belgian sub-dataset (ethnic groups with $n > 10$ observations (13))

	Spearman's Rho (ρ)	<i>p-value</i>
v8 vs. Individualism-Collectivism	-0.429	0.188
v9 vs. Individualism-Collectivism	-0.782	0.007
v8 vs. Flexibility-Monumentalism	-0.338	0.309
v9 vs. Flexibility-Monumentalism	-0.691	0.023

Next, we estimate cross-classified multilevel models to properly control for other potential confounding variables such as gender or education level of the respondent. Here, we consider the entire dataset rather than just the Belgian cases for several reasons. First, following the literature review, we can expect an effect of cultural values related both to the country of observation, as well as related to the ethnic group. Estimating these effects requires the analyses to look 'beyond' just Belgium. In addition, considering the multiple levels of the variables under investigation (cultural values are at the macro-level, and PSM is measured at the individual-level), a multilevel model is required to correctly account for this. This, however, requires sufficient observations at both the higher level as well as the individual-level. Limiting the analysis to the Belgian data subset results in ethnic groups with very few observations as only 13 ethnic groups have more than 10 observations.

Table 5 contains the results of the relevant models. Compared to the null-model, the AIC and BIC drop substantially, indicating a much stronger fit with the added predictors. After accounting for individual characteristics and the cultural context of the country of observation, the culture of an individual's country of origin – the proxy used for the cultural values of the ethnic group - continues to have a statistically significant, independent effect on their motivation to have a job that helps other people, and that is useful to society. Specifically, for every one-standard-deviation¹¹ increase in the Individualism score of the culture associated with the ethnic background, the likelihood of valuing a job that "helps other people" drops by 16.2% (OR 0.838), and the likelihood of valuing a job that is "useful to society" drops by 16.3% (OR 0.837). In sum, as a culture becomes more individualistic, people from that background are less likely to consider prosocial job aspects related to PSM as "very important".

Conversely, for the context of the country of observation, flexibility/monumentalism has a significant effect. After accounting for individual characteristics and the cultural context of the ethnic group, the culture of the country an individual resides in has an independent effect on their motivation to be useful to society. Specifically, for every one-standard-deviation increase in the Flexibility score of the country, the odds of an individual rating 'being useful to society' as 'very important' decrease by 43.1% and 33.9% (OR 0.569 and 0.661 for *helping others* and *being useful to society* respectively).

The level of formal education an individual has achieved too appears to be a consistent predictor of PSM. The models show a step-wise positive effect: as educational attainment increases, so does the likelihood of valuing work that helps others and is useful to society. For instance, compared

¹¹ For the statistical analysis, the cultural scores were standardized, complicating interpretation. Referring back to Figure 14, Morocco, for instance, scores roughly negative one standard deviation.

to individuals with no formal education, those with an upper secondary education are 47.5% more likely to rate helping others as "very important" (OR 1.475) and 72.4% more likely to prioritize being useful to society (OR 1.724). This effect becomes even more pronounced at the highest levels of education, with individuals holding an upper-level tertiary degree being 60% more likely to value helping others (OR 1.6) and more than twice as likely (a 132.8% increase) to value being useful to society (OR 2.328). This suggests that higher education may cultivate or attract individuals with a stronger motivation for public service.

Gender emerges as a significant predictor for both PSM indices: After controlling for all other variables, men are consistently less likely than women to rate these job aspects as "very important." Specifically, the odds of men valuing a job that "helps other people" are 25.9% lower than for women (OR 0.741). Similarly, the odds of men prioritizing a job that is "useful to society" are 15.9% lower than for women (OR 0.841), suggesting that public service motivations are more pronounced among women.

Finally, working in the public sector is strongly associated with valuing prosocial aspects of a job. Compared to their private sector counterparts, public sector employees are 32.6% more likely to consider a job that helps other people as "very important" (OR 1.326). The effect is even stronger for the importance of having a job useful to society, with public employees being 37.2% more likely to prioritize this aspect of their work (OR 1.372).

The findings for the two contexts (country of observation and ethnic origin country) suggest two potential effects. First, as the country of observation (*i.e.*, *the country of residence*) has a strong and significant effect on PSM, a socializing effect may be at play. Individuals living in more flexible (less monumental) societies are less likely to consider being useful to society to be a top priority in employment – even accounting for individual differences. This finding suggests a process of acculturation (Berry et al., 1997) or assimilation (Gordon, 1964). Specifically, residence in a nation with a more "flexible" cultural orientation—which prioritizes pragmatism and situational adaptation over the absolute principles often found in "monumentalist" cultures—is correlated with a lower likelihood of prioritizing prosocial motivations in one's career.

However, this does not necessarily imply a complete assimilation or rejection of one's origin culture; rather, the persistence of the ethnic origin effect alongside the host country effect suggests a complex interplay, potentially pointing towards a form of bicultural value adoption. Even when living in a cultural context that differs in values from those of one's ancestors, the emphasis on the collective versus the self that is learned from this background appears to have enduring effects, shaping this specific, society-focused aspect of their work motivation. This effect is independent of the acculturation pressures from the country of observation.

In sum, the hypothesis that ethnic minority underrepresentation in the public sector stems from a divergence in cultural values **finds little support in this analysis**. On the contrary, the evidence suggests an opposite tendency: on average, individuals from the underrepresented groups studied here possess cultural backgrounds often associated with higher public service motivation. Therefore, **ceteris paribus**, public sector employment should hold a greater, not lesser, attraction for them compared to Belgian ethnic majority members. This shifts the explanatory focus away from a 'cultural deficit' model and strongly suggests that other factors are more likely drivers of their underrepresentation.

Table 5 Cross-Classified Multilevel Logistic Regression Models on Public Service Motivation

	<i>"How important is a job that allows someone to help other people?"</i>			<i>"How important is a job that is useful to society?"</i>		
	null model	full model		null model	full model	
		OR	[95% CI]		OR	[95% CI]
Fixed effect estimates						
(Intercept)		0.271*	[0.217, 0.339]		0.235*	[0.184, 0.299]
age (scaled)		0.931*	[0.905, 0.958]		0.985	[0.958, 1.014]
DEGREE - Primary school (elementary education)		1.336	[1.112, 1.606]		1.491*	[1.242, 1.79]
DEGREE - Lower secondary		1.452*	[1.229, 1.715]		1.602*	[1.357, 1.891]
DEGREE - Upper secondary		1.475*	[1.25, 1.742]		1.724*	[1.461, 2.034]
DEGREE - Post secondary, non-tertiary		1.577*	[1.327, 1.875]		1.812*	[1.525, 2.153]
DEGREE - Lower level tertiary, first stage		1.588*	[1.340, 1.882]		1.935*	[1.634, 2.291]
DEGREE - Upper level tertiary		1.6*	[1.341, 1.908]		2.328*	[1.954, 2.774]
TYPORG2 - Public employer		1.326*	[1.252, 1.405]		1.372*	[1.296, 1.453]
gender - male		0.741*	[0.704, 0.78]		0.841*	[0.800, 0.885]
ind_col_country		1.004	[0.826, 1.221]		0.917	[0.736, 1.143]
flex_mom_country		0.569*	[0.457, 0.707]		0.661*	[0.519, 0.843]
ind_col_ethnic group		0.838*	[0.764, 0.919]		0.837*	[0.755, 0.927]
ethn_flex_ethnic group		1.041	[0.916, 1.183]		0.924	[0.804, 1.061]
Random effects (standard deviation)						
ethnic group	0.213		0.097	0.298		0.167
country of observation	0.662		0.429	0.644		0.489
Model fit						
AIC	47191		36648.3	47978.0		37407.1
BIC	47217		36782.9	48003.9		37541.6
N (individuals)	41,660		41,660	41,660		41,660
N (ethnic groups)	157		157	155		155
N (countries)	33		33	33		33

Note: DEGREE: ref. = "No formal education"; TYPEORG2: ref. = private sector; gender: ref. = women; *p < .05; p < .01, *p<.001

4.4 Ethnic differences in the perceptions of signals in job advertisements for the public sector

4.4.1 Introduction and literature findings

Recruitment involves a whole toolkit of strategies for attracting suitable applicants (Griffeth et al. 2014). Job postings, for instance, continue to be a cornerstone of this activity, serving as a primary channel to connect with prospective hires by outlining the nature of the organization and the position itself. As a form of corporate messaging, these ads inevitably help to shape the company's public image. That image, in turn, heavily sways the outcome of the advertisement, influencing who applies and in what numbers (G. Ryan et al., 2000). This makes job advertisements a key method to attract the 'right' kind of candidates for organizations.

As suggested by signalling theory, job advertisements are not solely neutral messages but contain various signals (Celani & Singh, 2011). These signals enable job seekers to create an image of the organization and the position advertised, as well as allow them to evaluate to what extent their own profile and values are in alignment (fit). Job seekers are affected to how they perceive their fit within an organization, based on a similarity of values or other characteristics (Cable & Judge, 1996). Intrinsic, extrinsic, or prosocial job attribute signals are commonly discussed as having an impact on candidate decision making (Asseburg et al., 2020; Vogel et al., 2024). These signals align with established theoretical frameworks, including self-determination theory in a broader sense, and public service and prosocial motivation specifically within public sector literature (Esteve et al., 2016; Mergel et al., 2021). Job attributes play a crucial role in increasing the attractiveness of a job position to prospective employees (Rynes & Barber, 1990). Alongside organizational attractiveness and reputation, they are a key factor in the formation of applicants' job pursuit intentions (Lievens & Highhouse, 2003; Ployhart & Kim, 2013). Research suggests that job seekers are more likely to respond positively to employment attributes that corresponds with their values and motivations. As such, job attributes signalling congruency with such motivations can help the candidate assess their fit with the job and positively affect job attractiveness when they trigger congruent basic motivations of the applicants.

In public sector recruitment literature, one of the go-to frameworks for these motivations has been public service motivation (PSM – see also WP2.1) (Jakobsen et al., 2023; Korac et al., 2019). Whereas empirical research has validated some of its claims, increasingly calls are made to investigate multi-incentive settings and motivations for public sector employment – including intrinsic and extrinsic attributes (Asseburg et al., 2020; Perry & Vandenabeele, 2015) – as job applicants may be motivated by a range of other motivations as well.

Intrinsic and extrinsic attributes, correspond to two broad categories of motivation explaining human behavior captured by the framework of self-determination theory (Deci et al., 1989; Gagné & Deci, 2005). While primarily a psychological needs-based framework, self-determination theory distinguishes intrinsic from extrinsic motivations. Intrinsic motivation points to motivation due to some inherent satisfaction, interest, or enjoyment. In other words, the motivation to do something is in the activity itself rather than “because of external prods, pressures, or rewards” (R. M. Ryan & Deci, 2000). Applied to the context of employment, this translates into being motivated to a job because of the inherent satisfaction of the work itself. Individuals motivated this way may be less triggered by signals on salary or benefits but are rather attracted by the work itself. On the other hand, extrinsic motivation signifies doing something “in order to attain some separable outcome” (R. M. Ryan & Deci, 2000). Most – if not all - employment relations rely to some extent on this type of motivation, as indicated by there being tangible incentives such as salary (increases), bonus systems, social prestige, as well as the

prospect of furthering one's career (Asseburg et al., 2020). In other words, the reward for an activity - like a job - is separate from the action itself.

Empirical research suggests mixed findings on the relationship between intrinsic and extrinsic motivations, and public sector job pursuit. In a comparative study covering 26 countries, Van de Walle, Steijn, and Jilke (2015) establish how extrinsic rather than intrinsic motivation is correlated to people's intention to work in the public sector. Similarly, Lewis and Frank (2002) find that job security – an extrinsic motivation – is the primary driver for working governmental jobs. In a study on Korean college students, Lee and Choi (2016) find evidence that job security is more important than PSM and prosocial behavior in explaining drivers for public sector job pursuit intentions. Finally, Ritz and Waldner (2011) show that for public sector attraction, job prospects and security, as well as high salary are the main drivers.

Using British longitudinal data, Georgellis, Iossa, and Tabvuma (2011), however, reach the opposite conclusion, finding that public sector attraction is explained by intrinsic rather than extrinsic rewards. Park and Word (2012) too stress the importance of intrinsic rather than extrinsic motivation as a key driver in the public sector labour market. A South Korean study (J. Kim, 2018) finds that intrinsically motivated public servants were less likely to leave their organizations, and experience lower levels of burnouts compared to extrinsically motivated ones. Vogel and colleagues pose that due to the ubiquity of extrinsic motivational signals in job advertisements in the labour market, the effect of extrinsic motivation – although important in experiments – may be overestimated (Vogel et al., 2024). These and similar studies suggest the importance of considering both intrinsic and extrinsic motivation of job candidates for public sector recruitment.

Finally, alongside intrinsic and extrinsic motivations are prosocial motivations, a form of other-regarding motivations (Mergel et al., 2021). Several empirical studies support claims of prosocial drivers for public sector job pursuits. Vogel and colleagues (2024), in an ambitious attempt to combine experimental data with big data quantitative text analysis, find little effect of intrinsic and extrinsic incentives in real-world data but argue that prosocial signals can attract more qualified candidates, arguable because prosocial signals align with what candidates assume public sector employment entails (Leisink & Steijn, 2008).

Aside from job attribute signals, recruitment and selection related signals too can affect perceived organizational and job attractiveness in the eyes of prospective candidates. Some of these can be of relevance for public sector recruitment which is typically characterized by rather inflexible procedural requirements and a strong focus on merit-based hiring. Set in a bureaucratic context, this has led to typically high levels of formalization and standardization in the recruitment procedures (Chen & Rainey, 2014). Signals of formalized recruitment in the public sector have been linked to lower application intentions (Sievert et al., 2022), suggesting potential negative perceptions of this particular characteristic in public sector job advertisements. On the other hand, formalized recruitment may also signal stronger adherence to merit and process-based recruitment, potentially increasing the perception of overall recruitment fairness and procedural justice.

Formalization and standardization can be considered a positive as well as a negative mechanism. Chen and Rainey (2014) argue in favour of a positive perspective, illustrating how it can aid organizations in attracting the right talent as well as offer a means of protection such as worker safety and procedural justice. This perspective, however, can be juxtaposed with one in which formalization takes the form of administrative burden. Closely related to (and sometimes caused by) formalization, administrative burden is an "individual's experience of policy implementation as onerous" (Burden et al., 2012) such as submitting documentation unnecessary for private sector recruitment (e.g.,

educational or language certificates) or procedural dissatisfaction regarding the recruitment process (Linos & Riesch, 2020). Such requirements or additional steps can lead to additional burden for applicants, leading to increased learning, psychological, and compliance costs for the candidate (Moynihan et al., 2015).

Finally, signals pertaining diversity climate potentially may affect prospective candidates for public sector positions. Diversity climate refers to “employees’ behavior and attitudes that are grounded in perceptions of the organizational context related to [...] minorities” (Mor Barak et al., 1998) p 83. More specifically, it concerns perceptions of the extent to which employees with a minority status are incorporated into the workplace (McKay et al., 2007). Scholarship covering this topic emerged roughly thirty years ago (Ward-Bartlett et al., 2023) with the interactional model of cultural diversity (IMCD), theorizing how employee attitudes and performance were affected by an organization’s diversity climate (Cox, 1994). Since then, research has identified two key contributions of diversity climate, namely promoting equal employment opportunities by reducing or eliminating discrimination as well as utilizing workforce diversity to ensure a competitive advantage (Hatter et al., 2024) and has been shown to be positively related with job-related outcomes (Hofhuis et al., 2012).

From the perspective of Social Identity Theory (Tajfel and Turner, 2004 (1986)), diversity climate signals may serve as contextual cues that confirm an organization’s commitment to the value of diverse social identities (Yeung, 2019). *Identity affirmation*, or the perception that because of the value these organizations put on diversity, the job seeker’s identities are likewise confirmed as well, may explain how such signals may be attractive for ethnic minorities. The interactional model of cultural diversity posits how employees’ responses to diversity are formed by dynamic interactions between an individual’s lived experience as well as structural elements. The model argues how diversity climate can influence both affective outcomes such as job satisfaction and organizational commitment, as well as on the job achievements through an interaction across the individual, the (inter)group, and the organisational level (Cox, 1994). Moreover, belonging uncertainty (Walton & Cohen, 2007) may explain why ethnic minorities have lower job-pursuit intentions for organizations that value diversity to a lesser extent. This uncertainty about one’s social connections particularly affects minorities due to their perceived underrepresentation may affect their motivation to apply even without direct discrimination or prejudice. Highlighting the importance of social belonging.

In the context of the public sector, the particulars of diversity can be at odds with each other. The strictly merit based nature of bureaucracy would argue in favour of a colourblind-approach, emphasizing equal treatment while ignoring group differences. However, this approach tends to be perceived more beneficial by ethnic majority members (Jansen et al., 2016) and has been shown to increase implicit bias among majority members, potentially mask discrimination, as well as induce pressure to conform to the dominant norms (Apfelbaum et al., 2012). On the other hand, multiculturalism, which emphasizes recognition and celebrates group differences, is related to higher minority engagement and perceptions of safety (Plaut et al., 2009; Russell Pascual et al., 2024) although may come at a cost of perceived tokenism among minority candidates (Leibbrandt & List, 2018; Nichols et al., 2023), as well as perceived cues of exclusion for majority members (Ballinger & Crocker, 2021).

4.4.2 Ethnic differences in signal perception

Due to the efficiency of within-subject survey experiments, 3,780 observations are recorded with a limited sample of 756 respondents. First, we specify our main models per outcome. These models clarify the overall effect of predictor variables on the outcomes. The predictor variables are ethnic

origin, and the three signal variables. The models also include level 2 control variables such as age, gender, education level, PSM (Public Service Motivation) score, preference of employment (public or private), and whether the respondent has family employed in the public sector. In all three models, lower education level has a significant negative effect, suggesting that compared to respondents with a higher education, those with a lower education background on average are less attracted to both organisation and job, and are less inclined to pursue the job. PSM has a significant positive effect in all three models, suggesting that higher PSM is associated with higher degrees of organisational and job attractiveness, as well as job pursuit intentions. On average, women are significantly less inclined to pursue the job compared to men, although organisational and job attractiveness are not significant.

Next, we add interactions between each ethnic origin and the three signals in order to estimate whether the signals have an outcome significantly different between the origin group and the reference category Table 6. Significant findings for a particular origin group suggest the effect of the signals in the job ads statistically differs from those of the reference category (Western-Europe). For *organisational attractiveness*, the interaction of prosocial job attributes and the Sub-Sahara African origin is marginally significant ($p = .066$), as well as the interaction of the signal colour-blindness (diversity climate) and the North African group ($p = .072$). For the outcome *job attractiveness*, the interaction of prosocial job attributes and the Sub-Sahara African origin is significant ($p = .036$), as well as the interaction of no standardised recruitment processes and the North African origin group ($p = .043$). Finally, the interaction model for *job pursuit intentions* suggests that only the interaction with the North African origin group and the signal colour-blindness is marginally significant ($p = .082$).

Finally, we ‘probe’ the interactions by means of a simple effects analysis (Hayes & Little, 2018), with conservative post-hoc significance tests (Tukey's HSD Test and Bonferroni) to test the effect sizes within the origin groups (Table 7). *Organisational attractiveness* remains marginally significant only for the interaction of diversity climate and the North African group. The estimate suggests that candidates of North African origin, confronted with a job ad with a signal of multiculturalism, on average, rated the attractiveness of the organisation .406 point higher compared to job ads containing signals of colour-blindness.

For the second outcome variable, *job attractiveness*, both Bonferroni and Tukey post-hoc tests suggest a marginally significant effect for the interactions discussed above. Specifically, for the Sub-Saharan African group, job attractiveness is rated marginally significantly higher when confronted with intrinsic job signals than with prosocial ones in job ads (.86 points). In addition, the North African group again rate job ads with multicultural signals .406 point higher compared to those containing signals of a colourblind diversity climate.

Finally, the outcome variable *job pursuit intentions* suggest a similar picture of marginally significant effects for both Bonferroni and Tukey tests. Candidates with a Sub-Saharan African background rate job ads with intrinsic job attribute signals .86 points higher compared to those with prosocial job attribute signals, whereas the North African group rate job ads with multicultural signals .41 point higher compared to those containing signals of a colourblind diversity climate.

Table 6 Linear Mixed-Effects Models Predicting 1) organisational attractiveness, 2) job attractiveness, 3) job pursuit intentions

		Model 1: Organizational Attractiveness		Model 2: Job Attractiveness		Model 3: Job Pursuit Intentions	
		Main effects model 1a	Interaction model 1b	Main effects model 2a	Interaction model 2b	Main effects model 2a	Interaction model 2b
Intercept (constant)		1.653*** (0.146)	1.649*** (0.146)	1.680*** (0.148)	1.676*** (0.148)	1.716*** (0.157)	1.714*** (0.157)
Main effects	Job Attribute (ref=intrinsic)						
	Extrinsic	-0.038 (0.036)	-0.031 (0.038)	-0.029 (0.037)	-0.019 (0.039)	-0.046 (0.037)	-0.042 (0.039)
	Prosocial	-0.011 (0.036)	-0.009 (0.037)	-0.005 (0.037)	-0.010 (0.038)	-0.004 (0.037)	-0.008 (0.039)
	Standardization (ref=standardized)						
	No standardization	0.010 (0.030)	-0.003 (0.031)	0.019 (0.031)	0.003 (0.032)	0.003 (0.031)	-0.005 (0.033)
	Diversity Climate (ref=multicultural)						
	Colorblind	-0.010 (0.030)	-0.005 (0.031)	-0.015 (0.031)	-0.008 (0.033)	-0.013 (0.031)	-0.011 (0.033)
	Origin group (ref=Western Europe)						
	North Africa	0.082 (0.193)	0.241 (0.288)	0.199 (0.196)	0.130 (0.296)	0.072 (0.208)	0.039 (0.304)
	Sub-Saharan Africa	0.241 (0.245)	0.588† (0.354)	0.158 (0.248)	0.534 (0.363)		
	Eastern and South Eastern Europe					0.279 (0.178)	0.439† (0.251)
	Missing	0.051 (0.228)	-0.239 (0.317)	0.229 (0.231)	-0.085 (0.324)	0.182 (0.245)	-0.071 (0.335)
Interaction effects	Sub-Saharan Africa x Prosocial		-0.721† (0.392)		-0.852* (0.406)		
	North Africa x No standardization				0.441* (0.218)		
	North Africa x Colorblind		-0.402† (0.223)				-0.404† (0.232)
	Missing x No standardization		0.466† (0.265)		0.497† (0.273)		
	Missing x Colorblind		0.424† (0.238)		0.412† (0.246)		0.508* (0.247)
Controls	Gender (ref = male)						
	female	-0.094† (0.054)	-0.094† (0.054)	-0.096† (0.055)	-0.097† (0.055)	-0.116* (0.058)	-0.117* (0.058)
	non-binary	-1.014† (0.524)	-0.995† (0.525)	-0.915† (0.531)	-0.859 (0.533)	-0.967† (0.564)	-0.950† (0.565)
	education (ref = higher educated)						
	lower educated	-0.171** (0.055)	-0.172** (0.055)	-0.149** (0.055)	-0.149** (0.055)	-0.174** (0.059)	-0.173** (0.059)
	PSM	0.442*** (0.032)	0.443*** (0.032)	0.413*** (0.033)	0.415*** (0.033)	0.415*** (0.035)	0.416*** (0.035)
	sector preference (ref = public sector)						
	private sector	-0.064 (0.074)	-0.063 (0.074)	-0.035 (0.075)	-0.033 (0.075)	-0.059 (0.080)	-0.058 (0.080)
	no preference	-0.189** (0.067)	-0.189** (0.067)	-0.190** (0.068)	-0.187** (0.068)	-0.194** (0.072)	-0.193** (0.072)
Model Fit	Marginal R ²	0.135	0.138	0.114	0.119	0.112	0.116
	N (Observations)	3730	3730	3730	3730	3730	3730
	N (Respondents)	746	746	746	746	746	746

Note: Unstandardized coefficients from linear mixed-effects models. Standard errors are in parentheses. Non-significant variables are not included in the table, but controlled for in the models.
† $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed tests).

Table 7 Simple Effects of signals on 1) organisational attractiveness, 2) job attractiveness, 3) job pursuit intentions, by origin region

	Origin group	Signal comparison	Estimate	SE	Z-Ratio	P-value (Tukey)	P-value (Bonferroni)
1. Organizational Attractiveness	Sub-Saharan Africa	intrinsic - extrinsic	0.437	0.307	1.425	0.328	0.463
	Sub-Saharan Africa	intrinsic - prosocial	0.730	0.390	1.870	0.147	0.184
	Sub-Saharan Africa	extrinsic - prosocial	0.293	0.411	0.711	0.757	1.000
	North Africa	multicultural - colorblind	0.406	0.221	1.837	0.066	0.066
2. Job Attractiveness	Sub-Saharan Africa	intrinsic - extrinsic	0.442	0.317	1.391	0.346	0.493
	Sub-Saharan Africa	intrinsic - prosocial	0.863	0.404	2.136	0.083	0.098
	Sub-Saharan Africa	extrinsic - prosocial	0.421	0.426	0.989	0.584	0.968
	North Africa	multicultural - colorblind	0.406	0.221	1.837	0.066	0.066
3. Job Pursuit Intentions	Sub-Saharan Africa	intrinsic - extrinsic	0.442	0.317	1.391	0.346	0.493
	Sub-Saharan Africa	intrinsic - prosocial	0.863	0.404	2.136	0.083	0.098
	Sub-Saharan Africa	extrinsic - prosocial	0.421	0.426	0.989	0.584	0.968
	North Africa	multicultural - colorblind	0.415	0.230	1.804	0.071	0.071

4.5 Actual fairness of selection procedures (*inflow*)

4.5.1. Literature findings

First, we reviewed literature on the fairness of selection tests administered for federal selection, namely, the situational judgment test, work sample tests (like the file test and the in-basket test), and cognitive ability tests (like the abstract, numeric and verbal ability tests).

Situational Judgement Test. In situational judgement tests (SJT's), candidates are presented with relevant work situations and several possible reactions to the work situation. Candidates must indicate how appropriate the reactions are and/or choose how they would react. Meta-analytic results showed differences in scores between ethnic minority and majority individuals, as illustrated by a moderate d of .38 (Whetzel et al., 2008). However, recently, Herde et al. (2020) found that SJT's had much larger subgroup score differences ($d = .66$) than initially expected, since earlier research often focused on job incumbents instead of actual applicants. Generally, SJT's still generate moderate subgroup differences and have lower adverse impact compared to cognitive ability tests ($d = .72$ for moderately complex jobs; Cabrera & Nguyen, 2001; Pulakos & Schmitt, 1996; Weekley & Jones, 1999). Moreover, video-based SJT's and game-based SJT's further decrease adverse impact compared to paper-and-pencil SJT's (Chan & Schmitt, 1997; Leutner et al., 2021) and the method of scoring (De Leng et al., 2018; McDaniel et al., 2011), as well as the response format (Arthur Jr. et al., 2014) can also influence adverse impact. A recent meta-analysis of Sackett et al. (2023) shows that SJTs have moderate subgroup differences and that these differences depended on the type of SJT. Specifically, knowledge-based SJTs showed somewhat larger Black-White mean differences ($d = 0.39$) compared to behaviourally-based SJTs ($d = 0.34$). In terms of the validity: recent results show that SJT performance correlated positively with job performance's ($r = .26$, both knowledge-based and behaviourally-based SJTs), which is comparable to the predictive value of conscientiousness and interest tests, but lower than that of structured interviews, work sample tests, and cognitive ability tests. In sum, SJTs offer a reasonable balance between predictive validity and adverse impact: Whereas SJTs are less valid than previously thought, the moderate adverse impact makes them an interesting assessment tool.

In the federal government, SJT's are used on all organisational levels, with text-based SJT's on levels A, B and C and video-based SJT's on level D. Constructs measured with the SJT are the 'key competences': teamwork, service-orientation, reliability, self-development and achieving objectives, complemented by certain specific competences for each level (for example: managing people on level A and B, but not on C and D). It is rather unclear whether SJTs at BOSA are behaviourally-based (i.e., how one would react in a situation) rather than knowledge-based (i.e., what the best reactions are in a given situation according to someone) given that the response instructions are not clearly knowledge-based nor behaviourally-based (i.e., *Wat is je oordeel over deze reactie?*).

Work Sample Tests. Work sample tests are tests "in which the applicant performs a selected set of actual tasks that are physically and/or psychologically similar to those performed on the job" (Ployhart et al., 2005, p. 538). Examples of work sample tests are in-basket tests, role play exercises or technical exercises (e.g., electricity repair when for the job of electrician). In their meta-analysis, Roth et al. (2008) found that work samples tests were not as 'resistant' to adverse impact as initially thought. That is, up until then, work sample tests were known to have moderate differences in scores between ethnic majority vs. minority groups ($d = .52$), yet smaller when compared to, for instance, those of cognitive ability tests ($d = .72$ for moderately complex jobs; Roth et al., 2001). Roth et al. (2008)'s meta-analysis showed that this was only the case for specific work sample tests such as role-plays. In-basket and technical work sample tests were found to convey much larger group-differences ($d = .74$), more

comparable to those of cognitive ability tests. Indeed, Bobko & Roth (2013) also found a large subgroup difference of .74 in their review when investigating a sample of actual applicants instead of job incumbents. Finally, the construct measured within work samples also affects group differences such that work samples measuring cognitive skills and job knowledge resulted in larger subgroup differences ($d = .80$), while work samples measuring social skills (e.g., communication and interpersonal skills) resulted in lower adverse impact (small d 's ranging from .23 until .27). In terms of predictive validity, correlations between work sample scores and job performance were recently found to be .33 (Sackett et al., 2023), which is higher than the recent corrective correlations for cognitive ability ($r = .23$).

The federal government uses two work sample tests, namely an *in-basket test* and *file test*. First, the *in-basket tests* are administered at all organisational levels and measure the following competences: Information integration and decision making (level A); Information analysis and problem solving (levels B and C); Handling information and structuring work (level D). Candidates are presented with different emails (with information) that they receive in their inbox and need to start working on (e.g., structuring and prioritizing). Secondly, the federal government's *file test* (*test dossierbehandeling/test de traitement de dossiers*) on levels, A, B and C is also an example of a work sample test. The file test is similar to the in-basket test, but instead of a series of emails, candidates are now presented with different documents (or files) to prioritize and follow-up (see also Table 2). Competences of interest are information integration and decision making (level A) and information analysis and problem solving (levels B and C). Finally, the federal government also administers an *administrative test* (levels B, only), which is a technical test, measuring candidates' ability to review and organise administrative data. Specifically, candidates must review columns of data and indicate the number of errors in these columns.

Cognitive ability tests. Organisations often include cognitive ability tests in their selection procedures to measure candidates' abilities in terms of abstract, numerical, verbal and logical reasoning. Cognitive ability tests are generally known to generate large group-differences ($d = .72$ for moderately complex jobs; (P. L. Roth et al., 2001) Roth et al., 2001) and hence risk adverse impact for ethnic minority groups (McFarland & Kim, 2021; Ployhart & Holtz, 2008; Pyburn Jr et al., 2008; Robie et al., 2017; P. L. Roth et al., 2001; Woods & Patterson, 2024). Woods and Patterson (2024) note that these findings hold true in other contexts than the USA alone, for example, in Europe (Guenole et al., 2003; Lucey & Saguil, 2020; McManus et al., 2013; Stegers-Jager et al., 2015). Cognitive ability test were long seen as one of the strongest predictors of job performance (Schmidt & Hunter, 1998; Schmitt, 2014), making the 'validity-diversity dilemma' of this selection tool all the more complex and difficult for organisations. However, in their recent, critical review, Woods and Patterson (2024) argue that the use of cognitive testing is no longer as justified as originally assumed, based on evidence that their validity has been overrated by earlier studies. Indeed, recent review studies (Sackett, 2021; Sackett et al., 2023) found that while earlier cognitive ability – job performance correlations were around .51, corrected correlations are around .31, which is substantially lower than previously reported. The reason for this remarkable difference is that previous meta-analytic studies overcorrected for range restriction, which is unnecessary in cross-sectional designs (as in most validation studies). However, cognitive ability still remains the most predictive predictor of job performance for people without prior experience. When candidates do possess prior experience, the use of work samples (correlation of .33) is advised.

Apart from the above-mentioned studies on general cognitive ability (i.e., general mental ability, including abstract, verbal, numerical and logical aspects of intelligence), studies have also investigated subgroup differences for verbal ability tests in particular and found that verbal ability tests generated the larger subgroup differences compared to non-verbal cognitive ability measures (Vanderpool &

Catano, 2008). However, verbal ability test in particular are often used in selection procedures for civil service jobs (McFarland & Kim, 2021; Sackett et al., 2001), considering the relevance of the language requirement, embedded in legislation (see for instance: *Koninklijk besluit van 18 juli 1966 houdende coördinatie van de wetten op het gebruik van de talen in bestuurszaken/Arête royal du 18 juillet 1966 portant coordination des lois sur l'emploi des langues en matière administrative*). Pulakos and Schmitt (1996) found that selection procedures that include verbal ability testing might complement the procedure with other, non-verbal tests (e.g., SJT, interviews) to decrease adverse impact and increase test validity at the same time. Alternatively, different verbal ability tests can be used that test candidates' verbal skills in an applied and relevant context using different formats, for example, a video-based writing test (Pulakos & Schmitt, 1996).

The federal government uses cognitive ability tests on all organisational levels by means of (a) abstract, (b) numerical, and (c) verbal reasoning tests. These tests measure candidates' (a) ability to derive rules from a set of abstract data and apply them to new information, (b) accuracy and speed of making connections between pieces of numerical information and (c) accuracy and speed of making connections between pieces of verbal information, respectively.

4.5.2. Adverse impact calculations¹²

While previous sections focused on early factors such as job advertisement signals and applicant motivation, subsequent phases, such as the application (*inflow*) stage, also warrant examination. Here, we first investigate the actual fairness of selection procedures (**WP 3.1**) via the adverse impact of commonly used tests and assessments within federal public services against ethnic minority applicants. All tests, except for the numerical reasoning test ($N = 125$), show significant differences in selection rates, with ethnic minorities passing significantly less often than ethnic majorities. These differences were particularly observed among men (not women) and applicants older than 50 years (not younger applicants). Table 8 provides an overview of the details discussed below.

Situational Judgement Test. Selection ratios of ethnic minority ($SR = .44$) and ethnic majority ($SR = .57$) candidates were significantly different from each other, $z = 13.88$, $p < .001$, $\chi^2(1) = 219.79$, $p < .001$ (two-tailed¹³); Fisher's exact test = 220.08 $p < .001$ (two-tailed). The adverse impact ratio of .79 was just below .80, indicating the presence of adverse impact in both a statistical and practical sense, when considering the total sample. The adverse impact ratio of men ($AIR = .76$) was significantly different than that of women ($AIR = .80$), $z = -6.34$, $p < .001$, but no significant differences were found in adverse impact ratio between candidates older ($AIR = .79$) and younger ($AIR = .79$) than 50 years, $z = .23$, $p = .82$.

File test. Selection ratios of ethnic minority ($SR = .43$) and ethnic majority ($SR = .55$) candidates were significantly different from each other, $z = 14.79$, $p < .001$, $\chi^2(1) = 218.86$, $p < .001$, Fisher's exact test = 219.28, $p < .001$. The adverse impact ratio of .78 was smaller than .80, indicating the presence of adverse impact in both a statistical and practical sense for the total sample. The adverse impact ratio of men ($AIR = .75$) was significantly different than that of women ($AIR = .80$), $z = -8.86$, $p < .001$, but no significant differences were found in adverse impact ratio between candidates older ($AIR = .77$) and younger ($AIR = .78$) than 50 years, $z = 1.30$, $p = .194$.

¹² We investigated selection rates and adverse impact ratios taking into consideration the test outcome (pass/fail) instead of test scores' subgroup differences (i.e., d-scores) given that actual test scores were (a) difficult to retrieve and (b) difficult to interpret in a proper way, due to differential norms that were used for different groups of applicants.

¹³ Reported p-values for both the Chi-square test and Fisher's exact test are all two-tailed.

In-basket test. Selection ratios of ethnic minority (SR = .63) and ethnic majority (SR = .73) candidates were significantly different from each other, $z = 3.34$, $p < .001$, $\chi^2(1) = 11.56$, $p = .003$, Fisher's exact test = 11.58 $p = .003$. Yet, the adverse impact ratio of .86 was larger than .80, indicating absence of adverse impact in a practical sense in the total sample. Adverse impact ratios in the gender- and age-subgroups were all larger than .80. When comparing subgroup AIRs, adverse impact ratio of men (AIR = .82) was significantly different than that of women (AIR = 1.00), $z = -6.46$, $p < .001$, but no significant differences were found in adverse impact ratio between candidates older (AIR = .85) and younger (AIR = .85) than 50 years, $z = -0.05$, $p = .960$.

Abstract reasoning test. Selection ratios of ethnic minority (SR = .43) and ethnic majority (SR = .56) candidates were significantly different from each other, $z = 12.51$, $p < .001$, $\chi^2(1) = 156.54$, $p < .001$, Fisher's exact test = 220.08 $p < .001$. The adverse impact ratio of .76 was smaller than .80, indicating the presence of adverse impact in both a statistical and practical sense. The adverse impact ratio of men (AIR = .71) was significantly different than that of women (AIR = .79), $z = -10.24$, $p < .001$, and the adverse impact ratio of candidates older (AIR = .56) and younger (AIR = .78) than 50 years, were significantly different from each other, $z = 15.26$, $p < .001$.

Numerical reasoning test. Selection ratios of ethnic minority (SR = .43) and ethnic majority (SR = .57) candidates were not significantly different from each other, $z = 1.46$, $p = .15$, $\chi^2(1) = 2.64$, $p = .27$, Fisher's exact test = 2.61 $p = .28$, while the adverse impact ratio of .76 was smaller than .80. This indicates absence of adverse impact in a statistical sense, while presence of adverse impact in a practical sense. The adverse impact ratio of men (AIR = .89) was significantly different than that of women (AIR = .63), $z = -3.44$, $p = .001$, and the adverse impact ratio of candidates older (AIR = .39) and younger (AIR = .68) than 50 years, were significantly different from each other, $z = -4.11$, $p < .001$.

Verbal reasoning test. Selection ratios of ethnic minority (SR = .43) and ethnic majority (SR = .70) candidates were significantly different from each other, $z = 10.61$, $p < .001$, $\chi^2(1) = 114.43$, $p < .001$, Fisher's exact test = 111.11 $p < .001$. The adverse impact ratio of .79 was smaller than .80, indicating the presence of adverse impact in both a statistical and practical sense. The adverse impact ratio of men (AIR = .61) was not significantly different than that of women (AIR = .61), $z = .01$, $p = .99$, but the adverse impact ratio of candidates older (AIR = .39) and younger (AIR = .64) than 50 years, were significantly different from each other, $z = 6.29$, $p < .001$.

Administrative test. Selection ratios of ethnic minority (SR = .42) and ethnic majority (SR = .55) candidates were significantly different from each other, $z = 6.84$, $p < .001$, $\chi^2(1) = 51.15$, $p < .001$, Fisher's exact test = 51.22 $p < .001$. The adverse impact ratio of .77 was smaller than .80, indicating the presence of adverse impact in both a statistical and practical sense. The adverse impact ratio of men (AIR = .75) was significantly different than that of women (AIR = .78), $z = -2.04$, $p = .04$, and the adverse impact ratio of candidates older (AIR = .69) and younger (AIR = .77) than 50 years, were significantly different from each other, $z = -4.67$, $p < .001$.

Table 8 Selection Rates and Adverse Impact-ratios according to the Four-Fifth Rule for federal selection tests

Test		Selection Rates (% passed)					AI ratios (4/5th rule)				
		Total	Male	Female	<50y	>50y	Total	Male	Female	<50y	>50y
Situational Judgment Test											
	Ethnic minority	.44	.44	.45	.46	.36					
	Ethnic majority	.57	.57	.56	.58	.46	.79	.76	.80	.79	.79
File Test											
	Ethnic minority	.43	.41	.45	.45	.31					
	Ethnic majority	.55	.54	.56	.57	.40	.78	.75	.80	.78	.77
In-basket Test											
	Ethnic minority	.63	.58	.81	.64	.55					
	Ethnic majority	.73	.71	.81	.75	.65	.86	.82	1.00	.85	.85
Abstract Reasoning											
	Ethnic minority	.43	.42	.43	.45	.21					
	Ethnic majority	.56	.60	.54	.58	.38	.76	.71	.79	.78	.56
Numerical Reasoning											
	Ethnic minority	.43	.54	.33	.40	.29					
	Ethnic majority	.57	.61	.53	.60	.75	.76	.89	.63	.68	.39
Verbal Reasoning											
	Ethnic minority	.43	.43	.42	.46	.20					
	Ethnic majority	.70	.70	.69	.72	.52	.61	.61	.61	.64	.39
Administrative Test											
	Ethnic minority	.42	.40	.44	.45	.26					
	Ethnic majority	.55	.53	.56	.58	.38	.77	.75	.78	.77	.69

4.6 Perceived fairness of selection procedures (*inflow*)

In addition to the actual fairness of selection procedures, *perceived fairness* of commonly used tests and assessments within federal public services (WP 3.2) too can be examined.

4.6.1 Literature findings

Regardless of selection tests' *actual fairness*, candidates might *perceive* selection tests as either fair or not. That is, regardless of whether a selection test is unfair (e.g., has an adverse impact for certain groups), it can be perceived by candidates as either fair or unfair, based on procedural aspects (i.e., *How was I treated during the selection process?*), distributive aspects (i.e., *Do I think the test outcome is fair?*) and discriminatory aspects (i.e., *Did I feel discriminated during this test?*). These perceptions of candidates regarding selection tests have been thoroughly studied and are referred to in the literature as *applicant reactions*. Such perceptions are important as they might affect applicants' test-taking motivation, organisational attractiveness, and intentions to accept job offers (McCarthy et al., 2017).

The aim of the present study is to investigate applicant reactions toward frequently used federal selection tests and compare this to actual fairness (adverse impact) of these selection tests. In 2004, a meta-analysis by (Hausknecht et al., 2004) investigated applicant reactions toward different selection tests and reported that work sample tests were more favored than cognitive ability tests. In 2010, Anderson and colleagues conducted a similar meta-analysis across 17 countries. Results were in line with the earlier findings: Candidates prefer work sample tests to cognitive ability tests. While these consistent results guide our expectations regarding applicant reactions to different federal selection tests, we here discuss some recent findings regarding applicants' reactions to those categories of tests that are also being used for federal selection procedures (e.g., McCarthy et al., 2017).

Situational Judgement Tests and Work Sample Test. Candidates typically react positively to situational tests that reflect work tasks (Van Iddekinge et al., 2023). SJTs and work sample tests (like an in-basket, file test, administrative test) are examples of situational tests who present candidates with high fidelity situations. Indeed, SJT's typically receive positive candidate reactions (Lievens & Motowidlo, 2016). Recent technological innovations, such as video SJT's and Virtual Reality SJT's might further improve applicant reactions (Kanning et al., 2006; Oostrom et al., 2010; Rijdsdijk et al., 2021), due to increased job relatedness or fidelity. Work sample tests include, by definition, relevant work situations, and are also very well-received by candidates (Anderson & Witvliet, 2008; Cascio, 2006; Hoang et al., 2012; P. Roth et al., 2008). McCarthy et al. (2017) further showed that tests can be even more positively received by applicants when more information is offered about the testing to candidates that convey transparency, respect and reassurance.

Cognitive ability tests. Cognitive ability tests are not candidates' favourite type of selection test when compared to situational tests. Face validity evaluations are typically low for this type of test (Anderson et al., 2010). Recent studies mostly investigate how cognitive ability tests can be made more interestingly by means of new technologies such as video's, gamification or game-based assessments (Collmus & Landers, 2019; Leutner et al., 2023; Ohlms et al., 2024, 2024; Simons et al., 2023). However, within the context of the federal selection procedure, these technologies are not (yet) incorporated. Therefore, it is expected that applicant perceptions of cognitive ability tests will be rather low, especially when compared to applicants' reactions towards situational tests (like SJTs).

Ethnic minorities' reactions to tests. Little is known about how applicants' ethnicity might moderate their perceptions of selection procedures (McCarthy et al., 2017). Yet, in times of labour

market shortages, it is crucial to reach and attract as many qualified applicants as possible, including those from (ethnic) minority backgrounds. Hence why applicant reactions matter, also those from ethnic minority groups and why we investigated ethnic differences in applicant reactions in the context of federal selection procedures.

In general, there is little evidence for ethnic subgroup differences in applicant reactions towards situational tests, such as SJTs (Golubovich & Ryan, 2021). This might be explained by lower ethnic subgroup differences in actual test performances. However, when investigating ethnic differences in applicant reactions toward selection tests, cognitive ability tests provide an interesting case, given the large subgroup differences in tests scores that applicants might know about (see also section on ‘actual fairness’) or lower test scores they might have experienced themselves in previous selection procedures. Hence, it can be expected that ethnic minority applicants (vs. ethnic majority applicants) experience (a) higher levels of test anxiety, and (b) lower levels of general fairness/procedural justice, driven by candidates’ causal attributions (e.g., Oostrom & De Soete, 2016). Indeed, according to the Applicant Attribution-Reaction Theory (AART; Ababneh et al., 2014; Oostrom & De Soete, 2016; Ployhart & Harold, 2004), candidates may attribute their prior (actual/perceived) test performances to different causes, which may influence how they feel about the selection procedure. When compared to ethnic majorities, Oostrom and De Soete (2016) showed that ethnic minority applicants had a stronger tendency to attribute their previous test experiences and perceived test performance externally, namely to the fairness of a cognitive ability testing procedure. Ethnic minority applicants, therefore, may be expected to have a more external and uncontrollable attribution style to certain selection tests, for instance, cognitive ability tests, which show considerably higher fail rates for ethnic minorities compared to ethnic majorities, and situational tests (like SJTs, in-basket test, file test, administrative tests). Together with candidates’ perceived performance, this external and uncontrollable attribution style can be expected to further decrease the perceived fairness of these tests. This, in turn, could affect dropout rates and affect future application intentions (Oostrom & De Soete, 2016).

4.6.2 Perceived fairness of selection procedures

Time 1: Immediate Post-Test Reactions

Table 10 presents means for the outcome (or dependent) variables of interest for each type of test separately and Figure 16 through Figure 22 present graphs for the outcome variables of interest but divided for ethnic minority and ethnic majority candidates. Table 9 shows the correlation table for all independent and dependent variable studied (across the 7 tests).

For each type of test, we conducted multivariate analysis of variance (MANOVA) to simultaneously investigate effects of our independent variables (perceived test performance, actual test performance (or test result), rejection sensitivity, self-efficacy), interactions of these variables with ethnicity, and finally, interactions of ethnicity, age and gender on the following outcome variables: *Procedural justice*, *General fairness*, *Perceived test discrimination related to ethnicity/age/gender*, *Test Anxiety*, *Test Motivation*, *Attitude toward Tests and Test Ease*

Situational Judgement Test. The combination of outcome variables was significantly affected by all four IVs, namely: expected performance, actual performance (pass/fail), rejection sensitivity, and self-efficacy, in the expected direction. No significant effect of ethnicity on the outcome variables was found, $\Lambda = .995$, $F(9, 2742) = 1.45$, $p = .16$, based on Willks’ Lambda. In terms of moderating effects with ethnicity, we observed a significant three-way interaction between ethnicity, age and gender on the outcome variables, $\Lambda = .985$, $F(27, 8232) = 1.52$, $p = .036$. Older male ethnic minority candidates

perceived significantly more age discrimination than older, female ethnic minority candidates, $b = .01$, $SE = .00$, $t(2750) = 2.26$, $p = .024$. Older female ethnic minority candidates had higher levels of test anxiety than older, male ethnic minority candidates, $b = -.01$, $SE = .01$, $t(2750) = -2.11$, $p = .035$, when both compared to their younger counterparts

File test. For the file test, results were similar to those of the SJT. That is, expected performance, actual performance (pass/fail), rejection sensitivity, and self-efficacy significantly affected the combination of outcome variables in the expected direction. Again, no significant effect of ethnicity on the outcome variables was found, $\Lambda = .994$, $F(9, 1183) = .77$, $p = .65$. We observed a significant three-way interaction between ethnicity, age and gender on the outcome variables, $\Lambda = .960$, $F(27, 3555) = 1.82$, $p = .006$. Older female ethnic minority candidates perceived significantly lower levels of procedural justice, $b = .01$, $SE = .00$, $t(1191) = 2.97$, $p = .003$, and were significantly less motivated for test taking, $b = .01$, $SE = .00$, $t(1191) = 2.23$, $p = .026$, than older male ethnic minority candidates, when compared to their younger counterparts.

In-basket test. While expected result, rejection sensitivity, and self-efficacy significantly affected the combination of outcome variables in the expected direction, no such effects were found for candidates' actual result, $\Lambda = .905$, $F(9, 96) = 1.12$, $p = .35$. Moreover, no significant effect of ethnicity on the outcome variables was found, $\Lambda = .855$, $F(9, 1183) = 7.10$, $p = .08$. However, significant interactions were found between ethnicity and self-efficacy, $\Lambda = .744$, $F(9, 1183) = 3.67$, $p < .001$, as well as between ethnicity and expected result, $\Lambda = .803$, $F(9, 1183) = 2.61$, $p = .01$, on the outcome variables. First, ethnic majority candidates who expected to have a good result on the test perceived significantly lower procedural justice, $b = -.84$, $SE = .35$, $t(1191) = -2.38$, $p = .019$, and lower general fairness, $b = -1.63$, $SE = .58$, $t(1191) = -2.79$, $p = .006$, compared to ethnic minority candidates who expected to have a good result on the test. Interestingly, the positive effects of one's self-efficacy on procedural justice, $b = -.43$, $SE = .15$, $t(1191) = -2.86$, $p = .005$, and test motivation, $b = -.91$, $SE = .21$, $t(1191) = -4.36$, $p < .001$, were both larger for ethnic minority candidates compared to ethnic majority candidates. Further, we observed a significant three-way interaction between ethnicity, age and gender on the outcome variables, $\Lambda = .659$, $F(27, 294) = 1.56$, $p = .041$, such that older male ethnic minority candidates experienced less discrimination related to ethnicity, $b = .04$, $SE = .02$, $t(1191) = 2.26$, $p = .026$, as well as age, $b = .06$, $SE = .02$, $t(1191) = 2.73$, $p = .007$, than older male ethnic majority candidates, compared to their younger counterparts.

Abstract reasoning test. Remarkably, none of the expected IVs were significantly related to the outcome variables, and no significant effect of ethnicity on the outcome variables was found, $\Lambda = .86$, $F(9, 76) = 1.41$, $p = .20$. We did, however, find a significant interaction effect between ethnicity and self-efficacy on the outcome variables, $\Lambda = .81$, $F(9, 76) = 2.61$, $p = .01$. That is, a higher self-efficacy lowered test anxiety, but only for ethnic minority candidates, $b = 1.76$, $SE = .49$, $t(84) = -3.60$, $p < .001$.

Numerical reasoning test. Of our IVs, only self-efficacy significantly affected the outcome variables, $\Lambda = .617$, $F(9, 44) = 3.04$, $p = .007$. No effects of ethnicity were found, $\Lambda = .825$, $F(9, 44) = 1.04$, $p = .43$, nor were any interaction effects with ethnicity, either.

Verbal reasoning test. All IVs significantly affected the combination of outcome variables, with the exception of one's actual result, $\Lambda = .946$, $F(9, 209) = 1.33$, $p = .23$. No effect of ethnicity was found, $\Lambda = .975$, $F(9, 209) = .60$, $p = .80$, and there were no significant interaction effects with ethnicity, either.

Administrative test. Of all IVs, only expected result and self-efficacy significantly affected the combination of outcome variables. No effect of ethnicity was found, $\Lambda = .970$, $F(9, 215) = .74$, $p = .68$, and there were no significant interaction effects with ethnicity.

Table 9 Descriptives and Correlations of Time 1 variables

	<i>M</i>	<i>SD</i>	<i>N</i>	<i>01.</i>	<i>02.</i>	<i>03.</i>	<i>04.</i>	<i>05</i>	<i>06.</i>	<i>07.</i>	<i>08.</i>	<i>09.</i>	<i>10.</i>	<i>11.</i>	<i>12.</i>
01. Procedural justice	3.78	0.56	5256	--											
02. General test fairness	3.44	1.10	5256	.503	--										
03. Perceived ethnic discrimination	1.41	0.83	5256	-.309	-.079	--									
04. Perceived gender discrimination	1.38	0.81	5256	-.306	-.085	.898	--								
05. Perceived age discrimination	1.47	0.90	5256	-.308	-.130	.775	.806	--							
06. Test anxiety	2.98	1.28	5256	-.066	-.187	0.010	0.002	.035*	--						
07. Test motivation	4.2	0.90	5256	.412	.244	-.217	-.220	-.221	-.038	--					
08. Test attitude	3.08	1.21	5256	.400	.409	.047	.043	-0.008	-.072	.327	--				
09. Perceived test performance	3.06	0.86	5256	.279	.252	0.010	0.007	-.028*	-.078	.212	.326	-			
10. Actual test performance ¹	0.87	0.33	5256	.049	-0.020	-.097	-.093	-.072	.065	.040	-.052	.050	--		
11. Type of candidate ²	0.3	0.46	5153	-.135	-.126	0.018	0.026	.038	.051	-.089	-.145	-.054	.082	--	
12. Tenure	9.88	9.63	5256	-0.021	-0.013	.037	.039	.052	.078	.031*	0.021	-.055	-0.001	.068	--
13. Self Efficacy	2.85	0.89	5256	.331	.337	.052	.051	0.004	-.105	.177	.464	.624	0.018	-.101	-.052
14. Rejection sensitivity	3.91	1.02	5256	.043	-.034*	-.099	-.102	-.064	-.048	.204	-.052	.031*	0.015	0.018	-0.012
15. Ethnic identification	3.37	1	4662	.119	.078	-0.017	-0.026	-0.023	.034*	.108	.041	.071	.036*	-0.022	0.022
16. Ethnic origin ³	0.23	0.43	5256	-.053	0.019	.104	.077	.071	-.109	-0.016	.061	.052	-.141	-.063	-.108
17. Gender ⁴	0.65	0.48	5165	.031*	-.039	-.067	-.072	-.049	.040	.074	-.052	-.137	-0.007	.029*	-0.0228
18. Age	34.33	9.77	5256	-.027*	0.001	.064	.056	.071	0.017	0.021	.029*	-.043	-.059	0.020	.879
19. Level ⁵	3.31	1.28	5256	-.113	-.075	-.065	-.079	-.044	-.133	-.084	-.249	-.051	.061	-.007	-.190
20. Language ⁶	0.58	0.49	5256	-.071	.127	.077	.074	.060	-.571	.015	-.040	.025	-.131	-.093	-.084

Table 9 Descriptives and Correlations of Time 1 variables (Continued)

	<i>M</i>	<i>SD</i>	<i>N</i>	13.	14.	15.	16.	17.	18.	19.	20.
01. Procedural justice	3.78	0.56	5256								
02. General test fairness	3.44	1.10	5256								
03. Perceived ethnic discrimination	1.41	0.83	5256								
04. Perceived gender discrimination	1.38	0.81	5256								
05. Perceived age discrimination	1.47	0.90	5256								
06. Test anxiety	2.98	1.28	5256								
07. Test motivation	4.2	0.90	5256								
08. Test attitude	3.08	1.21	5256								
09. Perceived test performance	3.06	0.86	5256								
10. Actual test performance ¹	0.87	0.33	5256								
11. Type of candidate ²	0.3	0.46	5153								
12. Tenure	9.88	9.63	5256								
13. Self Efficacy	2.85	0.89	5256	--							
14. Rejection sensitivity	3.91	1.02	5256	-.036	--						
15. Ethnic identification	3.37	1.00	4662	.057	.064	--					
16. Ethnic origin ³	0.23	0.42	5256	.084	0.003	-0.007	--				
17. Gender ⁴	0.65	0.48	5165	-.137	.040	-0.026	-0.021	--			
18. Age	34.33	9.77	5256	-.041	0.005	0.000	0.002	-.067	--		
19. Level ⁵	3.31	1.28	5256	-.117	.099	-.017	.087	.022	-.062	--	
20. Language ⁶	0.58	0.49	5256	.063	.136	-.067	.191	-.010	.031*	.293	--

Note. ¹Actual test performance: 0 = fail, 1 = pass. ²Type of candidate: 0 = extern, 1 = intern. ³Ethnic origin: 0 = Belgian; 1 = non-Belgian ethnic minority.

⁴Gender: 0 = man, 1 = women. ⁵Level: 1 = D, 2 = C, 3 = B, 4 = A1, 5 = A2. ⁶Language: 0 = Flemish/Dutch, 1 = French. * $p < .05$. $p < .01$.

Table 10 Mean and Standard Deviations of Applicant Reactions towards Tests at Time 1 (N = 5,256)

T1 (N = 5,256)	PROCEDURAL JUSTICE	GENERAL FAIRNESS	PERCEIVED TEST DISCRIMINATION RELATED TO:			TEST ANXIETY	TEST MOTIVATION	TEST ATTITUDE	TEST EASE
TEST			<i>Ethnicity</i>	<i>Age</i>	<i>Gender</i>				
SJT (N = 3,219)	3.78 (.53)	3.41 (1.11)	1.37 (.81)	1.43 (.88)	1.34 (.88)	2.98 (1.29)	4.21 (.90)	2.97 (1.23)	3.46 (.97)
File Test (N = 1,262)	3.81 (.61)	3.54 (1.10)	1.45 (.85)	1.51 (.92)	1.43 (.83)	2.86 (1.29)	4.23 (.90)	3.27 (1.18)	3.19 (.97)
In-basket Test (N = 123)	3.77 (.69)	3.21 (1.00)	1.33 (.70)	1.37 (.75)	1.29 (.67)	3.52 (1.07)	4.02 (1.04)	3.06 (1.16)	3.33 (.91)
Abstract Reasoning (N = 101)	3.65 (.55)	2.95 (1.20)	1.24 (.68)	1.39 (.89)	1.26 (.74)	3.50 (1.17)	3.91 (1.01)	2.67 (1.14)	3.72 (.80)
Numerical Reasoning (N = 70)	3.69 (.54)	3.33 (1.11)	1.67 (.90)	1.69 (.97)	1.61 (.92)	2.46 (.94)	4.07 (.82)	3.27 (1.26)	3.47 (.91)
Verbal Reasoning (N = 236)	3.72 (.53)	3.61 (.98)	1.62 (.90)	1.66 (.92)	1.59 (.85)	2.32 (.90)	4.30 (.88)	3.44 (1.13)	3.11 (.86)
Administrative Test (N = 245)	3.82 (.57)	3.39 (.97)	1.49 (.93)	1.62 (1.03)	1.48 (.90)	3.76 (1.03)	4.10 (.91)	3.32 (1.09)	3.38 (.95)
TOTAL	3.78 (.56)	3.44 (1.1)	1.41 (.83)	1.38 (.81)	1.47 (.90)	2.98 (1.28)	4.20 (.90)	3.08 (1.21)	3.38 (.97)

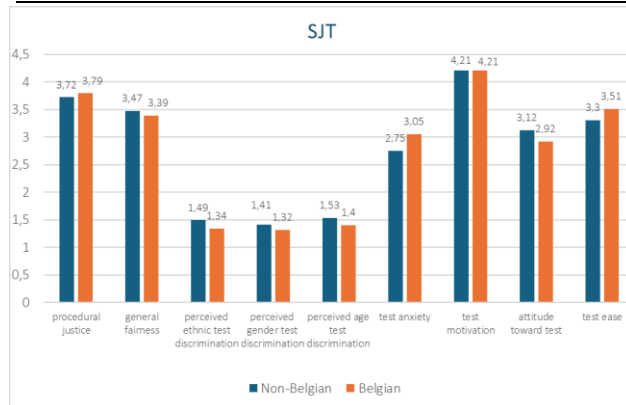


Figure 16 Ethnic Minority – Majority Differences in Applicant Reactions (SJT)

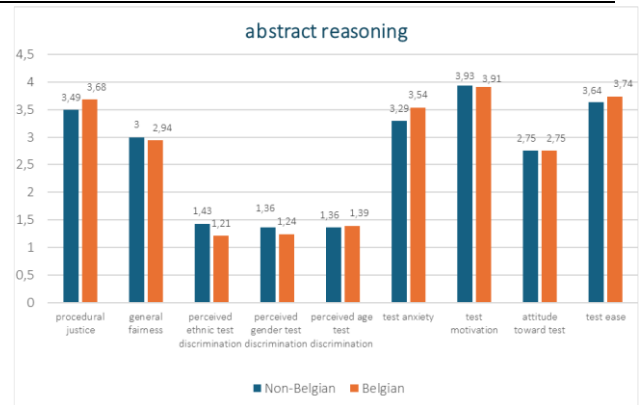


Figure 19 Ethnic Minority – Majority Differences in Applicant Reactions (abstract reasoning)

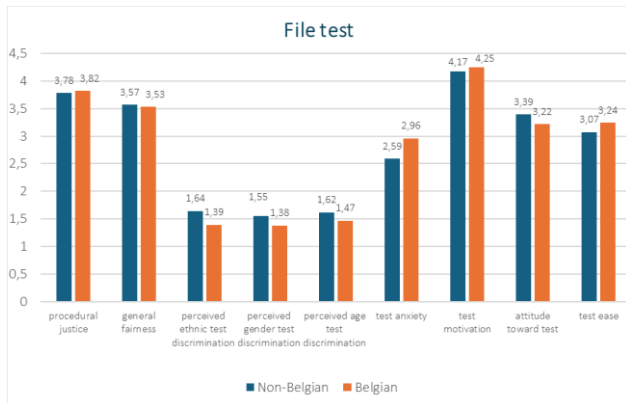


Figure 17 Ethnic Minority – Majority Differences in Applicant Reactions (File test)

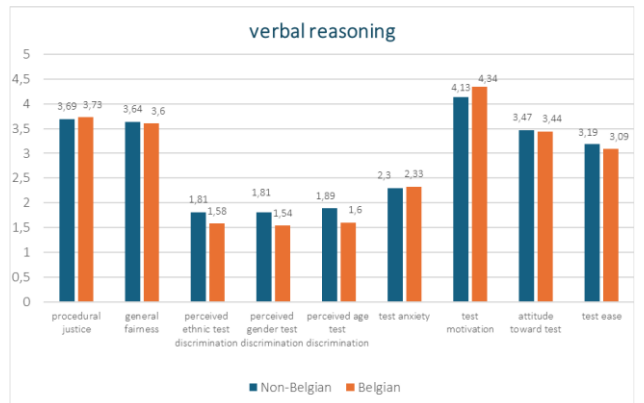


Figure 20 Ethnic Minority – Majority Differences in Applicant Reactions (verbal reasoning)

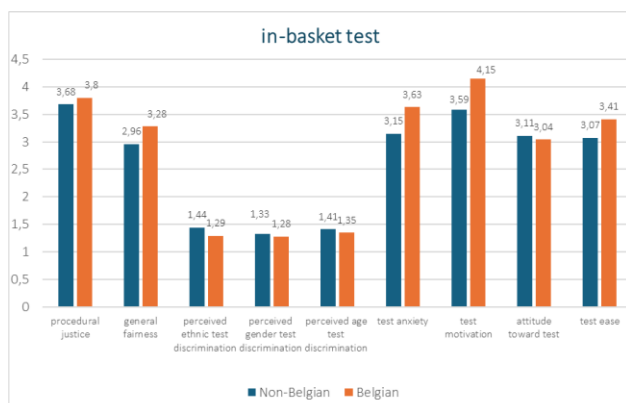


Figure 18 Ethnic Minority – Majority Differences in Applicant Reactions (in-basket test)

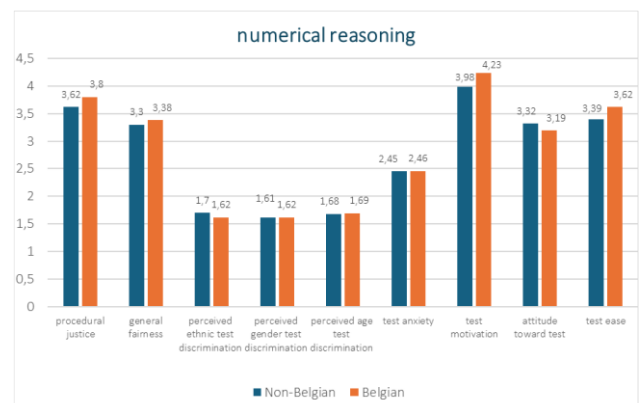


Figure 21 Ethnic Minority – Majority Differences in Applicant Reactions (numerical reasoning)

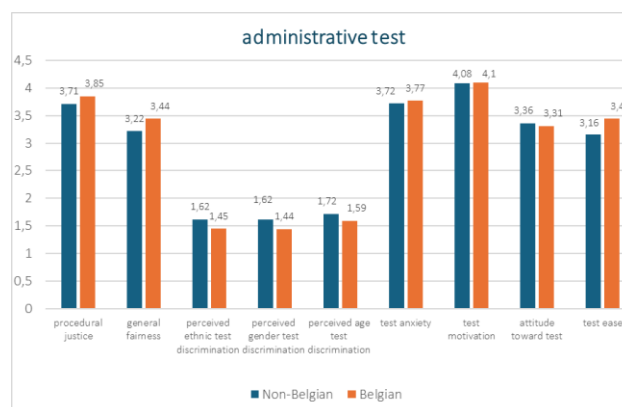


Figure 22 Ethnic Minority – Majority Differences in Applicant Reactions (administrative test)

Time 2: Post-Feedback Reactions

Procedural justice. First, we tested a moderated mediation model with perceived performance as independent variable (IV), the three attribution styles (i.e., locus of causality, stability, and personal control) as potential mediators, procedural justice of the selection procedure as the outcome variable, and ethnicity and gender as potential moderators.

Results showed that firstly, there was a significant effect of perceived performance on procedural justice, $F(4, 279) = 18.75, p < .001$, but no interaction effects with ethnicity, $F(4, 279) = .54, p = .70$, ethnicity and gender, $F(4, 279) = .47, p = .76$, or ethnicity and age, $F(4, 282) = 0.25, p = .88$. This means that, independent of one's ethnicity, gender and age, people who perceived performed well on the test(s) administered at Time 1, gave a higher rating on procedural justice at Time 2.

In terms of the mediating effects of type of attribution style, while an internal locus of causality had a positive effect on procedural justice, $b = .06, SE = .02, t(288) = 3.42, p < .001$, no such effects were found for stability, $b = .01, SE = .02, t(288) = 0.47, p = .64$, or personal control, $b = .04, SE = .02, t(288) = 1.92, p = .06$.

We thus proceeded to test the moderated mediation through internal locus of causality and found a significant indirect effect of perceived performance on procedural justice. No moderating effects were found of ethnicity and gender, $IMM = .02, SE = .02, 95\%CI [-.03, .08]$, and ethnicity and age, $IMM = .00, SE = .00, 95\%CI [-.003, .003]$, so when removing the moderators from the model, we established a significant, positive indirect effect from perceived performance through internal causality to procedural justice, $b = .04, SE = .01, 95\%CI [.02, .06]$.

This means that, irrespective of ethnicity, age and gender, candidates who were more satisfied with of how well they performed on the (previous) test(s), are more positive regarding (external factors such as) the selection procedure, which can be partly explained by a higher internal (vs. external) locus of causality.

Distributive justice. Subsequently, we tested a moderated mediation model with perceived performance as the independent variable, the three attribution styles (i.e., locus of causality, stability and personal control) as potential mediators, distributive justice of the selection procedure as outcome variable, and ethnicity and gender/age as potential moderators.

We found a significant effect of perceived performance on distributive justice, $F(4, 287) = 31.88, p < .001$, but no interaction effects with ethnicity, $F(4, 287) = .96, p = .43$, gender, $F(4, 287) = 2.00, p = .14$, or age, $F(4, 290) = 0.48, p = .62$. This means that, independent of one's ethnicity, gender and age, people who perceived they performed good on the test, gave a higher rating on distributive justice.

In terms of the mediating effects of type of attribution style, we found that both an internal locus of causality, $b = .11, SE = .04, t(288) = 3.05, p = .003$, and personal control, $b = .11, SE = .04, t(288) = 2.56, p = .01$, had a positive effect on distributive justice, but no such effects were found for stability, $b = .04, SE = .04, t(288) = 0.99, p = .32$. We thus proceeded to test the moderated mediation through internal locus of causality and personal control and found a significant indirect effect of perceived performance on distributive justice through internal locus of causality and personal control.

No moderating effects were found of ethnicity and gender, $IMM = .05, SE = .05, 95\%CI [-.05, .15]$, and ethnicity and age, $IMM = -.00, SE = .00, 95\%CI [-.01, .01]$ on internal locus of causality. Further, No moderating effects were found of ethnicity and gender, $IMM = .00, SE = .00, 95\%CI [-.003, .003]$, and ethnicity and age, $IMM = .00, SE = .00, 95\%CI [-.001, .01]$ on personal control. We removed the moderators from the model, and we established a significant, positive indirect effect from perceived

performance through internal causality, $b = .10$, $SE = .03$, 95%CI [.03, .15], and personal control, $b = .10$, $SE = .04$, 95%CI [.02, .18], to distributive justice.

This means that, irrespective of ethnicity, age and gender, candidates who have were more satisfied with how well they performed on the test(s), are more positive regarding the distributive justice of the selection procedure (i.e., fairness of its outcome), which can be partly explained by a higher internal (as opposed to external) locus of causality and a more personal (as opposed to external) feeling of control.

Conclusion

In sum, the two-wave survey on applicant reactions revealed generally favourable perceptions of the selection tests and the selection process. At Time 1, applicants rated both perceived procedural justice and general fairness of the tests as above average, along with high scores for test motivation, test attitude, and perceived test ease. Importantly, ratings for perceived discrimination based on ethnicity were below average and showed no significant difference between ethnic minority and ethnic majority applicants across any of the tests. The analyses on the Time 2 follow-up survey found that applicants who were more *satisfied with their test performances* subsequently evaluated the selection procedure higher on both procedural and distributive justice. This relationship was partially explained by an individual's attribution style (internal locus of control and causality). Critically, these effects were not moderated by ethnicity, nor by age or gender, leading to the conclusion that both ethnic minority and majority applicants expressed little concern regarding potential unfairness during the federal public service selection process.

Table 11 Descriptives and Correlations of Time 2 variables

	<i>M</i>	<i>SD</i>	<i>N</i>	01.	02.	03.	04.	05.	06.	07.	08.	09.	10.	11.	12.	13.	14.	15.	16.
01. Perceived Performance	3.11	1.218	376	--															
02. Internal locus of causality	5.25	1.938	376	.488	--														
03. Internal controllability	5.045	1.688	376	.365	.661	--													
04. Stability	4.78	1.66	376	.124*	.312	.496	--												
05. Procedural justice	3.685	0.52	376	.459	.464	.433	.160	--											
06. Distributive justice	3.28	1.28	376	.523	.535	.454	.180	.586	--										
07. Perceived ethnic discrimination	1.35	0.84	376	-.165	-.225	-.157	-.148	-.262	-.280	--									
08. Perceived gender discrimination	1.34	0.77	376	-.153	-.195	-.144	-0.095	-.255	-.267	.760	--								
09. Perceived age discrimination	1.36	0.79	376	-.139	-.211	-.151	-0.059	-.287	-.292	.699	.750	--							
10. Ethnic origin ¹	0.21	0.41	376	-.109*	-0.037	0.029	-0.014	0.019	0.004	0.093	0.024	0.045	--						
11. Gender ²	1.66	0.50	376	0.029	-0.062	-0.047	0.037	-0.044	-0.092	0.068	0.036	0.039	-0.075	--					
12. Age	36.82	10.17	376	-.156	-0.016	-0.051	-0.014	-0.056	-0.021	0.074	0.027	0.001	0.054	-0.075	--				
13. Type candidate ³	0.31	0.46	362	-0.062	-0.118	-0.119	-0.051	-0.207	-0.244	0.055	-0.016	0.025	-0.086	-0.016	0	--			
14. Tenure	12.36	9.94	376	-0.079	0.009	-0.044	-0.008	-0.043	0.015	0.032	0.015	0.011	-0.098	-0.098	1	0.069	--		
15. Level ⁴	3.48	1.25	376	-0.002	-0.096	0.016	.111*	-0.043	-.116*	-0.040	-0.006	-0.037	-0.001	0.050	-.113*	-0.025	-.181	--	
16. Language ⁵	.52	.50	376	-0.017	-0.047	-0.005	-0.046	0.052	0.077	0.025	0.064	-0.055	.186	-0.045	0.096	-0.098	0.011	.164	--

Note. ¹Ethnic origin: 0 = Belgian; 1 = non-Belgian ethnic minority. ²Gender: 0 = man, 1 = women. ³Type of candidate: 0 = extern, 1 = intern. ⁴Level: 1 = D, 2 = C, 3 = B, 4 = A1, 5 = A2. ⁵Language: 0 = Flemish/Dutch, 1 = French. * $p < .05$. $p < .01$.

4.6.3 Conclusions quantitative analyses: Perceived and actual fairness of selection procedures

The conclusions for section 4.5 and 4.6 dealing with actual and perceived fairness of **the selection tests** are summarized in Table 2. First, as regards the actual test fairness, there was evidence for adverse impact of all tests. Note, however, that mixed findings were found for the in-basket test (i.e., presence of adverse impact in a statistical sense, while absence of adverse impact in a practical sense) and the numerical reasoning test (i.e., absence of adverse impact in a statistical sense, while presence of adverse impact in a practical sense), which might in case of the numerical test be explained by a smaller sample size. The *perceived* test fairness (procedural justice, general fairness, and perceived discrimination) by ethnic minority applicants did not differ significantly from that of ethnic majorities. In sum, although selection rates were (significantly) lower for ethnic minority compared to majority applicants for all tests, applicants did not perceive any of the tests as unfair.

Furthermore, as regards **the selection procedure**, applicants who were satisfied with how they performed on the selection tests (i.e., perceived performance) evaluated the selection procedure higher on both procedural and distributive justice, especially when they attributed their test performance to their own competencies (i.e., internal attribution) and when they felt in control of their performance (i.e., perceived controllability). Note that ethnic background did not play a role (i.e., had no moderating effect on these relationships).

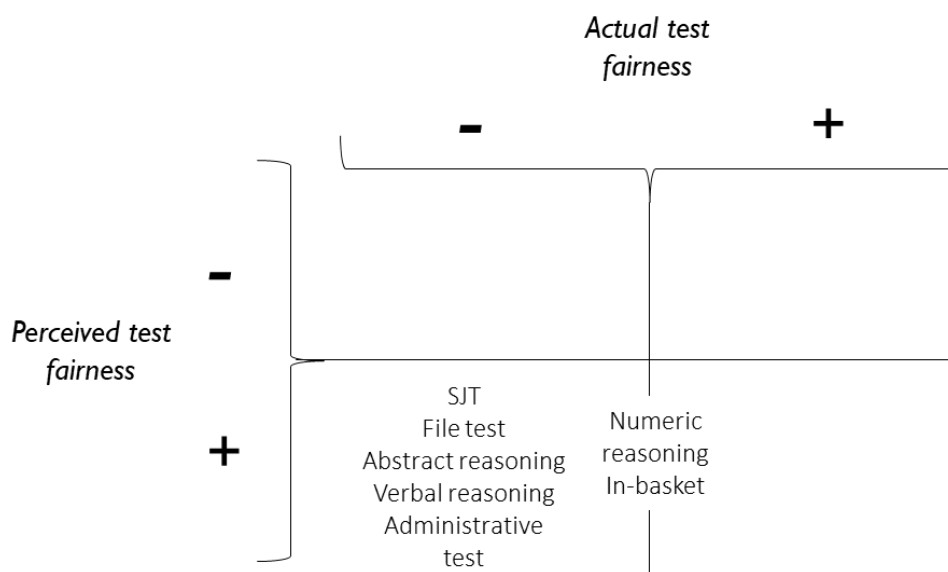


Figure 23 Actual and perceived fairness of test administered by FOD BOSA

4.6.4 Focus groups

Theme 1: Overall Adverse Impact. Of the seven selection instruments, only one test passed the .80 AI criterium of fair testing (Council of the European Union, 2000; Equal Employment Opportunity Commission, 1978). When presented with an overview of the AI statistics per test, job applicants and field experts offered different reactions. Experts were “not surprised” by the disadvantages faced by ethnic minority applicants because the results align with their knowledge from scientific studies. Conversely, applicants reported being “surprised” and “shocked” as they underestimated the AI. This corresponds with the location of most federal selection tests within the four-quadrant model, namely actual test unfairness present, but no perceived test unfairness by candidates (see Figure 23).

Applicants' positive perceptions of test fairness could be understood by minorities' beliefs that computerised testing ensures fair treatment. However, being presented with AI statistics, the minority members of the focus group were confronted with systematic, indirect and subtle instances of unequal treatment that they were previously unaware of (and surprised about)

Theme 2: High Adverse Impact of Verbal Reasoning Tests. The verbal reasoning test had the highest AI with a selection ratio (ethnic minority applicant/ethnic majority applicant) of .61, whereas tests are considered fair starting from ratios of .80 and higher. Experts explain the high AI from these tests as the result of the language nuances required to pass the ambiguous test and its items and themes being unrelated to jobs. However, experts also state that verbal reasoning tests are rarely used in practice and are reserved for specific occupations. Moreover, when administered, their results are rarely "a deciding factor". The experts are aware of the test's high AI and acknowledge the difficult trade-off between tests' performance (i.e., predictive validity) and fairness (i.e., AI). Job applicants perceived the high AI of verbal reasoning tests as the result of actual differences in cultural content and language competencies. The minority candidate believed that interpretations of the tests' AI as a manifestation of "racism", although commonly stated, are incorrect. Instead, it was suggested that applicants with ethnic minority backgrounds need "more time" to achieve the competency levels of ethnic majorities (i.e., human capital deficit hypothesis).

When discussing the role of language in testing, an emerging sub-theme centred around the formal language tests from the federal government. The experts explained that verbal reasoning tests are fundamentally different from language tests. Whereas the former is a means to measure the reasoning abilities of an applicant, the latter provides a certification for applicants' language proficiency. Both applicants and experts agree that the use of language tests is often unnecessary. As an example: an ethnic majority applicant whose mother tongue is Dutch has had to obtain Dutch language certification because they obtained an English Master's degree. The experts explained that examples like this one are the result of outdated legislation and that they most likely are a source of drop-out.

"For us, it is also a source of frustration because these additional language tests are more likely to cause us to lose good candidates..." — Selection expert

Theme 3: Perceived Fairness of Cognitive Ability Tests. In the literature, it has been established that cognitive ability tests have a high AI (Roth et al., 2001; Sackett et al., 2022). In our analyses, this is also the case for abstract reasoning (.76), numerical reasoning (.76) and verbal reasoning (.61). As a result of their adverse impact, the literature argues that ethnic minority applicants, aware of the disadvantage, experience test anxiety when required to complete cognitive tests (Oostrom & De Soete, 2016). In our focus group, however, applicants did not recognise themselves in this literature finding. Instead, the minority applicant from the panel reported very little test anxiety in computerised testing because they believed the lack of face-to-face contact with interviewers guaranteed equal treatment, as demonstrated by the following excerpt:

"There [in computerised testing], they don't see that I have a different ethno-cultural background. This made me feel safe. They were only going to consider my abilities." — Ethnic minority applicant

According to the majority applicant, however, selection instruments induced test anxiety, albeit for another reason than unequal treatment. In their experience, when test content is irrelevant and a test impacts hiring decisions, they experience test anxiety; despite perceiving oneself as competent for the job. Note that these different reactions from ethnic minority-majority panel participants corroborate with applicant perceptions measured at Time 1 as part of WP3.2 (i.e., $r_{T1_test\ anxiety - ethnic\ origin} = -.11, p < .01$; see Table 9).

“That fear [test anxiety] actually lies in the fact that the test content requires me to do things that, to my understanding, are not relevant to positions I am applying for ... You know you have the right background, but then you’re being asked about things that I find irrelevant. When this is the case, you’ve lost me [as an applicant], I’m very firm about that. ... Maybe I should be a bit more confident in that area [testing] too.” — Ethnic majority applicant

Theme 4: Future of Testing. When evaluating future improvements of selection tests, several of the themes we identified in the literature were spontaneously brought up in the focus group. In addition to job irrelevant test content (cf. supra), applicants voiced their dissatisfaction with the competency system based on their level of education. Applicants perceive it as confusing that competency measures are linked to education levels in the sense that they are unsure whether to perform to their (maximum) ability on tests or according to the prescribed level of education required for a job.

“I prepared for the test [SJT] using a guidebook in which the competencies were explained by level of degree. I found it bizarre that I had to stay within the ‘bachelor limits’ because I have not been able to continue studying because of personal socio-economic reasons. [...] In other jobs, my character and assertiveness have been very different from the [prescribed] ‘bachelor limits’ [...] During the test, I didn’t have to choose [the response options] that I thought were fitting. [...] I am happy to have been accepted, but I’m also afraid to do things that are not allowed according to my level.” — Ethnic minority applicant

Selection experts explain the limited job representativeness of selection tests as being the result of test development’s economic costs and the unique character of each job within the federal government’s organisations (e.g., similar jobs but with a national or international orientation). Administering test batteries, the experts consistently make trade-offs in terms of content representativity for a specific job, in which applicants will recognise scenarios from their vocational experience, and broader testing modules that applicable across a wide array of occupations.

Besides improving test representativeness, applicants propose adopting multi-media support in testing. More concretely, they believe that including video or audio versions of testing materials will improve the clarity of the assignment. As an example, having an item read out loud helps interpreting ambiguous sentences. A final improvement suggested by the applicants is gamification. They perceive gamified tests as more representative for measures of individual characteristics (e.g., risk-aversion: applicants press a button to inflate a digital balloon. The longer participants press on the inflate button, the higher the risk the balloon will pop). This positive appraisal of gamified tests is in line with earlier research (Georgiou et al., 2019). Although both ethnic minority and majority applicants are unconcerned with potential unfairness when taking selection tests, and a minority applicant explains adverse impact as being the result of cultural differences and the competency deficit hypothesis all applicants feel that tests lack job relevance. In the words of an ethnic minority applicant:

“There [in computerised testing], they don’t see that I have a different ethno-cultural background. This made me feel safe. They were only going to consider my abilities.”

Furthermore, the applicants as well as selection professional express strong opinions on test requirements in terms of language, educational degree and competencies. The selection professional states how *“it is also a source of frustration because these additional language tests are more likely to cause us to lose good candidates...”*

A summary of major findings from the focus group can be found in Table 12.

Table 12 Results from focus groups, discussing findings of WP3

1. Overall Adverse Impact	2. High Adverse Impact of Verbal Reasoning Tests	3. Perceived Fairness of Cognitive tests	4. Future of Testing
<i>Job Candidates</i> - surprised, underestimated AI - perceived testing as unrelated to minority status (EMI)	<i>Job Candidates</i> - explained by: (a) language competencies (b) cultural competencies	<i>Job Candidates</i> - high test anxiety: low job representativity of tests jeopardises self-efficacy (EMA) - low test anxiety: tests do not require face-to-face interaction, curbing overt discrimination (EMI)	<i>Job Candidates</i> - sources of frustration: (a) low job representativity of tests (b) competencies per level of education - improvements: (a) multimedia: clarification (b) gamification: job representativity
<i>Professionals</i> - in line with expectations and research	<i>Professionals</i> - explained by: (a) language nuances (b) item themes - test use: (a) rare (b) little weight <i>Job Candidates and Professionals</i> - Policies on language certification^a; source of: (a) incomprehension (b) frustration (c) candidate drop-out	<i>Professionals</i> - trade-off between validity and AI	<i>Professionals</i> - job representativity explained by: (a) economic cost of test development (b) heterogeneity in job content

Notes. Abbreviations used: AI (Adverse Impact), EMI (Ethnic Minority Candidate), EMA (Ethnic Majority). ^a Language tests are distinct from verbal reasoning tests: they are a means of certification for candidates' language levels – often for languages other than their mother tongue.

4.7 Analysis of career success and advancement (throughflow)

WP4 investigated career success and advancement of both ethnic minorities and majorities.

Career success (*organisational throughflow*) of ethnic minority/majority employees has been considered to an even lesser extent than their actual/perceived fairness of hiring procedures (*inflow*). In general, career success is defined as the ‘*positive psychological, work-related outcome or achievement one accumulates as a result of work experience*’ (Seibert et al., 1999, p. 417) and encompasses both ‘objective’ and ‘subjective’ career success (Abele & Spurk, 2009). *Objective career success* is typically operationalized by occupational titles, promotion and income growth (Apers et al., 2019) and was considered in **WP4.1**. *Subjective career success*, on the other hand, refers to self-referenced experiences of success, operationalized by career satisfaction measures and was considered in **WP4.2**. Addressing the goal for a diverse/inclusive federal workforce that reflects the current Belgian population, the present study investigates career success of ethnic majority (Belgian) versus ethnic minority (non-Belgian) federal job incumbents in both an objective and subjective way.

4.7.1 Literature findings

Objective Career Success

Objective career success refers to one’s career success in terms of measurable and more objective indicators. In the literature on objective career success, it is operationalized as either the amount of promotions/upward career steps (Ebner & Paul, 2023; Gara Bach Ouerdian & Mansour, 2019; Golden & Eddleston, 2020), salary (Apers et al., 2019; Blickle et al., 2018; Elsey et al., 2022; Giraud et al., 2019; Sammarra et al., 2013; Volmer et al., 2018) or occupational prestige (Hirschi et al., 2021; Kirchmeyer, 2006). Hence, as a first KPI of objective career success, we measured *amounts of promotions*.

Given that career success is changing/growing in its nature, scholars have touched on this temporal aspect by measuring ‘*promotion rate*’ (i.e., ratio of promotions on organisational tenure; Kirchmeyer, 2002) instead of number of promotions. This operationalisation is particularly useful in a government context that tends to employ people in a sustainable way (e.g., as reflected by our study sample with an average organisational tenure of 9 to 12 years). Hence, as a second way of measuring employees’ objective career success, we calculated *promotion rates*, defined by the amount of upward career steps divided by their tenure in the federal government.

Alternatively, scholars have included the temporal nature of one’s career more directly by means of longitudinal study designs (Apers et al., 2019; Judge et al., 2000) or by incorporating the ‘time’ aspect by measuring ‘*promotion speed*’ as well. For instance, Weng (2010) incorporated ‘promotion speed’ by means of three self-reported items (e.g., ‘My promotion speed in my present organisation is fast’). Building on this idea, but with the aim of obtaining a more objective measure, we operationalized promotion speed based on Kumazawa (2010), who measured it by looking at the ‘time before a promotion step’.

In sum, based on the literature, we decided to operationalize federal employees’ objective career success by means of three KPI’s, namely *promotions*, *promotion rate* and *promotion speed*. We aimed to compare these measures between ethnic majority and minority employees within the federal government.

Subjective Career Success

Recently, literature stresses the importance of one's more personal career evaluation or *subjective career success* (Spurk et al., 2019). Objective and subjective career success are two distinct ways of measuring one's career success (Abele & Spurk, 2009; Stumpf & Tymon, 2012) and might be differently related to career outcomes and determinants. Ng and colleagues (2005) for example, investigated effects of a wide range of factors affecting objective versus subjective career success.

Career satisfaction. In their review, subjective career success was investigated in terms of career satisfaction. Indeed, earlier research on the subject tends to view subjective career success as a unidimensional construct (e.g., 'How satisfied are you with your career?'; Greenhaus et al., 1990). Gradually, scholars have adopted a multidimensional interpretation of the concept, meaning that many different aspects of one's career are included in the concept, for example: external compensation, intrinsic fulfillment, growth and development (Shockley et al., 2016; W. Zhou et al., 2013). These studies are, however, situated in different locations and cultures (e.g., China versus the USA in the abovementioned examples). Considering the cultural embeddedness of career aspects, Briscoe et al. (2021) developed/validated a cross-culturally representative measure of subjective career success including different aspects that are suitable to measure among people of different ethnic origins. Interestingly, their subjective career construct encompasses both the aspect of importance as well as the aspect of achievement. Hence, for each career success aspects, the satisfaction aspect can be investigated in relation to the importance that one individual holds for this aspect.

Career Resource Dimensions

Hirschi et al. (2018) defined four key 'Career Resource Dimensions' that determine both objective and subjective career success. First, *Human Capital Career Resources* resembles one's skills and competences, both in a general and job-related sense. According to the 'contest mobility perspective', these resources are primordial to one's career success, since employees have an 'equal chance to advance' and only differ in terms of these human capital aspects (Judge et al., 2010; Ng et al., 2005; R. H. Turner, 1960). Hirschi et al. (2018), however, defined three more resources important for one's career success. *Environmental Career Resources*, for instance, reflect social or organisational support and opportunities. Relatedly, the 'organisational sponsorship perspective' acknowledges these factors and predicts that employees that receive more access to opportunities/resources through sponsoring/mentoring will have higher career success (Dossinger et al., 2019; Ng et al., 2005; R. H. Turner, 1960). Further, *Motivational Career Resources* present another key determinant of career success, including aspects such as one's career confidence and clarity. Finally, *behavioural tendencies* related to one's career preparation present the final key career resource according to Hirschi et al. (2018), which is referred to as *Career Management Behaviours*.

Ethnic origin and Career Success

Socio-demographic factors such as ethnic origin might also affect objective and subjective career success (Ng et al., 2005). An investigation into career paths of employees of different ethnic origins might be warranted given the underrepresentation of people of foreign origin at the higher echelons of organisations (Groeneveld & Meier, 2022; UNIA & Federale Overheidsdienst Werkgelegenheid, Arbeid en Sociaal Overleg, 2020). While this underrepresentation indicates lower career success among ethnic minority federal employees, an investigation into different operationalisations of career success might uncover the situation more in-depth: Do we observe lower promotion rates and/or lower promotion speed among ethnic minority employees? Based on meta-

analytic results (Ng et al., 2005), we expect that people with foreign ethnic origin would have lower objective career success in terms of salary (given their positions in organisations), but not in terms of number of promotions. Interestingly, subjective career success would be lower among ethnic minority employees as well. In their recent review, Avery et al., (2023) point to different ways that ethnic minority employees might be disadvantaged in their careers compared to ethnic majority employees. For instance, differential performance management and salary growth, differential selection into higher leadership positions, and administrative inconsistencies that lead to lower procedural justice on the job. Interestingly, the notion of procedural justice relates to our work package 3.2 in which we investigate whether procedural justice *in the selection process* differs between ethnic minority and majority employees (Bauer et al., 2001).

When investigating the influence of ethnic origin on career success, there are additional factors that should be included, since they might affect the relationship with outcome variables. For instance, in addition to measuring people's ethnic origin by means of categories (e.g., Belgian; North African; East Asian), it should also be acknowledged that certain individuals might feel more strongly connected with their ethnic origin than others (Phinney & Ong, 2007) and that this might affect the relationship between one's ethnic origin and outcomes such as career success. We measured one's ethnic identification or how strongly one's ethnic origin is part of their (social) identity (J. C. Turner et al., 1994). To that end, we measured people's identification with their ethnic origin in terms of two dimensions: (a) how committed are they towards their ethnic origin and (b) how much do they explore or seek information relevant to one's ethnic origin. People of foreign origins in Belgium differ in the way in which they have adopted the Belgian culture or adhere to their culture of origin. These behavioural differences, or extent of acculturation, might affect employees' career success, since acculturation was found to influence employees' wellbeing and behaviour at work (Beirens & Fontaine, 2011) as well as the amount of career barriers one perceives and one's career self-efficacy (Holloway-Friesen, 2018; Mejia-Smith & Gushue, 2017).

Further, employees with a foreign ethnic origin might have different beliefs about how people of Belgian origin think about them. Employees who believe others think negatively about their origin, might feel and behave differently than those who hold more positive beliefs. Meta-stereotypes or the beliefs about the stereotypes that Belgian employees hold about employees of foreign origin (Vorauer et al., 1998) shape people's self-beliefs (Koçak et al., 2023; Owuamalam & Zagefka, 2014) and might even lead to self-sabotaging behaviours in the form of not applying for certain positions. In the context of promotions, negative meta-stereotypes for women can indeed lead to lower intention to apply for the promotion (Koçak & Derous, 2022). Hence, meta-stereotype conscientiousness or the extent to which one is conscious of the stereotypes that people hold of their ethnic minority group (K. M. Ryan et al., 2015) might affect career success and is measured in the current study.

Employees with a foreign ethnic origin might have also *experienced* negativity/discrimination related to their ethnicity at work (McCord et al., 2018) or during selection procedures (see WP3 of this report). Indeed, perceived workplace discrimination and perceived selection discrimination are both higher among ethnic minority than ethnic majority candidates/employees and both shape individuals' attitudes toward the organisation and job satisfaction (Abu Ghazaleh, 2012; Doaty, 2022). Similarly, ethnic minority employees might have experienced discrimination during past promotion procedures (Avery et al., 2023; Becton et al., 2008), which might also affect their objective or subjective career success.

Not only ethnicity, but also other individual difference factors might affect career success. In general, individuals with higher core self-evaluations are likely to experience higher objective/subjective

career success. That is, a general believe in oneself increases the chances of promotion and career satisfaction (Judge, 2009; Stumpp et al., 2010). Moreover, certain individuals might be more highly sensitive to rejection than others, which creates a certain rejection threat/fear of rejection that inhibits their persistence, for instance, in terms of career success. Additionally, Ghazaleh Abu (2012) found that perceived selection discrimination is affected by a combination of core self-evaluations/rejection sensitivity and ethnic origin such that ethnic minority candidates who also experience low core self-evaluations or are also sensitive to rejection experience even more discrimination. Similarly, core self-evaluations, rejection sensitivity and ethnic origin might interact in predicting career success.

In response to perceptions of discrimination at work or during selection/promotion procedures, ethnic minority employees might have experienced allyship by other employees, meaning that people belonging more privileged groups in society work to end oppression against disadvantaged groups (Selvanathan et al., 2020). Successful allyship, as opposed to ineffective or performative allyship (Collier-Spruel & Ryan, 2022; Kutlaca & Radke, 2023), is expected to help (the cause of) disadvantaged group members by providing feelings of hope for social change and might hence influence how ethnic minority employees perceive the way in which they are treated in the workplace, for instance, in terms of their career success/discrimination.

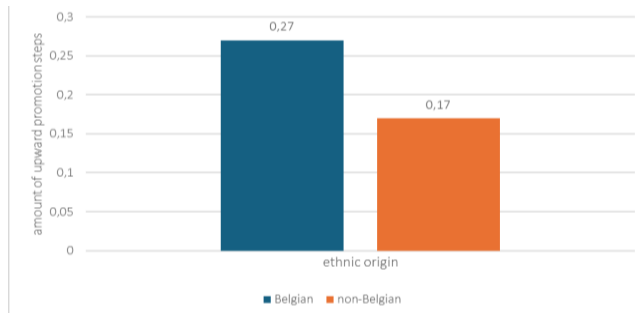
Building on cognitions (e.g., meta-stereotypes) and earlier experiences (e.g., perceived discrimination/allyship), employees of foreign ethnic origin might also anticipate specific career barriers related to their ethnic origin. Indeed, perceived career barriers are higher among ethnic minorities (Lipshits-Braziler & Tatar, 2012). These career barriers can lower employees' career adaptability (Li et al., 2023), self-efficacy (S. L. Wright et al., 2014) and hence career success as well (Santos, 2016).

Finally, according to the 'organisational sponsorship perspective', organisational and supervisor support play a primordial role in one's career success. Indeed, Ward et al. (2022) found that organisational environment, trust in supervisors and organisational justice were predictors of career satisfaction in the public sector. For ethnic minority employees, this organisational/supervisor support in terms of diversity/inclusion is an additional aspect that might add to one's career success. That is, Villotti et al. (2019) showed that organisational climate matters and when organisations recognise and value cultural differences rather than trying to reduce them, this positively affects ethnic minority employees' job satisfaction and lowers their intention to quit. Moreover, the role of supervisors/leadership as well as the perceived equity of procedures in organisations is also expected to affect work outcomes and satisfaction (Aryee et al., 2002; Guillaume et al., 2014; Nishii, 2013).

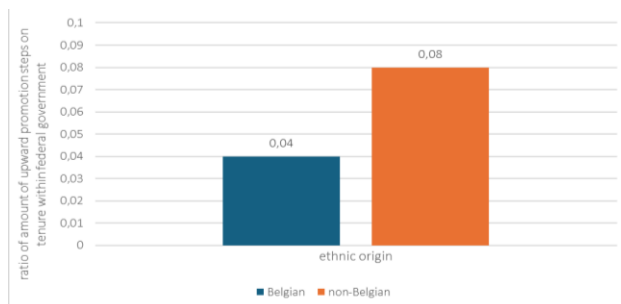
4.7.2 Objective career success: KPIs

We examined objective career success (**WP 4.1**) through the promotion opportunities for ethnic minority employees versus majorities through three KPIs. Compared to ethnic majority employees, ethnic minority employees experienced significantly fewer promotions but achieved promotions at a significantly faster rate. The time (in *months*) until their first, second, and third promotions was significantly shorter for ethnic minorities compared to ethnic majorities.

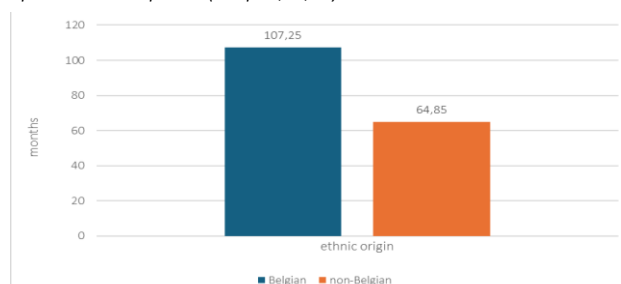
Please note that wordings in this report might differ from language used within the federal government. Vertical or upward movements within level A are usually referred to as promotions (*promotie/promotion*), while upward movements are referred to as advancement (*bevordering/accession*) within the federal context. In the current report, we refer to both types of upward movement as 'promotion'. Horizontal movements in which no change in/between level can be distinguished are not included as 'promotion'.

number of promotions

Despite the overall career success being mixed, the data reveals a significant disparity in the total number of promotions obtained. Ethnic minority employees obtained significantly fewer promotions ($M = 0.17$, $SD = 0.48$) compared to ethnic majority employees ($M = 0.27$, $SD = 0.59$), a statistically significant difference, $\beta = -0.493$, $p < 0.001$.

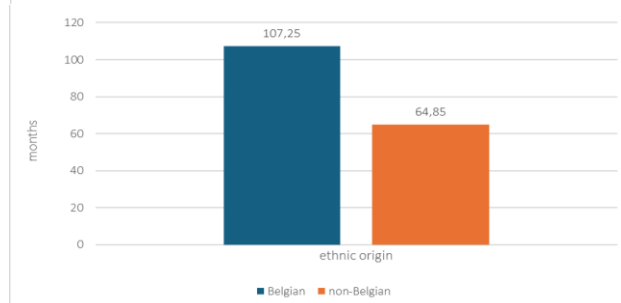
promotion rate

Conversely, after controlling organisational tenure, ethnic minority employees demonstrated a significantly higher promotion rate ($M = 0.008$, $SD = 0.11$) compared to their ethnic majority counterparts ($M = 0.004$, $SD = 0.02$), a statistically significant finding, $F(1,9582) = 9.45$, $p = .002$.

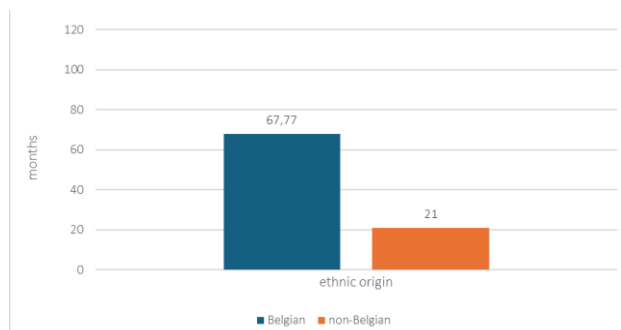
promotion speed (step 1, 2, 3)

In terms of promotion speed, ethnic minority employees achieved promotions at a significantly faster rate across the first three promotion milestones:

Upward Promotion: The time until the first promotion was significantly shorter for ethnic minority employees ($M = 64.85$ months, $SD = 66.59$) compared to ethnic majority employees ($M = 107.25$ months, $SD = 96.41$), $\beta = -0.407$, $p < 0.001$.



Upward Promotion: Ethnic minority employees also achieved their second promotion faster ($M = 69.88$ months, $SD = 75.67$) than ethnic majority employees ($M = 75.74$ months, $SD = 58.87$), $\beta = -0.163$, $p < 0.001$.



Third Upward Promotion: This pattern was most pronounced for the third promotion, where ethnic minority employees reached this milestone significantly faster ($M = 21.00$ months, $SD = 36.37$) than ethnic majority employees ($M = 65.77$ months, $SD = 66.35$), $\beta = -0.730$, $p < 0.001$.

Figure 24 Number of promotions, promotion rate and speed for Belgian vs. non-Belgian origin employees

4.7.3 Subjective career success and advancement (throughflow)

The final investigation of the career stage (*throughflow*) focuses on subjective career success among ethnic minority and majority employees (**WP 4.2**). We focus on job incumbents' perceptions of (a) *career resources*, (b) *career barriers*, (c) *organisational DEI* (Diversity, Equity, and Inclusion), and (d) *ethnicity-related individual characteristics*. The first three factors can be influenced by the organisation, whereas the latter is more personal in nature. Additionally, the relationship between objective career success (number of promotions) and these factors was explored. Table 13 and Table 14 present the descriptives and zero-order correlations of study variables on (a) the total sample and (b) ethnic minorities, only.

Table 13 Descriptives and zero-order correlations of study variables on the total sample (N = 1953)

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
01.SCS	3.57	0.60	--																	
02.Ethnic origin	0.11	0.32	0.001	--																
03.Gender	0.60	0.49	.091	.044*	--															
04.Age	44.16	10.99	-.048*	-.131	-.065	--														
05.Educational level	3.36	0.80	.049*	0.027	-0.025	-.146	--													
06.Occ.expertise	3.81	0.80	.146	0.002	-.077	.137	.173	--												
07.Hum capital devel.	3.52	0.86	.170	.065	.057	-.119	.067	.229	--											
08.Org.career support	2.812	1.14	.481	0.027	.051	-.154	-0.004	-.071	.100	--										
09.Soc.career support	3.50	0.83	.359	.055	.108	-.326	0.035	.133	.282	.431	--									
10.Mentoring	3.074	0.94	.554	0.000	.044*	-.198	.124	.091	.159	.647	.532	--								
11.Career clarity	3.73	0.89	.165	.041*	-0.022	.119	-0.019	.297	.338	.058	.247	.117	--							
12.Career preparation	2.11	0.88	.125	0.012	0.019	-.071	.052	0.038	.443	.155	.181	.169	.187	--						
13.Barriers_ethnicity	1.72	0.94	-.221	.199	-0.026	.070	-.085	-0.008	0.012	-.198	-.171	-.218	.047*	0.032	--					
14.Barriers_age	2.40	1.09	-.262	0.003	0.036	.250	-0.020	-0.014	-.065	-.257	-.275	-.277	-0.019	-0.032	.477	--				
15.Barriers_gender	2.03	1.10	-.202	.092	.185	-0.034	.065	0.008	-0.016	-.177	-.098	-.151	-0.024	0.021	.597	.564	--			
16.Org. diversity	3.34	0.75	.232	-.137	0.028	-0.036	-0.032	.049*	.056*	.240	.150	.257	.088	.065	-.143	-.121	-.118	--		
17.Org. inclusion	3.734	0.67	.243	-.141	-0.026	0.005	-.054*	0.018	.084	.260	.152	.257	.079	.062	-.223	-.161	-.201	.636	--	
18.Org. equity	3.024	1.10	.427	0.029	0.027	-.127	0.026	-.059	.070	.624	.289	.509	0.013	.119	-.191	-.261	-.198	.280	.342	--
19.Supervis.support DEI	3.74	0.90	.359	-0.036	0.010	-.070	-0.009	0.034	.095	.313	.252	.394	.094	.072	-.208	-.200	-.203	.420	.516	.384

Note. N= 1953. ¹Ethnic origin: 0 = Belgian; 1 = non-Belgian ethnic minority. ²Gender: 0 = man, 1 = women. ³Level: 1 = D, 2 = C, 3 =B, 4 = A, * p < .05. p < .01.

Table 14 Descriptives and zero-order correlations of study variables on the ethnic minority sample

	<i>M</i>	<i>SD</i>	<i>N</i>	1	2	3	4	5	6	7	8	9	10	11
01.SCS	3.54	0.64	270	--										
02.Ethnic identification	3.30	0.96	270	.185	--									
03.Gender	0.68	0.46	248 ¹	0.052	0.063	--								
04.Age	40.28	10.11	270	-0.010	-0.121	-0.061	--							
05.Educational level	3.43	0.77	250 ¹	0.035	-0.043	-0.007	0.041	--						
06.Acculturation_Adaptation	4.40	0.58	270	.167	-0.111	-0.025	.150*	0.042	--					
07.Acculturation_Maintenance	3.25	0.93	270	.249	.689	0.008	-.239	-0.032	-0.073	--				
08.Metastereotyped beliefs	2.81	0.96	270	-.235	.196	-0.008	-0.070	-0.069	-.226	.234	--			
09.Core Self-efficacy	3.45	0.61	270	.346	.164*	0.001	0.066	-0.050	0.115	.162*	-.173	--		
10.Perceived Career Discrimin.	2.33	1.46	270	-0.109	.163*	-0.024	.163*	-0.097	-0.032	.174	.556	-0.045	--	
11.Experienced allyship	2.90	1.02	270	0.095	.253	0.004	-0.080	0.046	-0.090	.266	.212	-0.079	0.106	--

Note. ¹ N for M and SD is 247; N for Spearman correlations of gender and educational level is based on pairwise deletion; N for Pearson correlations on listwise deletion. ²Ethnic origin: 0 = Belgian; 1 = non-Belgian ethnic minority. ³Gender: 0 = man, 1 = women. ⁴ Educational level: 1 = D, 2 = C, 3 =B, 4 = A. * p < .05. p < .01

Career resources

Career resources encompass human capital factors (occupational expertise, human capital development), environmental resources (organisational career support, social career support, mentoring), motivational factors (career confidence, career clarity/goals), and career management behaviours (e.g., career preparation/coaching).

A multiple regression analysis ($n = 1,953$) was conducted to examine career resources, job incumbents' demographics as well as their interaction as potential antecedents of subjective career success (SCS). First, a likelihood ratio test was conducted to evaluate the overall fit of the regression model predicting subjective career success (SCS). The model was statistically significant compared to the intercept-only model, $\chi^2(20) = 915.10$, $p < .001$, indicating that the set of variables reliably improved model fit. See Table 15 for the results.

Second, findings showed that subjective career success (SCS) did not depend on employees' ethnic origin nor their occupational level (A-B-C-D). However, older job incumbents ($B = 0.005$, $SE = 0.001$, $\chi^2(1) = 16.95$, $p < .001$) and women reported higher levels of SCS ($B = -0.093$, $SE = 0.020$, $\chi^2(1) = 22.87$, $p < .001$). Subjective career success (SCS) was positively related to organisational career support ($B = 0.174$, $SE = 0.034$, $\chi^2(1) = 26.74$, $p < .001$), mentoring ($B = 0.179$, $SE = 0.042$, $\chi^2(1) = 18.44$, $p < .001$), occupational expertise ($B = 0.140$, $SE = 0.039$, $\chi^2(1) = 13.25$, $p < .001$).

Except for the marginal interaction effect of ethnic origin with occupational expertise ($p = 0.066$) and career preparation ($p = 0.078$), ethnic background did not moderate these relationships significantly ($ps > .05$), indicating that career resources influence career success similarly across ethnic groups.

Table 15 Regression of SCS on Career Resources and Demographics ($n = 1,953$)

Variables	B	SE	Wald χ^2	df	p
(Intercept)	1.612	0.2047	61.991	1	<.001
Occupational Expertise	0.14	0.0386	13.247	1	<.001
Human Capital development	0.03	0.0402	0.561	1	0.454
Organisational Career Support	0.174	0.0336	26.737	1	<.001
Social Career Support	0.001	0.0422	0.001	1	0.974
Mentoring	0.179	0.0416	18.443	1	<.001
Career Clarity	0.01	0.0351	0.087	1	0.767
Career Preparations	0.041	0.0357	1.339	1	0.247
Ethnic origin (Belgian vs. other)	0.191	0.2063	0.853	1	0.356
Gender (men vs. women)	-0.093	0.0195	22.865	1	<.001
Age	0.003	0.001	12.403	1	<.001
Education Level (DCBA)	0.003	0.0122	0.06	1	0.806
Ethnicity x Occupational Expertise	-0.075	0.0407	3.39	1	0.066
Ethnicity x Human Capital	0	0.0424	0	1	0.992
Ethnicity x Organisational Career Support	-0.049	0.0357	1.903	1	0.168
Ethnicity x Social Career Support	0.023	0.0445	0.271	1	0.603
Ethnicity x Mentoring	0.064	0.0443	2.085	1	0.149
Ethnicity x Career Clarity	0.031	0.0373	0.691	1	0.406
Ethnicity x Career Preparations	-0.067	0.038	3.097	1	0.078

Note. Reference categories were non-Belgian origin (Ethnic origin) and women (Gender).

Career barriers

Career barriers refer to whether employees perceive their ethnic background, gender, or age as obstacles to career progression.

A multiple regression analysis ($n = 2,286$) was conducted to examine career barriers, job incumbents' demographics as well as their interaction as potential antecedents of subjective career success (SCS). The likelihood ratio test showed that the model was statistically significant compared to the intercept-only model, $\chi^2(10) = 256.08$, $p < .001$. See Table 16 for the results.

Subjective career success (SCS) did not depend on employees' ethnic origin but was higher for men ($B = -0.14$, $SE = 0.05$, Wald $\chi^2(1) = 7.28$, $p = .007$), older employees ($B = 0.008$, $SE = 0.003$, Wald $\chi^2(1) = 8.08$, $p = .004$), and those in higher job levels ($B = 0.06$, $SE = 0.015$, Wald $\chi^2(1) = 15.31$, $p < .001$). Only gender-related barriers negatively affected subjective career success ($B = -0.05$, $SE = 0.017$, Wald $\chi^2(1) = 8.68$, $p = .003$), barriers related to ethnicity or age did not.

The negative effect of age-related barriers was stronger among older employees ($B = -0.002$, $SE = 0.001$, Wald $\chi^2(1) = 4.88$, $p = .027$). Other interactions were non-significant ($ps > .05$), suggesting that the impact of ethnic barriers and gender barriers did not affect subjective career success (SCS) across ethnic and gender groups, respectively.

Table 16 Regression of SCS on Career Barriers and Demographics ($n = 2,286$)

Variables	B	SE	Wald χ^2	df	p-value
(Intercept)	3.513	0.1585	491.021	1	< .001
Barriers_ethnicity	-0.047	0.0307	2.383	1	.123
Barriers_age	0.003	0.0492	0.003	1	.958
Barriers_gender	-0.05	0.017	8.676	1	.003
Ethnic origin (Belgian vs. other)	-0.019	0.0847	0.049	1	.825
Gender (men vs. women)	-0.141	0.0524	7.28	1	.007
Age	0.008	0.0028	8.076	1	.004
Education Level (DCBA)	0.06	0.0152	15.312	1	< .001
Ethnicity $\times$ Barriers_ethnicity	-0.007	0.034	0.041	1	.839
Gender $\times$ Barriers_gender	-0.004	0.0245	0.031	1	.860
Age $\times$ Barriers_age	-0.002	0.001	4.876	1	.027

Note. Reference categories were non-Belgian origin (Ethnic origin) and women (Gender).

Organisational DEI-climate

Organisational DEI climate refers to Diversity, Equity, and Inclusion initiatives and perceptions of supervisors' DEI efforts within the organisation.

A multiple regression analysis ($n = 1,953$) was conducted to examine perceptions of organisational DEI initiatives (i.e., organisational DEI climate), job incumbents' demographics as well as their interaction as potential antecedents of subjective career success (SCS). The likelihood ratio test showed that the model was statistically significant compared to the intercept-only model, $\chi^2(12) = 536.57$, $p < .001$. See Table 17 for the results.

Subjective career success was higher for women ($B = -0.083$, $SE = 0.024$, Wald $\chi^2(1) = 11.94$, $p < .001$) and those at higher job levels ($B = 0.049$, $SE = 0.015$, Wald $\chi^2(1) = 10.63$, $p = .001$), with no observed differences related to *ethnic origin*. Perceptions of *organisational diversity* related positively

to subjective career success (SCS) but this effect was marginally significant ($B = 0.097$, $SE = 0.056$, Wald $\chi^2(1) = 2.97$, $p = .085$) and -therefore- should be interpreted with caution ($.05 < p < .10$). Perceptions of *organisational inclusion* were not significantly associated with subjective career success (SCS). Perceptions of *organisational equity* ($B = 0.146$, $SE = 0.039$, Wald $\chi^2(1) = 14.47$, $p < .001$) were positively associated with subjective career success (SCS).

Observing *supervisors implementing DEI principles* was also (marginally) linked to higher subjective career success ($B = 0.080$, $SE = 0.041$, Wald $\chi^2(1) = 3.70$, $p = .054$), with this effect being marginally stronger for employees with a non-Belgian ethnic background ($B = 0.075$, $SE = 0.045$, Wald $\chi^2(1) = 2.78$, $p = .096$) ($.05 < p < .10$, interpret cautiously). Except for the marginal interaction effect of ethnic origin with supervisors' support for DEI ($p = .096$), ethnic origin and its interactions with organisational climate variables were non-significant ($ps > .05$), suggesting that the impact of organisational DEI-climate on subjective career success (SCS) did not differ substantially across ethnic groups in this sample.

Table 17 Regression of SCS on Organisational DEI climate and Demographics ($n = 2,286$)

Variables	B	SE	Wald χ^2	df	p-value
(Intercept)	2.248	0.1651	185.376	1	<.001
Organisational diversity	0.097	0.0562	2.971	1	0.085
Organisational inclusion	0.018	0.0683	0.068	1	0.794
Organisational equity	0.146	0.0385	14.466	1	<.001
Supervisory support for DEI	0.08	0.0414	3.7	1	0.054
Ethnic origin (Belgian vs. other)	-0.058	0.1689	0.117	1	0.732
Gender (men vs. women)	-0.083	0.0241	11.939	1	<.001
Age	0.001	0.0011	1.683	1	0.195
Education Level (DCBA)	0.049	0.0149	10.626	1	0.001
Ethnic origin x Organisational diversity	-0.06	0.0605	0.988	1	0.32
Ethnic origin x Organisational inclusion	-0.036	0.0733	0.243	1	0.622
Ethnic origin x Organisational equity	0.039	0.0404	0.908	1	0.341
Ethnic origin x Supervisory support for DEI	0.075	0.0449	2.775	1	0.096

Note. Reference categories were non-Belgian origin (Ethnic origin) and women (Gender).

Ethnicity-related individual characteristics

Ethnicity-related individual characteristics include acculturation, meta-stereotyped beliefs, perceived career discrimination, experienced ethnic allyship, and core self-efficacy. These variables were only measured among job incumbents with a non-Belgian (or ethnic majority) background.

A multiple regression analysis ($n = 247$) was conducted to examine ethnicity-related individual characteristics, job incumbents' demographics as well as their interaction as potential correlates of subjective career success (SCS). The likelihood ratio test showed that the model was statistically significant compared to the intercept-only model, $\chi^2(16) = 87.32$, $p < .001$.

Subjective career success (SCS) related significantly positively to job level (with higher job levels perceiving more subjective career success, $B = 0.136$, $SE = 0.045$, Wald $\chi^2(1) = 8.99$, $p = .003$) but not to job incumbents' ethnic identification with their ethnic ingroup (other than Belgian), age, and gender. Acculturation (subscale 'cultural adaptation', $B = 0.414$, $SE = 0.172$, Wald $\chi^2(1) = 5.80$, $p = .016$) and

perceived allyship ($B = 0.324$, $SE = 0.118$, Wald $\chi^2(1) = 7.57$, $p = .006$) were positively related to subjective career success (SCS), while negative meta-stereotyped beliefs about one's ethnic ingroup were (marginally) negatively related to SCS ($B = -0.258$, $SE = 0.142$, Wald $\chi^2(1) = 3.32$, $p = .069$). The Acculturation subscale 'cultural maintenance', perceived career discrimination, and core self-efficacy were not significantly related to subjective career success ($ps > .05$).

Interaction effects of these ethnicity-related individual factors with the strength of ethnic ingroup identification revealed that a stronger identification with one's ethnic group weakened the positive impact of allyship on subjective career success ($B = -0.080$, $SE = 0.035$, Wald $\chi^2(1) = 5.32$, $p = .021$) as well as acculturation – subscale 'cultural adaptation' ($B = -0.092$, $SE = 0.052$, Wald $\chi^2(1) = 3.17$, $p = .075$) on subjective career success. Note that the latter had a marginally significant effect only ($.05 < p < .10$). The other interactions of the ethnicity-related individual difference variables with ethnic identification were non-significant ($ps > .05$), suggesting that the impact of ethnicity-related individual difference factors on subjective career success (SCS) did not differ substantially in terms of the strength of ethnic identification with one's non-Belgian ethnic roots.

Table 18 Regression of SCS on Ethnicity-related Individual Factors and Demographics (n = 247)

<i>Variables</i>	<i>B</i>	<i>SE</i>	<i>Wald χ^2</i>	<i>df</i>	<i>p-value</i>
(Intercept)	-0.293	1.1356	0.067	1	0.796
Acculturation_Adaptation	0.414	0.1718	5.797	1	0.016
Acculturation_Maintenance	0.178	0.135	1.746	1	0.186
Metastereotyped beliefs (origin)	-0.258	0.1415	3.315	1	0.069
Perceived Career Discrimination (origin)	-0.039	0.104	0.139	1	0.709
Experienced Allyship	0.324	0.1177	7.569	1	0.006
Core Self-Efficacy	0.225	0.2412	0.866	1	0.352
Ethnic identification	0.52	0.3422	2.309	1	0.129
Gender (men vs. women)	-0.124	0.0747	2.764	1	0.096
Age	0.001	0.0037	0.058	1	0.81
Education Level (DCBA)	0.136	0.0454	8.988	1	0.003
Ethnic identification × Acculturation_Adaptation	-0.092	0.0515	3.169	1	0.075
Ethnic identification × Acculturation_Maintenance	-0.003	0.0396	0.007	1	0.933
Ethnic identification × Metastereotyped beliefs (origin)	0.028	0.0426	0.423	1	0.516
Ethnic identification × Perceived Career Discrimination (origin)	0.012	0.0301	0.167	1	0.683
Ethnic identification × Experienced Allyship	-0.08	0.0347	5.317	1	0.021
Ethnic identification × Core Self-Efficacy	0.012	0.0697	0.031	1	0.861

Note. Reference categories were non-Belgian origin (Ethnic origin) and women (Gender).

4.7.5 Focus group

Theme 1: Expectations and reflections on career outcomes. Experts in the federal government suspect that ethnic minority employees will experience more career barriers while receiving fewer promotions. Additionally, a coaching expert suspected that underemployment could be an underestimated issue among ethnic minority employees. Indeed, as explained by another expert, ethnic minority representation declines at higher levels in organisations. More specifically, they

perceive ethnic minority workers become increasingly rare in “A4, A5 and mandate functions” – which might be indicative of a glass ceiling. As an explanation, the focus group puts forth the disadvantage ethnic minorities already face in hiring procedures. However, regarding subjective career success, the experts are more hesitant to express expectations because they believe there are a variety of intrapersonal drivers such as “intrinsic motivation”. In the survey, employees report equal subjective career success ($M_{\text{Majority}} = 3.55$, $SD_{\text{Majority}} = .60$; $M_{\text{Minority}} = 3.55$, $SD_{\text{Minority}} = .63$). Being presented with these figures, the experts interpret these findings as being the result of an *inclusive culture* in the federal government. However, the experts also raise concerns that ethnic minority members’ current subjective career success might not be predictive of future perceptions. According to them, current career satisfaction does not inform us of future career barriers and promotion trajectories. Indeed, result showed differences in objective career success such that ethnic minority employees are disadvantaged.

Theme 2: Interventions for career success. The experts agree that the federal government offers plenty of career opportunities in terms of work-life balance and internal mobility. Additionally, the experts are convinced that efforts are being made on the sensibilization of diversity and inclusion. Nevertheless, in the discussion, two barriers emerge as sub-themes. First, the federal government houses many different organisations (e.g., penitentiary institution, administrative departments) which results in an enormous heterogeneity of organisational management, policies, and cultures. In turn, this heterogeneity does not allow for homogenous government-wide career planning. A second and related barrier is that when career planning is organised at the level of units within the federal government, it becomes dependent on the goodwill of individual managers, who primarily wish to retain workers and are hesitant to facilitate career mobility (especially in times of shortages on the labour market, i.e., war on talent).

According to the experts, the way forward consists of introducing *systematic career planning*, because at present, this is far too dependent on the assertiveness of individual employees. Lowering the threshold for career planning would be advantageous for all employees, and ethnic minorities in particular. From a more practical perspective, the experts agree that formalizing career planning meetings, e.g., once every two to three years, would be valuable. Additionally, the federal government could support management in setting up such meetings by organizing workshops (e.g., on giving feedback) and embedding regular career meetings in employment regulation. As such, the experts advocate a *transversal career plan* allowing individual organisations to tailor it to their contexts.

Next, we complement the experts’ perspective with employee ratings of career advancement interventions we identified in the literature (Leimon et al., 2011; and Figure 25 below). Hence, we find that many measures/interventions taken by the federal government are also perceived as valuable by job incumbents (e.g., work-life balance, work leave and flexibility arrangements). Accordingly, the three most important interventions to advance the careers of ethnic minority members are related to work-life balance ($M = 4.18$, $SD = .94$), actions promoting workers’ self-efficacy ($M = 4.06$, $SD = .95$), and a more structural career planning ($M = 3.74$, $SD = .99$). Interestingly, when compared to the responses of ethnic minority workers, there are four domains of actions where ethnic majority members underestimate the value of interventions. More concretely, results from a multivariate variance analysis, followed by a series of one-way ANOVAs showed that ethnic minority workers attributed a relatively higher importance to *role modelling*, $F(1, 2317) = 9.25$, $p = .002$, partial $\eta^2 = .004$, *structural career planning*, $F(1, 2317) = 7.32$, $p = .007$, partial $\eta^2 = .003$, *organisational culture*, $F(1, 2317) = 9.88$, $p = .002$, partial $\eta^2 = .004$, and *work-life balance* interventions, $F(1, 2317) = 4.27$, $p = .039$, partial $\eta^2 = .002$, than ethnic majority workers.

Taken together, these quantitative findings match the experts' discussion. While the literature offers various potent interventions to improve the career outcomes of ethnic minorities, both in our quantitative and qualitative data, the structural organisation of career planning in which individual employees are empowered shows promise. Furthermore, these interventions could be most effective when taking place in an inclusive organisational climate where workers can maintain a healthy work-life balance.

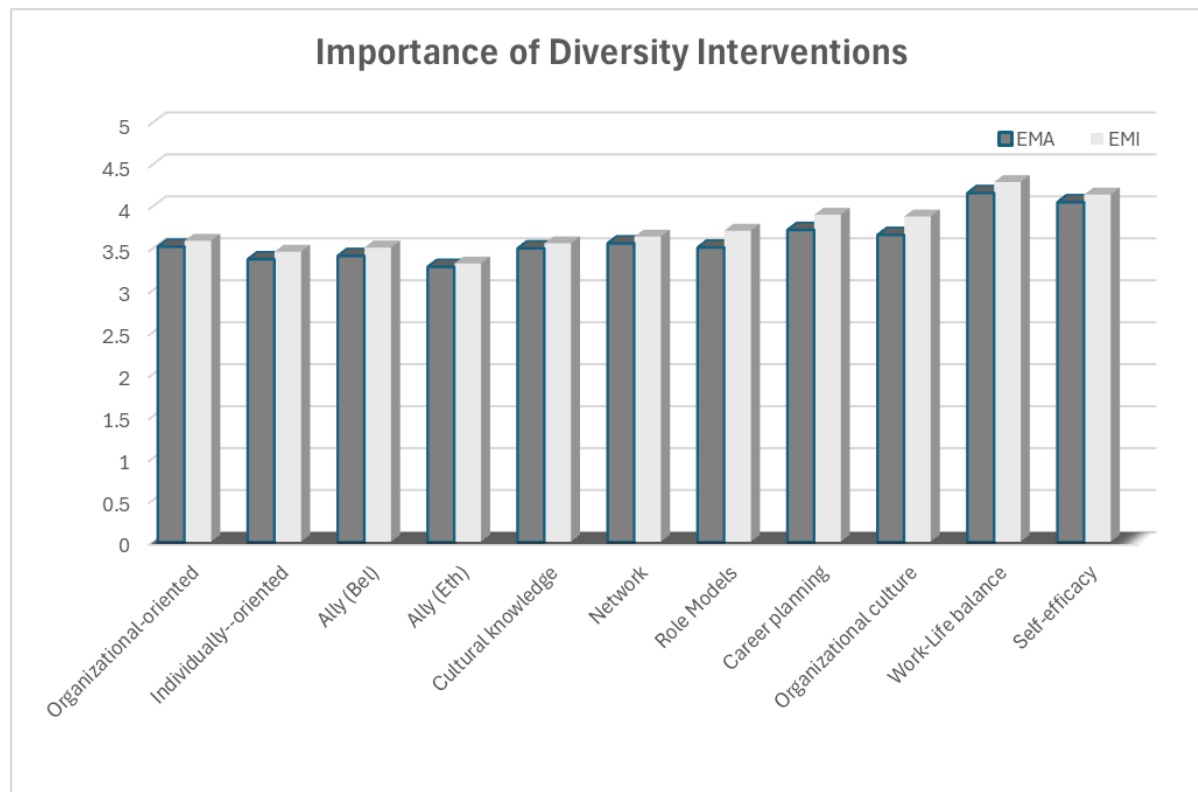


Figure 25 Perceived importance of diversity interventions for career planning for ethnic minority workers (EMI) and ethnic majority workers (EMA)

Note. The light-grey bars represent career interventions valued by ethnic minority workers and the dark-grey bars represent career interventions valued by ethnic majority employees.

4.7.6 Conclusion

Subjective career success was higher for women than for men and tended to be lower for older employees but was not related to employees' ethnic background. Subjective career success was positively associated with organisational support, mentoring, occupational expertise, ethnic acculturation, and perceived allyship, and negatively related to age-related career barriers and meta-stereotyped beliefs about ethnicity. Some of these factors can be influenced by organisational policies and practices.

5. DISSEMINATION AND VALORISATION

- Closing event on the 21st of October 2025, in which we discussed the project findings and their policy implications with 19 representatives from FOD BOSA, FOD WASO, the Flemish government and Unia.
- Conference presentation by Dr. Aylin Koçak about the WP4 survey results, at the *Annual Political Science Workshops of the Low Countries* (22nd Politicologenetmaal) on the 13th of June 2024 – Maastricht University.
- Lecture by Bert Leysen on diversity in the federal public services at the EUPAN Summer School in coordination with FOD BOSA on the 15th of May 2024.

6. PUBLICATIONS

No publications yet, but several working papers are in preparation in order to submit them to academic journals.

7. ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to the many individuals and organizations who made this research possible.

First and foremost, we extend our deepest thanks to the Belgian Science Policy Office (Belspo) for their generous funding of this endeavour and for their belief in the project. We are likewise profoundly grateful to the FPS Strategy and Support (BOSA) for their co-funding and invaluable guidance throughout the project. Their support in data collection and in providing access to critical data was indispensable. It is no exaggeration to say that this project would not have been possible without their extensive collaboration and support.

Our gratitude also extends to the SPF WASO for graciously allowing us to use their algorithms for WP 1.2. We also wish to acknowledge the Crossroads Bank for Social Security for their active and sustained support. Their assistance in troubleshooting dataset inconsistencies over several months was critical to our progress, and we are thankful for the generous use of their facilities.

Finally, we thank our academic partners for their insightful intellectual support and creativity, which greatly enriched the quality of this work.

8. REFERENCES

- Ababneh, K. I., Hackett, R. D., & Schat, A. C. H. (2014). The Role of Attributions and Fairness in Understanding Job Applicant Reactions to Selection Procedures and Decisions. *Journal of Business and Psychology*, 29(1), 111–129.
- Abele, A. E., & Spurk, D. (2009). The longitudinal impact of self-efficacy and career goals on objective and subjective career success. *Journal of Vocational Behavior*, 74(1), 53–62.
- Abraham, L., Hallermeier, J., & Stein, A. (2024). Words matter: Experimental evidence from job applications. *Journal of Economic Behavior & Organization*, 225, 348–391.
- Abu Ghazaleh, N. (2012). *Perceived discrimination*. University of Amsterdam.
- Ahmed, P., Feliciano, C., & Emigh, R. J. (2007). Internal and external ethnic assessments in eastern Europe. *Social Forces*, 86(1), 231–255.
- Aiman-Smith, L., Bauer, T. N., & Cable, D. M. (2001). Are You Attracted? Do You Intend to Pursue? A Recruiting Policy-Capturing Study. *Journal of Business & Psychology*, 16(2), 219–238.
- Akaliyski, P., Welzel, C., Bond, M. H., & Minkov, M. (2021). On “Nationology”: The Gravitational Field of National Culture. *Journal of Cross-Cultural Psychology*, 52(8–9), 771–793.
- Alanis, J. M., & Ryan, A. M. (2023). More than visas and paperwork: Post-hire HRM consideration of immigrant employees. In: Dianna L. Stone, Kimberly M. Lukaszewski, Julio C. Canedo, Brian Murray, & Jams, H. Dulebohn (Eds). *The plight of stigmatized groups in organizations* (pp. 109-140). Information Age Publishing Inc.
- Anderson, N. (2011). Perceived Job Discrimination: Toward a model of applicant propensity to case initiation in selection. *International Journal of Selection and Assessment*, 19(3), 229–244.
- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant Reactions in Selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18(3), 291–304.
- Anderson, N., & Witvliet, C. (2008). Fairness Reactions to Personnel Selection Methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, 16(1), 1–13.
- Apers, C., Lang, J. W. B., & Derous, E. (2019). Who earns more? Explicit traits, implicit motives and income growth trajectories. *Journal of Vocational Behavior*, 110, 214–228.
- Apfelbaum, E. P., Norton, M. I., & Sommers, S. R. (2012). Racial Color Blindness: Emergence, Practice, and Implications. *Current Directions in Psychological Science*, 21(3), 205–209.
- Arthur Jr., W., Glaze, R. M., Jarrett, S. M., White, C. D., Schurig, I., & Taylor, J. E. (2014). Comparative evaluation of three situational judgment test response formats in terms of construct-related validity, subgroup differences, and susceptibility to response distortion. *Journal of Applied Psychology*, 99(3), 535–545.
- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695–716.
- Aryee, S., Budhwar, P. S., & Chen, Z. X. (2002). Trust as a mediator of the relationship between organizational justice and work outcomes: Test of a social exchange model. *Journal of Organizational Behavior*, 23(3), 267–285.
- Aspinall, P. J. (2002). Collective Terminology to Describe the Minority Ethnic Population: The Persistence of Confusion and Ambiguity in Usage. *Sociology*, 36(4), 803–816.
- Aspinall, P. J. (2012). Answer Formats in British Census and Survey Ethnicity Questions: Does Open Response Better Capture ‘Superdiversity’? *Sociology*, 46(2), 354–364.

- Asseburg, J., Hattke, J., Hensel, D., Homberg, F., & Vogel, R. (2020). The Tacit Dimension of Public Sector Attraction in Multi-Incentive Settings. *Journal of Public Administration Research and Theory*, 30(1), 41–59.
- Asseburg, J., & Homberg, F. (2020). Public Service Motivation or Sector Rewards? Two Studies on the Determinants of Sector Attraction. *Review of Public Personnel Administration*, 40(1), 82–111.
- Asseburg, J., Homberg, F., & Vogel, R. (2018). Recruitment messaging, environmental fit and public service motivation: Experimental evidence on intentions to apply for public sector jobs. *International Journal of Public Sector Management*, 31(6), 689–709.
- Atzmüller, C., & Steiner, P. M. (2010). Experimental Vignette Studies in Survey Research. *Methodology*, 6(3), 128–138.
- Auspurg, K., & Hinz, T. (2014). *Factorial Survey Experiments*. SAGE Publications.
- Auspurg, K., Hinz, T., Liebig, S., & Sauer, C. (2014). The Factorial Survey as a Method for Measuring Sensitive Issues. In *Improving Survey Methods*. Routledge.
- Avery, D. R., Hall, A. V., Preston, M., Ruggs, E. N., & Washington, E. (2023). Is Justice Colorblind? A Review of Workplace Racioethnic Differences Through the Lens of Organizational Justice. *Annual Review of Organizational Psychology and Organizational Behavior*, 10(Volume 10, 2023), 389–412.
- Avery, D. R., Volpone, S. D., Stewart, R. W., Luksyte, A., Hernandez, M., McKay, P. F., & Hebl, M. (Mikki) R. (2013). Examining the Draw of Diversity: How Diversity Climate Perceptions Affect Job-Pursuit Intentions. *Human Resource Management*, 52(2), 175–193.
- Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, 22, 273–285.
- Ballinger, T., & Crocker, J. (2021). Understanding Whites’ perceptions of multicultural policies: A (non)zero-sum framework? *Journal of Personality and Social Psychology*, 120(5), 1231–1260.
- Barber, A. E. (1998). *Recruiting employees: Individual and organizational perspectives*. Sage Publications.
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the Selection Procedural Justice Scale (SPJS). *Personnel Psychology*, 54(2), 387–419.
- Becton, J. B., Feild, H. S., Giles, W. F., & Jones-Farmer, A. (2008). Racial differences in promotion candidate performance and reactions to selection procedures: A field study in a diverse top-management context. *Journal of Organizational Behavior*, 29(3), 265–285.
- Beirens, K., & Fontaine, J. R. J. (2011). Somatic and Emotional Well-Being Among Turkish Immigrants in Belgium: Acculturation or Culture? *Journal of Cross-Cultural Psychology*, 42(1), 56–74.
- Benditt, L. (2015). Race, gender, and public-sector work: Prioritizing occupational values as a labor market privilege. *Research in Social Stratification and Mobility*, 42, 73–86.
- Berry, J. W., Poortinga, Y. H., & Pandey, J. (1997). *Handbook of Cross-cultural Psychology*. John Berry.
- Blickle, G., Schütte, N., & Wihler, A. (2018). Political will, work values, and objective career success: A novel approach – The Trait-Reputation-Identity Model. *Journal of Vocational Behavior*, 107, 42–56.
- Bobko, P., & Roth, P. L. (2013). Reviewing, Categorizing, and Analyzing the Literature on Black–White Mean Differences for Predictors of Job Performance: Verifying Some perceptions and Updating/Correcting Others. *Personnel Psychology*, 66(1), 91–126.
- Bond, M. H., & Lun, V. M.-C. (2014). Citizen-making: The role of national goals for socializing children. *Social Science Research*, 44, 75–85.

- Booth, M. Z., Gerard, J. M., Deom, G. M., & Frey, C. J. (2022). Deconstructing adolescent ethnicity: A longitudinal analysis of ethnic-racial identity development and self-categorization. *Journal of Adolescence*, 94(3), 366–379.
- Boven, N., Exeter, D., Sporle, A., & Shackleton, N. (2020). The implications of different ethnicity categorisation methods for understanding outcomes and developing policy in New Zealand. *Kotuitui-New Zealand Journal of Social Sciences online*, 15(1), 123–139.
- Breaugh, J., Ritz, A., & Alfes, K. (2018). Work motivation and public service motivation: Disentangling varieties of motivation and job satisfaction. *Public Management Review*, 20(10), 1423–1443.
- Briscoe, J. P., Kaše, R., Dries, N., Dysvik, A., Unite, J. A., Adeleye, I., Andresen, M., Apospori, E., Babalola, O., Bagdadli, S., Çakmak-Otluglu, K. Ö., Casado, T., Cerdin, J.-L., Cha, J.-S., Chudzikowski, K., Dello Russo, S., Eggenhofer-Rehart, P., Fei, Z., Gianecchini, M., ... Zikic, J. (2021). Here, there, & everywhere: Development and validation of a cross-culturally representative measure of subjective career success. *Journal of Vocational Behavior*, 130, 103612.
- Brown, S. D., Unger Hu, K. A., Mevi, A. A., Hedderson, M. M., Shan, J., Quesenberry, C. P., & Ferrara, A. (2014). The Multigroup Ethnic Identity Measure-Revised: Measurement invariance across racial and ethnic groups. *Journal of Counseling Psychology*, 61(1), 154–161.
- Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2012). The Effect of Administrative Burden on Bureaucratic Perception of Policies: Evidence from Election Administration. *Public Administration Review*, 72(5), 741–751.
- Burton, J., Nandi, A., & Platt, L. (2010). Measuring ethnicity: Challenges and opportunities for survey research. *Ethnic and Racial Studies*, 33(8), 1332–1349.
- Cable, D. M., & Edwards, J. R. (2004). Complementary and Supplementary Fit: A Theoretical and Empirical Integration. *Journal of Applied Psychology*, 89(5), 822–834.
- Cable, D. M., & Judge, T. A. (1996). Person–Organization Fit, Job Choice Decisions, and Organizational Entry. *Organizational Behavior and Human Decision Processes*, 67(3), 294–311.
- Cabrera, M. A. M., & Nguyen, N. T. (2001). Situational Judgment Tests: A Review of Practice and Constructs Assessed. *International Journal of Selection and Assessment*, 9(1–2), 103–113.
- Callister, P., Didham, R., Potter, D., & Blakely, T. (2007). Measuring ethnicity in New Zealand: Developing tools for health outcomes analysis. *Ethnicity & Health*, 12(4), 299–320.
- Cascio, W. F. (2006). Global Performance Management Systems. In *Handbook of Research in International Human Resource Management*. Edward Elgar Publishing.
- Celani, A., & Singh, P. (2011). Signaling theory and applicant attraction outcomes. *Personnel Review*, 40(2), 222–238.
- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82(1), 143–159.
- Chandra, K., & Wilkinson, S. (2008). Measuring the effect of “ethnicity.” *Comparative Political Studies*, 41(4–5), 515–563.
- Chen, C.-A., & Rainey, H. G. (2014). Personnel formalization and the enhancement of teamwork: A public–private comparison. *Public Management Review*, 16(7), 945–968.
- Cheong, C. W. H., & Sinnakkannu, J. (2014). Ethnic diversity and firm financial performance: Evidence from Malaysia. *Journal of Asia-Pacific Business*, 15(1), 73–100.

- Chimienti, M., Guichard, E., Bolzman, C., & Le Goff, J.-M. (2021). How can we categorise 'nationality' and 'second generation' in surveys without (re)producing stigmatisation? *Comparative Migration Studies*, 9(1).
- Chor, D., Pereira, A., Pacheco, A. G., Santos, R. V., Fonseca, M. J. M., Schmidt, M. I., Duncan, B. B., Barreto, S. M., Aquino, E. M. L., Mill, J. G., Molina, M. C. B., Giatti, L., Almeida, M. C. C., Bensenor, I., & Lotufo, P. A. (2019). Context-dependence of race self-classification: Results from a highly mixed and unequal middle-income country. *PLOS ONE*, 14(5).
- Cieciuch, J. (2017). Exploring the complicated relationship between values and behaviour. In S. Roccas & L. Sagiv (Eds.), *Values and behavior: Taking a cross-cultural perspective* (pp. 237–248). Springer.
- Cieciuch, J., Davidov, E., & Algesheimer, R. (2016). The stability and change of value structure and priorities in childhood: A longitudinal study. *Social Development*, 25(3), 503–527.
- Collier-Spruel, L. A., & Ryan, A. M. (2022). Are all allyship attempts helpful? An investigation of effective and ineffective allyship. *Journal of Business and Psychology*, 39(1), 83–108.
- Collmus, A. B., & Landers, R. N. (2019). Game-framing to improve applicant perceptions of cognitive assessments. *Journal of Personnel Psychology*, 18(3), 157–162.
- Cox, T. (1994). *Cultural diversity in organizations: Theory, research and practice*. Berrett-Koehler Publishers.
- Croll, P. R., & Gerteis, J. (2019). Race as an open field: Exploring identity beyond fixed choices. *Sociology of Race and Ethnicity*, 5(1), 55–69.
- Day-Vines, N. L., Brodar, J. R., Hicks, D., Fernandez-Korto, E. B., Garcia, C., & Jones, K. (2022). An investigation of the relationship between school counselor trainees' broaching behavior and their racial identity attitudes. *Journal of Counseling and Development*, 100(1), 3–13.
- De Leng, W. E., Stegers-Jager, K. M., Born, M. Ph., & Themmen, A. P. N. (2018). Influence of response instructions and response format on applicant perceptions of a situational judgment test for medical school selection. *BMC Medical Education*, 18(1), 282.
- De Zwart, F. (2012). Pitfalls of top-down identity designation: Ethno-statistics in the Netherlands. *Comparative European Politics*, 10(3), 301–318.
- Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology*, 74(4), 580–590.
- DeHart-Davis, L., Marlowe, J., & Pandey, S. K. (2006). Gender dimensions of public service motivation. *Public Administration Review*, 66(6), 873–887.
- Dembosky, J. W., Haval, A. M., Haas, A., Hambarsoomian, K., Weech-Maldonado, R., Wilson-Frederick, S. M., Gaillot, S., & Elliott, M. N. (2019). Indirect estimation of race/ethnicity for survey respondents who do not report race/ethnicity. *Medical Care*, 57(5), e28–e33.
- Deros, E. (2020). Van boe-boe machine tot sociale media: Evidence-based werven en selecteren. In *HRM heden & morgen: Evidence-based practice and practice-based evidence* (pp. 74–142). Vakmedianet.
- Deros, E., & Pepermans, R. (2019). Gender discrimination in hiring: Intersectional effects with ethnicity and cognitive job demands. *Archives of Scientific Psychology*, 7(1), 40–49.
- Ding, F., & Riccucci, N. M. (2023). How does diversity affect public organizational performance? A meta-analysis. *Public Administration*, 101(4), 1367–1393.
- Doaty, J. (2022). *Perceptions of workplace discrimination: A closer look* [Doctoral thesis, University of South Florida].

- Dobewall, H., & Strack, M. (2014). Relationship of Inglehart's and Schwartz's value dimensions revisited. *International Journal of Psychology*, 49(4), 240–248.
- Dossinger, K., Wanberg, C. R., Choi, Y., & Leslie, L. M. (2019). The beauty premium: The role of organizational sponsorship in the relationship between physical attractiveness and early career salaries. *Journal of Vocational Behavior*, 112, 109–121.
- Douglass, S., & Umaña-Taylor, A. J. (2015). A brief form of the Ethnic Identity Scale: Development and empirical validation. *Identity*, 15(1), 48–65.
- Ebner, K., & Paul, K. I. (2023). The Career Decision-Making Adaptability Score: Exploration and validation of its predictive power for subjective and objective career success. *International Journal for Educational and Vocational Guidance*, 23(3), 635–658.
- Elling, R. C., & Harris, K. (2021). Difference in difference: Language, geography, and ethno-racial identity in contemporary Iran. *Ethnic and Racial Studies*, 44(13), 2315–2334.*
- Else-Quest, N. M., & Hyde, J. S. (2016). Intersectionality in quantitative psychological research: II. Methods and techniques. *Psychology of Women Quarterly*, 40(3), 319–336.
- Elsey, V., Van der Heijden, B., Smith, M. A., & Moss, M. (2022). Examining the role of employability as a mediator in the relationship between psychological capital and objective career success amongst occupational psychology professionals. *Frontiers in Psychology*, 13.
- Erikson, E. H. (1968). *Identity: Youth and crisis*. W. W. Norton & Company.
- Esteve, M., Urbig, D., Van Witteloostuijn, A., & Boyne, G. (2016). Prosocial behavior and public service motivation. *Public Administration Review*, 76(1), 177–187.
- Evers, A., Lucassen, W., Meijer, R., & Sijsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. NIP/COTAN.
- Figueiredo Filho, D. B., Paranhos, R., Rocha, E. C. D., Batista, M., Silva Jr., J. A. D., Santos, M. L. W. D., & Marino, J. G. (2013). When is statistical significance not significant? *Brazilian Political Science Review*, 7(1), 31–55.
- Fog, A. (2021). A Test of the Reproducibility of the Clustering of Cultural Variables. *Cross-Cultural Research*, 55(1), 29–57.
- Fog, A. (2023). Two-Dimensional Models of Cultural Differences: Statistical and Theoretical Analysis. *Cross-Cultural Research*, 57(2–3), 115–165.
- Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4), 331–362.
- Gans, H. J. (1979). Symbolic ethnicity: The future of ethnic groups and cultures in America. *Ethnic and Racial Studies*, 2(1), 1–20.
- Gara Bach Ouerdian, E., & Mansour, N. (2019). The relationship of social capital with objective career success: The case of Tunisian bankers. *Journal of Management Development*, 38(2), 74–86.
- Gardner, D. M., & Alanis, J. M. (2020). Together we stand: Ally training for discrimination and harassment reduction. *Industrial and Organizational Psychology*, 13(2), 196–199.
- Gardner, D. M., & Ryan, A. M. (2020). What's in it for you? Demographics and self-interest perceptions in diversity promotion. *Journal of Applied Psychology*, 105(9), 1062–1072.
- Georgellis, Y., Iossa, E., & Tabvuma, V. (2011). Crowding out intrinsic motivation in the public sector. *Journal of Public Administration Research and Theory*, 21(3), 473–493.
- Georgiou, K., Gouras, A., & Nikolaou, I. (2019). Gamification in employee selection: The development of a gamified assessment. *International Journal of Selection and Assessment*, 27(2), 91–103.

- Ghazaleh Abu, N. (2012). *Perceived discrimination: Why applicants and employees expect and perceive discrimination*. University of Amsterdam.
- Gibbs, B. G., Shafer, K., & Dufur, M. J. (2015). Why infer? The use and misuse of population data in sport research. *International Review for the Sociology of Sport*, 50(1), 115–121.
- Gibbs, B. G., Shafer, K., & Miles, A. (2017). Inferential statistics and the use of administrative data in US educational research. *International Journal of Research & Method in Education*, 40(2), 214–220.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review*, 18(4), 694–734.
- Giraud, L., Bernard, A., & Trinchera, L. (2019). Early career values and individual factors of objective career success: The case of the French business graduates. *Career Development International*, 24(4), 350–382.
- Golden, T. D., & Eddleston, K. A. (2020). Is there a price telecommuters pay? Examining the relationship between telecommuting and objective career success. *Journal of Vocational Behavior*, 116, 103348.
- Golubovich, J., & Ryan, A. M. (2021). Performance on video-based situational judgment test items: Simulated interracial interactions. *Journal of Business and Psychology*, 36(4), 693–711.
- Gordon, M. M. (1964). *Assimilation in American life: The role of race, religion, and national origins* (28th print). Oxford University Press.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108.
- Gravetter, F. J., & Wallnau, L. B. (2012). *Essentials of statistics for the behavioral sciences* (8th ed.).
- Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal*, 33(1), 64–86.
- Griffeth, R. W., Tenbrink, A., & Robinson, S. (2014). Recruitment sources: A review of outcomes. In K. Y. T. Yu & D. M. Cable (Eds.), *The Oxford handbook of recruitment* (pp. 215–250). Oxford University Press.
- Groeneveld, S., & Meier, K. J. (2022). Theorizing status distance: Rethinking the micro theories of representation and diversity in public organizations. *Public Administration Review*, 54(2), 29.
- Guenole, N., Englert, P., & Taylor, P. J. (2003). Ethnic group differences in cognitive ability test scores within a New Zealand applicant sample. *New Zealand Journal of Psychology*, 32, 49–54.
- Guillaume, Y. R. F., Dawson, J. F., Priola, V., Sacramento, C. A., Woods, S. A., Higson, H. E., Budhwar, P. S., & West, M. A. (2014). Managing diversity in organizations: An integrative model and agenda for future research. *European Journal of Work and Organizational Psychology*, 23(5), 783–802.
- Harewood, J., & Rosenfield, M. (2021). Defining race and ethnicity in optometry. *Ophthalmic and Physiological Optics*, 41(4), 659–662.
- Hassan, H. A., & Ahmad, A. B. (2021). The relationship between Islamic work ethic and public service motivation. *Administration & Society*, 28, 1–15.
- Hatter, K. B., Avery, D. R., & McKay, P. F. (2024). Organizational diversity climate: Recent scholarship and what's missing. *Current Opinion in Psychology*, 60, 101926.
- Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology*, 57(3), 639–683.
- Hayes, A. F., & Little, T. D. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). The Guilford Press.

- Heck, R. H., & Thomas, S. L. (2020). *An introduction to multilevel modeling techniques: MLM and SEM approaches* (4th ed.). Routledge.
- Herde, C. N., Lievens, F., Jackson, D. J. R., Shalfrooshan, A., & Roth, P. L. (2020). Subgroup differences in situational judgment test scores: Evidence from large applicant samples. *International Journal of Selection and Assessment*, 28(1), 45–54.
- Hiemstra, A. M. F., Derous, E., Serlie, A. W., & Born, M. P. (2013). Ethnicity effects in graduates' résumé content. *Applied Psychology*, 62(3), 427–453.
- Highhouse, S., Lievens, F., & Sinar, E. F. (2003). Measuring attraction to organizations. *Educational and Psychological Measurement*.
- Hirschi, A., Johnston, C. S., De Fruyt, F., Ghetta, A., & Orth, U. (2021). Does success change people? Examining objective career success as a precursor for personality development. *Journal of Vocational Behavior*, 127, 103582.
- Hirschi, A., Nagy, N., Baumeler, F., Johnston, C. S., & Spurk, D. (2018). Assessing key predictors of career success: Development and validation of the Career Resources Questionnaire. *Journal of Career Assessment*, 26(2), 338–358.
- Hirschman, A. O. (1964). The paternity of an index. *The American Economic Review*, 54(5), 761.
- Hoang, T. G., Truxillo, D. M., Erdogan, B., & Bauer, T. N. (2012). Cross-cultural examination of applicant reactions to selection methods: United States and Vietnam. *International Journal of Selection and Assessment*, 20(2), 209–219.
- Hofhuis, J., Van Der Rijt, P. G. A., & Vlug, M. (2016). Diversity climate enhances work outcomes through trust and openness in workgroup communication. *SpringerPlus*, 5(1), 714.
- Hofhuis, J., Van Der Zee, K. I., & Otten, S. (2012). Social identity patterns in culturally diverse organizations: The role of diversity climate. *Journal of Applied Social Psychology*, 42(4), 964–989.
- Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture*, 2(1).
- Holloway-Friesen, H. (2018). Acculturation, enculturation, gender, and college environment on perceived career barriers among Latino/a college students. *Journal of Career Development*, 45(2), 117–131.
- Holmes, O., Jiang, K., Avery, D. R., McKay, P. F., Oh, I.-S., & Tillman, C. J. (2021). A meta-analysis integrating 25 years of diversity climate research. *Journal of Management*, 47(6), 1357–1382.
- Houston, D. J. (2014). Public service motivation in the post-communist state. *Public Administration*, 92(4), 843–860.
- Hox, J. J., & Roberts, J. K. (Eds.). (2011). *Handbook of advanced multilevel analysis*. Routledge, Taylor & Francis Group.
- Inglehart, R. (Ed.). (2004). *Human beliefs and values: A cross-cultural sourcebook based on the 1999–2002 values surveys* (1st ed.). Siglo XXI.
- Inglehart, R., & Welzel, C. (2010). Changing mass priorities: The link between modernization and democracy. *Perspectives on Politics*, 8(2), 551–567.
- ISSP Research Group. (2017). *International Social Survey Programme: Work Orientations IV – ISSP 2015 (Version 2.1.0) [Dataset]*. GESIS Data Archive.
- Jacobs, D., Swyngedouw, M., Hanquinet, L., Vandezande, V., Andersson, R., Horta, A. P. B., Berger, M., Diani, M., Ferrer, A. G., Giugni, M., Morariu, M., Pilati, K., & Statham, P. (2009). The challenge of

measuring immigrant origin and immigration-related ethnicity in Europe. *Journal of International Migration and Integration / Revue de l'Intégration et de la Migration Internationale*, 10(1), 67–88.

Jakobsen, M., Løkke, A.-K., & Keppeler, F. (2023). Facing the human capital crisis: A systematic review and research agenda on recruitment and selection in the public sector. *Public Administration Review*, 83(6), 1635–1648.

Jansen, W. S., Vos, M. W., Otten, S., Podsiadlowski, A., & Van Der Zee, K. I. (2016). Colorblind or colorful? How diversity approaches affect cultural majority and minority employees. *Journal of Applied Social Psychology*, 46(2), 81–93.

Jasinskaja-Lahti, I., Maehoenen, T. A., & Ketokivi, M. (2012). The dynamics of ethnic discrimination, identities and outgroup attitudes: A pre–post longitudinal study of ethnic migrants. *European Journal of Social Psychology*, 42(7), 904–914.

Jenkins, R. (2008). *Rethinking ethnicity* (2nd ed.). SAGE Publications.

Johfre, S. S., Saperstein, A., & Hollenbach, J. A. (2021). Measuring race and ancestry in the age of genetic testing. *Demography*, 58(3), 785–810.

Judge, T. A. (2009). Core self-evaluations and work success. *Current Directions in Psychological Science*, 18(1), 58–62.

Judge, T. A., Bono, J. E., & Locke, E. A. (2000). Personality and job satisfaction: The mediating role of job characteristics. *Journal of Applied Psychology*, 85(2), 237–249.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The Core Self-Evaluations Scale: Development of a measure. *Personnel Psychology*, 56(2), 303–331.

Judge, T. A., Klinger, R. L., & Simon, L. S. (2010). Time is on my side: Time, general mental ability, human capital, and extrinsic career success. *Journal of Applied Psychology*, 95(1), 92–107.

Kanning, U. P., Grewe, K., Hollenberg, S., & Hadouch, M. (2006). From the subjects' point of view. *European Journal of Psychological Assessment*, 22(3), 168–176.

Karakas, S., & Özbilgin, M. F. (2019). Reflections on definitions, methods, challenges of and ways forward for ethnic counting in Europe. *International Perspectives on Equality, Diversity and Inclusion*, 6, 15–34.

Keppeler, F., & Papenfuß, U. (2020). Employer branding and recruitment: Social media field experiments targeting future public employees. *Public Administration Review*, 81(4), 763–775.

Kiazad, K., Kraimer, M. L., Seibert, S. E., & Sargent, L. (2020). Understanding organizational embeddedness and career success: Who and what you know. *Journal of Organizational Behavior*, 41(7), 678–693.

Kim, J. (2018). The contrary effects of intrinsic and extrinsic motivations on burnout and turnover intention in the public sector. *International Journal of Manpower*, 39(3), 486–500.

Kim, S. (2017). National culture and public service motivation: Investigating the relationship using Hofstede's five cultural dimensions. *International Review of Administrative Sciences*, 83(1_suppl), 23–40.

Kirchmeyer, C. (2002). Gender differences in managerial careers: Yesterday, today, and tomorrow. *Journal of Business Ethics*, 37(1), 5–24.

Kirchmeyer, C. (2006). The different effects of family on objective career success across gender: A test of alternative explanations. *Journal of Vocational Behavior*, 68(2), 323–346.

Koçak, A., & Derous, E. (2022). *Onderzoek naar de doorstroom van vrouwen naar topkader binnen de Vlaamse overheid: Rapport met resultaten van een kwalitatieve interviewstudie en een experimentele studie bij vrouwen binnen de Vlaamse overheid*.

Koçak, A., Derous, E., Born, M. Ph., & Duyck, W. (2023). What (not) to add in your ad: When job ads discourage older or younger job seekers to apply. *International Journal of Selection and Assessment*, 31(1), 92–104.

Koçak, A., Rooman, C., Duyck, W., & Derous, E. (2022). Mind the ad: How personality requirements affect older and younger job seekers' job attraction. *Journal of Occupational & Organizational Psychology*, 95(3), 718–737.

Korac, S., Saliterer, I., & Weigand, B. (2019). Factors affecting the preference for public sector employment at the pre-entry level: A systematic review. *International Public Management Journal*.

Kukutai, T., Thompson, V., & McMillan, R. (2015). Whither the census? Continuity and change in census methodologies worldwide, 1985–2014. *Journal of Population Research*, 32(1), 3–22.

Kumazawa, R. (2010). Promotion speed and its effect on attrition of Navy-enlisted personnel: Addressing heterogeneity in high school credentials. *Applied Economics*, 42(20), 2563–2576.

Kutlaca, M., & Radke, H. R. M. (2023). Towards an understanding of performative allyship: Definition, antecedents and consequences. *Social and Personality Psychology Compass*, 17(2), e12724.

Lachowsky, N. J., Saxton, P. J. W., Dickson, N. P., Hughes, A. J., Jones, R. G., Clark, T. C., Ho, E., Summerlee, A. J. S., & Dewey, C. E. (2020). Ethnicity classification systems for public health surveys: Experiences from HIV behavioural surveillance among men who have sex with men. *BMC Public Health*, 20(1).

Lee, G., & Choi, D. L. (2016). Does public service motivation influence college students' intention to work in the public sector? Evidence from Korea. *Review of Public Personnel Administration*, 36(2), 145–163.

Leibbrandt, A., & List, J. A. (2018). Do equal employment opportunity statements backfire? Evidence from a natural field experiment on job-entry decisions (No. w25035). *National Bureau of Economic Research*.

Leimon, A., Moscovici, F., & Goodier, H. (2011). *Coaching women to lead: A systematic approach to coaching women for success*. Routledge.

Leisink, P., & Steijn, B. (2008). Recruitment, attraction, and selection. In J. L. Perry & A. Hondeghem (Eds.), *Motivation in public management: The call of public service* (pp. 118–135). Oxford University Press.

Leivas, P. H. S., & dos Santos, A. M. A. (2018). Horizontal inequality and ethnic diversity in Brazil: Patterns, trends, and their impacts on institutions. *Oxford Development Studies*, 46(3), 348–362.

Leong, F. T. L. (Ed.). (2014). *Career development and vocational behavior of racial and ethnic minorities* (0 ed.). Routledge.

Leslie, L. M., Bono, J. E., Kim, Y. (Sophia), & Beaver, G. R. (2020). On melting pots and salad bowls: A meta-analysis of the effects of identity-blind and identity-conscious diversity ideologies. *Journal of Applied Psychology*, 105(5), 453–471.

Leutner, F., Codreanu, S.-C., Brink, S., & Bitsakis, T. (2023). Game-based assessments of cognitive ability in recruitment: Validity, fairness and test-taking experience. *Frontiers in Psychology*, 13.

Leutner, F., Codreanu, S.-C., Liff, J., & Mondragon, N. (2021). The potential of game- and video-based assessments for social attributes: Examples from practice. *Journal of Managerial Psychology*, 36(7), 533–547.

- Lewis, G. B., & Frank, S. A. (2002). Who wants to work for the government? *Public Administration Review*, 62(4), 395–404.
- Leysen, B., Verhaeghe, P.-P., & Derous, E. (2025). Even grounds: Assessing discrimination in recruitment in the public sector. *International Journal of Public Sector Management*, 1–19.
- Li, T., Tien, H.-L. S., Gu, J., & Wang, J. (2023). The relationship between social support and career adaptability: the chain mediating role of perceived career barriers and career maturity. *International Journal for Educational and Vocational Guidance*, 23(2), 319–336.
- Lie, J. (2008). *Zainichi (Koreans in Japan): Diasporic nationalism and postcolonial identity*. University of California Press.
- Liebler, C. A., & Hou, F. (2020). Churning races in Canada: Visible minority response change between 2006 and 2011. *Social Science Research*, 86.
- Lievens, F., & Highhouse, S. (2003). The relation of instrumental and symbolic attributes to a company's attractiveness as an employer. *Personnel Psychology*, 56(1), 75–102.
- Lievens, F., & Motowidlo, S. J. (2016). Situational Judgment Tests: From Measures of Situational Judgment to Measures of General Domain Knowledge. *Industrial and Organizational Psychology*, 9(1), 3–22.
- Linos, E. (2018). More Than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police. *Journal of Public Administration Research and Theory*, 28(1), 67–85.
- Linos, E., & Riesch, N. (2020). Thick Red Tape and the Thin Blue Line: A Field Study on Reducing Administrative Burden in Police Recruitment. *Public Administration Review*, 80(1), 92–103.
- Lippens, L., Vermeiren, S., & Baert, S. (2023). The state of hiring discrimination: A meta-analysis of (almost) all recent correspondence experiments. *European Economic Review*, 151, 104315.
- Lipshits-Braziler, Y., & Tatar, M. (2012). Perceived career barriers and coping among youth in Israel: Ethnic and gender differences. *Journal of Vocational Behavior*, 80(2), 545–554.
- Lucey, C. R., & Saguil, A. (2020). The Consequences of Structural Racism on MCAT Scores and Medical School Admissions: The Past Is Prologue. *Academic Medicine*, 95(3), 351.
- Luzzo, D. A., & McWhirter, E. H. (2001). Sex and Ethnic Differences in the Perception of Educational and Career-Related Barriers and Levels of Coping Efficacy. *Journal of Counseling & Development*, 79(1), 61–67.
- Macan, T. H., Avedon, M. J., Paese, M., & Smith, D. E. (1994). The effects of applicants' reactions to cognitive ability tests and an assessment center. *Personnel Psychology* 47, 715–738
- Maehler, D. B., Zabal, A., & Hanke, K. (2019). Adults' Identity in Acculturation Settings: The Multigroup Ethnic & National Identity Measure (MENI). *Identity*, 19(4), 245–257.
- Martínez Guzmán, J. P., Jordan, M. M., & Joyce, P. G. (2023). Towards inclusive public administration systems: Public budgeting from the perspective of critical race theory. *Public Administration*, 1–18.
- McAuley, E., Duncan, T. E., & Russell, D. W. (1992). Measuring Causal Attributions: The Revised Causal Dimension Scale (CDSII). *Personality and Social Psychology Bulletin*, 18(5), 566–573.
- McCandless, S., Bishu, S. G., Gómez Hernández, M., Paredes Eraso, É., Sabharwal, M., Santis, E. L., & Yates, S. (2022). A long road: Patterns and prospects for social equity, diversity, and inclusion in public administration. *Public Administration*, 100(1), 129–148.
- McCarthy, J. M., Bauer, T. N., Truxillo, D. M., Anderson, N. R., Costa, A. C., & Ahmed, S. M. (2017). Applicant Perspectives During Selection: A Review Addressing “So What?,” “What’s New?,” and “Where to Next?” *Journal of Management*, 43(6), 1693–1725.

- McCord, M. A., Joseph, D. L., Dhanani, L. Y., & Beus, J. M. (2018). A meta-analysis of sex and race differences in perceived workplace mistreatment. *Journal of Applied Psychology*, 103(2), 137–163.
- McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology*, 96(2), 327–336.
- McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW), 72:1-72:23.
- McFarland, L. A., & Kim, Y. (2021). An examination of the relationship between applicant race and accrued recruitment source information: Implications for applicant withdrawal and test performance. *Personnel Psychology*, 74(4), 831–861.
- McKay, P. F., Avery, D. R., & Morris, M. A. (2008). Mean Racial-Ethnic Differences in Employee Sales Performance: The Moderating Role of Diversity Climate. *Personnel Psychology*, 61(2), 349–374.
- McKay, P. F., Avery, D. R., Tonidandel, S., Morris, M. A., Hernandez, M., & Hebl, M. R. (2007). Racial Differences in Employee Retention: Are Diversity Climate Perceptions the Key? *Personnel Psychology*, 60(1), 35–62.
- McManus, I., Dewberry, C., Nicholson, S., Dowell, J. S., Woolf, K., & Potts, H. W. (2013). Construct-level predictive validity of educational attainment and intellectual aptitude tests in medical student selection: Meta-regression of six UK longitudinal studies. *BMC Medicine*, 11(1), 243.
- Mejia-Smith, B., & Gushue, G. V. (2017). Latina/o College Students' Perceptions of Career Barriers: Influence of Ethnic Identity, Acculturation, and Self-Efficacy. *Journal of Counseling & Development*, 95(2), 145–155.
- Mergel, I., Bellé, N., & Nasi, G. (2021). Prosocial Motivation of Private Sector IT Professionals Joining Government. *Review of Public Personnel Administration*, 41(2), 338–357.
- Minkov, M. (2017). A revision of Hofstede's model of national culture: Old evidence and new data from 56 countries. *Cross Cultural & Strategic Management*, 25(2), 231–256.
- Minkov, M., & Kaasa, A. (2020). A test of Hofstede's model of culture following his own approach. *Cross Cultural & Strategic Management*, 28(2), 384–406.
- Minkov, M., & Kaasa, A. (2021). A Test of the Revised Minkov-Hofstede Model of Culture: Mirror Images of Subjective and Objective Culture across Nations and the 50 US States. *Cross-Cultural Research*, 55(2–3), 230–281.
- Minkov, M., & Kaasa, A. (2022). Do dimensions of culture exist objectively? A validation of the revised Minkov-Hofstede model of culture with World Values Survey items and scores for 102 countries. *Journal of International Management*, 28(4), 100971.
- Minkov, M., & Kaasa, A. (2024). Aligning Schwartz's model of culture with that of Minkov-Hofstede. *International Journal of Cross Cultural Management*, 24(1), 129–148.
- Mor Barak, M. E., Cherin, D. A., & Berkman, S. (1998). Organizational and personal dimensions in diversity climate: Ethnic and gender differences in employee perceptions. *The Journal of Applied Behavioral Science*, 34(1), 82–104.
- Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen–state interactions. *Journal of Public Administration Research and Theory*, 25(1), 43–69.
- Mutz, D. C. (2011). *Population-based survey experiments*. Princeton University Press.

- Nelissen, P. T. J. H., Hülshager, U. R., van Ruitenbeek, G. M. C., & Zijlstra, F. R. H. (2017). Lending a helping hand at work: A multilevel investigation of prosocial motivation, inclusive climate, and inclusive behavior. *Journal of Occupational Rehabilitation*, 27(3), 467–476.
- Neumark, D. (2016). *Experimental research on labor market discrimination (Working Paper Series No. 22022)*. National Bureau of Economic Research.
- Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and subjective career success: A meta-analysis. *Personnel Psychology*, 58(2), 367–408.
- Nichols, A. D., Axt, J., Gosnell, E., & Ariely, D. (2023). A field study of the impacts of workplace diversity on the recruitment of minority group members. *Nature Human Behaviour*, 7(12), 2212–2227.
- Nishii, L. H. (2013). The benefits of climate for inclusion for gender-diverse groups. *Academy of Management Journal*, 56(6), 1754–1774.
- O’Connell, H. A., Bratter, J. L., & Casarez, R. S. (2022). One drop on the move: Historical legal context, racial classification, and migration. *Ethnic and Racial Studies*, 45(5), 809–828.
- Ohlms, M. L., Voigtländer, E., Melchers, K. G., & Kanning, U. P. (2024). Is gamification a suitable means to improve applicant reactions and convey information during an online test? *Journal of Personnel Psychology*, 23(4), 169–178.
- Olsen, J. E., & Martins, L. L. (2016). Racioethnicity, community makeup, and potential employees’ reactions to organizational diversity management approaches. *Journal of Applied Psychology*, 101(5), 657–672.
- Oostrom, J. K., Born, M. Ph., Serlie, A. W., & Van Der Molen, H. T. (2010). Effects of individual differences on the perceived job relatedness of a cognitive ability test and a multimedia situational judgment test. *International Journal of Selection and Assessment*, 18(4), 394–406.
- Oostrom, J. K., & De Soete, B. (2016). Ethnic differences in perceptions of cognitive ability tests: The explanatory role of self-serving attributions. *International Journal of Selection and Assessment*, 24(1), 14–23.
- Outtz, J. L. (Ed.). (2010). *Adverse impact: Implications for organizational staffing and high-stakes selection*. Routledge.
- Owuamalam, C. K., & Zagefka, H. (2014). On the psychological barriers to the workplace: When and why metastereotyping undermines employability beliefs of women and ethnic minorities. *Cultural Diversity and Ethnic Minority Psychology*, 20(4), 521–528.
- Pandey, S. K., & Stazyk, E. (2008). Antecedents and correlates of public service motivation. In J. L. Perry & A. Hondeghem (Eds.), *Motivation in public management: The call of public service* (pp. 101–117). Oxford University Press.
- Park, S. M., & Word, J. (2012). Driven to serve: Intrinsic and extrinsic motivation for public and nonprofit managers. *Public Personnel Management*, 41(4), 705–734.
- Parola, H. R., Harari, M. B., Herst, D. E. L., & Prysmakova, P. (2019). Demographic determinants of public service motivation: A meta-analysis of PSM–age and –gender relationships. *Public Management Review*, 21(10), 1397–1419.
- Perry, J. L. (2000). Bringing society in: Toward a theory of public service motivation. *Journal of Public Administration Research and Theory*, 10(2), 471–488.
- Perry, J. L., & Vandenabeele, W. (2015). Public service motivation research: Achievements, challenges, and future directions. *Public Administration Review*, 75(5), 692–699.
- Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. *Public Administration Review*, 50(3), 367–373.

- Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences*. Blackwell Publishing.
- Phinney, J. S. (1992). The Multigroup Ethnic Identity Measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, 7(2), 156–176.
- Phinney, J. S., & Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54(3), 271–281.
- Piatak, J. S., Sowa, J. E., Jacobson, W. S., & McGinnis Johnson, J. (2020). Infusing public service motivation (PSM) throughout the employment relationship: A review of PSM and the human resource management process. *International Public Management Journal*, 24(1), 86–105.
- Pickett, R. E. M., Saperstein, A., & Penner, A. M. (2019). Placing Racial Classification in Context. *Socius*, 5.
- Plaut, V. C., Thomas, K. M., & Goren, M. J. (2009). Is Multiculturalism or Color Blindness Better for Minorities? *Psychological Science*, 20(4), 444–446.
- Ployhart, R. E., & Harold, C. M. (2004). The Applicant Attribution-Reaction Theory (AART): An Integrative Theory of Applicant Attributional Processing. *International Journal of Selection and Assessment*, 12(1–2), 84–98.
- Ployhart, R. E., & Holtz, B. C. (2008). The Diversity–Validity Dilemma: Strategies for Reducing Racioethnic and Sex Subgroup Differences and Adverse Impact in Selection. *Personnel Psychology*, 61(1), 153–172.
- Ployhart, R. E., & Kim, Y. (2013). Strategic Recruiting. In K. Y. T. Yu & D. M. Cable, *The Oxford Handbook of Recruitment* (pp. 5–20). Oxford University Press.
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2005). *Staffing Organizations: Contemporary Practice and Theory* (3rd ed.). CRC Press.
- Posner, D. (2004). Measuring ethnic fractionalization in Africa. *American Journal of Political Science*, 48(4), 849–863.
- Pulakos, E. D., & Schmitt, N. (1996). An Evaluation of Two Strategies for Reducing Adverse Impact and Their Effects on Criterion-Related Validity. *Human Performance*, 9(3), 241–258.
- Pyburn Jr, K. M., Ployhart, R. E., & Kravitz, D. A. (2008). The Diversity–Validity Dilemma: Overview and Legal Context. *Personnel Psychology*, 61(1), 143–151.
- Rademakers, R., & van Hoorn, A. (2021). Ethnic switching: Longitudinal evidence on prevalence, correlates, and implications for measuring ethnic segregation. *Journal of Development Economics*, 152.
- Rainey, H. G. (1982). Reward Preferences among Public and Private Managers: In Search of the Service Ethic. *The American Review of Public Administration*, 16(4), 288–302.
- Ratuvu, S. (Ed.). (2019). *The Palgrave Handbook of Ethnicity*. Springer Singapore.
- Rechlin, A. M. (2016). *The development and validation of the Mentoring Functions Measure—ProQuest* [Doctoral thesis, Colorado State University].
- Rijsdijk, J., Born, M. P., Koch, B. P. N., & Hiemstra, A. M. F. (2021). De validering van een virtual reality situationele beoordelingstest (VR-SJT) voor leidinggeven. *Gedrag & Organisatie*, 34(1), 107–139.
- Ritz, A., & Brewer, G. A. (2013). Does Societal Culture Affect Public Service Motivation? Evidence of Sub-national Differences in Switzerland. *International Public Management Journal*, 16(2), 224–251.
- Ritz, A., & Waldner, C. (2011). Competing for Future Leaders: A Study of Attractiveness of Public Sector Organizations to Potential Job Applicants. *Review of Public Personnel Administration*, 31(3), 291–316.
- Robie, C., Christiansen, N. D., Hausdorf, P. A., Murphy, S. A., Fisher, P. A., Risavy, S. D., & Keeping, L. M. (2017). International comparison of group differences in general mental ability for immigrants versus non-immigrants. *International Journal of Selection and Assessment*, 25(4), 347–359.

- Roccas, S., & Sagiv, L. (2010). Personal Values and Behavior: Taking the Cultural Context into Account. *Social and Personality Psychology Compass*, 4(1), 30–41.
- Rokeach, M. (1973). *The Nature of Human Values*. The Free Press.
- Roth, P., Bobko, P., McFARLAND, L., & Buster, M. (2008). Work Sample Tests in Personnel Selection: A Meta-Analysis of Black–White Differences in Overall and Exercise Scores. *Personnel Psychology*, 61(3), 637–661. <https://doi.org/10.1111/j.1744-6570.2008.00125.x>
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer Iii, F. S., & Tyler, P. (2001). Ethnic Group Differences in Cognitive Ability in Employment and Educational Settings: A Meta-Analysis. *Personnel Psychology*, 54(2), 297–330.
- Rotter, M., Nigg, C. R., Renda, G. A., & Novotny, R. (2016). Acculturation is Not Related to Physical Activity Stage of Change for Children in Hawai'i. *Journal of Medicine & Public Health : A Journal of Asia Pacific Medicine & Public Health*, 75(2), 35–41.
- Russell, D. (1982). The Causal Dimension Scale: A measure of how individuals perceive causes. *Journal of Personality and Social Psychology*, 42(6), 1137–1145.
- Russell Pascual, N., Kirby, T. A., & Begeny, C. T. (2024). Disentangling the nuances of diversity ideologies. *Frontiers in Psychology*, 14, 1293622.
- Ryan, G., Gubern, M., & Rodriguez, I. (2000). Recruitment advertising: The marketing-human resource interface. *International Advances in Economic Research*, 6(2), 354–364.
- Ryan, K. M., King, E. B., & Finkelstein, L. M. (2015). Younger workers' metastereotypes, workplace mood, attitudes, and behaviors. *Journal of Managerial Psychology*, 30(1), 54–70.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67.
- Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860.
- Rynes, S. L., & Barber, A. E. (1990). Applicant Attraction Strategies: An Organizational Perspective. *Academy of Management Review*, 15(2), 286–310.
- Sackett, P. R. (2021). Reflections on a Career Studying Individual Differences in the Workplace. *Annual Review of Organizational Psychology and Organizational Behavior*, 8, 1–18.
- Sackett, P. R., Berry, C. M., Lievens, F., & Zhang, C. (2023). Correcting for range restriction in meta-analysis: A reply to Oh et al. (2023). *Journal of Applied Psychology*, 108(8), 1311–1315.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, 56(4), 302–318.
- Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2022). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology*, 107(11), 2040–2068.
- Sagiv, L., & Roccas, S. (2017). What Personal Values Are and What They Are Not: Taking a Cross-Cultural Perspective. In S. Roccas & L. Sagiv (Eds.), *Values and behavior: Taking a cross cultural perspective* (pp. 3–14). Springer.

- Sammarra, A., Profili, S., & Innocenti, L. (2013). Do external careers pay-off for both managers and professionals? The effect of inter-organizational mobility on objective career success. *The International Journal of Human Resource Management*, 24(13), 2490–2511.
- Santos, G. G. (2016). Career barriers influencing career success: A focus on academics' perceptions and experiences. *Career Development International*, 21(1), 60–84.
- Saperstein, A. (2006). Double-Checking the Race Box: Examining Inconsistency between Survey Measures of Observed and Self-Reported Race. *Social Forces*, 85(1), 57–74.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.
- Schmitt, N. (2014). Personality and Cognitive Ability as Predictors of Effective Performance at Work. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 45–65.
- Schneider, B. (1987). The People Make The Place. *Personnel Psychology*, 40(3), 437–453.
- Schneider, S. L., & Heath, A. F. (2020). Ethnic and cultural diversity in Europe: Validating measures of ethnic and cultural background. *Journal of Ethnic and Migration Studies*, 46(3), 533–552.
- Schuck, A. M. (2021). Motivations for a career in policing: Social group differences and occupational satisfaction. *Police Practice and Research*, 22(5), 1507–1523. <https://doi.org/10.1080/15614263.2020.1830772>
- Schwartz, S. H. (2012). An Overview of the Schwartz Theory of Basic Values. *Online Readings in Psychology and Culture*, 2(1).
- Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, 103(4), 663–688.
- Seibert, S. E., Crant, J. M., & Kraimer, M. L. (1999). Proactive personality and career success. *Journal of Applied Psychology*, 84(3), 416–427.
- Selvanathan, H. P., Lickel, B., & Dasgupta, N. (2020). An integrative framework on the impact of allies: How identity-based needs influence intergroup solidarity and social movements. *European Journal of Social Psychology*, 50(6), 1344–1361.
- Shiao, J. L. (2019). When (In)Consistency Matters: Racial Identification and Specification. *Socius*, 5.
- Shockley, K. M., Ureksoy, H., Rodopman, O. B., Poteat, L. F., & Dullaghan, T. R. (2016). Development of a new scale to measure subjective career success: A mixed-methods study. *Journal of Organizational Behavior*, 37(1), 128–153.
- Sievert, M., Vogel, D., & Feeney, M. K. (2022). Formalization and Administrative Burden as Obstacles to Employee Recruitment: Consequences for the Public Sector. *Review of Public Personnel Administration*, 42(1), 3–30.
- Sievert, M., Vogel, D., Reinders, T., & Ahmed, W. (2020). The Power of Conformity in Citizens' Blame: Evidence from a Survey Experiment. *Public Performance & Management Review*, 43(1), 53–80.
- Simons, A., Wohlgenannt, I., Zelt, S., Weinmann, M., Schneider, J., & vom Brocke, J. (2023). Intelligence at play: Game-based assessment using a virtual-reality application. *Virtual Reality*, 27(3), 1827–1843.
- Singha, K. (2017). Understanding ethnicity-based autonomy movements in India's northeastern region. *Nationalities Papers*, 45(4), 687–706.
- Sniderman, P. M., & Grob, D. B. (1996). Innovations in Experimental Design in Attitude Surveys. *Annual Review of Sociology*, 22, 377–399.

- Spurk, D., Hirschi, A., & Dries, N. (2019). Antecedents and Outcomes of Objective Versus Subjective Career Success: Competing Perspectives and Future Directions. *Journal of Management*, 45(1), 35–69.
- Stegers-Jager, K. M., Steyerberg, E. W., Lucieer, S. M., & Themmen, A. P. N. (2015). Ethnic and social disparities in performance on medical school selection criteria. *Medical Education*, 49(1), 124–133.
- Stumpf, S. A., & Tymon, W. G. (2012). The effects of objective career success on subsequent subjective career success. *Journal of Vocational Behavior*, 81(3), 345–353.
- Stumpp, T., Muck, P. M., Hülshager, U. R., Judge, T. A., & Maier, G. W. (2010). Core Self-Evaluations in Germany: Validation of a German Measure and its Relationships with Career Success. *Applied Psychology*, 59(4), 674–700.
- Swargiary, K., & Roy, K. (2022). Literacy rate in India in 2022. *Academia: An International Multidisciplinary Research Journal*, 12(8), 87–93.
- Tajfel, H., & Turner, J. C. (2004). The Social Identity Theory of Intergroup Behavior. In J. T. Jost & J. Sidanius (Eds.), *Political Psychology* (pp. 276–293). Psychology Press.
- Taylor, J., & Deane, F. P. (2002). Development of a Short Form of the Test Anxiety Inventory (TAI). *The Journal of General Psychology*, 129(2), 127–136.
- Thorsteinson, T. J., & Ryan, A. M. (1997). The effect of selection ratio on perceptions of the fairness of a selection test battery. *International Journal of Selection and Assessment*, 5(3), 159–168.
- Tonidandel, S., Quiñones, M. A., & Adams, A. A. (2002). Computer-adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology*, 87(2), 320–332.
- Truxillo, D. M., Bauer, T. N., & Sanchez, R. J. (2001). Multiple Dimensions of Procedural Justice: Longitudinal Effects on Selection System Fairness and Test-Taking Self-Efficacy. *International Journal of Selection and Assessment*, 9(4), 336–349.
- Turner, J. C., Oakes, P. J., Haslam, S. A., & McGarty, C. (1994). Self and Collective: Cognition and Social Context. *Personality and Social Psychology Bulletin*, 20(5), 454–463.
- Turner, R. H. (1960). Sponsored and Contest Mobility and the School System. *American Sociological Review*, 25(6)
- Umaña-Taylor, A. J., Yazedjian, A., & Bámaca-Gómez, M. (2004). Developing the Ethnic Identity Scale Using Eriksonian and Social Identity Perspectives. *Identity*, 4(1), 9–38.
- UNIA & Federale Overheidsdienst Werkgelegenheid, Arbeid en Sociaal Overleg. (2020). *Socio-economische Monitoring—Arbeidsmarkt en origine 2019* (No. 4; p. 320).
- UNIA & FOD WASO. (2022). Socio-Economic Monitoring 2022: Labour market and origin. UNIA, FOD WASO. <https://www.unia.be/en/publications-statistics/publications/socio-economic-monitoring-2022-labour-market-and-origin>
- Van Caeneghem, J. (2019). *Legal aspects of ethnic data collection and positive action: The Roma minority in Europe*. Springer.
- Van de Walle, S., Steijn, B., & Jilke, S. (2015). Extrinsic motivation, PSM and labour market characteristics: A multilevel model of public sector employment preference in 26 countries. *International Review of Administrative Sciences*, 81(4), 833–855.
- Van Hooft, E. A. J., Born, M. Ph., Taris, T. W., & Van der Flier, H. (2006). Ethnic and Gender Differences in Applicants' Decision-Making Processes: An Application of the Theory of Reasoned Action. *International Journal of Selection and Assessment*, 14(2), 156–166.

- Van Iddekinge, C.H., Lievens, F. and Sackett, P.R. (2023), "Personnel selection: A review of ways to maximize validity, diversity, and the applicant experience", *Personnel Psychology*, 76 (2), 651–686.
- Vandenabeele, W., & Jager, S. (2020). Government Calling Revisited: A Survey-Experiment on the Moderating Role of Public Service Motivation in Assessing Employer Attractiveness. *Frontiers in Psychology*, 11.
- Vandenabeele, W., Ritz, A., & Neumann, O. (2018). Public Service Motivation: State of the Art and Conceptual Cleanup. In E. Ongaro & S. Van Thiel (Eds.), *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 261–278). Palgrave Macmillan UK.
- Vandenabeele, W., & Van de Walle, S. (2008). International Differences in Public Service Motivation: Comparing Regions across the World. In J. L. Perry & A. Hondeghem (Eds.), *Motivation in Public Management* (pp. 223–244). Oxford University PressOxford.
- Vanderpool, M., & Catano, V. M. (2008). Comparing the Performance of Native North Americans and Predominantly White Military Recruits on Verbal and Nonverbal Measures of Cognitive Ability. *International Journal of Selection and Assessment*, 16(3), 239–248.
- Vandiver, B. J., Cross, W. E., Fhagen-Smith, P. E., Worrell, F. C., Swim, J., & Caldwell, L. (2012). *Cross Racial Identity Scale [Dataset]*.
- Veerman, G.-J., & Platt, L. (2021). School composition and multiple ethnic identities of migrant-origin adolescents in the Netherlands. *Ethnic and Racial Studies*, 44(16), 106–125.
- Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, 30(6), 1024–1054.
- Villotti, P., Stinglhamber, F., & Desmette, D. (2019). The Influence of Multiculturalism and Assimilation on Work-Related Outcomes: Differences Between Ethnic Minority and Majority Groups of Workers. *Psychologica Belgica*, 59(1), 246–268.
- Vogel, D., Doring, M., & Sievert, M. (2024). Motivational signals in public sector job advertisements and how they relate to attracting and hiring candidates. *Public Management Review* 26(10), 2868–2900.
- Volmer, J., Orth, M., & Wolff, H.-G. (2018). Multidimensional Networking Behavior in Germany and China: Measurement Invariance and Associations With Objective Career Success. *Journal of Career Assessment*, 26(4), 678–696.
- Vorauer, J. D., Main, K. J., & O’Connell, G. B. (1998). How do individuals expect to be viewed by members of lower status groups? Content and implications of meta-stereotypes. *Journal of Personality and Social Psychology*, 75(4), 917–937.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96.
- Ward, A.-K., Beal, D. J., Zyphur, M. J., Zhang, H., & Bobko, P. (2022). Diversity climate, trust, and turnover intentions: A multilevel dynamic system. *Journal of Applied Psychology*, 107(4), 628–649.
- Ward-Bartlett, A.-K., Ravlin, E., & Park, J. E. (2023). How genuine is your diversity climate? A new typology highlighting the emergence of specious diversity climates. *Journal of Organizational Behavior*, 44(9), 1301–1319.
- Waters, M. C. (1990). *Ethnic options: Choosing identities in America*. University of California Press.
- Weekley, J. A., & Jones, C. (1999). Further Studies of Situational Tests. *Personnel Psychology*, 52(3), 679–700.
- Weng, Q. X. (2010). Career Growth Study: Scale development and validity test. *Management Review*, 22(10), 22–31.

- Weske, U., Ritz, A., Schott, C., & Neumann, O. (2020). Attracting future civil servants with public values? An experimental study on employer branding. *International Public Management Journal*, 23(5), 677–695.
- Westover, J. H., & Taylor, J. (2010). International differences in job satisfaction: The effects of public service motivation, rewards and work relations. *International Journal of Productivity and Performance Management*, 59(8), 811–828.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup Differences in Situational Judgment Test Performance: A Meta-Analysis. *Human Performance*, 21(3), 291–309.
- Williams, M., & Husk, K. (2013). Can we, should we, measure ethnicity? *International Journal of Social Research Methodology*, 16(4), 285–300.
- Woo, H., & Kim, S. (2024). Public service motivation and public sector employment in Korea. *Public Administration Review*, n/a(n/a), 1–16.
- Woods, S. A., & Patterson, F. (2024). A critical review of the use of cognitive ability testing for selection into graduate and higher professional occupations. *Journal of Occupational and Organizational Psychology*, 97(1), 253–272.
- Woolverton, G. A., & Marks, A. K. (2023). “I Just Check ‘Other’”: Evidence to Support Expanding the Measurement Inclusivity and Equity of Ethnicity/Race and Cultural Identifications of US Adolescents. *Cultural Diversity & Ethnic Minority Psychology*, 29(1), 64–73.
- Worrell, F. C., & Gardner-Kitt, D. L. (2006). The Relationship Between Racial and Ethnic Identity in Black Adolescents: The Cross Racial Identity Scale and the Multigroup Ethnic Identity Measure. *Identity*, 6(4), 293–315.
- Worrell, F. C., Mendoza-Denton, R., & Wang, A. (2019). Introducing a New Assessment Tool for Measuring Ethnic-Racial Identity: The Cross Ethnic-Racial Identity Scale–Adult (CERIS-A). *Assessment*, 26(3), 404–418.
- Wright, B. E., & Pandey, S. K. (2011). Public Organizations and Mission Valence: When Does Mission Matter? *Administration & Society*, 43(1), 22–44.
- Wright, S. L., Perrone-McGovern, K. M., Boo, J. N., & White, A. V. (2014). Influential Factors in Academic and Career Self-Efficacy: Attachment, Supports, and Career Barriers. *Journal of Counseling & Development*, 92(1), 36–46.
- Xu, D., & Zhang, Y. (2022). Identifying ethnic occupational segregation. *Journal of Population Economics*, 35(3), 1261–1296.
- Yamada, A.-M., Marsella, A. J., & Yamada, S. Y. (1998). The Development of the Ethnocultural Identity Behavioral Index: Psychometric Properties and Validation with Asian Americans and Pacific Islanders. *Asian American and Pacific Islander Journal of Health*, 6(1), 35–45.
- Yao, E. S., Bullen, P., Meissel, K., Tiatia, J., Fleming, T., & Clark, T. C. (2022). Effects of Ethnic Classification on Substantive Findings in Adolescent Mental Health Outcomes. *Journal of Youth and Adolescence*, 51(8), 1581–1596.
- Yao, E. S., Meissel, K., Bullen, P., Carr, P. A., Clark, T. C., & Morton, S. M. B. (2021). Classifying multiple ethnic identifications: Methodological effects on child, adolescent, and adult ethnic distributions. *Demographic Research*, 44, 481–512.
- Yao, E. S., Meissel, K., Bullen, P., Clark, T. C., Atatoa Carr, P., Tiatia-Seath, J., Peiris-John, R., & Morton, S. M. B. (2022). Demographic discrepancies between administrative-prioritisation and self-prioritisation of multiple ethnic identifications. *Social Science Research*, 103.
- Yeung, E. (2019). *On Diversity Climate in Organizations* [Doctoral thesis]. University of Waterloo.

Yip, T., Wang, Y., Mootoo, C., ace, & Mirpuri, S. (2019). Moderating the Association Between Discrimination and Adjustment: A Meta-Analysis of Ethnic/Racial Identity. *Developmental Psychology*, 55(6), 1274–1298.

Yoon, E. (2011). Measuring ethnic identity in the Ethnic Identity Scale and the Multigroup Ethnic Identity Measure-Revised. *Cultural Diversity and Ethnic Minority Psychology*, 17(2), 144–155.

Zhou, W., Sun, J., Guan, Y., Li, Y., & Pan, J. (2013). Criteria of Career Success Among Chinese Employees: Developing a Multidimensional Scale With Qualitative and Quantitative Approaches. *Journal of Career Assessment*, 21(2), 265–277.

Zhou, Y. (2019). Question of ethnic group formulation in the Chinese census. *China Population and Development Studies*, 3.