IWH Halle Institute for Economic Research
Member of the Leibniz Association

Discussion Papers

No. 12
May 2023

Discrimination in Universal Social Programs? A Nationwide Field Experiment on Access to Child Care

Henning Hermes, Philipp Lergetporer, Fabian Mierisch, Frauke Peter, Simon Wiederhold

Authors

Henning Hermes
ifo Institute Munich

Philipp Lergetporer
Technical University of Munich, TUM School of Management, Heilbronn, and ifo Institute Munich

Fabian Mierisch
KU Eichstätt-Ingolstadt, Ingolstadt School of Management

Frauke Peter
German Centre for Higher Education Research and Science Studies (DZHW)

Simon Wiederhold
Halle Institute for Economic Research (IWH) – Member of the Leibniz Association, Department of Structural Change and Productivity, Martin Luther University Halle-Wittenberg, and ifo Institute Munich
E-mail: simon.wiederhold@iwh-halle.de
Tel +49 345 7753 840

The responsibility for discussion papers lies solely with the individual authors. The views expressed herein do not necessarily represent those of IWH. The papers represent preliminary work and are circulated to encourage discussion with the authors. Citation of the discussion papers should account for their provisional character; a revised version may be available directly from the authors.

Comments and suggestions on the methods and results presented are welcome.

IWH Discussion Papers are indexed in RePEc-EconPapers and in ECONIS.

Editor

Halle Institute for Economic Research (IWH) – Member of the Leibniz Association

Address: Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany
Postal Address: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

Discrimination in Universal Social Programs? A Nationwide Field Experiment on Access to Child Care*

First version: 29.05.2023
This version: 04.03.2024

Abstract

Although explicit discrimination in access to social programs is typically prohibited, more subtle forms of discrimination prior to the formal application process may still exist. Unveiling this phenomenon, we provide the first causal evidence of discrimination against migrants seeking child care. We send emails from fictitious parents to > 18, 000 early child care centers across Germany, inquiring about slot availability and application procedures. Randomly varying names to signal migration background, we find that migrants receive 4.4 percentage points fewer responses. Replies to migrants contain fewer slot offers, provide less helpful content, and are less encouraging. Exploring mechanisms using three additional treatments, we show that discrimination is stronger against migrant boys. This finding suggests that anticipated higher effort required for migrants partly drives discrimination, which is also supported by additional survey and administrative data. Our results highlight that difficult-to-detect discrimination in the pre-application phase could hinder migrants' access to universal social programs.

Keywords: child care, discrimination, field experiment, inequality, program take-up

JEL classification: C93, J13, J15, J18, I24

* We thank Martin Abel, Mallory Avery, Vojtěch Bartoš, Stefan Bauernschuster, Peter Bergman, Aislinn Bohren, Alexander Cappelen, Alex Chan, Mathias Ekström, Ruben Enikolopov, Katharina Hartinger, James Heckman, Krysztof Karbownik, Boon Han Koh, Katrine Løken, Giovanni Peri, Olga Stoddard, Mirco Tonin, Bertil Tungodden, George Wu and seminar participants at the ASSA, VfS, EALE, AFE, SOLE, RES, In_equality Conference Konstanz, Applied Young Economists Webinar, CESifo Area Confernce on Education Economics, DICE Düsseldorf, FAIR NHH Bergen, KU Leuven, ifo Institute Munich, KU Eichstätt-Ingolstadt, and Chicago Booth School of Business for valuable comments and suggestions. Lukas Seubert provided excellent research assistance. We also thank David Broneske and Nikhilkumar Italiya for their support in implementing the natural language processing model. We gratefully acknowledge funding from the Technical University of Munich, the DICE Düsseldorf, the KU Eichstätt-Ingolstadt (ProFOR+ funding No. U060230003FL 4/3), and the ifo Institute Munich. The study was preregistered in the AEA RCT Registry (AEARCTR-0007389). IRB approval was obtained from the Ethics Commission of the Department of Economics, University of Munich (Project 2020-17).

1. Introduction

Not all eligible individuals participate in social programs. Disadvantaged groups, in particular, exhibit notably lower take-up rates, a pattern observed across many countries (Currie, 2006; Landersø and Heckman, 2017; Ko and Moffitt, 2022). Existing explanations for incomplete program take-up have primarily focused on factors associated with the beneficiaries, such as expected lack of benefits, high costs, and stigma of program participation (e.g., Bertrand et al., 2000; Bitler et al., 2003; Bhargava and Manoli, 2015), along with information deficits and difficulties navigating complex application processes (e.g., Finkelstein and Notowidigdo, 2019; Dynarski et al., 2021). While these factors can significantly impact program participation, they do not fully explain low take-up rates of disadvantaged groups (Ko and Moffitt, 2022).

In this paper, we focus on the behavior of program managers, particularly discriminatory practices, as an explanation for the low program take-up of disadvantaged groups. Legislation commonly prohibits discrimination in the admission decisions to social programs on the basis of ethnicity, race, and similar characteristics (see, e.g., United States Congress, 1964). At the same time, the complicated application processes for theses programs (e.g., Kleven and Kopczuk, 2011; Bettinger et al., 2012) require applicants to be well-informed, which often involves contacting program managers prior to applying. We argue that program managers may use this informal pre-application phase to subtly discourage certain individuals from applying, thereby influencing the composition of the applicant pool in a way they consider favorable.

To study this subtle form of discrimination, we conduct a nationwide field experiment within the context of Germany's early child care system.[1] Similar to many social programs worldwide, disadvantaged groups such as migrants are severely underrepresented in early child care in Germany. In fact, enrollment rates for migrant and native families differ by a factor of two: Only 21% of migrant children compared to 42% of native children are enrolled in early child care (Education Report, 2020). This gap is especially stunning in light of the very similar demand for early child care among migrant and native parents (Jessen et al., 2020) and very low fees (Felfe and Lalive, 2018). Discrimination in the final admission decision is unlikely to be a reason for this migrant-native gap in enrollment, as there is a legal ban on discrimination in child care access (BMJ, 2006) and the law grants every child the right to a child care slot from one year of age, so eligibility

[1]Our definition of discrimination follows Bertrand and Duflo (2017), p. 310: "Members of a minority group [...] are *treated differentially* (less favorably) than members of a majority group with otherwise identical characteristics in similar circumstances."

is universal and unambiguously determined (BMFSFJ, 2013). Yet, there are various reasons to suspect *pre-application discrimination* to occur. First, the complex application process for a child care slot frequently prompts parents to contact child care centers for information before applying, potentially allowing managers to discriminate during this informal stage. Second, demand for early child care slots typically exceeds supply, necessitating that managers choose among candidates. Third, as is generally the case for social programs, market-clearing is not achieved through the price mechanism.[2] For these reasons, the German early child care system is an ideal setting to examine pre-application discrimination in social programs.

In the experiment, we send emails from fictitious parents drafted according to a set of real parental email messages to $N = 18,663$ early child care centers in Germany. The emails ask if there are slots available and how to apply.[3] In the emails, we randomly vary whether or not a parent's name signals a migration background (Bertrand and Mullainathan, 2004).[4]

We find a large negative effect of migrant sender names on response rates. Emails signaling that the sender is a migrant receive 4.4 percentage points (pp) fewer replies from child care center managers than emails signaling that the sender is a native. With a response rate of 71% for native senders, this translates to a scaled effect of -6.2%. This effect size is relatively large compared to other recent correspondence studies. For instance, estimated black-white response gaps in U.S.-based studies varying sender names range from 2 to 4 pp (see, e.g., Bergman and McFarlin, 2018; Giulietti et al., 2019; Kline et al., 2022). Hence, our results demonstrate that migrants in Germany face substantial discrimination during the pre-application phase for child care.

Since we received more than 12,500 email replies from child care centers, we can also analyze the *content* of the written responses. The most important dimension of the email content is whether parents are actually offered a slot in the contacted child

[2]Germany's child care system functions as a social program rather than a conventional market. The government not only established a legal entitlement to child care slots but also heavily subsidizes slots, with non-profit providers running 97% of child care centers (Education Report, 2020).

[3]To inform our design, we surveyed $N = 447$ child care center managers. We find that (i) managers are regularly contacted by parents interested in securing a child care slot (on average, 14 times per week), (ii) almost all managers regularly receive emails from parents (85%), (iii) prior email contact is often important for parents to ultimately receive a slot (stated by 40% of managers), and (iv) the two most frequently asked questions are about available slots and on how to apply (see Appendix E.1 for details).

[4]We chose the most common German (*native*) and Turkish (*migrant*) names based on German name registry data, and validated our name selection in additional pretests (see Appendix E.2). Turkish migrants are particularly relevant in the German context, as they are the largest and geographically most widespread migrant group, and are strongly underrepresented in early child care (see Appendix F).

care center. Based on manual ratings from five independent reviewers, we find that replies to a migrant sender are significantly less likely to contain actual slot offers than replies to a native sender. Unconditional on receiving a response, the treatment effect amounts to -1.1 pp, which corresponds to -23% relative to the mean offer rate for natives. Conditional on receiving a response, the effect is only slightly smaller (-1.0 pp or -20%). While these offers are not final in the sense that they typically require further action by the parents (e.g., visiting the child care center), this result suggests that discrimination against migrants in the early child care system is not limited to information provision, but also extends to the allocation of slots.

Migrant-native gaps on other email content dimensions are consistent with the evidence for slot offers: Both conditional and unconditional on receiving a response, replies to migrant senders are shorter, contain fewer offers to join a waiting list, contain less helpful content, and are rated as less encouraging and generally less appealing in substance and tone. Effect sizes are large, up to 18% relative to the control group mean. Overall, our results show that pre-application discrimination against migrants in the early child care system exists at both the extensive margin (response rates) and the intensive margin (content of the email).

Next, we investigate potential mechanisms underlying discrimination by child care center managers against migrant families. For this mechanism analysis, we leverage the randomization of three additional elements of our email: (i) the education background of the email sender (with an email signature indicating that the sender holds a higher education degree), (ii) the gender of the email sender, and (iii) the gender of the child.

The effects of *parental* characteristics on discrimination against migrants are limited. Adding the higher education signal to the email does not significantly affect the migrant-native gaps in response rates or slot offers. We therefore consider it unlikely that the observed discrimination against migrants is due to child care center managers discriminating against a (perceived) lower educational background of migrant parents. Likewise, although discrimination against migrants tends to intensify when the email sender is male, this effect is not consistent across outcomes.

In contrast, discrimination against migrants substantially varies with the gender of the *child*. In line with prior evidence that boys from disadvantaged backgrounds are more disruptive in (pre)school than girls (Bertrand and Pan, 2013; Gilliam et al., 2016), we find that the migrant-native gap in responses and slot offers is significantly higher if parents mention a son rather than a daughter in their email. This finding suggests that child care center managers consider the expected effort associated with the child when deciding whether and how to respond to email inquiries. Below, we provide several pieces of evidence supporting the notion that center managers' concerns regarding the effort or the resources required to educate migrant children may drive discrimination against them.

First, the general public believes that migrant children require more effort to be educated in early child care. Asking a representative sample of German adults ($N \approx 4,800$) about reasons for migrants' unequal chances in the early child care system, half of the respondents attributed this inequality to child care center managers' perceptions of higher workload associated with migrant families (this was also the most frequently selected reason, see Appendix E.3). Second, discrimination correlates meaningfully with the resources available to child care centers for educating migrant children. Linking our experimental data to various administrative data sources, we find that discrimination decreases with (i) financial incentives for centers to take in migrant children, (ii) the staff-to-child-ratio, and (iii) the budget per capita in the municipality of the center. Finally, using a causal forest analysis (Athey and Wager, 2019), we affirm the relevance of child care centers' resources in explaining managers' discriminatory behavior, relative to other potential channels. This data-driven approach shows that the resources available to child care centers are, by a considerable margin, the most important predictor of treatment effect heterogeneity.

The robustness of our findings is confirmed by a variety of additional analyses. Treatment effects are robust to using (i) randomization inference, (ii) corrections for multiple hypothesis testing, (iii) and Probit estimations. We also estimate models with zip code fixed effects, comparing child care centers that are arguably similar in (unobserved) characteristics.[5] Furthermore, while our baseline results for the email content are based on a manual coding of the email replies by a team of five research assistants ("reviewers") blind to the treatment, employing an alternative approach that classifies email content based on a supervised machine-learning algorithm yields very similar results.

Our paper is the first to document substantial discrimination against migrants seeking child care. Thereby, we contribute to several strands of the literature. First, we add to the literature that examines the reasons for imperfect take-up of social programs, especially among disadvantaged groups (see, e.g., Currie, 2006; Ko and Moffitt, 2022, for reviews). While previous research has largely focused on beneficiary-related factors, our findings indicate that discriminatory behavior of program managers may as well con-

[5]We cannot include child care center fixed effects since we sent only one email to each center. We made this design choice to minimize both the burden on the centers and the risk of detection (Bertrand and Duflo, 2017).

tribute to the underrepresentation of disadvantaged groups in such programs. Complex admission processes to social programs often require applicants to seek pre-application information, and our study demonstrates that this process allows for substantial discrimination. This subtle form of discrimination has especially harmful effects in environments where slots are rationed and enrollment processes are decentralized and opaque, granting individual managers considerable control over admissions — a scenario prevalent across numerous social programs, including child care systems worldwide (see Table A1; Mocan, 2007; Spiess, 2008; Hermes et al., 2021).[6] Importantly, discrimination in informal pre-application phases is very difficult to detect due to the lack of recorded data on informal inquiries, undermining the enforcement of existing anti-discrimination laws.[7]

Furthermore, we add to the literature on discrimination in the education system (e.g., Alesina et al., 2018; Carlana, 2019; Alan et al., 2023; Lavy et al., 2022). While previous studies have documented discrimination in the admission processes of schools (Bergman and McFarlin, 2018; de Lafuente, 2021; Olsen et al., 2022; Bell and Jilke, 2024) and universities (Dynarski et al., 2018; Arcidiacono et al., 2022), we are the first to show the existence of discrimination already in the earliest stage of the education system, i.e., in the admission to early child care, providing a possible reason for the underrepresentation of migrant children in child care systems worldwide (OECD, 2018; Hussar et al., 2020; Jessen et al., 2020). Such underrepresentation is a major concern, as it can put migrant children on a worse educational trajectory with persistent negative impacts over their life course (e.g., Heckman et al., 2010).

Finally, we contribute to the literature that uses correspondence studies for detecting discrimination (for reviews, see Baert, 2018; Neumark, 2018). In particular, we demonstrate substantial discrimination at both the extensive margin (response rates) and intensive margin (email content), indicating that focusing solely on call-back rates — a common practice in correspondence studies — may underestimate the true extent of discrimination. Few previous studies have also examined email response content, for instance, in terms of length or cordiality (Hemker and Rink, 2017; Giulietti et al., 2019; Bergman and

[6]Substantial evidence shows that disadvantaged individuals often lack crucial information needed to navigate the complex application processes for programs for which they are eligible (see, e.g., Dynarski and Scott-Clayton, 2006; Aizer, 2007; Bettinger et al., 2012). This underscores the significant role that withholding relevant application information plays in exacerbating inequality in take-up of social programs. In the context of early child care, for instance, Hermes et al. (2021) demonstrate that especially disadvantaged parents commonly, but incorrectly, believe that they are limited to applying to only one child care center.

[7]The significance of discrimination in the initial *formal* stage of multi-phase admission processes has also been shown in hiring experiments (Bohren et al., 2022) or foster care placements (Baron et al., 2024).

McFarlin, 2018). Our approach, utilizing a wide range of content outcomes, both objective (such as response length) and subjective (such as perceived encouragement), and employing machine-learning to validate human coding, highlights the multidimensional nature of discrimination in pre-application interactions. Moreover, our findings of discrimination in *actual slot offers* addresses the frequent critique that observed outcomes in correspondence studies (e.g., callbacks to job applications) lack direct economic impact and serve merely as proxies for the outcome of interest (e.g., actual job offers, see Bertrand and Duflo, 2017).

The remainder of this paper is structured as follows. Section 2 provides information on the institutional background of the early child care market in Germany. Section 3 describes our experimental design and introduces the data. Section 4 presents our empirical strategy. Section 5 reports our results and various robustness checks. Section 6 presents the mechanism analysis and Section 7 concludes.

2. Institutional Background

In this section, we discuss key features of the German child care system that are potentially conducive to discrimination. As shown in Table A1, these institutional features are similar in several other developed countries.

Early child care provision. Child care provision in Germany is universal, targeting all children before they enter school (at the age of six). Child care is available for children at two distinct age groups: (i) under the age of three years (early child care or *Krippe*) and (ii) between three and six years (*Kindergarten*). Each child has a legal entitlement to a child care slot from the age of one year onward. Early child care is heavily publicly subsidized, with the public sector paying about three-quarters of the total cost (Spiess, 2013). Parents pay very low child care fees (on average 250 EUR per month, equivalent to 10% of the average income), and lower-income families are eligible for fee reductions or even exemptions (Felfe and Lalive, 2018). Compared to other countries, the quality of early child care is relatively high and homogeneous across Germany, for example, in terms of group sizes or staff-to-child ratios (Felfe and Lalive, 2018).

Despite the fact that child care is coined universal, its utilization is far from universal. On average, 34% of children under the age of three are enrolled in early child care. Attendance rates increase substantially with age, from only 1% for children under the age of one to 55% for children aged two to three years (Education Report, 2020). More than 90% of children attend *Kindergarten*, so that almost all children have attended child care by the time they start school. As a consequence, the relevant margin is not *whether* children have access to child care, but rather *when* they have access, in particular, whether they enroll into early child care. Past research has shown that earlier enrollment in child care can have pronounced positive effects on child development (Drange and Havnes, 2019).

Like in many other countries, a key characteristic of the German child care market is the rationing of slots, as parental demand exceeds supply. Rationing is especially severe for parents with a migration background. While there is almost no difference between native and migrant parents in the wish to enroll their child in early child care (Jessen et al., 2020), actual enrollment differs substantially: Only 21% of children with a migration background are enrolled, compared to 42% of native children (Education Report, 2020). From a legal standpoint, the disparity in enrollment rates between migrants and natives should not stem from discrimination against migrants. In Germany, the denial of access to child care based on migration background is unlawful, as the General Equal Treatment Act explicitly forbids discrimination on grounds of race, ethnic origin, and other attributes, in accordance with constitutional law (BMJ, 2006). However, whether discrimination occurs in practice is an empirical question, which we address in this paper.

Organization and funding of early child care. In Germany, child care is part of the child and youth welfare system under the responsibility of the federal government, but the actual implementation of child care provision takes place at the municipality level. As such, these decentralized, local child care markets are very heterogeneous and differ in prices for child care, fee reductions, application procedures and deadlines, and admission criteria. Child care centers are mostly government-funded, with very low fees for parents (Alt et al., 2019). Across regions, there are substantial differences in the amount of resources and the staff-to-child ratios in early child care centers. Moreover, in nine out of 16 federal states, centers receive additional funding to care for children with a migration background. These additional funds are mainly intended to promote children's language skills.

In Germany, 65% of child care centers are operated by charitable organizations (run by churches, trusts, parents, etc.), 32% are operated by municipalities, and only 3% are run by for-profit organizations or companies (see Education Report, 2020). Child care centers are mostly small in size (typically serving to 25–75 children, see DJI, 2021), and there is little competition between centers (Spiess, 2008). Importantly, in almost all German child care centers, managerial tasks — such as communicating with parents and allocating slots — are performed by single center manager (79%) or a team of manager and deputy manager (13%), who are released from child care responsibilities to perform these tasks (DJI, 2021).

Enrollment process in early child care. Due to the decentralized structure of the child care market, each center typically has its own admission process, resulting in an unstructured and individualized application process for families. Additionally, there are no mandatory, standardized criteria for child care center managers how to prioritize when allocating slots, and no accountability system to track enrollment decisions.[8] Therefore, the application process for child care in Germany is complex and can differ for each child care center, which implies that information about the application process is crucial for parents to successfully enroll their child. Acquiring a slot is likely more difficult for parents with a migration background, because the process of searching and applying for child care is resource-intensive (e.g., in terms of networks, social capital, and time). Therefore, obtaining information about how, when, and where to apply for child care is particularly important for migrant parents, who are less likely to have such information compared to native parents.[9]

For these reasons, we expect that child care centers (i.e., the supply side of the child care market) play a significant role in explaining the substantial gap in early child care enrollment between migrants and natives. The lack of both a centralized accountability system and clear admission criteria, combined with decentralized decision-making at the child care center level, grants center managers considerable discretion in slot allocation.

3. Experimental Design and Data

To investigate discrimination in the German child care market, we conduct an email correspondence study (for recent overviews, see Baert, 2018; Neumark, 2018).[10] We send emails over a three-day period in March 2021, and collect responses for a period of more

[8]In our nationwide survey, 86% of child care center managers report that the center manager is responsible for enrollment decisions, while less than 10% indicate that admission criteria of the provider or the municipality play a role for these decisions (see Appendix E.1).

[9]For instance, in a sample of >600 parents with young children in Hermes et al. (2021), only 61% of migrant parents know that they have a legal entitlement to child care once the child turns one year old, compared to 85% of native parents. Similarly, only 77% of migrant parents (compared to 94% of native parents) are aware that low-income families are eligible for child care fee reductions or waivers.

[10]Because it is not possible to obtain informed consent from study participants, the barriers for ethical approval of correspondence studies are high. We received IRB approval and clearance in terms of compliance with German data protection rules (DSGVO) from the ethics committee of LMU Munich (reference 2020-17).

than two months. In our emails, we experimentally vary the names of the fictitious parents (Bertrand and Mullainathan, 2004; Bertrand and Duflo, 2017) and three additional features to learn about mechanisms (a signature with a higher education signal, sender gender, and child gender; details below). We then record whether we receive any response to our email and also analyze the content of the responses received.

3.1. Email Design

To inform our study design, we first conducted qualitative interviews and a nationwide online survey with child care center managers (see Appendix E.1 for details). The data reveal that center managers are frequently contacted by parents during the pre-application phase for a slot in early child care: on average, they receive about 14 such inquiries per week. Accordingly, 85% of managers indicate that they regularly receive emails from parents. Requests for open slots (89%) and how to apply (60%) are the two most common questions to center managers. Finally, almost 40% of managers state that email contact with parents is an important prerequisite for actual enrollment, highlighting the relevance of email requests for the allocation of (rationed) slots in child care.

We drafted the email based on (i) the information from our survey with center managers and (ii) a set of 12 real, anonymized email messages from parents which we obtained confidentially from a child care center. We identified similarities in these emails in terms of sentence structure, length, and content for our email. Based on the center manager survey, we selected the two questions that are most frequently asked by parents: whether there is an open slot at a child care center and how to apply for a slot. Next, we recruited a sample of online workers ($N = 200$) to rate the degree of realism of our fictitious email (see Appendix E.2 for further information). Reassuringly, 80% of survey participants rated the email as realistic or very realistic, whereas only 4% rated the email as not realistic. In addition, the high overall response rate from child care centers in our study indicates that our efforts to design a realistic email were successful.

Figure 1 presents the general email template; specific examples are shown in Appendix Figures A1 and A2. The email indicates that the child is one year and five months old. The parent then asks (i) if the child care center has a slot for the child in about nine months (i.e., when the child is older than two years; some centers do not accept children before this age cut-off) and (ii) how to apply for a slot.

Note that we intentionally designed our email to be free of grammatical errors, typos, or other mistakes that might (statistically) be correlated with sender characteristics. Since we expect such formal errors to be more likely to occur among migrants (e.g., due to a

Figure 1: Email Template with Randomized Information Highlighted

> Dear Sir or Madam,
>
> We are looking for a child care slot for our [son/daughter] starting in January 2022. [He/she] is now 1 year and 5 months old.
>
> Do you still have a slot available? How can we apply for a slot?
>
> Thank you!
>
> Sincerely,
> [Name]
>
> ---
> [Name], Bachelor of Arts (FH)
> Email: [Name]@ ...
> ---

Notes: Figure shows the email sent from fictitious parents to child care centers, translated from the original German version. Text marked in grey is randomized and differs by version of the email. We randomized the gender of the child (2 variations), the name of the parent (16), and whether or not we include an email signature (2). The signature indicates that the parent holds a bachelor's degree from a University of Applied Sciences, which is the most common higher education degree in Germany. We sent a total of 64 ($2 \times 16 \times 2$) different versions of the email to child care centers.

lower language proficiency), any treatment effects should be interpreted as a lower bound estimate for the discrimination experienced by actual migrants.[11]

3.2. Treatment Variations

To signal whether the parents seeking child care had a native or migration background, we experimentally varied the name of the email sender (Bertrand and Mullainathan, 2004). We chose Turkish names to indicate migration background because Turkish immigrants are by far the largest and geographically most dispersed migrant group in Germany (around 13% of all migrants are Turkish; for additional information, see Appendix F). People with a Turkish migration background have often lived in Germany for several decades, as the major emigration waves from Turkey to Germany took place in the 1970s. At the same time, children from Turkish families are strongly underrepresented in early

[11] Note that in our survey of online workers, we do not find significant differences in realism ratings of emails sent by fictitious migrants compared to those sent by fictitious natives.

Table 1: Names Used to Signal Native and Turkish Migration Background

	German	**Turkish**
Male	Andreas Sebastian	Ömer Hüseyin
Female	Stefanie Christina	Eylül Fatma
Surname	Schmidt Müller	Yildirim Öztürk

child care. In fact, the enrollment rate for Turkish children is 12%, which is substantially lower than for migrant children overall (Jessen et al., 2020).

We selected names that most clearly signal a native or Turkish migration background in the following way: First, we chose the most common names for native Germans living in Germany and for Turkish migrants living in Germany, respectively.[12] Second, we used our survey of online workers to elicit which country of origin respondents associated with different first names and surnames. Finally, among the names for which at least 90% of respondents indicated a German or Turkish background, we selected two first names for each gender and two surnames for "German" and "'Turkish" senders (see Table 1).[13]

In total, we sent emails from 16 fictitious parents (four female first names, four male first names, each with two different surnames). We created an email account for each of the 16 fictitious parent names (e.g., Andreas Schmidt, Fatma Öztürk) following the pattern of name.surname0528@.... Each account sent about 1,400 emails (see Section 3.3).[14]

To learn about the mechanisms of discrimination, we vary three additional features in the email: (i) whether or not the email contains a signature with a higher education signal

[12]We used the online portal of the Society for German Language (GfdS) to identify the most common names from the cohort born in 1986.

[13]Using data from a public website providing ratings for first names, we can also rule out that any of the names selected for our study are outliers in terms of their popularity ratings.

[14]In one of the 16 accounts, there was a spelling mistake: While the *email address* was correctly shown as oemer.yildirim0528@..., the *sender name* and, depending on the treatment, the email signature displayed "Yildirm" instead of "Yildirim" as surname. The response rate to this account was lower than for the other accounts (about 20 pp on average), most likely driven by the spelling mistake. This is evidence that child care center managers payed close attention to all aspects of the email, including the names displayed for the sender. In our regressions, we always include a dummy for this email account. As expected, treatment effects become even stronger if we do not control for the account. Results are virtually identical if we exclude emails sent from this account from the estimation sample.

(Giulietti et al., 2019), (ii) the gender of the email sender, indicated by the name, and (iii) the gender of the child, specified as a 'son' or 'daughter' with corresponding pronouns in the email (see Figure 1).

First, by including a signature with a higher education signal, we aimed to fix child care center managers' beliefs about the educational background of the sender. For instance, if Turkish names were associated with lower levels of education, center managers might be more likely to discriminate against such senders simply because of the (perceived) lower educational background. To test this possibility, a random subset of the sent emails included an email signature indicating that the sender has obtained a higher education degree (Giulietti et al., 2019).[15] For the signature, we used the most frequently obtained higher (tertiary) education degree in the German population, a Bachelor of Arts from a University of Applied Sciences (HRK, 2021). Since approximately 30% of individuals in the relevant age group (30–34 years) in Germany hold this university degree, our treatment effectively conveys an above-average level of education while still representing a sizeable portion of the population. Using a higher degree, such as a Ph.D., which is held by only 2% of the population, would not have been as representative. Through our survey of online workers, we verified that respondents were able to recall the information from the email signature indicating that the sender possessed a higher education degree (see Appendix E.2).

Second, we vary the gender of the sender because gender roles are influenced by cultural aspects (e.g., Alesina et al., 2013), and child care center managers may infer different gender roles from the gender of the email sender, depending on their migration background. Furthermore, we conjecture that child care center managers might have different beliefs about the likelihood of (mothers') participation in the labor market based on their migration background (Estrada, 2018). Both of these aspects could subsequently influence response behavior.

Third, child-rearing practices are influenced by cultural aspects (see, e.g., Doepke et al., 2019). Consequently, child care center managers may hold different beliefs about the potential cost of educating children with and without a migration background. These perceptions could be gender-specific, as is the case for black and white children in the US, where preschool educators perceive black boys as more disruptive than girls (Gilliam et al., 2016). Randomizing the gender therefore allows for an indirect test of whether

[15] Indicators for families' socioeconomic background other than education, for example, occupation or income, are very rarely mentioned in actual emails, so they can hardly be signaled in emails in an unobtrusive way.

child care center managers' beliefs about the effort to educate migrant children drives discrimination against them. Moreover, compared to other potential indicators of the effort required to educate a child, such as information about conduct or behavioral issues, including the child's gender in parental email inquiries is quite natural, thereby reducing the risk of sending emails that may appear suspicious or unnatural.

In total, by combining the 16 names (including sender's gender), the higher education signal, and the gender of the child, we send a total of 64 different emails.

3.3. Sample

For the sampling, we use a comprehensive and commercially available data set of child care centers with email addresses, covering the vast majority of all child care centers in Germany. Because we focus on early child care, we restrict the full sample to centers that enroll children below the age of three years. We further exclude about 2,000 centers that share an email address with another center to minimize the risk of detection. We end up with a sample of 22,458 child care centers for our field experiment, representing about 60% of all early child care centers in Germany (Destatis, 2020).

We sent a total of $N = 22,458$ emails to an equal number of distinct early child care centers. Of these emails, $N = 3,795$ were not delivered to the recipient (mostly because the email address was no longer valid).[16] Our final analysis sample consists of $N = 18,663$ delivered emails. We received a total of $N = 12,547$ responses, resulting in an average response rate of 67.2%.[17]

We use stratified randomization to generate the treatment groups. We construct strata by federal state, level of urbanization of the county where the center is located (i.e., predominantly urban, intermediate, or predominantly rural), and provider type (i.e., public, ecclesiastical, and other types of child care centers, such as centers run by for-profit providers or by parental initiatives).[18]

Table B1 shows that our randomization was successful. Despite the large sample size, we cannot detect statistically significant differences between the characteristics of the baseline group (native sender) and those of the experimental group (migrant sender) in

[16]The likelihood of an email not being delivered is independent of treatment status (see Table B2).

[17]In addition, we received around 400 emails (~2%) which we could not link to an observation in our database. These cases are mostly emails that have been forwarded to youth welfare offices or other child care centers, which then responded to our initial request, and therefore could not be matched to initial emails.

[18]In total, we use 131 strata in our randomization in which we randomized the the migrant treatment and the higher education signal in a 2×2 design. For both sender and child gender, we use a plain randomization independent of the stratification.

a total of 33 comparisons. All differences are also economically negligible in size. In Table B2, we repeat the same analysis in the sample of all emails sent (including bounced emails), again showing that the stratified randomization has been successful. The table further indicates that emails that could not be delivered due to incorrect or outdated email addresses are randomly distributed across groups. Randomization was also successful for the additional treatments, i.e., higher education signal, sender gender, and child gender (Tables B3–B5).

Table B1 also provides summary statistics for the analytical sample (see Appendix D for detailed definitions of all variables). The average child care center has a maximum capacity of 68 children. All centers in our sample cater to children below age 3, most (93%) also cater to children between 3 and 6 years, and only a few (9%) offer afternoon care for children older than 6 years. One-quarter of centers are operated by ecclesiastical providers, and one-fifth are public child care centers. Regarding regional characteristics, 44% of the centers are located in predominantly urban counties, 18% in predominantly rural counties, and 37% in counties that are neither classified as urban nor rural ("intermediate"). The average share of migrants in a county is 23%.

For our mechanism analysis in Section 6, we additionally use regional information on whether the federal state provides additional funding for child care centers for enrolling migrant children, staff-to-child ratios, and budget per capita (i.e., tax income per capita in a municipality). About 87% of the centers are located in a federal state that provides financial incentives to cater to migrant children, the average staff-to-child ratio is about 1:8, and the average budget per capita is EUR 1070.

4. Empirical Strategy

4.1. Estimation

Our main specification estimates treatment effects by regressing the outcome of interest (see Section 4.2) on randomized treatment indicators using ordinary least squares (OLS) models. In order to increase precision and to account for slight imbalances between treatment and control groups, we include a vector of preregistered control variables in our main specification:

$$Y_{ij} = \beta_0 + \beta_1 Migrant_j + \mathbf{X}_{ij}\mu + \varepsilon_{ij} \tag{1}$$

Here, Y_{ij} is the outcome of interest for a fictitious parent j sending an email to the child care center i. $Migrant_j$ is our main treatment variable, taking a value of one if the name of the parent j signals a Turkish migration background, and zero if it signals a native background. X_{ij} is a vector of control variables that includes child care center characteristics (i.e., center's maximum capacity and indicators of whether the child care center also has a kindergarten or afternoon care for school-aged children), the share of migrants in the municipality in which the contacted child care center is located, and strata fixed effects (i.e., interactions between provider type, urban class, and federal state). ε_{ij} is an idiosyncratic error term.[19] Based on the randomized research design, the causal effect of the migrant treatment on outcome Y_{ij} is given by β_1.

For the mechanism analysis, we also estimate the following regression model:

$$\begin{aligned} Y_{ij} = \; & \gamma_0 \; + \; \gamma_1 Migrant_j \; + \; \gamma_2 Mechanism_j \\ & + \; \gamma_3 Migrant_j \times Mechanism_j \; + \; \mathbf{X}_{ij}\nu \; + \; \upsilon_{ij} \end{aligned} \tag{2}$$

Here, $Mechanism_j$ can be (i) an indicator variable equal to one if the email contains a signature indicating that the sender has obtained a higher education degree, and zero if no such education signal is included (education signal), (ii) an indicator for a male email sender (sender gender), and (iii) an indicator for an email asking for child care slot for a son (child gender). We add the same vector of controls as in equation (1). In this specification, γ_1 shows the migrant treatment effect with no higher education signal, for female senders, and for daughters, respectively. γ_2 indicates the effect of sending a higher education signal (vs. no signal), being a male sender (vs. a female sender), and looking for child care for a son (vs. for a daughter), respectively, always for native senders. γ_3 indicates whether the effect of any of the three mechanisms differs between natives and migrants.

Our findings are robust in a wide array of alternative model specifications, such as probit estimations and the inclusion of ZIP code fixed effects. Results are also similar when

[19] In Appendix D, we provide further details on variable definitions and data sources. There we also describe in detail how we handle missing observations. We have missings for a child care center's maximum capacity (6.4%), its provider type (29.4%), and the share of migrants in the center's municipality due to missing information about the center's location (7.0%). We impute missing values for maximum capacity and share of migrants by the sample mean of the next-higher higher regional level (i.e., county, or NUTS2 level). We further assign missing values for the provider type to the "else" category. To ensure that the imputed data are not affecting our results, all regressions include an indicator for each variable with missing observations that equals 1 for imputed values and 0 otherwise.

using randomization inference and are robust to multiple hypothesis testing corrections following List et al. (2019), Westfall and Young (1993), and Romano and Wolf (2005). These robustness checks are described in more detail in Section 5.3.

4.2. Outcome Measures

The main outcomes of our study are (i) a binary indicator for whether or not a contacted child care center replies to the email request and (ii) several dimensions of the content of the email response, described below.

We calculate the length of responses by using an automated method that counts the number of characters in the email body. For the other content dimensions, we employed a team of five research assistants ("reviewers"), each of whom rated all 12,547 email responses independently. The reviewers received a half-day training session and a manual that explained how to review each email and code responses. Importantly, reviewers were not informed about the purpose of the study or that there were randomized treatments. Furthermore, reviewers were blind to both the treatment assignment, i.e., the name of the sending parent, as well as the identity of the child care centers or managers, as we deleted all this information from the emails prior to the rating.

For each outcome dimension, we aggregate the five reviewer ratings into a binary indicator using simple majority rule (see Appendix G for details on the rating process and on the variable definitions).[20] To verify the robustness of our email content analysis, we show that results hold when using a supervised machine-learning algorithm to classify the content of email responses instead of human coding (see Section 5.3 and Appendix H).

To capture the various possible dimensions of discrimination that may arise in email responses, we defined the following set of content dimensions which were then rated by each reviewer:

Slot Offer. Our most important content dimension is whether the email response contains a slot offer to parents, as it allows us to uncover discrimination in actual enrollment decisions. Although the email requests only inquired about the availability of slots and the application process, 798 center managers (6.4% of answers) directly offered slots in their response email (we declined all slot offers within 24 hours).

Waiting List. Child care centers may also extend the option to place families on a waiting list instead of directly offering a slot. Being on a waiting list presents the potential for future enrollment, contingent on the availability of a slot at the center. Out of the total

[20] All results are robust to using alternative aggregation rules (see Section 5.3).

email responses received, 9,976 responses (79.51%) contained an offer for a waiting list spot. While placement on a waiting list moves families closer to receiving a slot offer, such offers are non-binding and will often not lead to actual enrollment.

Long Response. The length of the email response serves as a proxy for its informational value. Receiving information about the child care application process is particularly important for migrant parents, as they often lack such information (Hermes et al., 2021). We measure response length as the number of characters in the text body (excluding names, signatures, and email histories), and create a dummy indicating above-median response length for exposition (results are qualitatively similar for continuous length).

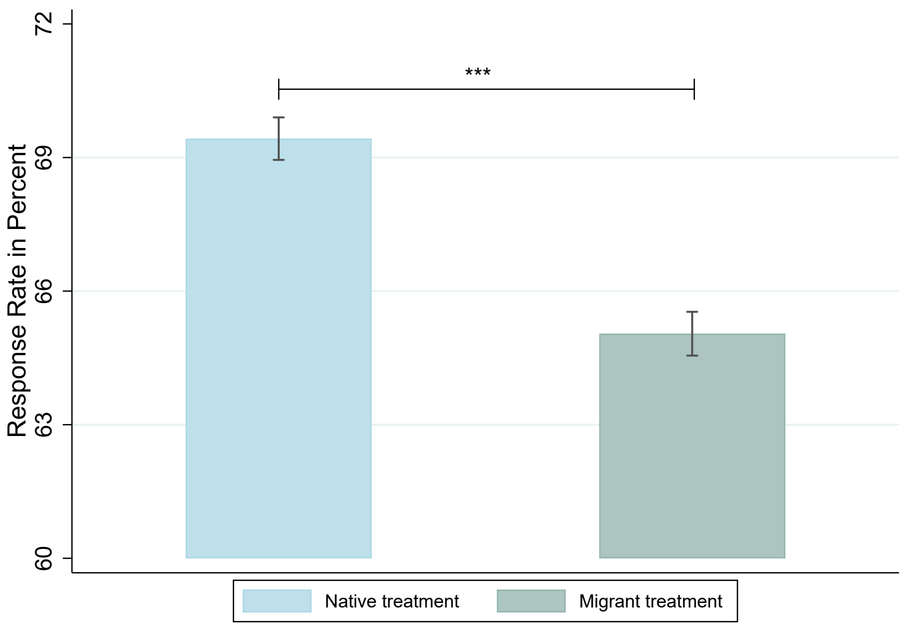
Helpful Content. We also consider the inclusion of practical information that supports the application process as another content dimension. This encompasses details such as a contact telephone number, a link to a registration portal or the center's website, mentions of alternative institutions, or an application form. This dimension further aims to measure whether the information provided enables senders to continue their search for a child care slot, even in instances where the initially contacted center has no available slots. Reviewers identify helpful content in 6,013 emails, accounting for 47.9% of responses.

Encouraging and Recommendation. The final two email content dimensions focus on the tone of the email response, recognizing that a negative tone might deter parents from applying for a slot at the responding center.[21] Given the challenge of objectively measuring the tone of unstructured textual emails, we depend on subjective evaluations by reviewers. They were tasked with determining (i) if the email response was encouraging and (ii) based on their overall impression, whether they would recommend a befriended couple seeking child care for their 1.5-year-old child to apply to the responding center.

An important question in analyzing treatment effects on content measures is whether outcomes should be considered unconditional on whether a child care center responds (i.e., non-response would be coded as zero), or conditional on response (i.e., non-response would be coded as missing). The advantage of unconditional content measures is that estimates are not biased by selection into response. However, the disadvantage is that any treatment effects observed on unconditional content measures may be attributable

[21] Discouraging emails could potentially prevent parents from continuing their search for a child care slot. To illustrate: In Hermes et al. (2021), a non-negligible share of parents actually believe that they can apply only to a single child care center or even only to the geographically closest center.

Figure 2: Treatment Effects on Response Rate



Notes: Figure shows response rates across treatment cells, based on multivariate OLS regressions shown in Table C1, Column (2). The left (blue) bar depicts response rates to emails from native senders and the right (green) bar depicts response rates to emails from migrant senders. The response rate difference between native and migrant senders is statistically significant at the 1%-level. Differences are tested with two-sided t-tests. Error bars indicate standard errors. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01.

solely to the impact of the treatment on response rates. Therefore, our content analysis in Section 5.2 presents both, treatment effects unconditional and conditional on response.

5. Results

5.1. Results for Response Rate

We present our main findings for response rates in Figure 2. The figure shows response rates to emails sent by parents with a native or a migrant name, estimated based on equation (1) (see also Table C1). Compared to the native control group, emails signaling migrant background received significantly fewer responses ($p < 0.001$). The treatment effect of -4.4 pp translates to a 6.2% decrease relative to the response rate in the native control group. Effects are sizable in comparison to, for example, black-white response gaps of about 2 to 4 pp in recent U.S.-based studies (Bergman and McFarlin, 2018; Giulietti et al., 2019; Kline et al., 2022). These results indicate that migrant parents searching for

Table 2: Treatment Effects on Response Content

	(1) Slot Offer	(2) Waiting List	(3) Long Resp.	(4) Helpful Content	(5) Encouraging	(6) Recommendation
Panel A (Unconditional)						
Migrant treatment	-0.011***	-0.043***	-0.066***	-0.036***	-0.028***	-0.064***
	(0.003)	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean (Native Sender)	0.049	0.566	0.467	0.346	0.157	0.419
Scaled Treatment Effect	-22.8	-7.6	-14.2	-10.5	-17.9	-15.3
N	18,663	18,663	18,663	18,663	18,663	18,663
Panel B (Conditional)						
Migrant treatment	-0.010**	-0.013*	-0.059***	-0.024***	-0.026***	-0.059***
	(0.004)	(0.007)	(0.009)	(0.009)	(0.007)	(0.009)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Scaled Treatment Effect	-20.2	-2.3	-12.7	-6.9	-16.4	-14.0
N	12,547	12,547	12,547	12,547	12,547	12,547

Notes: Table shows treatment effects on email content measures, based on multivariate OLS regressions. Outcome variables are defined as follows: Column (1): indicator for whether the contacted child care center offers a child care slot before the next turn cycle (August 2022); Column (2): indicator for whether the contacted child care center offers a spot on the waiting list; Column (3): indicator for whether the length of the email response, measured as the number of characters in the email body, is above median; Columns (4) and (5): indicators for whether a child care center responds with a "helpful content" or in an "encouraging" manner; Column (6): indicator for whether the reviewers would recommend the child care center to a befriended couple with a young child. In Panel A, outcome variables receive a value of zero for non-responses (e.g., a non-response is coded as no offer in Column (1)). In Panel B, non-responses are excluded from the estimation sample, so results are conditional on receiving a response. See Section 4.2 and Appendix G for a description of the email rating procedure. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). *Scaled treatment effect* expresses the treatment effect relative to the mean of the respective outcome in the control group of native senders in percent. Robust standard errors in parentheses. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01. We additionally report p-values based on randomization inference and correcting for multiple hypothesis testing in Table C2.

a child care slot face substantial discrimination at the "extensive margin," which refers to the lower rate of email responses they receive compared to native parents. Next, we investigate whether discrimination also exists at the "intensive margin" by examining the content of the email responses.

5.2. *Results for Email Content*

In this section, we investigate treatment effects on six binary dimensions of email content, based on manual ratings by five independent reviewers (see Section 4.2 for details): (i) whether parents are offered a slot (*Slot Offer*), (ii) whether parents are put on a waiting list (*Waiting List*), (iii) whether parents receive a response of above-median length (*Long Response*), (iv) whether the response included helpful information for parents (*Helpful Content*), (v) whether the response is rated as "encouraging" (*Encouraging*), and (vi) whether reviewers would recommend a befriended couple with a young child to apply to the responding center (*Recommendation*).

We detail our findings for these six content dimensions in Table 2. As the response rate is affected by the migrant treatment, there is non-random selection into response (see Section 4.2 for a discussion). We therefore estimate treatment effects in two ways: unconditional on response (including all contacted child care centers, coding content outcomes of non-responders as zero) and conditional on response (including only centers that provided a response, where content outcomes of non-responders are treated as missing).

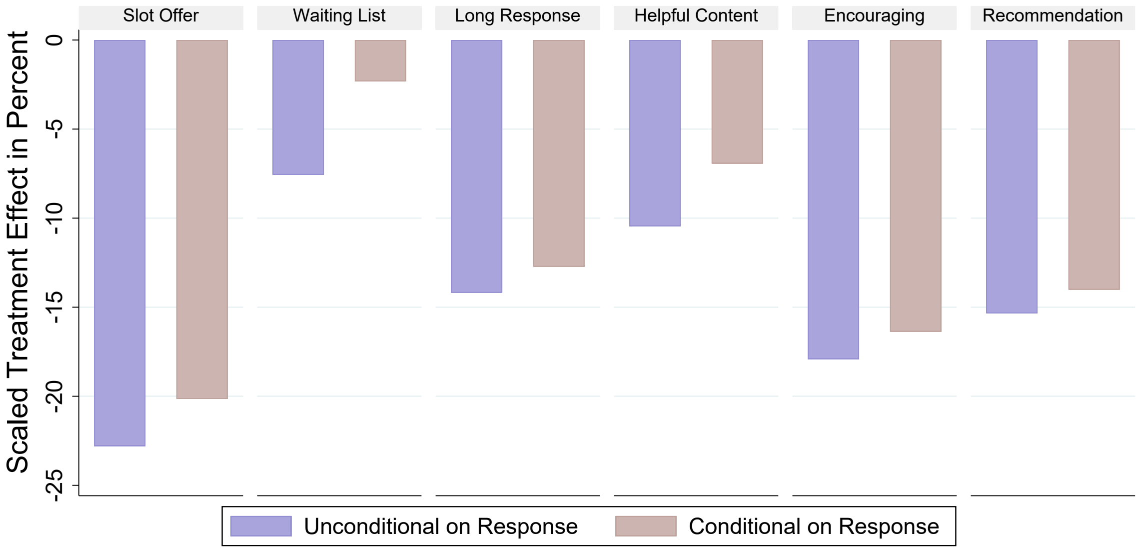
Panel A of Table 2 shows that there is a strong negative effect of the migrant treatment on all six dimensions of email content in the full sample (i.e., unconditional on response). All treatment effects are significant at the 1%-level. In particular, parents with a migrant name are 1.1 pp less likely to receive a slot offer, which represents a treatment effect of -23.1% when scaled by the control-group mean. Migrant parents are also 4.3 pp less likely to be put on a waiting list and 6.5 pp less likely to receive a response with above-median length, which corresponds to treatment effects of -7.5% and -14%, respectively.[22] Email responses to migrant parents also contain fewer helpful elements and are more negative in tone: They are 3.5 pp (10.2%) less likely to be rated as containing helpful content, 2.8 pp (17.5%) less likely to be rated as encouraging, and 6.3 pp (15.1%) less likely to induce raters to recommend applying to the responding center.

Panel B confirms that all treatment effects also hold in the subsample of centers that responded (i.e., conditional on response). Mechanically, all effects get smaller in size but remain significant at the 5%-level or better, except for the effect on being put on a waiting list ($p = .071$). Thus, migrant parents not only receive fewer responses, but the responses they do receive are also substantially worse in terms of their content.

Next, we assess the extent to which discrimination in content outcomes stems from the extensive or intensive margin. For this analysis, we express all treatment effects from Table 2 relative to the mean in the native control group (full sample), and compare their magnitude between the full sample and the subsample of responding centers. The results, displayed in Figure 3, reveal the following key insights: First, all scaled treatment effects (except for *Waiting List*) are large, amounting to at least 10% of the control group mean. Moreover, we observe the largest scaled treatment effects on arguably the most important outcome, as the likelihood of migrant parents receiving a slot offer is more than 20% lower than that of native parents, both unconditional and conditional on response.

[22] If we estimate the treatment effect on response length measured by the number of characters, we find a treatment effect of -43 characters (or 14.4% shorter emails, $p < .001$). Given that a German word has on average 5–7 characters, the size of the treatment effect corresponds to emails being about one sentence shorter.

Figure 3: Scaled Treatment Effects on Response Content



Notes: Figure shows scaled treatment effects for email content outcomes, based on the multivariate OLS models shown in Table 2. *Scaled treatment effect* expresses the treatment effect relative to the sample mean (unconditional on response) of the respective outcome in the native control group.

Second, we find that discrimination in content outcomes is mainly driven by receiving a worse response (i.e., the intensive margin) for all outcomes except for *Waiting List*. For instance, the scaled treatment effect on slot offers is -23.1% unconditional on response (purple bar), and only slightly decreases to -20.5% conditional on response (brown bar).

Overall, our findings demonstrate that child care centers not only discriminate against migrants by not responding to emails, but also through the content of the emails they send in response, which are shorter, contain less helpful information, and less encouraging than those to natives. This is a more subtle and obfuscated form of discrimination than simply not responding to emails (and thus withholding information). Most importantly, emails to migrants contain substantially fewer slot offers, so child care centers also discriminate in terms of actual enrollment into early child care.

5.3. Robustness Checks

Our results are robust in an extensive number of additional analyses. First, we use randomization inference and apply different procedures to correct for multiple hypothesis testing (see Table C2). Second, we estimate Probit regressions instead of linear probability models when analyzing response rates and email content outcomes (see Table C3). Third, to compare child care centers that are as homogeneous as possible in (unobservable) characteristics, we include zip-code level fixed effects ($k = 3,263$). Table C4 shows that our results are robust to using only within-zip-code variation; in fact, treatment effects even get slightly larger for most outcomes.

Furthermore, we test the robustness of our email content outcomes. To create the content measures, we aggregate the five reviewers' ratings from a four-point Likert scale into binary measures (see Appendix G). We then combine the individual binary ratings into one binary indicator for each content dimension by applying simple majority rule (i.e., we use the rating given by at least three reviewers). However, since this simple majority decision rule is somewhat arbitrary, we check if our results hold when using alternative aggregation rules. Table C5 reports the results for content measures when one, two, three, four, or all five reviewers rated an email as a slot offer, waiting list offer, etc. Results are strikingly consistent across aggregation methods, both in terms of magnitude (in particular, for the scaled treatment effects) and statistical significance (with only a few exceptions, all effects are significant at 5% or better).

In a similar vein, we also check whether our results are sensitive to the transformation of the reviewers' ratings elicited on four-point scales into a binary scale. In Table C6, we use the full variation in the ratings by constructing a standardized version of the four-point scale ratings.[23] Reassuringly, all treatment effects for the standardized content outcomes are negative, large, and statistically significant at the 1%-level.

Finally, although the reviewers were blind to the experimental design, treatment assignment, and identity of the child care centers and managers, there is still a possibility that they could have been subject to some form of (systematic) bias in their ratings. To address this potential concern, we collaborated with computer scientists to develop a BERT model.[24] This is a pre-trained natural language processing model, which we use for an alternative classification of the email content outcomes (i.e., slot offers and waiting list offers) (see Appendix H for details). Our results are robust to using this alternative, purely computational classification of content outcomes (see Table H1). In fact, the scaled treatment effects are remarkably close to those obtained from manual coding.[25]

[23]Following Kling et al. (2007), we first z-standardize each reviewer's four-point-scale rating, then create an equally weighted average of the standardized ratings across all five reviewers, and finally z-standardize the averages again.

[24]BERT stands for Bidirectional Encoder Representations from Transformers (see Devlin et al., 2018).

[25]When employing the computational classification approach, scaled treatment effects are -22.3% (unconditional) and -22.0% (conditional) (-8.1% and -3.0%) for slot offers (waiting list offers), compared to -22.8% and -20.2% (-7.8% and -2.3%) when using manual coding (see Table 2).

6. Exploring the Channels for Discrimination

To be able to address discrimination on the child care market, we need to better understand *why* it occurs. In this section, we explore potential underlying reasons for discrimination against migrant parents searching for a child care slot. In a first step, we focus on three additional treatments that allow us to pin down causal channels. In a second step, we further investigate the channel of (perceived) increased effort to educate children from migrant families, using a representative survey, a large array of administrative data, and a data-driven approach based on a causal forest analysis. We show all results for response rates and slot offers, as the latter is the email content outcome most closely linked to actual child care enrollment.

Education Signal. First, we examine whether discrimination might be based on child care center managers' beliefs about parents' educational background (see Table 3, Column (1)). As described in Section 3.2, we randomly varied whether the email contained a signature indicating that the sender has a higher education degree. However, even when we compare senders signaling higher education, the response rate to emails from migrants is 3.7 pp lower than to emails from natives ($p < 0.001$). While the treatment effect is somewhat larger for senders without a higher education signal (5.1 pp), the difference is not statistically significant, as indicated by the interaction term ($p = 0.309$). This pattern also holds when we investigate treatment effects on slot offers in Column (4). These results suggest that child care managers' lower response rates towards migrant parents cannot be attributed to their beliefs about the educational background of migrants and natives.

Sender Gender. Second, we consider the randomized gender of the email sender. The intuition behind this analysis is that the gender of a sender might send different signals about the role of women in native vs. migrant families, as gender roles might depend on the associated culture of a sender (e.g., Alesina et al., 2013). These differently perceived gender roles could in turn influence the reaction of child care center managers to email inquiries. For example, if child care center managers harbor stereotypes about traditional gender roles being more prevalent in Turkish families, they might respond differently to male than female senders with Turkish names. An inquiry from a Turkish father might be met with a more welcoming response as it signals that child care responsibilities are divided more equally than managers would have expected. Alternatively, a mail from a Turkish father might as well indicate to managers a family structure where fathers are the predominant decision-makers or that the Turkish mother has low German language skills, prompting the father to communicate with the child care center.

Table 3: Treatment Effect Heterogeneity on Response Rates and Slot Offers

	Response Rate			Slot Offer		
	(1)	(2)	(3)	(4)	(5)	(6)
Migrant treatment	-0.051***	-0.036***	-0.030***	-0.010**	-0.005	-0.004
	(0.010)	(0.009)	(0.010)	(0.004)	(0.004)	(0.004)
× Higher education signal	0.014			-0.003		
	(0.013)			(0.006)		
× Sender male		-0.018			-0.015**	
		(0.014)			(0.006)	
× Child male			-0.027**			-0.013**
			(0.014)			(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Migrant + Migrant × interaction	-0.037***	-0.053***	-0.057***	-0.013***	-0.019***	-0.018***
N	18,663	18,663	18,663	18,663	18,663	18,663

Notes: Table shows treatment effect heterogeneity on the response rate and slot offers, based on multivariate OLS regressions. Heterogeneity by: Columns (1) and (4), show treatment effects for including the higher education signal; Columns (2) and (5), show treatment effects for emails signaling a male sender; and Columns (3) and (6) sow treatment effects for emails signaling a male child. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Higher education signal* is an indicator variable taking a value of one if the email includes a signature that indicates a higher educational background of the sender, and zero if the email does not include a signature. *Sender male* is an indicator variable taking a value of one if the email indicates that the sender of the email is male/the father, and zero if the email indicates that the sender of the email is female/the mother. *Child male* is an indicator variable taking a value of one if the email indicates that the child of the sender is a boy, and zero if the email indicates that the child is a girl. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Robust standard errors in parentheses. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01. We additionally report p-values based on randomization inference and correcting for multiple hypothesis testing in Table C2.

We find that migrant fathers tend to be treated less favorably than migrant mothers, but the evidence is not fully conclusive. In Table 3, the migrant treatment effects for fathers are stronger than those for mothers for both response rates (Column (2)) and slot offers (Column (5)), albeit not being statistically significant for response rates ($p = 0.209$). These findings give some support to the idea that differing beliefs about gender roles in migrant families held by center managers affect their response behavior. For female senders, the probability to receive a slot offer does not even differ significantly between migrants and natives, while we observe a distinct disadvantage for migrants when the mail originates from a male sender.[26]

Child Gender. Third, we examine whether the effects of the migrant treatment are influenced by the gender of the child ('daughter' vs. 'son', see Figure 1). Randomly varying child gender was motivated by evidence from the United States, suggesting that

[26]Investigating sender gender differences for natives, we show in other work that response rates do not differ between mothers and fathers, while mothers receive less positive responses (Hermes et al., 2023).

preschool educators perceive *boys* as more disruptive (Gilliam et al., 2016). Similarly, in our context, child care center managers might perceive migrant boys as more disruptive or, more generally, requiring more effort and being more costly to educate, leading to differential email responses based on the child's gender. Indeed, we find that the migrant-native gap widens substantially for emails sent on behalf of boys, by 2.7 pp for responses ($p = 0.049$) and 1.3 pp for slot offers ($p = 0.021$) (Table 3, Columns (3) and (6)).[27]

In sum, Table 3 shows that child care center managers' beliefs about education background or gender roles in migrant families do not appear to be systematically related to the level of discrimination, but that center managers discriminate significantly stronger against migrant boys. Combined with the evidence showing that disadvantaged boys in preschool are (perceived as) particularly difficult to educate (Bertrand and Pan, 2013; Gilliam et al., 2016), this observation makes it plausible that the expected effort associated with the enrollment of a child is a relevant determinant in the decision of child care center managers about whether and how to reply to emails sent by parents. We therefore hypothesize that higher perceived effort required to educate migrant children could be a major reason for discrimination.

Representative Survey. As a first test of this hypothesis, we were able to include a question in a large-scale survey ($N \approx 4800$) representative of the (adult) German population (the "Inequality Barometer" run by the University of Konstanz, see Appendix E.3 for details). In this survey, we asked participants: "According to a recent scientific study, Turkish parents have lower chances of successfully applying for child care slots compared to German parents. How would you explain these lower chances for Turkish parents?" The participants could then select multiple answers (or provide open text responses). Reassuringly, the most frequent reason (selected by half of the participants) was the statement that child care centers perceive Turkish families as representing a higher workload for them. In other words, a large fraction of the German population believes that increased effort and higher workload is a key explanation for the discrimination against migrants in the child care market, strongly supporting our hypothesis above.[28]

[27]Further analyses based on triple-interaction models reveal that the effects for migrant fathers and migrant boys are additive, i.e., there is no significant interaction effect for emails sent by migrant fathers on behalf of migrant boys.

[28]Hermes et al. (2024) provides more information about the survey. Interestingly, the second most frequently selected answer was that child care centers would make sure not to enroll too many Turkish children, as this would be what "many parents want". In line with the idea that the decision-making of managers is partially driven by what they perceive to be the parents' preferences, we observe that

Resources and Incentives. To lend further support to the hypothesis that discrimination against migrants in early child care reflects higher effort associated with migrant children, we link our experimental data to administrative data at three different levels of aggregation: (municipality, county, and federal state; see Budde and Eilers, 2014; INKAR, 2021). These data provide various indicators of resources available to child care centers, which potentially help them to cope with the (perceived) increased effort of enrolling migrant children (e.g., due to language barriers or additional organizational tasks related to food or other requirements). In Table 4, we present evidence on how discrimination varies with regional resource characteristics, based on the model in equation (2). While these results are descriptive in nature, given the lack of exogenous variation in the resource variables, they lend credence to the notion that the causal evidence presented in Table 3 indeed reflects (perceived) higher cost to educate migrant children.

In fact, the premise that migrant children might incur higher educational costs for child care centers is already reflected in the public funding structures supporting these institutions. In Germany, nine out of 16 federal states provide additional resources to child care centers to educate migrant children, mainly in the form of additional funding or staff. If the cost channel is relevant for discrimination, these incentives should reduce discrimination against migrants. Our findings support this hypothesis. Table 4 indicates stronger discrimination in federal states without financial incentives to educate migrant children. The interaction is economically important for both response rates (Column (1)) and slot offers (Column (5)), but is statistically significant at conventional levels only for response rates ($p < 0.001$; $p = 0.115$ for slot offers).

While financial incentives for catering to migrant children is directly linked to child care centers' incurred cost of educating migrant children, these incentives only vary at the federal state level ($k = 16$). For the remainder of Table 4, we thus consider resource indicators that vary at a finer regional level. We first consider the average staff-to-child ratio in the county ($k = 400$) where the respective center is located. If centers can afford more care workers relative to the number of children in care, we expect them to be more adept at managing any extra requirements that might emerge from accommodating migrant children. Indeed, Columns (2) and (6) show that centers in counties with a higher staff-to-child ratio tend to discriminate less ($p = 0.009$ for response rates and $p = 0.025$ for slot offers).

---

discrimination against migrants is stronger in constituencies with a higher far-right-wing vote share. Results are available upon request.

Table 4: Heterogeneity by Regional Characteristics

	Response Rate				Slot Offers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Migrant treatment	-0.124***	-0.047***	-0.044***	-0.044***	-0.029**	-0.011***	-0.009***	-0.010***
	(0.020)	(0.007)	(0.008)	(0.008)	(0.012)	(0.003)	(0.003)	(0.003)
× Migrant incentive	0.088***				0.020			
	(0.021)				(0.013)			
× Staff-to-child ratio (std.)		0.018***				0.008**		
		(0.007)				(0.004)		
× Budget per capita (std.)			0.015				0.014**	
			(0.011)				(0.006)	
× Resource index (std.)				0.032***				0.012***
				(0.008)				(0.004)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	16,281	16,281	16,281	16,281	16,281	16,281	16,281	16,281

Notes: Table shows treatment effect heterogeneity on the response rate and slot offers, based on multivariate OLS regressions. We estimate the regressions on the sample for which all observations have complete information in all variables (see Appendix D). *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. In Columns (1) and (5), *Migrant incentive* indicates whether the federal state ($k = 16$) of the child care center provides an additional financial incentive for taking up migrant children; in Columns (2) and (6), *Staff-to-child ratio* measures the average ratio of pedagogical staff to the number of slots in a child care center in a county ($k = 400$); in Columns (3) and (7), we use the *Budget per capita* (tax income) of a municipality ($k = 3,707$); in Columns (4) and (8), we create a *Resource index*, combining all three heterogeneities, i.e., the migrant incentive, the staff-to-child ratio, and budget per capita. The resource index is computed using the loadings on a single factor derived from a factor analysis. Staff-to-child ratio, budget per capita, and resource index are standardized with mean = 0 and standard deviation = 1. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Robust standard errors in parentheses. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$. We additionally report p-values based on randomization inference and correcting for multiple hypothesis testing in Table C2.

Yet, county-level variables might still mask substantial heterogeneity in resources available to individual child care centers. To proxy resources for child care centers on an even more granular level, we leverage data on municipalities' financial capacity ($k = 3,707$). The idea behind this analysis is that richer municipalities will, on average, allocate more resources to child care centers. Measuring a municipality's financial capacity by its tax income per capita, Columns (3) and (7) of Table 4 show that child care centers located in municipalities with larger budgets tend to discriminate less against migrants ($p = 0.156$ for response rates and $p = 0.029$ for slot offers).

Finally, to reduce measurement error and to mitigate multiple testing concerns, we also employ a factor analysis to combine all three resource variables into a single *resource index* reflecting the overall resources of a child care center.[29] The results for this resource index, depicted in Columns (4) and (8) of Table 4, show a consistent pattern: Centers with a better resource endowment discriminate less, both for response rates ($p < 0.001$) and

[29]Results are robust to using an index that gives the same weight to each of the three variables (Kling et al., 2007).

for slot offers ($p = 0.006$). Thus, this analysis supports the hypothesis that discrimination against migrant families in searching and applying for early child care is — at least partly — driven by perceived higher cost to educate migrant children.

Causal Forest. We also complement our previous analysis with a purely data-driven approach. Specifically, we employ a causal forest analysis, which identifies variables that are important in predicting Conditional Average Treatment Effects (CATEs). CATEs describe the expected difference in outcomes for treated vs. untreated individuals from a particular subgroup, conditional on a set of covariates. As covariates, we use the three additional treatment variations (education signal, sender gender, and child gender), the resource index, and the full vector of control variables from equation (1) (see Appendix I for an explanation of the approach and details on its implementation). The findings from the causal forest analysis further corroborate the hypothesis that the resources at the disposal of child care centers affect the extent of discrimination. This is evidenced by the resource index exhibiting by far the highest variable importance (0.40, nearly double that of any other variable) for predicting CATEs.

The collective evidence consistently suggests that the anticipated higher costs of educating migrant children are an important factor driving discrimination in the child care market. This conclusion is based on a combination of three additional treatments analyzing causal mechanisms, a representative survey on reasons for discrimination, heterogeneity analyses (by various resource indicators varying at the level of child care centers, municipalities, counties, and federal states), and a data-driven causal forest approach. Consequently, policymakers may be able to address discrimination in early child care by ensuring adequate resources and financial incentives for the education of migrant children.

7. Conclusion

We provide the first causal evidence that migrants are discriminated against when searching for a slot in early child care. Using a randomized email correspondence study with a large, nationwide sample of child care centers in Germany, we find that emails from fictitious migrant parents have a 4.4 pp (6.2%) lower chance of receiving a response compared to emails from native parents. Importantly, discrimination is not only present at the extensive margin of receiving an email response, but also at the intensive margin of email content, including the likelihood of receiving a slot offer.

As is the case in many social programs worldwide, the process of applying for a child care slot is complex, leading parents to seek information from child care centers before submitting an official application. We show that discrimination occurs at this difficult-to-observe pre-application phase, for which, in contrast to formal applications or admission decisions, typically no data are recorded. This discrimination might further explain the well-established problem of disadvantaged groups being underrepresented in social programs for which they are eligible, especially in contexts where discrimination in official admission decisions is explicitly prohibited by law (Currie, 2006; Ko and Moffitt, 2022).

The documented discrimination likely contributes to the large gap in early child care enrollment rates between migrants and natives, thereby reinforcing existing inequalities of opportunities for disadvantaged children, who could greatly benefit from attending child care in both the short and long term (Cornelissen et al., 2018; Felfe and Lalive, 2018; García et al., 2020). Importantly, the adverse effects of discrimination in the child care market are likely to extend beyond children themselves. In particular, mothers of migrant children may also suffer, as access to child care is an important prerequisite for integration into the labor market (see, e.g., Baker et al., 2008; Bauernschuster and Schlotter, 2015) and strongly improves within-household gender equality (Hermes et al., 2024). The far-reaching consequences of limited child care access for children, mothers, and society at large highlight the need to expand child care capacities and implement structural improvements that promote equitable access to universal early child care.

References

Aizer, A. (2007). Public Health Insurance, Program Take-up, and Child Health. *Review of Economics and Statistics 89*(3), 400–415.

Alan, S., E. Duysak, E. Kubilay, and I. Mumcu (2023). Social Exclusion and Ethnic Segregation in Schools: The Role of Teachers' Ethnic Prejudice. *Review of Economics and Statistics 105*(5), 1039–1054.

Alesina, A., M. Carlana, E. La Ferrara, and P. Pinotti (2018). Revealing Stereotypes: Evidence from Immigrants in Schools. Working Paper w25333, National Bureau of Economic Research.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the Origins of Gender Roles: Women and the Plough. *Quarterly Journal of Economics 128*(2), 469–530.

Alt, C., B. Gedon, S. Hubert, K. Hüsken, and K. Lippert (2019). DJI-Kinderbetreuungsreport 2018 - Inanspruchnahme und Bedarfe bei Kindern bis 14 Jahren aus Elternperspektive - ein Bundesländervergleich. DJI 2019, Deutsches Jugendinstitut, München.

Arcidiacono, P., J. Kinsler, and T. Ransom (2022). Asian American Discrimination in Harvard Admissions. *European Economic Review 144*, 104079.

Athey, S. and S. Wager (2019). Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies 5*(2), 37–51.

Baert, S. (2018). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. In S. M. Gaddis (Ed.), *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, pp. 63–77. Springer International Publishing.

Baker, M., J. Gruber, and K. Milligan (2008). Universal Child Care, Maternal Labor Supply, and Family Well-being. *Journal of Political Economy 116*(4), 709–745.

Baron, E. J., J. Doyle, Joseph J, N. Emanuel, P. Hull, and J. Ryan (2024). Discrimination in Multi-Phase Systems: Evidence from Child Protection. *The Quarterly Journal of Economics*, qjae007.

Bauernschuster, S. and M. Schlotter (2015). Public Child Care and Mothers' Labor Supply—Evidence from Two Quasi-experiments. *Journal of Public Economics 123*, 1–16.

Bell, E. and S. Jilke (2024). Racial Discrimination and Administrative Burden in Access to Public Services. *Scientific Reports 14*(1), 1071.

Bergman, P. and I. McFarlin (2018). Education for All? A Nationwide Audit Study of School Choice. Working Paper w25396, National Bureau of Economic Research.

Bertrand, M. and E. Duflo (2017). Field Experiments on Discrimination. *Handbook of Economic Field Experiments 1*, 309–393.

Bertrand, M., E. F. P. Luttmer, and S. Mullainathan (2000, 08). Network Effects and Welfare Cultures. *Quarterly Journal of Economics 115*(3), 1019–1055.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review 94*(4), 991–1013.

Bertrand, M. and J. Pan (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics 5*(1), 32–64.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment. *Quarterly Journal of Economics 127*(3), 1205–1242.

Bhargava, S. and D. Manoli (2015). Psychological Frictions and the Incomplete Take-up of Social Benefits: Evidence from an IRS Field Experiment. *American Economic Review 105*(11), 3489–3529.

Bitler, M. P., J. Currie, and J. K. Scholz (2003). WIC Eligibility and Participation. *Journal of Human Resources 38*, 1139–1179.

BMFSFJ (2013). Achtes Sozialgesetzbuch - SGB VIII (Kinder- und Jugendhilfegesetz), §24. Bundesministerium für Familie, Senioren, Frauen und Jugend, Bundesgesetzblatt. BGBl. I S. 1666.

BMJ (2006). Allgemeines Gleichbehandlungsgesetz. Bundesministerium der Justiz, Bundesgesetzblatt. BGBl. I S. 1897.

Bohren, J. A., P. Hull, and A. Imas (2022). Systemic Discrimination: Theory and Measurement. Working Paper w29820, National Bureau of Economic Research.

Budde, R. and L. Eilers (2014). Sozioökonomische Daten auf Rasterebene: Datenbeschreibung der Microm-Rasterdaten. RWI Materialien 77, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen.

Carlana, M. (2019). Implicit Stereotypes: Evidence from Teachers' Gender Bias. *Quarterly Journal of Economics 134*(3), 1163–1224.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schoenberg (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy 126*(6), 2356–2409.

Currie, J. (2006). *The Take-up of Social Benefits*, pp. 80–148. Russell Sage Foundation.

de Lafuente, D. M. (2021). Cultural Assimilation and Ethnic Discrimination: An Audit Study with Schools. *Labour Economics 72*, 102058.

Destatis (2020). *Statistiken der Kinder- und Jugendhilfe.* German Federal Statistical Office, Wiesbaden.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, preprint:1810.04805.

DJI (2021). Fachkräftebarometer Frühe Bildung 2021. Weiterbildungsinitiative Frühpädagogische Fachkräfte, Autorengruppe Fachkräftebarometer, München.

Doepke, M., G. Sorrenti, and F. Zilibotti (2019). The Economics of Parenting. *Annual Review of Economics 11*(1), 55–84.

Drange, N. and T. Havnes (2019). Early Childcare and Cognitive Development: Evidence from an Assignment Lottery. *Journal of Labor Economics 37*(2), 581–620.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2018). Closing the Gap: The Effect of a Targeted, Tuition-free Promise on College Choices of High-achieving, Low-Income Students. Working Paper w25349, National Bureau of Economic Research.

Dynarski, S., C. Libassi, K. Michelmore, and S. Owen (2021). Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-income Students. *American Economic Review 111*(6), 1721–1756.

Dynarski, S. M. and J. E. Scott-Clayton (2006). The Cost of Complexity in Federal Student Aid: Lessons from Optimal Tax Theory and Behavioral Economics. *National Tax Journal 59*(2), 319–356.

Education Report (2020). *Bildung in Deutschland 2020 - Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt.* Autorengruppe Bildungsberichterstattung, Bielefeld: wbv Media GmbH.

Estrada, V. J. K. (2018). Migrant Women Labor-force Participation in Germany: Human Capital, Segmented Labor Market, and Gender Perspectives. IAB-Discussion Paper 12/2018, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Felfe, C. and R. Lalive (2018). Does Early Child Care Affect Children's Development? *Journal of Public Economics 159*, 33–53.

Finkelstein, A. and M. J. Notowidigdo (2019). Take-up and Targeting: Experimental Evidence from SNAP. *Quarterly Journal of Economics 134*(3), 1505–1556.

García, J. L., J. J. Heckman, D. E. Leaf, and M. J. Prados (2020). Quantifying the Life-cycle Benefits of an Influential Early-childhood Program. *Journal of Political Economy 128*(7), 2502–2541.

Gilliam, W. S., A. N. Maupin, C. R. Reyes, M. Accavitti, and F. Shic (2016). Do Early Educators' Implicit Biases Regarding Sex and Race Relate to Behavior Expectations and Recommendations of Preschool Expulsions and Suspensions. *Yale University Child Study Center 9*(28), 1–16.

Giulietti, C., M. Tonin, and M. Vlassopoulos (2019). Racial Discrimination in Local Public Services: A Field Experiment in the United States. *Journal of the European Economic Association 17*(1), 165–204.

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Yavitz (2010). The Rate of Return to the High Scope Perry Preschool Program. *Journal of Public Economics 94*(1-2), 114–128.

Hemker, J. and A. Rink (2017). Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices. *American Journal of Political Science 61*(4), 786–803.

Hermes, H., M. Krauß, P. Lergetporer, F. Peter, and S. Wiederhold (2024). Early Child Care, Maternal Labor Supply, and Gender Equality: A Randomized Controlled Trial. Mimeo.

Hermes, H., P. Lergetporer, F. Mierisch, G. Schwerdt, and S. Wiederhold (2024). Does Information about Inequality and Discrimination in Early Child Care Affect Policy Preferences? CESifo Working Paper No. 10925, CESifo.

Hermes, H., P. Lergetporer, F. Peter, F. Mierisch, and S. Wiederhold (2023). Males Should Mail? Gender Discrimination in Access to Childcare. In *AEA Papers and Proceedings*, Volume 113, pp. 427–431. American Economic Association.

Hermes, H., P. Lergetporer, F. Peter, and S. Wiederhold (2021). Behavioral Barriers and the Socioeconomic Gap in Child Care Enrollment. CESifo Working Paper No. 9282, CESifo.

HRK (2021). Statistische Daten zu Studienangeboten an Hochschulen in Deutschland. Statistiken zur Hochschulpolitik, Hochschulrektorenkonferenz, Berlin.

Hussar, B., J. Zhang, S. Hein, K. Wang, A. Roberts, J. Cui, M. Smith, F. B. Mann, A. Barmer, and R. Dilig (2020). The Condition of Education 2020. *National Center for Education Statistics (NCES 2020-144)*.

INKAR (2021). Indikatoren und Karten zur Raum- und Stadtentwicklung. Statistiken zur räumlichen Zusammensetzung, Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung (BBR), Bonn.

Jessen, J., S. Schmitz, and S. Waights (2020). Understanding Day Care Enrolment Gaps. *Journal of Public Economics 190*, 104252.

Kleven, H. J. and W. Kopczuk (2011). Transfer Program Complexity and the Take-up of Social Benefits. *American Economic Journal: Economic Policy 3*(1), 54–90.

Kline, P., E. K. Rose, and C. R. Walters (2022). Systemic Discrimination Among Large US Employers. *Quarterly Journal of Economics 137*(4), 1963–2036.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental Analysis of Neighborhood Effects. *Econometrica 75*(1), 83–119.

Ko, W. and R. A. Moffitt (2022). Take-up of Social Benefits. *Handbook of Labor, Human Resources and Population Economics*, 1–43.

Landersø, R. and J. J. Heckman (2017). The Scandinavian Fantasy: The Sources of Intergenerational Mobility in Denmark and the US. *Scandinavian Journal of Economics 119*(1), 178–230.

Lavy, V., E. Sand, and M. Shayo (2022). Discrimination Between Religious and Non-religious Groups: Evidence from Marking High-stakes Exams. *Economic Journal 132*, 2308–2324.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics 22*(4), 773–793.

Mocan, N. (2007). Can Consumers Detect Lemons? An Empirical Analysis of Information Asymmetry in the Market for Child Care. *Journal of Population Economics 20*, 743–780.

Neumark, D. (2018). Experimental Research on Labor Market Discrimination. *Journal of Economic Literature 56*(3), 799–866.

OECD (2018). Settling In 2018. Technical Report, OECD, Paris.

Olsen, A. L., J. H. Kyhse-Andersen, and D. Moynihan (2022). The Unequal Distribution of Opportunity: A National Audit Study of Bureaucratic Discrimination in Primary School Access. *American Journal of Political Science 66*(3), 587–603.

Romano, J. P. and M. Wolf (2005). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica 73*(4), 1237–1282.

Spiess, C. K. (2008). Early Childhood Education and Care in Germany: The Status Quo and Reform Proposals. *Zeitschrift für Betriebswirtschaftslehre 2008 67*, 1–20.

Spiess, C. K. (2013). Investments in Education: The Early Years Offer Great Potential. *DIW Economic Bulletin 3*(10), 3–10.

United States Congress (1964). Civil Rights Act of 1964. Public Law. S.1177, 114th Congress, Amended through Public Law 114-95, Enacted December 10, 2015.

Westfall, P. H. and S. S. Young (1993). *Resampling-based Multiple Testing: Examples and Methods for p-Value Adjustment*, Volume 279. John Wiley & Sons.

Online Appendix

for

"Discrimination in Universal Social Programs?
A Nationwide Field Experiment on Access to Child Care"

by

Henning Hermes, Philipp Lergetporer, Fabian Mierisch,
Frauke Peter & Simon Wiederhold

Appendix A. Treatment and Setting

Figure A1: Example for Email: Message on Behalf of a Son with Migration Background, No Higher Education Signal

Dear Sir or Madam,

We are looking for a child care slot for our son starting in January 2022.
He is now 1 year and 5 months old.

Do you still have a slot open? How can we apply for a slot?

Thank you!

Sincerely,
Eylül Yildirim

Figure A2: Example for Email: Message on Behalf of a Daughter with Native Background, with Higher Education Signal

Dear Sir or Madam,

We are looking for a child care slot for our daughter starting in January 2022.
She is now 1 year and 5 months old.

Do you still have a slot open? How can we apply for a slot?

Thank you!

Sincerely,
Sebastian Müller

Sebastian Müller, Bachelor of Arts (FH)
Email: Sebastian.Müller0528@gmail.com

Table A1: Cross-Country Comparison of Early Child Care Systems

Country	Migrant-native Enrollment Gap	Reduced Fees	Slots Rationed	Decentralized
France	Yes[a1]	Yes[b1]	Yes[c1]	Yes[d1]
Germany	Yes[a2]	Yes[b2]	Yes[c2]	Yes[d2]
Italy	Yes[a3]	Yes[b3]	Yes[c3]	No[d3]
Spain	Yes[a4]	Yes[b4]	Yes[c4]	No[d4]
UK	Yes[a5]	Yes[b5]	Yes[c5]	Yes[d5]
US	Yes[a6]	Yes[b6]	Yes[c6]	Yes[d6]

Notes: Table shows features of early child care systems of the five largest European countries (in terms of GDP) and the United States, in alphabetical order. Migrant-native Enrollment Gap: "Yes" if children with a migration background are underrepresented in early child care. Reduced fees: "Yes" if lower-income families are eligible for fee reductions or exemptions. Slots rationed: "Yes" if average demand for a child care slot exceeds average supply. Decentralized admission decision: "Yes'" if admission decisions are taken by individual child care centers.

Sources: [a1,a3,a4,a5] OECD (2018); [a2,c2,d2]Jessen et al. (2020); [a6] Cui et al. (2021); [b1,c1,b3,c3,b4,c4,b5,c5] Eurydice (2019); [b2] Felfe and Lalive (2018); [b6] OECD (2020); [c6] Malik et al. (2018); [d1] Expat (2022); [c3] Del Boca et al. (2016); [d4] Harvey (2022); [d5] Renfrewshire Council (2023); [d6] NYC Department of Education (2023).

Appendix B. Balancing

Table B1: Balancing Migrant Treatment (Analysis Sample)

	(1) Native treatment		(2) Migrant treatment		(2)-(1)	(2)-(1)
	Mean	(SD)	Mean	(SD)	Diff (mean)	p-Value
Email characteristics						
Higher education signal	0.497	(0.500)	0.501	(0.500)	0.004	0.563
Sender male	0.501	(0.500)	0.499	(0.500)	-0.002	0.764
Child male	0.502	(0.500)	0.507	(0.500)	0.005	0.516
Child care center characteristics						
Center's maximum capacity	68.854	(43.209)	68.713	(45.482)	-0.141	0.828
Kindergarten (age 3 – 6 years)	0.933	(0.249)	0.929	(0.257)	-0.005	0.220
Afternoon care (age >6 years)	0.094	(0.291)	0.098	(0.297)	0.004	0.364
Provider						
Church	0.250	(0.433)	0.246	(0.431)	-0.003	0.586
Else	0.571	(0.495)	0.575	(0.494)	0.004	0.556
Public	0.179	(0.384)	0.179	(0.383)	-0.001	0.884
Regional characteristics						
Urban class						
City	0.443	(0.497)	0.444	(0.497)	0.001	0.853
Intermediate	0.374	(0.484)	0.376	(0.484)	0.001	0.866
Rural	0.183	(0.386)	0.180	(0.384)	-0.003	0.652
Share of migrants (in percent)	23.424	(12.094)	23.492	(12.041)	0.068	0.710
Migrant incentive	0.872	(0.334)	0.873	(0.332)	0.002	0.730
Staff-to-child ratio	0.132	(0.024)	0.132	(0.024)	0.000	0.954
Budget per capita	1070.231	(492.779)	1069.892	(519.868)	-0.339	0.966
Resource index (std.)	0.002	(1.005)	-0.002	(0.995)	-0.004	0.808
State						
Baden Wurttemberg	0.149	(0.356)	0.151	(0.358)	0.002	0.675
Bavaria	0.144	(0.351)	0.138	(0.345)	-0.005	0.291
Berlin	0.067	(0.249)	0.064	(0.244)	-0.003	0.399
Brandenburg	0.026	(0.159)	0.025	(0.156)	-0.001	0.645
Bremen	0.007	(0.085)	0.007	(0.083)	-0.000	0.843
Hamburg	0.031	(0.173)	0.030	(0.170)	-0.001	0.761
Hesse	0.062	(0.242)	0.063	(0.242)	0.000	0.959
Mecklenburg-Western Pomeria	0.012	(0.109)	0.012	(0.107)	-0.001	0.713
Lower Saxony	0.081	(0.274)	0.083	(0.276)	0.001	0.730
North Rhine-Westphalia	0.236	(0.424)	0.240	(0.427)	0.004	0.489
Rhineland-Palatine	0.039	(0.193)	0.041	(0.198)	0.002	0.443
Saarland	0.015	(0.120)	0.015	(0.122)	0.000	0.835
Saxony	0.057	(0.232)	0.058	(0.234)	0.001	0.733
Saxony-Anhalt	0.012	(0.109)	0.012	(0.111)	0.000	0.866
Schleswig-Holstein	0.037	(0.189)	0.036	(0.185)	-0.001	0.629
Thuringa	0.025	(0.157)	0.026	(0.158)	0.000	0.887
Sent (N = 18,663)	9,313		9,350			

Notes: Table shows means and standard deviations of variables by treatment group. The analysis sample excludes "bounced" emails. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Diff* is the difference in the mean of the respective variable between the baseline group (native treatment and no higher education signal) and each of the other three experimental groups. We report p-values for two-sided t-tests of the null hypothesis that differences are equal to zero. For detailed variable descriptions, see Appendix D. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B2: Balancing Migrant Treatment (Sent Emails)

	(1) Native treatment		(2) Migrant treatment		(2)-(1)	(2)-(1)
	Mean	SD	Mean	SD	Diff (mean)	p-Value
Email characteristics						
Higher education signal	0.499	(0.500)	0.500	(0.500)	0.001	0.926
Sender male	0.500	(0.500)	0.500	(0.500)	-0.000	0.947
Child male	0.504	(0.500)	0.505	(0.500)	0.001	0.894
Child care center characteristics						
Center's maximum capacity	68.450	(42.325)	68.326	(44.045)	-0.125	0.829
Kindergarten (age 3 – 6 years)	0.934	(0.248)	0.930	(0.255)	-0.004	0.223
Afternoon care (age >6 years)	0.094	(0.293)	0.097	(0.296)	0.002	0.573
Provider						
Church	0.240	(0.427)	0.239	(0.427)	-0.001	0.870
Else	0.590	(0.492)	0.591	(0.492)	0.001	0.916
Public	0.170	(0.375)	0.170	(0.376)	0.000	0.962
Regional Characteristics						
Urban class						
City	0.426	(0.495)	0.426	(0.495)	0.000	0.988
Intermediate	0.382	(0.486)	0.382	(0.486)	-0.000	0.981
Rural	0.192	(0.394)	0.192	(0.394)	0.000	0.992
Share of migrants (in percent)	23.350	(12.046)	23.266	(11.967)	-0.084	0.613
Migrant incentive	0.878	(0.327)	0.878	(0.327)	0.000	0.996
Staff-to-child ratio	0.132	(0.024)	0.132	(0.024)	-0.000	0.709
Budget per capita	1065.651	(491.816)	1066.105	(504.454)	0.455	0.949
Resource index (std.)	-0.007	(0.984)	-0.010	(0.980)	-0.004	0.777
State						
Baden Wurttemberg	0.162	(0.369)	0.162	(0.369)	0.000	0.976
Bavaria	0.153	(0.360)	0.153	(0.360)	-0.000	0.995
Berlin	0.060	(0.238)	0.060	(0.238)	-0.000	0.997
Brandenburg	0.026	(0.158)	0.026	(0.158)	-0.000	0.998
Bremen	0.007	(0.084)	0.007	(0.083)	-0.000	0.935
Hamburg	0.028	(0.164)	0.028	(0.164)	0.000	0.969
Hesse	0.062	(0.241)	0.062	(0.241)	-0.000	0.975
Mecklenburg-Western Pomeria	0.012	(0.108)	0.012	(0.108)	-0.000	0.950
Lower Saxony	0.081	(0.273)	0.082	(0.274)	0.000	0.945
North Rhine-Westphalia	0.230	(0.421)	0.231	(0.421)	0.000	0.968
Rhineland-Palatine	0.041	(0.199)	0.041	(0.199)	0.000	0.976
Saarland	0.014	(0.116)	0.014	(0.117)	0.000	0.956
Saxony	0.053	(0.224)	0.053	(0.224)	-0.000	0.926
Saxony-Anhalt	0.012	(0.108)	0.012	(0.107)	-0.000	0.949
Schleswig-Holstein	0.035	(0.183)	0.034	(0.182)	-0.000	0.911
Thuringa	0.024	(0.154)	0.024	(0.155)	0.000	0.967
Bounces	0.171	(0.376)	0.167	(0.373)	-0.003	0.529
Sent (N = 22,458)	11,228		11,230			

Notes: Table shows means and standard deviations of variables by treatment group. The sent sample includes "bounced" emails. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. For variable definitions, see Appendix D. *Diff* is the difference in the mean of the respective variable between the baseline group (native treatment and no higher education signal) and each of the other three experimental groups. We report p-values for two-sided t-tests of the null hypothesis that differences are equal to zero. For detailed variable descriptions, see Appendix D. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B3: Balancing Higher Education Signal (Analysis Sample)

	(1) No higher education signal		(2) Higher education signal		(2)-(1)	(2)-(1)
	Mean	SD	Mean	SD	Diff (mean)	p-Value
Email characteristics						
Migrant treatment	0.499	(0.500)	0.503	(0.500)	0.004	0.563
Sender male	0.500	(0.500)	0.500	(0.500)	-0.000	0.971
Child male	0.503	(0.500)	0.506	(0.500)	0.003	0.697
Child care center characteristics						
Center's maximum capacity	68.496	(45.126)	69.072	(43.582)	0.576	0.376
Kindergarten (age 3 – 6 years)	0.931	(0.254)	0.931	(0.253)	0.000	0.967
Afternoon care (age >6 years)	0.095	(0.294)	0.096	(0.295)	0.000	0.917
Provider						
Church	0.249	(0.433)	0.247	(0.431)	-0.002	0.693
Else	0.571	(0.495)	0.575	(0.494)	0.003	0.634
Public	0.179	(0.384)	0.178	(0.383)	-0.001	0.865
Regional characteristics						
Urban class						
City	0.444	(0.497)	0.444	(0.497)	-0.000	0.967
Intermediate	0.374	(0.484)	0.376	(0.484)	0.001	0.837
Rural	0.182	(0.386)	0.181	(0.385)	-0.001	0.837
Share of migrants (in percent)	23.539	(12.095)	23.377	(12.039)	-0.162	0.350
Migrant incentive	0.875	(0.331)	0.870	(0.336)	-0.004	0.357
Staff-to-child ratio	0.133	(0.024)	0.132	(0.024)	-0.000	0.493
Budget per capita	1073.746	(501.556)	1066.375	(511.506)	-7.371	0.353
Resource index (std.)	0.008	(0995)	-0.008	(1.005)	-0.017	0.256
State						
Baden Wurttemberg	0.152	(0.359)	0.149	(0.356)	-0.003	0.559
Bavaria	0.140	(0.347)	0.142	(0.349)	0.002	0.670
Berlin	0.066	(0.248)	0.064	(0.245)	-0.002	0.625
Brandenburg	0.024	(0.153)	0.027	(0.162)	0.003	0.201
Bremen	0.007	(0.082)	0.007	(0.085)	0.000	0.716
Hamburg	0.030	(0.171)	0.031	(0.172)	0.001	0.841
Hesse	0.062	(0.242)	0.063	(0.243)	0.000	0.917
Mecklenburg-Western Pomeria	0.012	(0.108)	0.012	(0.108)	-0.000	0.961
Lower Saxony	0.082	(0.275)	0.082	(0.275)	-0.000	0.955
North Rhine-Westphalia	0.239	(0.426)	0.237	(0.425)	-0.002	0.763
Rhineland-Palatine	0.041	(0.197)	0.039	(0.194)	-0.001	0.625
Saarland	0.015	(0.120)	0.015	(0.122)	0.001	0.702
Saxony	0.056	(0.230)	0.059	(0.236)	0.003	0.408
Saxony-Anhalt	0.013	(0.112)	0.012	(0.108)	-0.001	0.562
Schleswig-Holstein	0.037	(0.189)	0.036	(0.185)	-0.002	0.579
Thuringa	0.025	(0.156)	0.026	(0.159)	0.001	0.690
Sent (N = 18,663)	9,343		9,320			

Notes: Table shows means and standard deviations of variables by treatment group. The analysis sample excludes "bounced" emails. *Higher education signal* is an indicator variable taking a value of one if the email includes a signature that indicates a higher educational background of the sender, and zero if the email does not include a signature. *Diff* is the difference in the mean of the respective variable between the baseline group (native treatment and no higher education signal) and each of the other three experimental groups. We report p-values for two-sided t-tests of the null hypothesis that differences are equal to zero. For detailed variable descriptions, see Appendix D. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table B4: Balancing Sender Male (Analysis Sample)

	(1) Sender female		(2) Sender male		(2)-(1)	(2)-(1)
	Mean	SD	Mean	SD	Diff (mean)	p-Value
Email characteristics						
Migrant treatment	0.502	(0.500)	0.500	(0.500)	-0.002	0.764
Higher education signal	0.500	(0.500)	0.499	(0.500)	-0.000	0.971
Child male	0.501	(0.500)	0.509	(0.500)	0.008	0.263
Child care center characteristics						
Center's maximum capacity	68.739	(47.291)	68.828	(41.225)	0.089	0.891
Kindergarten (age 3 – 6 years)	0.929	(0.256)	0.933	(0.250)	0.004	0.343
Afternoon care (age >6 years)	0.093	(0.291)	0.098	(0.297)	0.005	0.270
Provider						
Church	0.247	(0.431)	0.249	(0.432)	0.002	0.764
Else	0.573	(0.495)	0.573	(0.495)	0.000	0.968
Public	0.180	(0.384)	0.178	(0.382)	-0.002	0.696
Regional characteristics						
Urban class						
City	0.444	(0.497)	0.443	(0.497)	-0.001	0.840
Intermediate	0.374	(0.484)	0.375	(0.484)	0.001	0.878
Rural	0.181	(0.385)	0.182	(0.386)	0.000	0.946
Share of migrants (in percent)	23.372	(12.036)	23.544	(12.098)	0.173	0.346
Migrant incentive	0.872	(0.335)	0.874	(0.332)	0.002	0.651
Staff-to-child ratio	0.132	(0.024)	0.132	(0.024)	0.000	0.814
Budget per capita	1065.077	(496.546)	1075.069	(516.392)	9.992	0.208
Resource index (std.)	-0.007	(0.994)	0.004	(1.006)	0.015	0.319
State						
Baden Wurttemberg	0.155	(0.362)	0.145	(0.352)	-0.010*	0.055
Bavaria	0.138	(0.345)	0.144	(0.351)	0.006	0.279
Berlin	0.062	(0.241)	0.068	(0.252)	0.006*	0.074
Brandenburg	0.025	(0.155)	0.026	(0.161)	0.002	0.428
Bremen	0.008	(0.086)	0.007	(0.081)	-0.001	0.486
Hamburg	0.031	(0.173)	0.030	(0.170)	-0.001	0.736
Hesse	0.060	(0.238)	0.065	(0.246)	0.005	0.202
Mecklenburg-Western Pomeria	0.012	(0.111)	0.011	(0.105)	-0.001	0.458
Lower Saxony	0.084	(0.278)	0.080	(0.272)	-0.004	0.314
North Rhine-Westphalia	0.237	(0.425)	0.239	(0.426)	0.002	0.787
Rhineland-Palatine	0.038	(0.191)	0.042	(0.200)	0.004	0.202
Saarland	0.015	(0.123)	0.014	(0.119)	-0.001	0.631
Saxony	0.059	(0.235)	0.056	(0.231)	-0.002	0.513
Saxony-Anhalt	0.011	(0.106)	0.013	(0.114)	0.002	0.257
Schleswig-Holstein	0.038	(0.191)	0.035	(0.183)	-0.003	0.227
Thuringa	0.026	(0.161)	0.024	(0.154)	-0.002	0.354
Sent (N = 18,663)	9,333		9,330			

Notes: Table shows means and standard deviations of variables by treatment group. The analysis sample excludes "bounced" emails. *Sender male* is an indicator variable taking a value of one if the email indicates that the sender of the email is male/the father, and zero if the email indicates that the sender of the email is female/the mother. *Diff* is the difference in the mean of the respective variable between the baseline group (native treatment and no higher education signal) and each of the other three experimental groups. We report p-values for two-sided t-tests of the null hypothesis that differences are equal to zero. For detailed variable descriptions, see Appendix D. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01.

Table B5: Balancing Child Male (Analysis Sample)

	(1) Child female		(2) Child male		(2)-(1)	(2)-(1)
	Mean	SD	Mean	SD	Diff (mean)	p-Value
Email characteristics						
Migrant treatment	0.499	(0.500)	0.503	(0.500)	0.005	0.516
Higher education signal	0.498	(0.500)	0.501	(0.500)	0.003	0.697
Sender male	0.496	(0.500)	0.504	(0.500)	0.008	0.263
Child care center characteristics						
Center's maximum capacity	68.286	(45.163)	69.271	(43.557)	0.985	0.130
Kindergarten (age 3 – 6 years)	0.932	(0.252)	0.930	(0.255)	-0.002	0.630
Afternoon care (age >6 years)	0.097	(0.296)	0.094	(0.292)	-0.002	0.565
Provider						
Church	0.250	(0.433)	0.246	(0.431)	-0.003	0.595
Else	0.570	(0.495)	0.576	(0.494)	0.005	0.456
Public	0.180	(0.384)	0.178	(0.382)	-0.002	0.716
Regional characteristics						
Urban class						
City	0.447	(0.497)	0.441	(0.497)	-0.006	0.423
Intermediate	0.377	(0.485)	0.373	(0.484)	-0.003	0.627
Rural	0.177	(0.381)	0.186	(0.389)	0.009	0.100
Share of migrants (in percent)	23.575	(12.012)	23.344	(12.120)	-0.232	0.182
Migrant incentive	0.873	(0.333)	0.872	(0.334)	-0.001	0.857
Staff-to-child ratio	0.133	(0.024)	0.132	(0.024)	-0.000	0.408
Budget per capita	1073.261	(518.705)	1066.920	(494.348)	-6.342	0.424
Resource index (std.)	0.006	(0.992)	-0.006	(1.008)	-0.013	0.387
State						
Baden Wurttemberg	0.150	(0.357)	0.150	(0.357)	0.000	0.949
Bavaria	0.138	(0.345)	0.144	(0.351)	0.006	0.276
Berlin	0.065	(0.246)	0.066	(0.248)	0.001	0.805
Brandenburg	0.024	(0.153)	0.027	(0.162)	0.003	0.187
Bremen	0.008	(0.087)	0.006	(0.080)	-0.001	0.325
Hamburg	0.032	(0.176)	0.029	(0.167)	-0.003	0.209
Hesse	0.061	(0.239)	0.064	(0.246)	0.004	0.293
Mecklenburg-Western Pomeria	0.012	(0.110)	0.011	(0.106)	-0.001	0.537
Lower Saxony	0.085	(0.279)	0.079	(0.270)	-0.006	0.160
North Rhine-Westphalia	0.243	(0.429)	0.232	(0.422)	-0.011*	0.075
Rhineland-Palatine	0.039	(0.194)	0.041	(0.197)	0.002	0.578
Saarland	0.015	(0.123)	0.015	(0.120)	-0.001	0.687
Saxony	0.055	(0.228)	0.060	(0.238)	0.005	0.142
Saxony-Anhalt	0.011	(0.106)	0.013	(0.114)	0.002	0.264
Schleswig-Holstein	0.037	(0.189)	0.036	(0.185)	-0.001	0.653
Thuringa	0.024	(0.154)	0.026	(0.160)	0.002	0.365
Sent (N = 18,663)	9,242		9,421			

Notes: Table shows means and standard deviations of variables by treatment group. The analysis sample excludes "bounced" emails. *Child male* is an indicator variable taking a value of one if the email indicates that the child of the sender is a boy, and zero if the email indicates that the child is a girl. *Diff* is the difference in the mean of the respective variable between the baseline group (native treatment and no higher education signal) and each of the other three experimental groups. We report p-values for two-sided t-tests of the null hypothesis that differences are equal to zero. For detailed variable descriptions, see Appendix D. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01.

Appendix C. Robustness Checks

Table C1: Treatment Effect on Response Rate

	(1)	(2)
Migrant treatment	-0.044***	-0.044***
	(0.007)	(0.007)
Controls	No	Yes
Control Mean (Native Sender)	0.707	0.707
Scaled Treatment Effect	-6.3	-6.2
N	18,663	18,663

Notes: Table shows treatment effects on an indicator for whether or not a child care center responds to the email, based on multivariate OLS regressions. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). *Scaled treatment effect* expresses the treatment effect relative to the mean of the respective outcome in the control group of native senders in percent. Robust standard errors in parentheses. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01. We additionally report p-values based on randomization inference and correcting for multiple hypothesis testing in Table C2.

Table C2: Randomization Inference and Corrections for Multiple Hypothesis Testing

	Coefficient (1)	p-value Main Table (2)	Rand. Inference (3)	List-Shakih-Xu (4)	Westphal-Young (5)	Romano-Wolf (6)
Panel A: Content Outcomes - Unconditional (Table 2, Panel A)						
Slot Offer	-0.011***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Waiting list	-0.043***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Long response	-0.066***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Helpful content	-0.036***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Encouraging	-0.028***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Recommendation	-0.064***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Panel B: Content Outcomes - Conditional (Table 2, Panel B)						
Slot offer	-0.010**	**0.020**	**0.013**	**0.020**	**0.018**	**0.022**
Waiting list	-0.013*	**0.071**	**0.063**	**0.072**	**0.071**	**0.066**
Long response	-0.059***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Helpful content	-0.024**	**0.008**	**0.006**	**0.009**	**0.008**	**0.006**
Encouraging	-0.026***	**0.000**	**0.001**	**0.001**	**0.000**	**0.001**
Recommendation	-0.059***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
Panel C: Response Rate (Table 3)						
Migrant treatment	-0.044***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
× Higher education	0.014	0.309	0.297	0.314	0.309	0.320
× Sender male	-0.018	0.209	0.195	0.201	0.209	0.201
× Child male	-0.027**	**0.049**	**0.066**	**0.054**	**0.049**	**0.035**
Panel D: Slot Offer (Table 3)						
Migrant treatment	-0.011***	**0.000**	**0.000**	**0.001**	**0.001**	**0.001**
× Higher education	-0.003	0.606	0.626	0.587	0.606	0.608
× Sender male	-0.015**	**0.015**	**0.010**	**0.013**	**0.015**	**0.019**
× Child male	-0.013**	**0.021**	**0.030**	**0.037**	**0.022**	**0.018**
Panel E: Response Rate (Table 4)						
Migrant treatment						
× Migrant incentive	0.088***	**0.000**	**0.000**	**0.001**	**0.000**	**0.001**
× Staff-to-child ratio (std.)	0.018***	**0.009**	**0.004**	**0.007**	**0.004**	**0.019**
× Budget per capita (std.)	0.015	0.156	0.122	0.144	0.130	0.151
Panel F: Slot Offer (Table 4)						
Migrant treatment						
× Migrant incentive	0.020	0.115	0.134	0.139	0.136	0.141
× Staff-to-child ratio (std.)	0.008**	**0.025**	**0.019**	**0.017**	**0.022**	**0.033**
× Budget per capita (std.)	0.014**	**0.029**	**0.016**	**0.019**	**0.012**	0.120

Notes: Table shows p-values for our main results when using randomization inference and adjusting for multiple hypothesis testing. All p-values $< .10$ are printed in bold. For comparison, Column (1) displays coefficients and significance stars representing p-values from robust standard errors (* $p < .10$, ** $p < .05$, *** $p < .01$) as reported in the main tables. Column (2) shows the p-values as reported in the main tables. Randomization inference (RI) p-values in Column (3) are obtained from RI with 1,000 permutations, assigning the treatment status randomly within strata (using the Stata command 'ritest' by Heß (2017)). In Columns (4)—(6), we implement three different methods to correct for multiple hypothesis testing (controlling the family-wise error rates) using bootstrap resampling techniques. Column (4) uses the method by List et al. (2019), Column (5) the stepdown-approach by Westfall and Young (1993), and Column (6) the approach by Romano and Wolf (2005, 2016). The procedures by Westfall-Young (using the Stata command 'wyoung' by Julian Reif) and Romano-Wolf (using the Stata command 'rwolf' by Clarke et al. (2020)) account for the stratified randomization, that is, bootstrap samples are selected within each stratum. In Panels A and B, we correct for the multiple email content dimensions unconditional (conditional) on response. In Panels C and D, we correct for the multiple testing in the treatment effect heterogeneity by migrant treatment and the other three randomized treatments (i.e. higher education signal, the gender of the sender, and the gender of the child). In Panels E and F, we correct for the multiple testing in the treatment effect heterogeneity by migrant treatment and the resource correlation (i.e., migrant incentive, staff-to-child ratio, and budget per capita). Note that some corrected p-values are smaller than the original p-values because they are based on bootstrap methods. All control variables from the respective baseline specification are included.

Table C3: Treatment Effects on Response Rate and Response Content (Probit Regression)

	(1) Response Rate	(2) Slot Offer	(3) Waiting List	(4) Long Response	(5) Helpful Content	(6) Encouraging	(7) Recommendation
Migrant treatment	-0.127***	-0.135***	-0.112***	-0.173***	-0.104***	-0.127***	-0.174***
	(0.020)	(0.037)	(0.019)	(0.019)	(0.020)	(0.024)	(0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marginal effect	-0.044***	-0.011***	-0.043***	-0.066***	-0.036***	-0.027***	-0.064***
	(0.007)	(0.003)	(0.007)	(0.007)	(0.007)	(0.005)	(0.007)
N	18,617	17,678	18,652	18,654	18,633	18,623	18,634

Notes: Table shows treatment effects based on multivariate Probit regressions. Outcome variables are defined as follows: Column (1): indicator for whether or not a child care center responds to the email; Column (2): indicator for whether the contacted child care center offers a child care slot before the next turn cycle (August 2022); Column (3): indicator for whether the contacted child care center offers a place on the waiting list; Column (4): indicator for whether the length of the email response, measured as the number of characters in the email body, is above median; Columns (5) and (6): indicators for whether a child care center responds with a "helpful content" or in an "encouraging" manner; Column (7): indicator for whether the reviewers would recommend the child care center to a befriended couple with a young child. Outcome variables receive a value of zero for non-responses (e.g., a non-response is coded as no offer in Column (2)). See Section 4.2 and Appendix G for a description of the email rating procedure. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Marginal effects are reported in the bottom of the table. Robust standard errors in parentheses. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01.

Table C4: Treatment Effects on Response Rate and Response Content (Zip-Code Fixed Effects)

	(1) Response Rate	(2) Slot Offer	(3) Waiting List	(4) Long Response	(5) Helpful Content	(6) Encouraging	(7) Recommendation
Migrant treatment	-0.047***	-0.011***	-0.052***	-0.072***	-0.040***	-0.026***	-0.067***
	(0.008)	(0.003)	(0.009)	(0.009)	(0.008)	(0.006)	(0.008)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean (Native Sender)	0.707	0.049	0.566	0.467	0.346	0.157	0.419
Scaled Treatment Effect	-6.7	-23.4	-9.1	-15.3	-11.6	-16.6	-15.9
N	16,917	16,917	16,917	16,917	16,917	16,917	16,917

Notes: Table shows treatment effects on response rate and email content measures, based on multivariate OLS regressions. Outcome variables are defined as follows: Column (1): indicator for whether or not a child care center responds to the email; Column (2): indicator for whether the contacted child care center offers a child care slot before the next turn cycle (August 2022); Column (3): indicator for whether the contacted child care center offers a place on the waiting list; Column (4): indicator for whether the length of the email response, measured as the number of characters in the email body, is above median; Columns (5) and (6): indicators for whether a child care center responds with a "helpful content" or in an "encouraging" manner; Column (7): indicator for whether the reviewers would recommend the child care center to a befriended couple with a young child. All outcome variables receive a value of zero for non-responses (e.g., a non-response is coded as no offer in Column (1)). *Scaled treatment effect* expresses the treatment effect relative to the mean of the respective outcome in the control group of native senders in percent. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Controls also include zip-code fixed effects. Regressions leave out 1,746 observations without treatment variation within a zip-code area. Estimating our preregistered model on this smaller sample yields very similar results as in the full sample in Tables C1 and 2. Robust standard errors in parentheses. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C5: Treatment Effects for Different Definitions of Content Outcomes

	1 Reviewer		2 Reviewers		3 Reviewers		4 Reviewers		5 Reviewers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Slot Offer										
Migrant treatment	-0.023***	-0.023***	-0.014***	-0.012***	-0.011***	-0.010**	-0.008***	-0.007*	-0.006***	-0.005
	(0.004)	(0.006)	(0.003)	(0.005)	(0.003)	(0.004)	(0.003)	(0.004)	(0.002)	(0.003)
Control mean (Native sender)	0.103	0.103	0.060	0.060	0.049	0.049	0.041	0.041	0.028	0.028
Scaled treatment effect	-22.8	-22.1	-23.1	-21.0	-22.8	-20.2	-20.6	-16.3	-22.2	-18.0
Waiting List										
Migrant treatment	-0.054***	-0.022***	-0.051***	-0.021***	-0.043***	-0.013*	-0.039***	-0.010	-0.043***	-0.019**
	(0.007)	(0.005)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)
Control mean (Native sender)	0.650	0.650	0.608	0.608	0.566	0.566	0.537	0.537	0.494	0.494
Scaled treatment effect	-8.3	-3.4	-8.3	-3.4	-7.6	-2.3	-7.3	-1.9	-8.6	-3.8
Helpful Content										
Migrant treatment	-0.052***	-0.034***	-0.049***	-0.036***	-0.036***	-0.024***	-0.025***	-0.013	-0.012**	-0.003
	(0.007)	(0.009)	(0.007)	(0.009)	(0.007)	(0.009)	(0.007)	(0.009)	(0.006)	(0.008)
Control mean (Native sender)	0.483	0.483	0.410	0.410	0.346	0.346	0.276	0.276	0.183	0.183
Scaled treatment effect	-10.7	-7.1	-11.9	-8.8	-10.5	-6.9	-9.0	-4.8	-6.7	-1.7
Encouraging										
Migrant treatment	-0.056***	-0.042***	-0.042***	-0.033***	-0.028***	-0.026***	-0.017***	-0.018***	-0.009***	-0.008**
	(0.007)	(0.009)	(0.007)	(0.009)	(0.005)	(0.007)	(0.003)	(0.005)	(0.003)	(0.004)
Control mean (Native sender)	0.461	0.461	0.312	0.312	0.157	0.157	0.061	0.061	0.038	0.038
Scaled treatment effect	-12.1	-9.1	-13.5	-10.6	-17.9	-16.4	-28.5	-30.0	-24.5	-21.9
Recommendation										
Migrant treatment	-0.061***	-0.034***	-0.066***	-0.052***	-0.064***	-0.059***	-0.055***	-0.057***	-0.030***	-0.034***
	(0.007)	(0.006)	(0.007)	(0.008)	(0.007)	(0.009)	(0.006)	(0.009)	(0.005)	(0.007)
Control mean (Native sender)	0.636	0.636	0.521	0.521	0.419	0.419	0.276	0.276	0.124	0.124
Scaled treatment effect	-9.6	-5.3	-12.7	-10.0	-15.3	-14.0	-19.9	-20.7	-24.4	-27.3
Conditional on response	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	18,663	12,547	18,663	12,547	18,663	12,547	18,663	12,547	18,663	12,547

Notes: Table shows treatment effects on email content measures, based on multivariate OLS regressions. Columns (1) and (2) show migrant treatment effects on the respective content outcome when only one of the five reviewers rated the email as an offer, waiting list offer, including helpful content, encouraging, or recommendable. Analogously, Columns (3)-(10) report results when two, three, four, or all five reviewers considered the email as an offer, waiting list offer, helpful content, encouraging, or recommendable. Uneven columns show effects on unconditional on response and even columns show effects on conditional on response. *Scaled treatment effect* expresses the treatment effect relative to the mean of the respective outcome in the control group of native senders in percent. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Robust standard errors in parentheses. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Table C6: Treatment Effects on Standardized Response Content

	Slot Offer		Waiting List		Helpful Content		Encouraging		Recommendation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Migrant treatment	-0.055***	-0.040**	-0.086***	-0.033*	-0.077***	-0.048***	-0.081***	-0.063***	-0.132***	-0.118***
	(0.015)	(0.017)	(0.015)	(0.018)	(0.015)	(0.018)	(0.015)	(0.018)	(0.015)	(0.018)
Conditional on response	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	18,663	12,547	18,663	12,547	18,663	12,547	18,663	12,547	18,663	12,547

Notes: Table shows treatment effects on email content measures, based on multivariate OLS regressions. Email content outcomes are measured as the mean of the standardized ratings of all five reviewers, standardized with mean = 0 and standard deviation = 1. Uneven columns show effects on unconditional on response and even columns show effects on conditional on response. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Robust standard errors in parentheses. Significance levels: * $p < .10$, ** $p < .05$, *** $p < .01$.

Appendix D. Data Section

Table D1: Variable Definitions

Variable Name	Data Source	Variable Definition	Missing Values
Outcome Variables			
Response Rate	Experimental	Indicator variable that is equal to one if the child care center responded to sent email inquiry, zero otherwise	There are no missing values for this variable in our data set.
Slot Offer	Rating	Indicator variable that is equal to one if the email contains an offer for a child care slot before August 2022, zero otherwise.	There are no missing values for this variable in our data set.
Waiting List	Rating	Indicator variable that is equal to one if the email contains an offer to be placed on a waiting list, zero otherwise.	There are no missing values for this variable in our data set.
Long Answer	Rating	Indicator variable that is equal to one if the email has above median string length, zero otherwise.	There are no missing values for this variable in our data set.
Helpful Content	Rating	Indicator variable that is equal to one if it contains helpful information such as telephone numbers, zero otherwise.	There are no missing values for this variable in our data set.
Encouraging	Rating	Indicator variable that is equal to one if the email is perceived as encouraging, zero otherwise.	There are no missing values for this variable in our data set.
Recommendation	Rating	Indicator variable that is equal to one if the independent raters would recommend contacting the child care center to a befriended couple with a child aged 18 months, zero otherwise.	There are no missing values for this variable in our data set.
Randomized Treatments			
Migrant Treatment	Experimental	Indicator variable that is equal to one if the email sender's name signals a migration background and zero if the email sender's name signals a native background.	There are no missing values for this variable in our data set.
Higher Education Signal	Experimental	Indicator variable that is equal to one if the email includes a signature that indicates a higher education background of the sender and zero if the email does not include a signature.	There are no missing values for this variable in our data set.
Sender Male	Experimental	Indicator variable that is equal to one if the email includes a male name (Andreas, Sebastian, Hüseyin, and Ömer) as the sender and zero if the email includes a female name (Christina, Stefanie, Eylül, and Fatma).	There are no missing values for this variable in our data set.
Child male	Experimental	Indicator variable that is equal to one if the child mentioned in the email is a son (male child) and is zero if the email mentions a daughter (female child)	There are no missing values for this variable in our data set.
Oemer Yildirim Indicator	Experimental	Indicator variable that is equal to one if the emails were sent by the sending account with the spelling mistake in the displayed name, and zero otherwise.	There are no missing values for this variable in our data set.

(continued on next page)

Table D1: Continued

Variable Name	Data Source	Variable Definition	Missing Values
Stratification Variables			
Provider Type	Commercial data set	Categorical variable with three categories indicating the type of provider of the child care center. We categorize ecclesiastic providers such as the Catholic and Protestant church as well as non-profit providers that are close to the church (e.g., "Diakonie") as "church". All types of public providers are classified as "public", while all other providers (e.g, for-profit providers, clubs, parental initiatives, and missing values) are assigned to the "other" category.	Information on the provider type is missing for 5,494 (29.4%) observations.
Urban Class	Commercial data set	Categorical variable with three categories indicating the urban classification of the county, where the child care center is located. Data on the urban type of a county (referring to the year 2021) are taken from Eurostat (see Eurostat). The measure uses inhabitants per square kilometer to classify three degrees of urbanity of a county (i.e., NUTS3 region), namely "city", "intermediate", and "rural". In the first step, the approach defines "urban clusters" as contiguous grid cells of one square kilometer with a density of at least 300 inhabitants per square kilometer and a joint minimum population of 5,000. All other counties are classified as " rural areas". In a second step, a county is classified as "rural" if the share of the population living in rural areas is greater than 50%, as "intermediate" if the share of the population living in rural areas is between 20% and 50%, and "city" if the share of the population living in rural areas is less than 20%. Finally, the classification takes into account the presence of cities. A rural county containing a city of more than 200,000 inhabitants that make up at least 25% of the county's population is classified as "intermediate", and an "intermediate" county containing a city of more than 500,000 inhabitants that make up at least 25% of the county's population is classified as "city"(see also Dijkstra et al., 2019). The federal state is taken from the commercial data set.	There are no missing values for this variable in our data set.
State	Commercial data set	Categorical variable with 16 categories indicating the federal state in which the child care center is located.	There are no missing values for this variable in our data set.
Strata Fixed Effects	Experimental	Categorical variable with 131 categories indicating the strata the child care center belongs to. Theoretically, there could be 3*3*16 = 144 strata. In some cases, the number of observations within a stratum was small (< 8). Therefore, we decided to merge a total of 27 observations (out of 22,458) with the neighboring strata in terms of urban classification (i.e., we either merged "rural" with "intermediate" or "intermediate" with "city"). In total, we stratified on 131 strata in our randomization.	There are no missing values for this variable in our data set.

(continued on next page)

Table D1: Continued

Variable Name	Data Source	Variable Definition	Missing Values
Control Variables			
Center's Maximum Capacity	Commercial data set	Continues variable of the maximum number of children a child care center is allowed to enroll. We z-standardize the variable to have a mean of zero and a standard deviation of one.	Information on the variable is missing for 1,197 observations.
Afternoon Care (age >6 years)	Commercial data set	Indicator variable that is equal to one if the child care center also offers afternoon care for schoolchildren, zero otherwise.	There are no missing values for this variable in our data set.
Kindergarten (age 3 - 6 years)	Commercial data set	Indicator variable that is equal to one if the child care center also caters to children aged three to six years, zero otherwise.	There are no missing values for this variable in our data set.
Share of Migrants	RWI-GEO-GRID	Continuous variable of the share of persons with a migration background from the total population of the municipality in which the child care center is located. There are 3,707 municipalities in our sample. We z-standardize the variable to have a mean of zero and a standard deviation of one.	Information on the migrant share is missing for 1,299 (7.0%) of observations.
Variables for Heterogeneity Analysis			
Finances			
Migrant Incentive	State legislation	Indicator variable that is equal to one if the relevant laws of the federal state, in which the child care center is located, do specify any subsidies or other advantages for enrolling children with migration background in child care centers, zero otherwise. There are 9 out of 16 federal states in our sample that provide such an incentive.	We have no missing values for this variable.
Staff-to-child Ratio	INKAR	Continuous variable of the ratio of pedagogical staff to the number of slots in a child care center, measured at the level of the county of the child care center in 2019. We standardize the variable to have a mean of zero and a standard deviation of one.	Data are missing for 1,238 (6.6%) observations.
Budget per Capita	INKAR	Continuous variable of the available budget (taxing capacity) meaning the sum of all tax income of a municipality in which the child care center is located per capita in 2020. There are 3,707 municipalities in our sample. We z-standardize the variable to have a mean of zero and a standard deviation of one.	Information on the budget per capita is missing for 2,382 (12,76%) observations.
Resource Index	State legislation, INKAR	Continuous variable of a comprised index of the migrant incentive, staff-to-child-ratio and the budget per capita variable using the factor loading's on one factor from a factor analysis.	Information on the resource index is missing for 1,144 (6.1%) observations.

Notes: For RWI-GEO-GRID data see RWI webpage (Budde and Eilers, 2014). INKAR data is retrieved from INKAR webpage (INKAR, 2021). We retrieve voting data from the German Statistical Office (Bundeswahlleiter, 2021). Data on the slot rationing was provided by the KiBS (Kayed et al., 2023). For control variables we impute missing values at the next higher regional level (i.e. county, or NUTS2 level) before we include them into regressions.

Appendix E. Additional Surveys

Appendix E.1. Survey with Child Care Centers

In summer 2019, we conducted a series of interviews with six child care centers to further inform our study design. Based on these interviews, we then conducted a survey with $N = 447$ child care centers in September 2019. To recruit participants for the survey, we sent invitations to a subset of 6,000 randomly chosen early child care centers from the commercial data set and specifically asked the child care center manager to take part in the survey (these 6,000 centers are excluded from our main study). Importantly, in the present study, we did not contact these centers again. The invitation email specifically stated that only center managers have the necessary information to participate in the survey and should participate in the survey.

We distributed the invitation emails in two waves, spaced two weeks apart from each other, to account for differing German vacation schedules. One week after our invitation and four days after our initial reminder, we then sent two waves of additional reminders to take part in our survey. We ask center managers about the institutional background of the center (provider, size, and age), the demographics of the enrolled children (gender, age, and migration background), the criteria for preferential treatment during the enrollment process, the decision-maker for enrollment decisions, and the center's communication with parents (content of the parental requests and communication device).

In the survey, managers state that they receive an average of 14 inquiries from parents per week. Further, 85% of managers state that they regularly receive email requests from parents, most frequently asking about open slots and information on how to apply (the two questions we also ask in our email). Almost four out of ten (38%) of managers indicate that prior email contact is important for parents to ultimately receive a slot. Thus, emails are an essential communication tool between parents and child care centers. In addition, 86% of child care center managers report that the center manager is responsible for enrollment decisions (using the survey question "Who decides on the admission of a child in your facility?"). In contrast, less than 10% of respondents indicate that the rules of the provider or other institutions (such as other employees of the child care center or the city administration) play a role for these decisions.

Appendix E.2. Online Surveys

Online Surveys. To supplement our study, we conducted two online surveys using Clickworker.de, a German counterpart to Amazon MTurk, where individuals can take short surveys or complete other tasks for micro-payments. For both surveys, we only included participants who currently lived in Germany and spoke German as their native language. Additionally, participants had to be above the age of 20 and under the age of 50. Furthermore, we ensured that our samples reflected males and females as well as the age groups 20 to 35 and 36 to 50 years in equal shares.

Online Survey I. We conduced the first of these surveys in April 2019 with $N = 200$ online workers. We use the survey to validate the origin of the names in our study and to check the realism of our emails. We use the first survey to pre-select names for our experiment. To this end, survey participants had to state if they associate a migration background with the most common names for native Germans living in Germany and for Turkish migrants living in Germany. Participants repeat this exercise for a total of 16 surnames (eight German, eight Turkish) and 56 first names (14 for each gender and country pair). We retrieve names from the online portal of the Society for German Language (GfdS) to identify the most common names from the cohort born in 1986. Finally, we choose the four Turkish and German surnames as well as the four first names for each gender country pair that are most clearly associated with a native or migration background to be validated for their exact origin in the second survey.

Online Survey II. We conducted the second online survey in October 2019 with $N = 200$ online workers. In the survey, we asked participants to write the country of origin next to each of the names identified in the first survey. In our field experiment, we then include the names that were most clearly associated with a German or Turkish origin (at least 90% of respondents associated each of the chosen surnames with either a German or Turkish origin). We then showed participants the email for the field experiment, including the higher education signal, and asked participants if they think the email is realistic and if they perceived the information conveyed in the signature. Of the $N = 182$ participants that responded to the question, $N = 145$ (80%) see our email as realistic or very realistic and only $N = 8$ (4%) see our email as not or not at all realistic. Importantly, we cannot detect any differences in perceived realism based on the name of the email sender (signaling native or migration background). Moreover, more than 80% of the respondents correctly recalled that the sender has a higher education degree, based on the higher education signal included in the email signature.

Appendix E.3. Inequality Barometer

For exploring the hypothesis that a higher perceived effort required to educate migrant children could be a driver for discrimination, we were able to include a question into the representative panel survey *"Inequality Barometer"* administered by the University of Konstanz ($N = 4,822$). This panel survey explores the perception of inequality for a broad set of domains in the German population. It was conducted by the survey company Kantar Public in November 2022. The sample is representative of the (adult) German population in terms of gender, age, education, and region of residence (NUTS2). Furthermore, the survey company provides survey weights to ensure the representativeness of our sample for the overall German population.

The question we use to investigate potential reasons for discrimination was part of a larger module investigating public perceptions about migrant-native gaps in early child care (see Hermes et al., 2024). The question reads: "According to a recent scientific study, Turkish parents have lower chances of applying for child care slots than German parents. How would you explain these lower chances for Turkish parents? Assume that the applications of German and Turkish parents are equally good." Respondents could then select multiple of the following reasons (further, respondents could also provide answers in an open text field or select "don't know" or "not specified"):

(i) "Turkish parents are disadvantaged because of their cultural background."

(ii) "Child care centers assume that Turkish parents come with a greater workload, e.g., because of language barriers."

(iii) "Child care centers make sure that the share of Turkish children in the groups is not too large, accommodating what many other parents want."

Applying survey weights for the representativeness of the sample, 28.3% of of the German population say that Turkish parents are disadvantaged due to their cultural background, 50.3% state that child care centers assume that Turkish parents come with a greater workload, and 44.9% think that child care centers make sure that the share of Turkish children in the groups is not too large, accommodating what many parents want.

Appendix F. Turkish Migrants in Germany

Turkish immigrants are by far the largest and most regionally dispersed ethnic group in Germany. In 2019, there were approximately 1.5 million people of Turkish origin in Germany. This accounts for approximately 1.3 percent of the German population and 13.0 percent of all migrants in Germany. The second largest group are Polish immigrants (0.9% of the German population; about 7.4% of all migrants). Turkish migrants first came to Germany in the 1960s to expand Germany's labor force. The influx of people of Turkish origin was implemented based on an agreement with Turkey to recruit guest workers. In the 1970s, there was a second wave of migration from Turkey to Germany due to family reunification and political instability in Turkey. As a result, the majority (52,6%) of people with a Turkish migration background are second or third generation migrants with no personal migration experience (Bundesamt für Migration, 2019).

Despite the long history of Turkish migrants in Germany, people with Turkish migration backgrounds are less educated and earn less than native Germans. For instance, in the overall German population, about 7% of Turkish persons hold a university degree, while 17% of persons without migration background hold a university degree. Furthermore, the average monthly income per capita of persons in Germany without a migration background is 1,776 EUR, while persons with a Turkish migration background have an average monthly income per capita of 1,237 EUR (Destatis, 2021). Past evidence also shows that persons with Turkish migration background are discriminated against on the German labor market (Kaas and Manger, 2012).

Most important for our setting, children from Turkish migrants are substantially less often enrolled in early child care than children from German parents (11.7% vs. 33.2%). However, Turkish parents state that they want to enroll their child into early child care almost as often as German parents (40.2% vs. 43.9%, see Jessen et al., 2020). Note that Jessen et al. (2020) is the only data source that provides child care enrollment rates and demand information specifically for Turkish migrants in Germany; their data refer to the years 2012–2016.

Appendix G. Rating Procedure

Five reviewers evaluate each response from child care centers on the following five outcome dimensions:

1. Whether the email contains an offer for a slot before August 2022 (*Slot offer*).
2. Whether the email contains an offer to get a place on a waiting list (*Waiting list*).
3. Whether the email contains helpful information such as telephone numbers, links, or references to other institutions (*Helpful Content*).
4. Whether the email is perceived as encouraging (*Encouraging*).
5. Whether one would recommend contacting the child care center to a befriended couple with a child aged 1.5 years (*Recommendation*).

For the rating, we recruited five student assistants ("reviewers") from three different universities in Germany during February 2022. Three of the recruited reviewers were female, two were male. Moreover, two were currently in their Master's studies, while three were studying towards a Bachelor's degree. We trained the reviewers in a four-hour workshop in which we went through the rating criteria and jointly reviewed a set of practice emails. Additionally, we provided a handbook on how to code the different outcome dimensions and also supplied them with a rating tool to help them minimize technical errors while reviewing.

In the rating tool, reviewers answered a short survey for each of the 12,547 responses. Each survey question represented one of the outcome dimensions shown above. The respective email response text was visible at the top of the screen at all points in the survey. The questions had to be answered on a four-point Likert scale (except for the recommendation outcome, which was measured on a ten-point Likert scale) from "Clearly not ..." to "Clearly ...". We provided reviewers with completely anonymous emails, i.e., we deleted names of the center managers and the parents beforehand. Thus, reviewers were unaware of the treatment variation. Furthermore, we did not inform reviewers about the purpose of the study.

To use the reviewer results for our analysis, we create binary outcome measures in two steps. First, for each reviewer, we created binary measures by coding emails rated as "Clearly not ..." or "Somewhat not ..." as not a slot offer/waiting list offer/helpful content/encouraging. Similarly, we coded emails rated as "Somewhat ..." or "Clearly ..." as a slot offer/waiting list offer/helpful content/encouraging. For the recommendation dimension, we create the binary measure by combining answers between scale points one to five to “no recommendation”, and combined answers from scale points six to 10 to “recommendation”.

In the second step, we combine the individual ratings into one variable. We code a dimension as a slot offer/waiting list offer/helpful content/encouraging/recommendation, if the email was rated by three or more reviewers as such, zero otherwise. Results are robust to different specifications of the binary measure (see Table C5) and reviewer-specific use of the scales (see Table C6).

We compute measures of inter-rater reliability for each outcome by comparing the binary variables between reviewers. Depending on the inter-rater criteria, values between 0.61 and 0.8 are interpreted as substantial inter-rater reliability, and values between 0.81 and 1 are seen as almost perfect inter-rater reliability. The most relevant outcome is the slot offer. The percent agreement between the ratings for offers is 0.95 and the most restrictive measures of inter-rater reliability, such as the Cohen’s Kappa or the Krippendorff’s Alpha, are around 0.68. Hence, we interpret the rating for the slot offers as very reliable. Furthermore, the rating for the waiting list is very reliable, with values of 0.89 percent agreement and around 0.67 for Cohen’s Kappa and Krippendorff’s Alpha. The inter-rater reliability for the other outcomes is somewhat lower due to the more subjective nature of the outcomes. Still, with values for the percent agreement between 0.65 and 0.8, the reliability of the rating of the other outcomes is also substantial.

Appendix H. Natural Language Processing Outcome Classification

To classify email responses, we used the supervised Bidirectional Encoder Representations from Transformers (BERT) for sequence classification model proposed by Devlin et al. (2018) and adapted to German by Chan et al. (2020). To pre-train the model, we used 1,000 of our pre-classified emails. We chose observations for the training randomly and balanced on outcomes, such that the algorithm cannot take the likelihood of an outcome in the final sample into account. Also, we train the model on a set of anonymous emails, which do not contain the names of parents, child care centers, and center managers to not induce any biase into the classification. The test data ($N = 100$) of our model is randomly chosen from all other remaining observations. We therefore use a total of 1,100 observations to calibrate the model and classify the remaining 11,447 observations into outcome categories. To decide on the model parameters and to avoid over-fitting, we select the model epoch for which training loss is equal (or slightly higher) to test loss. Otherwise we use default settings for the model parameters (batch size: 100; Epochs: 10; Step p. Epoch: 15; Learning rate: 1e-05). Finally, we do not exclude any type of stop word from the analysis to avoid inducing any form of bias.

As natural language processing algorithms are not yet capable of reliably detecting emotions in texts (such as the level of encouragement), which are often communicated between the lines or require a deep understanding of context. Therefore, we are only able to classify slot offers and waiting list offers computationally, as both are explicitly mentioned in the email texts. Results from the BERT NLP algorithm classification are presented in Table H1. For both categories, our classification accuracy in the test data set has an F1 score of above 0.98.

Table H1: Treatment Effects for Outcomes Classified by Natural Language Processing

	Slot Offer (BERT)		Waiting List (BERT)	
	(1)	(2)	(3)	(4)
Migrant treatment	-0.015***	-0.015***	-0.046***	-0.017**
	(0.004)	(0.005)	(0.008)	(0.007)
Conditional on response	No	Yes	No	Yes
Controls	Yes	Yes	Yes	Yes
Control Mean (Native Sender)	0.066	0.066	0.565	0.565
Scaled Treatment Effect	-22.3	-22.0	-8.1	-3.0
N	17,563	11,447	17,563	11,447

Notes: Table shows treatment effects (estimated by multivariate OLS regressions) on selected email content outcomes classified by a BERT NLP algorithm. Column (1): indicator for whether the contacted child care center offers a child care slot before the next turn cycle (August 2022); Column (2): indicator for whether the contacted child care center offers a place on the waiting list. The sample excludes all data which is used for training and testing the NLP algorithm (1,100 observations). Uneven columns show effects unconditional on response and even columns show effects conditional on response. *Migrant treatment* is an indicator variable taking a value of one if the email sender's name signals a migration background, and zero if the email sender's name signals a native background. *Controls* include strata fixed effects, as well as characteristics of the contacted child care center and the municipality where it is located (see Section 4.1 for details). Robust standard errors in parentheses. Significance levels: * $p <$ .10, ** $p <$ .05, *** $p <$ .01.

Appendix I. Heterogeneity Based on Causal Forest

We employ a causal forest method (Wager and Athey, 2018; Athey and Wager, 2019) to narrow down the potential drivers for heterogeneous treatment effects of our migrant treatment on the response rate outcome. The approach estimates Conditional Average Treatment Effects (CATEs) of an outcome, defined as $E(Y_{1i} - Y_{0i}|X_i = x)$, where Y is the outcome of interest, and X is a vector of observable baseline characteristics. To estimate the CATE, the approach recursively builds decision trees that split the sample into subsets. For deciding how to split a decision tree, the algorithm chooses the variable that most enhances the accuracy of the treatment effect prediction (variables are chosen from a random subset of X, where the size of this subset is a predefined hyperparameter).

The algorithm records how many times it selected each variable to split the sample, i.e., how often the variable increased the prediction accuracy of the heterogeneous treatment effects. This number can be used to calculate the so-called "variable importance" (VI), which indicates the variable's explanatory power for treatment effect heterogeneity. This data-driven approach provides an indication of the extent to which each variable contributes to the estimation of the CATE, revealing how much the inclusion of a particular variable improves the prediction accuracy of the treatment effects in the causal forest model. As explained above, VI highlights the relevance of variables in predicting heterogeneous treatment effects within the context of the model but it does not imply any (significant) heterogeneous treatment effects in OLS estimations because the explanatory power might depend on the sample's conditioning on other variables (previous partitioning in the forest). Furthermore, the data-driven approach does not generate hypotheses explaining *why* a variable is important in estimating the CATE. Yet, the causal forest approach helps to identify variables for which there potentially is a heterogeneous treatment effect and helps narrowing down the set of potential contributors. Exploring treatment effect heterogeneity using OLS regressions often involves testing multiple hypothesis. By using a causal forest, we also aim to circumvent the potential issue of multiple hypothesis testing and focus on the variables with the largest VI for our investigation of treatment effect heterogeneity.

For the implementation of the causal forest, we follow the methodology outlined by Athey and Wager (2019) and utilize the R package *grf* developed by Tibshirani et al. (2018). The variables we use in the approach are the following: Higher Education Signal (0/1), Sender Male (0/1), Child Male (0/1), Provider type (public / ecclesiastical / other), Urban class (rural / suburban / urban), Center's maximum capacity (std.), Afternoon care for age >6 years (0/1), Kindergarten for age 3–6 years (0/1), Share of migrants (std.), and Resource Index (std.).

We therefore implement the causal forest with 9 variables and 17,425 observations, all of which have complete data.[30] Given our large sample size, we specify 10,000 trees to build the causal forest. Otherwise, we use the default settings of the *grf* package that also specifies an honest approach by splitting the sample into a training and test set to avoid overfitting.

In line with our hypothesis from Section 6, the variable with the by far highest VI in the CATE prediction is the resource index (VI = 0.40).[31] Apart from our treatments, only two other variables from the causal forest seem relevant, namely the share of migrants in a child care center's municipality (VI = 0.21) and the capacity of the child care center (VI = 0.16). For all other variables, variable importance is below 0.01. For completeness, we also analyzed treatment effect heterogeneity for these two variables in the same way and the same sample as in Table 4. For the share of migrants, we find no significant heterogeneity in treatment effects for response rates ($b = 0.010$, $p = .148$) but for slot offers ($b = 0.078$, $p = .019$). For the capacity of the child care center, we see that a one standard deviation larger center discriminates somewhat stronger on response rates ($b = -0.021$, $p = .006$), but there is no difference in discrimination for slot offers ($b = 0.003$, $p = .568$). Because of these inconsistent findings and because we lack a clear hypothesis as to why the size of child care centers should be related to the intensity of discrimination, we refrain from further interpreting these results.

[30]To not loose too many observations in this analysis, we impute missing values with the mean of the next higher regional level (i.e., the NUTS3 level). We drop 1,144 observations for which we do not have information on the location.

[31]We check if the ordering of VI in the estimation of the CATE is robust to running two separate causal forests for factor variables and z-standardized continuous variables. We find that the order within variable types stays robust. Similarly, if we use the three components of the resource index instead of the index itself, each component shows up with a high VI (migrant incentive: 0.12, staff-to-child-ratio: 0.10, budget per capita: 0.10).

References

Athey, S. and S. Wager (2019). Estimating Treatment Effects with Causal Forests: An Application. *Observational Studies 5*(2), 37–51.

Budde, R. and L. Eilers (2014). Sozioökonomische Daten auf Rasterebene: Datenbeschreibung der Microm-Rasterdaten. RWI Materialien 77, Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Essen.

Bundesamt für Migration (2019). *Migrationsbericht der Bundesregierung - Migrationsbericht 2019.* Bundesminsterium des Inneren.

Bundeswahlleiter, D. (2021). Ergebnisse der repräsentativen Wahlstatistik. Statistiken zur Bundestagswahl, Der Bundeswahlleiter, Wiesbaden.

Chan, B., S. Schweter, and T. Möller (2020). German's Next Language Model. *arXiv*, preprint:2010.10906.

Clarke, D., J. P. Romano, and M. Wolf (2020). The Romano–Wolf Multiple-hypothesis Correction in Stata. *The Stata Journal 20*(4), 812–843.

Cui, J., L. Natzke, and S. Grady (2021). Early Childhood Program Participation: 2019. First Look. Technical report, National Center for Education Statistics at IES, Washington DC.

Del Boca, D., C. Pronzato, and G. Sorrenti (2016). When Rationing Plays a Role: Selection Criteria in the Italian Early Childcare System. *CESifo Economic Studies 62*(4), 752–775.

Destatis (2021). *Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund.* German Federal Statistical Office, Wiesbaden.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv*, preprint:1810.04805.

Dijkstra, L., H. Poelman, and P. Veneri (2019). The EU-OECD Definition of a Functional Urban Area. Technical Report, EU-OECD.

Eurydice (2019). *Key Data on Early Childhood Education and Care in Europe.* Luxembourg: Publications Office of the European Union.

Expat (2022). How to get a Place in a Crèche for your Baby. `https://expat-in-france.com/place-creche/`. Accessed on March 27, 2023.

Felfe, C. and R. Lalive (2018). Does Early Child Care Affect Children's Development? *Journal of Public Economics 159*, 33–53.

Harvey, S. (2022). Daycare and Preschool in Spain. `https://www.expatica.com/es/living/family/preschool-spain-107654/`. Accessed on March 27, 2023.

Hermes, H., P. Lergetporer, F. Mierisch, G. Schwerdt, and S. Wiederhold (2024). Does Information about Inequality and Discrimination in Early Child Care Affect Policy Preferences? CESifo Working Paper No. 10925, CESifo.

Heß, S. (2017). Randomization Inference with Stata: A Guide and Software. *The Stata Journal 17*(3), 630–651.

INKAR (2021). Indikatoren und Karten zur Raum- und Stadtentwicklung. Statistiken zur räumlichen Zusammensetzung, Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumordnung (BBR), Bonn.

Jessen, J., S. Schmitz, and S. Waights (2020). Understanding Day Care Enrolment Gaps. *Journal of Public Economics 190*, 104252.

Kaas, L. and C. Manger (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review 13*(1), 1–20.

Kayed, T., J. Wieschke, and S. Kuger (2023). Der Betreuungsbedarf bei U3-und U6-Kindern. DJI-Kinderbetreuungsreport 2022. Studie 1 von 6.

List, J. A., A. M. Shaikh, and Y. Xu (2019). Multiple Hypothesis Testing in Experimental Economics. *Experimental Economics 22*(4), 773–793.

Malik, R., K. Hamm, L. Schochet, C. Novoa, and S. Workman (2018). America's Child Care Deserts in 2018. `https://www.americanprogress.org/issues/early-childhood/reports/2018/12/06/461643/americas-child-care-deserts-2018/`. Accessed on March 27, 2023.

NYC Department of Education (2023). Pre-K. `https://www.schools.nyc.gov/enrollment/enroll-grade-by-grade/pre-k`. Accessed on March 27, 2023.

OECD (2018). Settling In 2018. Technical report, OECD, Paris.

OECD (2020). Is Childcare Affordable? Policy Brief on Employment, Labour and Social Affairs. Technical Report, OECD, Paris.

Renfrewshire Council (2023). Apply for a Free Nursery or Childcare Place. Accessed on March 27, 2023.

Romano, J. P. and M. Wolf (2005). Stepwise Multiple Testing as Formalized Data Snooping. *Econometrica 73*(4), 1237–1282.

Romano, J. P. and M. Wolf (2016). Efficient Computation of Adjusted p-Values for Resampling-based Stepdown Multiple Testing. *Statistics & Probability Letters 113*, 38–40.

Tibshirani, J., S. Athey, S. Wager, R. Friedberg, L. Miner, and M. Wright (2018). grf: Generalized Random Forests (Beta). R package version 0.10. 1. *Rproject. org/package= grf*.

Wager, S. and S. Athey (2018). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

Westfall, P. H. and S. S. Young (1993). *Resampling-based Multiple Testing: Examples and Methods for p-Value Adjustment*, Volume 279. John Wiley & Sons.

Halle Institute for Economic Research –
Member of the Leibniz Association

Kleine Maerkerstrasse 8
D-06108 Halle (Saale), Germany

Postal Adress: P.O. Box 11 03 61
D-06017 Halle (Saale), Germany

Tel +49 345 7753 60
Fax +49 345 7753 820

www.iwh-halle.de

ISSN 2194-2188

The IWH is funded by the federal government and the German federal states.