

Studying ethnic discrimination in the Dutch labour market through the use of correspondence testing: In search of the appropriate method of analysis

Iris Andriessen, Eline Nievers, Laila Faulk & Jaco Dagevos
The Netherlands Institute for Social Research

Paper prepared for the IMISCOE conference on Interethnic Relations: Multidisciplinary Approaches 13 - 15 May, 2009, Lisbon, Portugal

WORK IN PROGRESS –PLEASE DO NOT QUOTE

Abstract

In the period may- december 2008 a total number of 1463 correspondence tests were performed to study ethnic discrimination in the Dutch labour market. A research design was used that allowed to distinguish between applicant characteristics (ethnic group, gender, starting vs experienced workers) and vacancy characteristics (sector, occupational level, customer contact). Three methods of analysis are applied for the analysis of correspondence testing throughout the literature. Our presentation will list the advantages and disadvantages of these methods. Furthermore, we will show how results may differ across these methods. The question of which method is considered the appropriate method of analysis in light of the research design is subsequently raised.

1. Correspondence tests as a means for measuring ethnic discrimination

Substantial evidence exists for (enduring) ethnic stratification in Western labour markets (Heath, 2007). Minority disadvantage may be due to group characteristics (resources), such that some groups are better qualified or more motivated to participate. Alternatively, social exclusion (discrimination) may be a cause. The topic has inspired many scholars to conduct research on ethnic minority disadvantage in the labour market. Most often this takes the form of decomposition analyses that calculate the size of net ethnic penalties, by controlling for relevant human capital characteristics. A major disadvantage of such methods is that they are unable to distinguish discrimination from unmeasured variables in the ethnic penalties (*omitted variable bias*). Alternatively, qualitative methods have been applied, either interviewing ethnic minorities about their experiences with discrimination in the labour market or interviewing employers about their preferences and selection methods. The qualitative methods offer an account of people's experiences, but these cannot be equaled with discriminatory practices. One reason is that people's experience may not coincide with 'actual' discrimination (one may feel discriminated against, whereas this was not the case; or one may have been discriminated against without recognizing this as discrimination), another reason is social desirability. Because of social norms in society against discrimination, people may feel reluctant to admit that they prefer hiring a native employee over an ethnic minority employee even though both are equally qualified. Vignette studies have been designed to minimize social desirability. Employers are asked to rank job candidates with different profiles according to their preferences. The profiles differ with regard to relevant labour market characteristics (e.g. work experience or age). By 'hiding' ethnic background in the profiles, it is possible to calculate the effect of ethnic background on employers' preferences. Although vignette studies reduce the risk of social desirability, their disadvantage is that they still measure hypothetical situations or at best intentions, whereas actual behaviour can differ substantially.

Correspondence tests are a type of field experiment, designed to measure employers' actual selection behaviour. Typically, two fictitious job candidates apply for the same (real) job vacancy. They are matched on the relevant labour market characteristics that are usually presented in job resumes (e.g. work experience, age, education etc), but they differ in ethnic background (one native candidate, one ethnic minority). The fictitious job candidates are presented to employers with job vacancies, who are unaware that they participate in a field experiment¹. Not informing employers of their participation is believed to result in employers' treating of the fictitious candidates as real candidates (transforming a would-be situation into a real situation), and is thus necessary to observe actual selection behaviour. A significant larger chance of being invited for a job interview for the native candidate indicates discrimination, since the fictitious job candidates are matched in relevant characteristics. Correspondence tests offer by their experimental design harder proof of ethnic discrimination than other methods. In addition, they may also answer more fine grained questions on the chances of ethnic minorities on the labour market. For example, how much are ethnic minorities' chances for getting an invitation for a job interview reduced for jobs with customer contact (as opposed to jobs without customer contact).

¹ This is not in line with the rule of informed consent. However, the importance of the topic in combination with the expectation that employers may alter their behaviour when aware of the experiment is generally considered a valid argument for not complying with this rule

2. Discrimination in the Netherlands?

Statistical analyses on the labour market position of ethnic minorities in the Netherlands have repeatedly shown continuing minority disadvantage. Compared to native Dutch, non-western ethnic minorities less often have paid work and fixed contracts, are more often unemployed and more often work in lower level occupations (Dagevos, 2007; Veenman en Bijwaard 2006). Also, the second generation is more often unemployed and less often has fixed contracts (Ours & Veenman, 2002; Dagevos 2007). Decomposition analyses show that the gross disadvantage cannot be explained by differences in age, work experience, or educational level (Andriessen & Dagevos, 2007). Even more finegrained analyses, controlling for GPA within educational tracks, results in a significant net disadvantage for non-western ethnic minorities (Andriessen & Dagevos, 2007).

Other sources also point to the occurrence of ethnic discrimination in the Dutch labour market: non-western minorities themselves report having experienced ethnic discrimination while applying for a job (Nievers, 2007). In a study among employers 9% reported that they would not under any circumstance hire an employee from an ethnic minority background, while 18% admitted to only hire an ethnic minority member when no suitable native candidate could be found (Kruisbergen & Veld, 2002).

The results from these various sources indicate that minority disadvantage can – in part – be due to ethnic discrimination in the labour market, but they offer no direct proof. It has been already over ten years that correspondence testing has been carried out in substantial numbers of job vacancies to test for ethnic discrimination in the Dutch labour market (Bovenkerk, 1995). In these tests, Moroccan or Surinamese fictitious candidates were matched with native Dutch candidates and applied for jobs on low and high occupational level. Application for a job of low occupational level was performed by telephone (fictitious candidates calling the employer); for jobs of high occupational level the application was sent by mail. The main results of this study are presented in Table I:

Table I: Net discrimination rate of Moroccan and Surinamese job candidates, 1993/1994

	Occupational level	Method	N	Net discrimination (%)
Moroccan males	Low	Telephone	175	32
Moroccan females	Low	Telephone	92	35
Surinamese males	Low	Telephone	102	40
Surinamese females	Low	Telephone	100	36
Surinamese males	High	Mail	157	18
Surinamese females	High	Mail	78	No discrimination

Source: Bovenkerk et al. (1995)

The results indicate that Moroccan and Surinamese candidates overall have a lower chance of being invited for a job interview. Comparing the results for high and low level occupations, discrimination seems to occur more often in low level jobs; typically the jobs that real ethnic minority job seekers mostly apply for. No substantial difference in the amount of discrimination was found between Moroccan and Surinamese candidates. This research has offered rather convincing proof that ethnic discrimination occurs to a fairly large extent in the Dutch labour market. However, the results are now more than ten years old. To be able to say anything about ethnic discrimination in the Dutch labour market, correspondence tests needed to be repeated.

3. Research design for large scale correspondence testing

The aim of performing correspondence test in the Dutch labour market was, firstly, to obtain information about the extent of ethnic discrimination, and secondly, to study how discrimination varies across applicant characteristics (e.g. gender, amount of work experience) and vacancy characteristics (jobs with or without customer contact). Applicant characteristics include: ethnic background (native Dutch vs. Moroccan, Turkish, Surinamese or Antillean), gender (male or female), work experience (none, a few years, substantial). Vacancy characteristics include: customer contact (yes or no), occupational level (low, intermediate or high), sector (government, finance, retail, health care or hotel and catering industry). The research design therefore systematically combines applicant and vacancy characteristics, as shown in Table 2.

Table 2. Research design

Row	Matched candidates ethnic background	Gender	Sector	Occupational level	Customer Contact
A	Native Dutch - Moroccan	Male	Government	High	No
B	Native Dutch - Moroccan	Male	Care	Intermediate	Yes
C	Native Dutch - Moroccan	Female	Recreation	Intermediate	Yes
D	Native Dutch - Moroccan	Female	<i>All five sectors</i>	Low	-
E	Native Dutch - Moroccan	Female	Finance	High	No
F	Native Dutch - Turkish	Female	Government	Intermediate	-
G	Native Dutch - Turkish	Female	Care	High	Yes
H	Native Dutch - Turkish	Male	<i>All five sectors</i>	Low	-
I	Native Dutch - Turkish	Male	Retail	Intermediate	Yes
J	Native Dutch - Turkish	Male	Finance	High	No
K	Native Dutch - Surinamese	Male	Government	Intermediate	-
L	Native Dutch - Surinamese	Male	<i>All five sectors</i>	Low	
M	Native Dutch - Surinamese	Female	Recreation	High	Yes
N	Native Dutch - Surinamese	Female	Retail	High	No
O	Native Dutch - Surinamese	Man	Finance	Intermediate	
P	Native Dutch – Antillean	Female	Government	High	-
Q	Native Dutch – Antillean	Female	<i>All five sectors</i>	Low	
R	Native Dutch – Antillean	Man	Recreation	Intermediate	Yes
S	Native Dutch – Antillean	Man	Retail	High	No
T	Native Dutch – Antillean	Female	Care	Intermediate	yes

For each row in table 2 (A – T) a specific (set of) job(s) was specified that matched with the pre-specified vacancy criteria (sector, occupational level and customer contact) and that met the criterion of demand for labour. The specified job needed to publish at least 12 vacancies each month in order to be able to perform a substantial amount of correspondence tests within the fieldwork time frame.

The specification of job-types before data collection, allowed for working with standardized resumes and application letters. Once we had selected the target occupations that would be used in the correspondence tests, resumes and application letters could be designed. Each

resume had an entry on personal information (e.g. name, contact information, date of birth, birth place), information on education (educational tracks, dates, diplomas, additional courses etc), work experience (where, when, what) and other activities (e.g. volunteer work). The Moroccan and Turkish applicants had a birthplace in the Netherlands, whereas the Surinamese and Antillean candidates, who are not easily recognised by their name only, had a birthplace in Surinam or the Antilles in their resumes. Their resumes showed that all relevant education had been taken in the Netherlands.

After designing two resumes and application letters for each occupation, attractiveness and equality of the matched resumes and application letters was tested in a panel of employers. We had five panels: one for each sector. Each panel existed out of five to six employers of selectors with extensive experience in selecting job candidates for the target occupations. The resumes and application letters were first rated on relevance for the job, attractiveness, completeness. The equality of the two matched resumes was rated next and discussed in detail. The panel discussion offered a wealth of information that was used to improve our resumes and ensure equivalence across job candidates.

Discussions with key informants furthermore revealed that jobseekers who apply for low occupational level jobs do so in various ways. For example, a part of these jobseekers send their resumes to the employer, another part applies for the job by calling the employer on the telephone and a last part presents themselves directly in-person to the employer. In the end it may matter how jobseekers go about to find jobs for their chances of being invited for a job interview (and thus for getting a job). Therefore, we decided to use two variations in applying for low occupational level jobs: sending applications by mail and applying by telephone. The third option (directly presenting oneself to the employer) was no option as it would yield serious methodological problems.

Telephone applications were used in half of the tests for low occupational level jobs; the other half consisted of applications that were sent to the employer by mail. Preparations for the tests by telephone followed initially the same pattern as the tests by mail: the first step was to design the resumes. These resumes were not tested for equivalence since they were much simpler in structure: each contained only a few years of education (same level, same direction, same number of years of duration, same degree for the matched candidates) and approximately two years of work experience. The next step for the tests by telephone was to select actors that would pretend to be job applicants. In total, 12 actors were selected to perform 200 tests in total. Following the research design in Table 2, the job applicants were native Dutch females, native Dutch males, Moroccan females, Turkish males, Surinamese males and Antillean females. For each 'type' of job applicant two actors were selected. The roles of Moroccan females and Turkish males were played out by native Dutch actors. Therefore, these actors had no accent or any other attribute that is associated with ethnic background, except for their name (which was chosen to clearly signal the target ethnic background). Thus, if employers would still distinguish between the native Dutch and the Turkish or Moroccan applicant, this would make a strong case of discrimination, as the only real difference between the fictitious candidates would be their names. The roles of the Surinamese and Antillean job candidates were played out by Surinamese and Antillean actors respectively. These actors would apply for the job by telephone using a light Surinamese or Antillean accent. It was decided to use accent as a marker of ethnicity because Surinamese and Antillean names may not be easily recognised as non-Dutch names.

After selection, the actors were trained to play their roles and to act equivalent to their matched partner. They were instructed to always be polite and motivated for the job. Also, application scenarios were designed and then trained with the actors. The scenarios

envisaged possible questions an employer might ask, and the appropriate response for a candidate of this level. Moreover, it was anticipated that an employer could ask the same questions to the matched fictitious applicants. Equal but distinct answers were developed that could be used by the actors. For example, a possible question could be: how much do you earn in your current job? The first actor would then answer by giving the salary per hour, the second actor would tell his month loan (which would be exactly equal to the answer of the first actor, albeit on a different level).

4. Field work

The preparation stage of the fieldwork resulted in equivalent resumes and application letters and in actors who had been trained to apply for a job in equivalent ways. The first step of the fieldwork was to find vacancies that matched with our criteria (Table 2). Internet was used as a main source to find job vacancies (job portals, employers' websites etc). When a suitable vacancy was found, the appropriate resumes and letters were randomly assigned to the native and non-native job candidate. The application letters were adapted to match the specific job (changing the paragraph that stated why an applicant would want to work for that specific company). Then the resumes and letters were provided with existing addresses, mobile numbers and e-mail addresses, and consequently sent to the employer.

The mobile numbers were attached to anonymous voicemail boxes that recorded the employers' messages. The addresses belonged to friends, family, colleagues etc across the Netherlands. They were asked to collect the employers' reactions weekly, and send them back to us. The email addresses were gmail and live accounts, especially made for the purposes of this research.

After sending the application, the data was fed into SPSS. Employers reactions were collected and registered into SPSS as well. Employers' reactions that were no clear invitations for a job interview or a clear rejection of the candidate (e.g. employers calling the fictitious candidate and asking to be called by the applicant) were contacted by e-mail. In the e-mail the candidate stated that they were not able to contact the employer by telephone (while being abroad on holiday for example), but that they checked their e-mail on a daily base, and that they would be happy to answer all requests through e-mail.

For the tests by telephone a series of about 13 vacancies of the same job type were selected. Two actors would come to the Netherlands Institute for Social Research to apply for these jobs. The two equivalent resumes were assigned in changing order to each of the actors (e.g. the native Dutch actors would start using resume I for the first vacancy, and used resume II for the second vacancy, the Surinamese actor then used resume II for the first vacancy, and resume I for the second vacancy). The actors applied for the same vacancy, and care was taken that the second actor called the employer at least one hour after the first actor had made the call. The order who was first to call the employer also changed for each vacancy (e.g. if the native Dutch actor made the first call for vacancy I, then the Surinamese actor would make the first call for vacancy II).

The actors would state their names very clearly to the employer, and than announced that they were interested in the job that was vacant. Conversation with the employers were mostly very brief. Quite often, the employer did not take a decision for inviting or rejecting the applicant immediately, but postponed taking the decision by (I) asking if he/she could call the applicant at another time, (II) by asking whether the applicant could call again later, or (III) by asking the applicant to send their resume. When the employer promised that they

would call back themselves, we would simply wait and see. For the second and third option, we chose to send the written resumes to the employer. All employers' reactions were recorded in SPSS.

All telephone calls were on speakerphone, such that the two researchers present and the matched actor could hear the entire conversation. After each phone call evaluation forms were filled out by the actor who applied for the job and the two researchers that were present (hence, three evaluation forms for each phone call). The evaluation form asked questions about the tone of the conversation (friendly to hostile), how much questions were asked, whether these questions were relevant, whether the applicant was discouraged to apply for the job.

In the period may- December 2008 in total 1463 tests (hence, 2926 applications) were carried out. Of these 1463 tests, 1371 were valid tests. A test was considered non-valid for the following reasons:

- Selection of job candidates was not performed by the employer themselves, but through a job selection agency
- The employers reaction was not a clear invitation or rejection
- The reaction period for the vacancy was already closed
- The employer was suspicious
- The resume and/or application letter contained an error

5. Methods of analysis

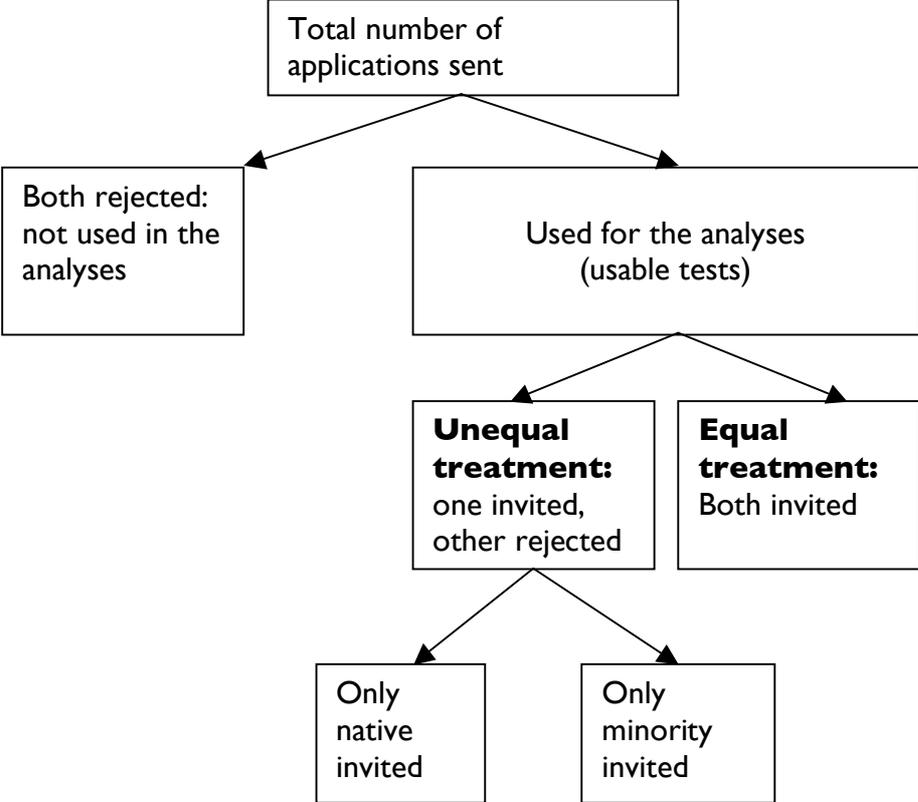
In principle, three methods of analysis can be applied to correspondence testing:

1. Chi-square tests (Bovenkerk, 1995)
2. Logistic regression on the level of tests
3. Logistic regression on the level of job applicants (Rooth, 2007)

Chi-square tests

This method has been applied in a number of studies for calculating the percentage of net-discrimination (e.g. Bovenkerk, 1995). In this method the tests where both applicants were rejected are considered ambiguous and therefore not used for analysis. The other categories (both are invited for a job interview, only one applicant is invited for an interview) are used for the analyses (see Figure 1).

Figure 1. Distinguishing between categories that will be used for analyses (Bovenkerk, 1995)



The net rate of discrimination is then calculated as the percentage of cases of unequal treatment divided by the total number of usable tests. One also may want to distinguish unequal treatment leading to a disadvantage for natives from unequal treatment resulting in disadvantage for ethnic minorities. More formally the formula reads:

$$\frac{N \text{ only native invited} - n \text{ only ethnic minority invited}}{\text{Number of usable tests}} \times 100$$

The percentage of discrimination thus calculated needs to be higher than a critical rate to rule out coincidence as a factor. The critical rate is calculated as

$$\sqrt{N} = 1.96\sigma / CR \sigma$$

- N= number of usable tests
- 1.96
- CR is the critical rate

The advantage of this method is that tests are analysed as a set, such that matched characteristics remain matched. The disadvantage is that normally half of the observations is not used for analyses; and that this method does not allow to control for any differences between the tests.

The second method does not suffer from these disadvantages. Again, the level of analysis is a test (two matched applications for the same vacancy). A test has four possible outcomes:

I) both candidates are rejected

II) both candidates are invited for a job interview

III) only the native applicant is invited, the ethnic minority applicant is rejected

IV) only the ethnic minority is invited, the native applicant is rejected

These outcomes may be used in a multinomial (all categories) or logistic (only native invited vs both rejected and both invited) regression analysis. The advantage of a regression analysis is that possible differences between tests (e.g. job type) can be controlled for in the analyses. The downside of this method is that a fairly large amount of tests is needed for the analyses, if one is interested in the effect of vacancy characteristics on the amount of discrimination.

The problem of the number of observations in the second method is counterbalanced in the third option. Instead of analysing the test as a single outcome, the third method of analysis takes the individual applicant as the level of analysis, thus multiplying the number of observation by two compared to the second method. The dependent variable is now the chance of being invited for an interview (logistic regression with adjusted standard errors). Control variables can be entered in the equation, The number of observations allows one more easily to calculate effects of vacancy characteristics, but this must be done through interaction effects.

Lastly, one may also think of the individual applicants and jobs (vacancies) as different levels. A multi-level approach may be combined with the second and third method of analysis.

The three methods result in different outcomes. The presentation at the IMISCOE conference will discuss the divergent outcomes and focus on the question of the appropriate method of analysis.