

# Names, Grades, and Metamorphosis: A Small-Scale Socio-onomastic Investigation into the Effects of Ethnicity and Gender-Marked Personal Names on the Pedagogical Assessments of a Grade School Essay

I. M. NICK

*Germanic Society for Forensic Linguistics, Flensburg, Germany*

In this small-scale, mixed-method investigation, the potential presence of school teacher and teacher trainees' name-based biases in reaction to schoolchildren's first names was investigated in two experiments. In the first, German school teachers were asked to qualitatively and quantitatively assess an authentic literary essay written by a monolingual native-speaking German schoolchild. To test for possible name biases, three nearly identical versions of the essay were prepared. The only difference between them was the first names of the child listed as having authored the essay (*Uwe*, *Achilleas*, and *Mustafa*). In the second experiment, the set of personal names was expanded to include female first names (i.e. *Heike*, *Athena*, and *Fatma*) and the study participants were volunteer teacher trainees. In both experiments, evidence for covert name biases was identified. On the basis of these and other findings, the article concludes with concrete suggestions for future research. Chief among these recommendations is a call for more collaborative action research between university faculty and students.

KEYWORDS ethnicity, gender, bias, teachers, Germany, action research.

## Introduction

The current study is an outgrowth of a collaborative research project conducted by the author and a group of undergraduate students from an introductory course on onomastics. During one of her lectures, the instructor offered a personal anecdote to illustrate

the effects of name-based bias and discrimination. As a child, the instructor attended an elementary school in Northern California where one of the subjects taught was cursive handwriting. This was one of the instructor's favorite subjects and, consequently, she invested a great deal of time and effort in completing her homework. Despite this fact, much to her disappointment and her fellow classmates' dismay, she repeatedly received below-average grades on her assignments. In comparison, "Terry," one of the most popular boys in her class, consistently received above-average grades for his handwriting homework, although he despised the subject and rarely spent more than a few minutes before class on completing his assignments. One day, the two classmates came up with what they felt would be a fun idea. The author wrote her name on Terry's homework and Terry wrote his name on hers. They then turned in their assignments and waited for the result. What happened was not only a surprise but a life-changing experience. For the very first time, Terry received a failing grade on his homework and the author received an A+. When the two classmates informed their teacher about their experiment, she blurted out in indignation: "Typical! The only way a nigger can do well is to cheat!" During the instructor's office hours, three students from the onomastic course explained that they wanted to replicate the instructor's childhood experiment for their end-of-semester project. Their reason for selecting this idea was their own personal experience with classroom discrimination against women and ethnic minorities in Germany. What began as a simple class project soon developed into a formal, collaborative, small-scale investigation. The results presented in this article are the culmination of this teacher-student action research (AR). For this investigation, AR is operationalized using the definition offered by Hine (2013, 151); namely, "action research" is a collaborative "process of systematic inquiry that seeks to improve social issues affecting the lives of everyday people;" with a collaborative research methodology that involves "planning, observing, and reflecting." As will be discussed later, this approach not only yielded important insights about name-based biases, but also provided the research team with a healthy means for processing the long-term effects of these prejudices.

## Past research on name-based biases

Decades of research conducted in education, economics, psychology, sociology, and linguistics have demonstrated that name-based biases can directly affect the level of attractiveness, morality, intelligence, affluence, power, and emotionality attributed to a name bearer (Erwin 1993; Garwood et al. 1980; Joubert 1983; Mehrabian 1992; Mehrabian and Piercy 1992; Savage and Wells 1948). The effects of these attributions have been particularly well documented for bearers of marked ethnic personal names across multiple contexts — from seeking housing (Ahmed et al. 2010; Bertrand and Mullainathan 2004; Bosch et al. 2010; Carpusor and Loges 2006; Deros et al. 2009; Feldman and Weseley 2013; Hanson and Hawley 2011); applying for employment (Baert et al. 2015; Bosch et al. 2010; Cotton et al. 2008; Pascual et al. 2015; Rooth 2010); to finding a partner (Gebauer et al. 2011). Collectively, this work has also demonstrated that name-based biases may be directly related to differential rates of acceptance and achievement for ethnic minorities with marked personal names. The effects of name bias have also been found in relation to gender marking (Mehrabian and Valdez 1990; Takiff et al. 2001). Personal names strongly associated with a male or female gender can significantly influence people's

perceptions, assessments, and decision making (Booth and Leigh 2010; Moss-Rascusin et al. 2012; Steinpreis et al. 1999). Investigators postulate that the mechanism at work here is identical to that triggered in reaction to ethnically marked names. In both cases, marked personal names seem to activate pre-existing, largely unconscious, stereotypes which in turn may yield measurable attitudinal and behavioral effects. As the author's personal anecdote illustrates and research confirms, such name-based "implicit biases"<sup>1</sup> may affect the lives of not only adult name bearers. Onomastic prejudices can also translate into concrete acts of discrimination for and against adolescents and children. For instance, it has been consistently found that young people who carry personal names perceived as desirable tend to be more favorably viewed by their peers than cohorts whose names are considered to be less desirable (Busse & Seeaydarian 1979; McDavid and Harari 1966; West and Shults 1976). Importantly, such name-based biases may affect both child-child as well as adult-child interactions. Where the latter relationship is concerned, much research has attested the presence of teacher biases in reaction to pupils' personal names. In 1973, Harari and McDavid found that teachers rated the quality of a schoolchild's essay much more highly when the child's name was associated with positive attributes. Similar findings were reported in 1976 by Garwood who found that children who bore names that elementary teachers had rated as desirable received higher scores on school achievement measures than peers whose names teachers had rated less favorably. Erwin and Caley (2011) also found a positive relationship between teachers' grading of student essays and their attractiveness ratings for pupils' personal names. The effect of teachers' name biases has even been detected in online instruction environments where face-to-face student-teacher contact is minimal to non-existent (Conaway and Bethune 2015). Investigations into name biases in the classroom have also explored teachers' reactions to marked ethnic names. Here again, clear evidence for onomastic prejudice and discrimination could be detected. Anderson-Clark et al. (2008) reported that elementary school teachers rated the achievement of fifth graders with marked African-American names lower than those written by pupils with Caucasian names. Sprietsma (2009, 2013) similarly found that teachers graded essays labelled with stereotypical German personal names significantly better than those essays featuring Turkish personal names. Considering the tremendous impact that teacher attitudes and evaluations can have upon students' immediate and future academic achievement (van Ewijk 2011; Peterson et al. 2009), continuing research into prejudicial assessment is imperative — particularly where vulnerable student populations are concerned.

## Women and ethnic minorities in Germany

According to the 2015 micro-census conducted by the German Federal Statistical Office (Destatis), women and ethnic minorities in Germany have continued to make significant gains in enrollment and graduation rates. Nevertheless, both groups still lag substantially behind their male native German cohorts. For example, in a Destatis press release (2014, para 6) it was reported that "15.5% of the population with a migrant background who are at least 15 years of age" had not completed their school education; by comparison, among German residents without an immigrant background, the dropout percentage was only 2.3%. The statistics are equally sobering for the female segment of German society. Although women make up half of the university student population in Germany,

in 2014, they made up only 11% of the nation's number of full professors (Destatis 2013). In the face of these and other findings, it should come as no surprise that women and ethnic minorities in Germany continue to report disproportionately low salaries, job security, and retirement pensions — all areas which are strongly related to educational attainment. It is therefore a social imperative to identify and eradicate those factors which thwart equal access and advancement in the German education system. This article focuses on one, all too often overlooked, barrier to academic equality: name-based biases in school teacher assessments.

## Research methodology

### *Experiment 1*

#### Study design

An authentic school essay was obtained from a local German schoolteacher. The topic of the essay was Franz Kafka's classic, "Die Verwandlung" ("The Metamorphosis"). The essay writer was a female, monolingual, German native-speaking pupil in the 11th grade. The essay had been awarded the numerical grade of 15.<sup>2</sup> This mark is the highest grade possible in the German school system and is equivalent to an A+ on the American grading scale. A typed version of the original essay was prepared to negate potential confounding biases in reaction to the pupil's handwriting such as (sub-)conscious judgments regarding the attractiveness, legibility, or accuracy of the pupil's script. The name of the original pupil was replaced by three new names: *Uwe*, *Mustafa*, and *Achilleas*. Thus, the only difference between the three versions was the pupil names appearing at the top of the essay. The three different versions of the essay were randomly sent to volunteer German teachers employed in public schools throughout a major metropolitan city located in the German federal state of North Rhine Westphalia. To avoid observer bias, the exact purpose of the study was not disclosed. Instead, participants were told that the investigation was part of a university course on pedagogy designed to show teacher trainees how to properly assess pupils' essays. With that justification, all participants were asked to provide a grade for the essay and a brief explanation for their assessment. Participants were assured that their assessments would be kept strictly anonymous. The accumulated numerical grades and accompanying explanations were then compared to determine whether the personal names assigned to the essay version had had a discernible effect on the assessments.

#### Findings

Two major findings were obtained in Experiment 1. The first pertained to the numerical assessments. Although the essay had been initially awarded 15 points, the highest grade possible on the German scale, the overall grades awarded by the study respondents were much lower with an average of 8.67 (i.e. between a C and a C+ on the US American grading scale). The second major finding involved the differences in the average grades given to each of the three essay versions. These cumulative differences are displayed in Figure 1.

As illustrated in Figure 1, there was a clear numerical difference in the assessments. The essays which were supposedly written by a pupil with the traditional male German name *Uwe* received an average of 10.7 points, the equivalent to a B– on the US American grading system. In the following excerpts from the teachers' commentary, a few justifications for the numerical assessment of Uwe's essay are provided:

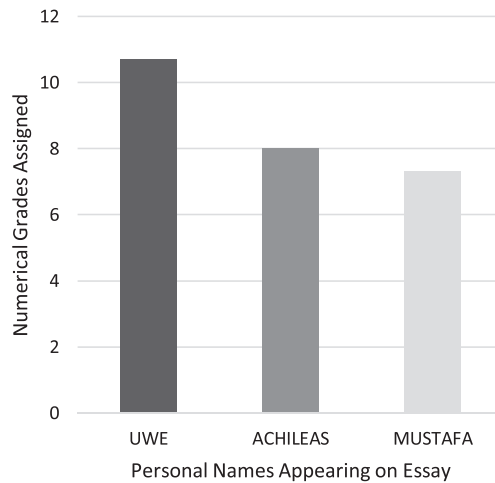


FIGURE 1 Average numerical grades assigned by teachers for each named version of the essay.

*Content:* The inner as well as the outer transformation [of the character] were given a detailed discussion. The student repeated some aspects rather often and the sentence structure could have been better. Nevertheless, all of the most essential points were covered.

[Sowohl auf die innere als auch auf die äußere Verwandlung ist ausführlich eingegangen worden. Manche Aspekte wiederholt er des Öfteren und die Struktur des Aufsatzes könnte noch besser sein. Aber alle wesentlichen Punkte sind genannt worden.]

*Language:* Few spelling errors [...] good sentence structure.

[Wenig Schreibfehler [...] guter Satzbau.]

By comparison, with a combined average of 7.65, the number of points awarded to the two versions with the ethnically marked names was far lower. As the Figure 1 bar chart also illustrates, a difference also emerged in the points given to the essays allegedly written by *Achilleas* and *Mustafa*. While the essays marked with the male Greek name received an average of 8.0 points, which corresponds to a C on the US American grading scale; the essays labelled with the male Middle Eastern name *Mustafa* were only given an average of 7.3 points. Not surprisingly, the commentaries the teachers provided on the essays with the ethnically marked names were commensurately critical. For example, one of the teachers who assessed an *Achilleas* version of the essay provided the following negative assessment:

“A linguistic analysis is practically non-existent; and the pupil quotes incorrectly. This performance is not even at an introductory course level.” [*Eine Sprachanalyse findet praktisch nicht statt, und der Schüler zitiert nicht korrekt. Auf Grundkurs-Niveau ist die Leistung nicht.*] The commentary received for the *Mustafa* versions was even more caustic. This observation is exemplified by the following respondent statement:

There is no text analysis. The pupil limits the discussion to the issue of the father’s transformation. With reference to this analysis, the quotations used are correct but there were also errors as well as evidence of a partial misunderstanding of the text. [There were] diverse errors in expression and tense selection, as well as a general lack of control in the use of literary terminology.

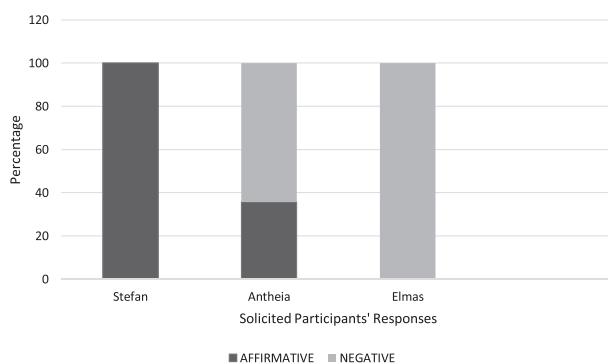


FIGURE 2 Percentage of affirmative and negative responses given by potential participants contacted, by name of student recruiter.

[Eine Textanalyse fehlt. Der Schüler beschränkt sich auf die Frage nach der Verwandlung des Vaters. Dazu überwiegend richtige Hinweise, aber auch Fehler, z.T. Missverständnisse. Verschiedene Fehler im Ausdruck, in der Tempusbildung, unsicherer Umgang mit der Fachterminologie.]

What is particularly remarkable about the above assessment, aside from its severity, is the fact that the teacher explicitly criticizes the pupil for having focused upon the father's transformation. However, in the directions appearing at the top of every essay, it is clearly stated that the pupils were to pay particular attention to the father's transformation. This means *Mustafa* was penalized for following the directions. Although dramatic, according to the chi-square test performed, the quantitative differences in assessment for the three different names did not reach a level of statistical significance [ $\chi^2_{\text{obt}} = 3.00$ ;  $\chi^2_{\text{crit}} = 7.82$ ;  $df = 3$ ;  $\alpha = 0.05$ ]. However, at  $\alpha = 0.50$ , the differences in grading would have reached the level of statistical significance. For the purposes of this study, the conventional *a priori* threshold was maintained at 0.05. An unexpected result of the first experiment pertained to the students' disproportionate success rates in participant recruitment. For consistency sake, the solicitation letters sent to the potential respondents were identical, save for one point: the name of the student featured at the bottom of the letters.<sup>3</sup> The variant response rates are illustrated in Figure 2.

As shown in Figure 2, the student recruiter with the traditional German male name of *Stefan* had a 100% affirmative response rate. By comparison, solicitation letters featuring the marked ethnic names of the two female student recruiters (i.e. *Antheia* and *Elmas*) had much lower positive response rates (i.e. 35.7% and 0.0%, respectively). What is more, on more than one occasion, teachers who had been sent solicitation letters labelled with the name *Elmas* wrote back indignant responses, demanding that the student learn to do her own work.

## Experiment 2

### Study design

In accordance with AR, the methodology and results of Experiment 1 were collectively reviewed and corresponding adjustments were made. The procedure followed in Experiment 2 was identical with the first, with four key differences. The first was to

expand the set of personal names to include both genders. This alteration was made to determine what, if any, effect the perceived gender of the essay writer might have upon the assessments. The second alteration was to substitute the German male name *Uwe* with *Thorsten*. The reason for this change was the collective sentiment among the native-speaking student co-researchers that *Uwe* was more typical of a much older name bearer and comparatively unusual for the test region. After group consultation, the following set of gendered names was chosen for Experiment 2: *Mustafa/Fatma*, *Achilleas/Athena*, and *Thorsten/Heike*. The third alteration involved the subjects solicited for the study. In reviewing the results of the initial experiment, it was postulated that the respondents' generation might have played a role in the findings. More specifically, the student co-researchers speculated that, in comparison to themselves, the teachers in Experiment 1 may have had fairly little direct contact with different ethnic groups during their formative years and this lack of experience may have affected their judgment. On the basis of this reasoning, for Experiment 2, the essays were circulated to volunteer university-level teacher trainees. The fourth and final procedural change involved the instructions provided to the participants. While, in Experiment 1, the subjects were only asked to provide grades and commentary; in Experiment 2, the participants were also asked to indicate their nationality, age, and gender. In addition, to ensure that the subjects had attended to the name of the essay writer, each participant was asked to write the name of the pupil on the answer sheets provided. With these procedural alterations, Experiment 2 sought to answer the following research questions: 1) Would the presence of a marked ethnic personal name significantly affect the teacher trainees' assessments? and 2) Would the essays labelled with a female personal name be differently assessed than those featuring a male personal name?

### Participants

All of the study participants had German nationality [ $n = 19$ ].<sup>4</sup> The average age of the participants was 24.79 with a range of 19 to 33. Presumably reflective of the fact that the teaching profession in Germany is still dominated by women, the subject pool also demonstrated a strong gender skew with 63.2% [12] of the respondents self-identifying as female and 36.8% [7] as male.

## Findings

### *Ethnicity marking*

A comparison of the grades assigned to the unmarked ethnic names (e.g. *Thorsten/Heike*) and the marked ethnic names (*Mustafa/Fatma*, *Achilleas/Athena*) revealed several differences. While the majority of the teacher trainees (42.1%) [8] gave the essays with *Thorsten* or *Heike* the equivalent of a B, more than half of these participants (57.9%) [11] who graded an essay featuring one of the marked ethnic names rated the work as deserving a C or even a D. The exact distribution of the grades for the marked ethnic names (MEN) and the unmarked ethnic names (UMEN) is shown in Figure 3.

As shown in Figure 3, the grades given to the MEN essays ranged between A and D. By contrast, the grades given to the UMEN essays ranged between B and D. Furthermore, while 11.1% [2] of the teacher trainees who had read an essay written by either *Heike* or *Thorsten* gave a grade equivalent to an A, none of the essays labelled with the names

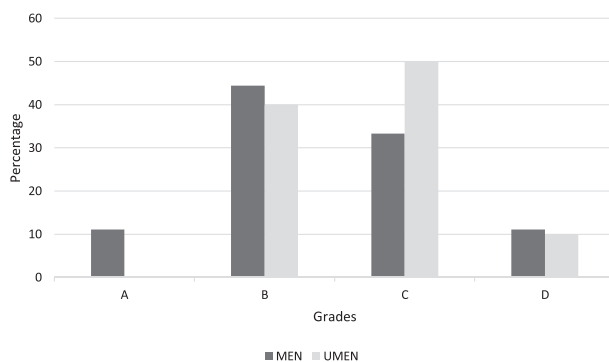


FIGURE 3 The percentage distribution of grades given to essays with marked ethnic names (MEN) and unmarked ethnic names (UMEN).

*Mustafa*, *Fatma*, *Achilleas*, or *Athena* were assessed at this level, *despite being absolutely identical*. The disparity in these evaluations is epitomized in the commentary provided for the essays labelled with *Thorsten* and *Mustafa*. At one end of the continuum, the *Thorsten* essay was described as being “very clear” and “understandable” with “only a few, very minor mistakes” and “very formal style,” despite the pupil’s tendency “to repeat key points” and the need for “more theoretical arguments.” The criticism received for the *Thorsten* essay was likewise fairly mild. At the other end of the continuum, the feedback given for the *Mustafa* version was extremely critical. One trainee made the following assessment: “Unfortunately, throughout the text, there were a variety of spelling and grammatical errors which should not be made in the 11th grade.” The critical reaction to the *Mustafa* essay was epitomized by the following appraisal offered by one of the teacher trainees: “[...] admittedly there were a few passages which flowed rather well in terms of the style and thoughts presented. *Given the quality of these sections, there is reason to believe that the essay may well have been plagiarized*” [emphasis added]. Here again, the pupil with a minority name was punished for following the directions and performing well. Despite such glaring quantitative and qualitative differences in the grades given to the UMEN and MEN essays, the chi-square test conducted did not detect a statistically significant difference [ $\chi^2_{\text{obt}} = 0.75$ ;  $\chi^2_{\text{crit}} = 7.82$ ;  $df = 3$ ;  $\alpha = 0.05$ ]. However, the results of the Pearson R test for correlation did reveal a moderately positive relationship between the respondents’ assessments and whether the name on the essay they received was ethnically marked or unmarked [ $r = 0.20$ ].

### *Gender marking*

The range of numerical grades assigned for the essays labelled with female and male personal names was nearly identical. An examination of the numerical averages for the two gendered groups also revealed a very minimal difference. Overall, the teacher trainees graded the essays with the male personal names only 0.93 times higher than the essays featuring a female personal name. This negligible difference was confirmed by the identical frequency distribution of essays awarded an A, B, C, or D or lower for both sets of gendered names. A review of the qualitative assessments yielded similar results. The only difference spied in the commentary related to the type of criticism offered.



While the essays with the male personal names tended to be deficit-oriented (i.e. lists of grammatical errors and detailed criticisms of deficient analyses), the essays labelled with the female personal names tended to be far more achievement-oriented, praising the points that the pupil had done particularly well.

## Discussion

Despite having received identical essays save for the name at the top, the respondents demonstrated a striking lack of consensus in their judgments of the quality and correctness of the essay with regard to not only the content, format, structure, and argumentation; but also the incidence and severity of the grammatical and orthographical errors. To a certain extent, these evaluative differences were to be expected. As considerable research in cognitive psychology has demonstrated, decision makers routinely demonstrate marked differences in the way they attend to, evaluate, and weight performance-related features (Dhami 2003; Gigerenzer and Gaissmaier 2011; Tversky and Kahneman 1974). Consequently, criteria such as spelling, punctuation, and grammar are only objective quality markers in theory; in practice, they can be and often are quite subjectively evaluated (Gamaroff 2000). Ultimately, the fact that inter-rater variation in the educator assessments was detected was not surprising. What was surprising was the fact that this variation was found to correlate, however mildly, with the presence (or absence) of ethnically marked names.

Given the small size [ $N = 28$ ] and the regional restriction of the samples drawn here, it would naturally be highly inappropriate to over-generalize the results of this non-representative investigation. Nevertheless, certain points can be made. The fact that neither the ethnicity nor the gender of the personal names was found to statistically significantly affect the numerical grades of either participant group was very encouraging. However, this does not mean to say that the teachers and teacher trainees in this study were free of name-based biases. Rather, this study, like many others (Anderson 2010; Baird 1998; van der Bergh et al. 2010; van Ewijk 2011; Laversuch 2011), may indicate that gender and culture biases have simply become more subtle than when the investigator was a child. This is a development that has been witnessed throughout modern European society. As Zick et al. (2011, 31) observed in an official report issued by Friederich-Ebert-Stiftung in Berlin: “In recent decades in Europe [...] strong social norms of tolerance [...] have become established, and increasingly inhibit open expression of prejudice [...] yet even where social norms of tolerance are widely shared, negative emotions towards particular groups often remain extant.” Like a mutating virus, contemporary bigotry can cause tremendous harm in its new, hard to detect, forms (e.g. undermining the confidence of others, relentlessly questioning their competence, public humiliation, and clandestine favoritism). Accordingly, the covert nature of modern prejudice does not automatically negate or diminish the potential damage done to either individual targets of discrimination or society at large. As Anderson (2010, 286) explains “even when equality of opportunity appears to exist, people’s subtle, covert, and often unconscious behavior undermines the reality of equal opportunity.”

In an effort to eliminate the real or suspected effects of name-based biases in assessments, many policy makers and researchers in education have advocated the use of blind assessments. Joe and Cowling (2009) even went so far as to suggest that student names

should be replaced with bar codes. Although such interventions would no doubt do much to combat the effects of name-based biases in academic assessments, on their own, they cannot solve the underlying problem of prejudice and discrimination in education. For this to happen, more open discussions about the short and long-term effects of prejudice and discrimination are needed. In addition, pre-existing policies to identify, educate, and, when necessary, punish repeat offenders must be enforced.

## Suggestions for future research

Despite the small sample size, the results of this investigation are in line with the large body of scientific evidence that establishes the existence and effects of name-based biases among educators (e.g. Anderson-Clark et al. 2008; Baird 1998; Kaiser 2010; Sprietsma 2013). Researchers interested in further investigating name-based biases would still be well advised to use larger, more gender-balanced samples drawn across a larger geographic area. Doing so would allow for more sophisticated statistical tests to determine whether the incidence, strength, or direction of teachers' assessments vary significantly with the evaluators' demography (e.g. age, ethnicity, gender, race, religion etc.) and/or the psychological attributes (e.g. openness, tolerance, agreeableness, and extraversion) as past research has indicated (Anderson et al. 2012; Joubert 1999; Lloyd 2013; Paludi and Strayer 1985; Steinpreis et al. 1999). Another suggestion for other researchers involves the name type investigated. This study, like many others before it, focused exclusively on the potential influence of first names. However, there is compelling evidence that surnames can also trigger name-based biases (Goldstein and Stecklov 2016; Pascual et al. 2015; Silberzahn and Uhlmann 2013). New and important insights may well be revealed by replicating this study using surnames either instead of or in addition to first names. Furthermore, given the established importance of perceived attractiveness, desirability, familiarity, and commonality in the formation and manifestation of personal name-based biases (Allen et al. 1941; Crisp et al. 1984; Ford et al. 1984; Pascual et al. 2015), it is recommended that other investigators take into account the potential effect of such variables. Doing so could help to disambiguate the findings obtained. For example, the fact that the essays labelled with *Mustafa* were comparatively negatively assessed may have simply been a manifestation of the perceived unattractiveness, undesirability, or uncommonness of this name. Considering the increasing number of studies that have attested to the presence of biases against Arabic-sounding names (Ahmed et al. 2010; Anderson et al. 2012; Gaddis and Ghoshal 2015; Rooth 2010; Widner and Chicoine 2011), it is also possible that the comparatively negative assessments of the *Mustafa* essay were, in whole or in part, a manifestation of xenophobic attitudes harbored against Muslims or Middle-Easterners. To help clarify this point additional research would be needed. The final recommendation to come from this investigation pertains to AR. As detailed in the introduction, this study was conducted by the investigator and a small team of undergraduates as part of an introductory university course on onomastics. As other scientists have observed (Bauman 1996; Hine 2013; Hong et al. 2007; Wilson 1995), the dual teacher-researcher role demanded by such projects can present many challenges. For example, to ensure that education and active participation of the students remained at the forefront, it was necessary to scale-down the complexity and the size of the study. In addition, the investigator was also charged with helping the students

process their sometimes exceedingly visceral reactions to the data. The fact that the differences obtained did not reach statistical significance did little to diminish the shock, disappointment, betrayal, and vindication that many of the students experienced when reading through the respondents' commentary. The strength of these reactions often required the investigator to step outside of the traditional role of an emotionally-removed scientist. This new position also provided the researcher with a unique opportunity to also confront her own experiences with institutional prejudice. After all, implicit bias in education is by no means limited to primary school interactions between teachers and pupils. In the final analysis, the heightened mutual, personal engagement altered the nature of the work but did not diminish its value. Indeed, for both sides of the collaboration, the advantages of this AR project far outweighed any of the disadvantages. For the teacher-investigator, despite the emotionally challenging dual role, it was extremely gratifying to watch students discover the importance of names and develop a real appreciation for the power (and limitations) of onomastic research. For the students, many of whom planned to become schoolteachers themselves, the realization that even unconscious personal prejudices can substantially impact professional judgments was transformational. The long-term effects of these benefits are by no means limited to the direct producers of AR projects. As Levin and Greenwood (2001, 103) observe, AR can help to move "institutions of higher learning towards becoming collective learning organizations engaged in improving society and [...] away from being redoubts of self-serving and autogenic academic activity." With these goals in mind, the final suggestion is that more onomastic scholars should invite their students to join them in collaborative AR projects.

## Acknowledgements

Special appreciation is given to F. Akkuş, U. Buschmann, P. Molochidis, and P. Schemensky for their tireless dedication in bringing this research project to fruition.

## Notes

<sup>1</sup> According to Greenwald et al. (1998, 1464), the psychological construct of an "implicit bias" refers to "actions or judgments that are under the control of automatically activated evaluation, without the performer's awareness."

<sup>2</sup> Of course, it is impossible to rule out the effect of bias in the first grading of this essay. Had this work been assessed by a different instructor, it might have been awarded a different grade. However, this very real potential for inter-rater variation only serves to

underscore the overall point that teacher assessments are susceptible to individual biases. (For more on this issue, see: Anderson-Clark et al. 2008; Baird 1998; Sprietsma 2013.)

<sup>3</sup> To maintain the privacy of the student recruiters, their real first names were exchanged for ethnolinguistic equivalents.

<sup>4</sup> The raw scores for the statistical percentages are provided in square brackets.

## Bibliography

- Ahmed, A., L. Andersson, and M. Hammarstedt. 2010. Can discrimination in the housing market be reduced by increasing the information about the applicants. *Land Economics* 86, no. 1: 79–90.
- Allen, L., V. Brown, L. Dickinson, and K. Pratt. 1941. The relation of first name preferences to their frequency in the culture. *Journal of Social Psychology* 14, no. 2: 279–293.
- Anderson, K. 2010. *Benign bigotry: The psychology of subtle prejudice*. Cambridge: Cambridge University Press.

- Anderson-Clark, T., R. Green, and T. Henley. 2008. The relationship between first names and teacher expectations for achievement motivation. *Journal of Language and Social Psychology* 27, no. 1: 94–99.
- Anderson, L., N. Jakobsson, and A. Kotsadam. 2012. A field experiment of discrimination in the norwegian housing market: Gender, class, and ethnicity. *Land Economics* 88, no. 2: 233–240.
- Baert, S., B. Cockx, N. Gheyle, and C. Vandamme. 2015. Is there less discrimination in occupations where recruitment is difficult? *ILR Review* 68, no. 3: 467–500.
- Baird, J. 1998. What's in name? Experiments with blind marking in a-level examinations. *Educational Research* 40, no. 2: 191–202.
- Bauman, J. 1996. Conflict or compatibility in classroom inquiry? One teacher's struggle to balance teaching and research. *Educational Researcher* 25, no. 7: 29–36.
- van der Bergh, L., E. Denessen, L. Hornstra, M. Voeten, and R. Holland. 2010. The implicit prejudiced attitudes of teachers relations to teacher expectations and the ethnic achievement gap. *American Educational Research Journal* 47, no. 2: 497–527.
- 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, no. 4: 991–1013.
- Booth, A., and A. Leigh. 2010. Do employers discriminate by gender? A field experiment in female-dominated occupations. *Economic Letters* 107, no. 2: 236–238.
- Bosch, M., M. Carnero, and L. Farré. 2010. Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional Science and Urban Economics* 40, no. 1: 11–19.
- Busse, T., and L. Seeaydarian. 1979. First Names and Popularity in Grade School Children. *Psychology in the Schools* 16, no. 1: 149–153.
- Carpusor, A., and W. Loges. 2006. Rental discrimination and ethnicity in names. *Journal of Applied Social Psychology* 36, no. 4: 934–952.
- Conaway, W., and S. Bethune. 2015. Implicit bias and first name stereotypes: What are the implications for online instruction? *Online Learning Journal* 19, no. 3: 162–178.
- Cotton, J., B. O'Neill, and A. Griffin, A. 2008. The 'name game': Affective and hiring reactions to first names. *Journal of Managerial Psychology* 23, no. 1: 18–39.
- Crisp, D., R. Apostol, and H. Luessenheide. 1984. The relationship of frequency and social desirability of first names with academic and sex role variables. *Journal of Social Psychology* 123, no. 1: 143–144.
- Derous, E., H. Nguyen, and A. Ryan. 2009. Hiring discrimination against arab minorities: Interactions between prejudice and job characteristics. *Human Performance* 22, no. 4: 297–320.
- Destatis. 2013. Female earnings-male earnings: How wide is the gap really? [https://www.destatis.de/EN/Publications/STATmagazin/EarningsLabourCosts/2013\\_03/2013\\_03GenderPayGap.html](https://www.destatis.de/EN/Publications/STATmagazin/EarningsLabourCosts/2013_03/2013_03GenderPayGap.html) (accessed May 31, 2016).
- Destatis. 2014. Proportion of women: Academic career. <https://www.destatis.de/EN/FactsFigures/SocietyState/EducationResearchCulture/Ins.html> (accessed June 7, 2016).
- Dhami, M. 2003. Psychological models of professional decision making. *Psychological Science* 14, no. 2: 175–180.
- Erwin, P. 1993. First names and perceptions of physical attractiveness. *The Journal of Psychology* 127, no. 6: 625–631.
- Erwin, P., and A. Caley. 2011. The influence of Christian name stereotypes on the marking of children's essays. *British Journal of Educational Psychology* 54, no. 2: 223–227.
- van Ewijk, R. 2011. Same work, lower grade? Student ethnicity and teachers' subjective assessments. *Economics of Education Review* 30, no. 5: 1045–1058.
- Feldman, M., and A. Weseley. 2013. Which name unlocks the door? The effect of tenant race/ethnicity on landlord response. *Journal of Applied Social Psychology* 43, no. S2: E416–E425.
- Ford, M., I. Miura, and J. Masters. 1984. Effects of social stimulus value on academic achievement and social competence: A reconsideration of children's first-name characteristics. *Journal of Educational Psychology* 76, no. 6: 1149–1158.
- Gaddis, S., and R. Ghoshal. 2015. Arab American housing discrimination, ethnic competition, and the contact hypothesis. *The ANNALS of the American Academy of Political and Social Science* 660, no. 1: 282–299.
- Gamaroff, R. 2000. Rater reliability in language assessment: The bug of all bears. *System* 28, no. 1: 31–53.
- Garwood, S. 1976. First-name stereotypes as a factor in self-concept and school achievement. *Journal of Educational Psychology* 68, no. 4: 482–487.
- Garwood, S., L. Cox, V. Kaplan, N. Wasserman, and J. Sulzer. 1980. Beauty in only 'name' deep: The effect of first-name on ratings of physical attraction. *Journal of Applied and Social Psychology* 10, no. 5: 431–435.
- Gebauer, J., M. Leary, and W. Neberich. 2011. Unfortunate first names: Effects of name-based relational devaluation and interpersonal neglect. *Social Psychology and Personality Science* 3, no. 5: 590–596.
- Gigerenzer, G., and W. Gaissmaier. 2011. Heuristic decision making. *Annual Review of Psychology* 62: 451–482.
- Goldstein, J., and G. Stecklov. 2016. From Patrick to John F.: Ethnic names and occupational success in the last era of migration. *American Sociological Review* 81, no. 1: 85–106.

- Greenwald, A., D. McGhee, and J. Schwartz. 1998. Measuring individual difference in implicit cognition. *Journal of Personality and Social Psychology* 74, no. 6: 1464–1480.
- Hanson, A., and Z. Hawley. 2011. Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *Journal of Urban Economics* 70, no. 2: 99–114.
- Harari, H., and J. McDavid. 1973. Name stereotypes and teachers' expectations. *Journal of Educational Psychology* 65, no. 2: 222–225.
- Hine, G. 2013. The importance of action research in teacher education programs. *Issues in Educational Research* 23, no. 2: 151–163.
- Hong, J., H. Chick, and J. Moss. 2007. Classroom researcher as teacher-researcher. *The Mathematics Educator* 10, no. 2: 1–26.
- Jae, H., and J. Cowling. 2009. Objectivity in grading: The promise of bar codes. *College Teaching* 57, no. 1: 51–55.
- Joubert, C. 1983. Unusual names and academic achievement. *Psychological Reports* 53, no. 1: 266.
- Joubert, C. 1999. Correlates of preferences for personal names with neo five factor inventory. *Psychological Reports* 84, no. 2: 575–581.
- Kaiser, Astrid. 2010. Vornamen: Nomen est omen? Vorerwartungen und Vorurteile in der Grundschule. *Schulverwaltung. Zeitschrift für Schulleitung und Schulaufsicht* 21, no. 2: 58–59.
- Laversuch, I. 2011. *Multiple language, variation, variety and use*. Aachen: Shaker.
- Levin, M., and D. Greenwood. 2001. Pragmatic action research and the struggle to transform universities into learning communities. In *Handbook of action research*, eds. P. Reason and H. Bradbury, 103–113. London: Sage.
- Lloyd, M. 2013. Gender factors in reviewer recommendations for manuscript publication. *Journal of Applied Behavior Analysis* 23, no. 4: 539–543.
- McDavid, J., and H. Harari. 1966. Stereotyping of names and popularity in grade-school children. *Child Development* 37, no. 2: 453–459.
- Mehrabian, A. 1992. Interrelationships among name desirability, name uniqueness, emotional characteristics connoted by names and temperament. *Journal of Applied Social Psychology* 22, no. 23: 1797–1798.
- Mehrabian, A., and M. Piercy. 1992. Positive or negative connotations of unconventionally and conventionally spelled names. *The Journal of Social Psychology* 133, no. 4: 445–451.
- Mehrabian, A., and P. Valdez. 1990. Basic name connotations and related sex stereotyping. *Psychological Reports* 66, no. 3: 1309–1310.
- Moss-Racusin, C., J. Dovidio, V. Brescoll, M. Graham, and J. Handelsman. 2012. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*. 109, no. 41: 16474–16479.
- Paludi, M., and L. Strayer. 1985. What's in an author's name? Differential evaluations of performance as a function of author's name. *Sex Roles* 12, no. 3/4: 353–361.
- Pascual, A., N. Guéguen, B. Vallée, M. Lourel, and O. Cosnefroy. 2015. First name popularity as a predictor of employability. *Names* 63, no. 1: 30–36.
- Peterson, T., K. Van Dam, and L. Wheeler. 2009. Who do we think we are. In *Proceedings of World Conference on E Learning in Corporations, Government, Healthcare, and Higher Education*, eds. T. Bastiaens, J. Dron and C. Xin, 2227–2237. Chesapeake, VA: AACE.
- Rooth, D. 2010. Automatic associations and discrimination in hiring: Real world evidence. *Labor Economics* 17, no. 3: 523–534.
- Savage, B., and F. Wells. 1948. A note on singularity in given names. *Journal of Social Psychology* 27, no. 2: 271–272.
- Silberzahn, R., and E. Uhlmann. 2013. It pays to be Herr Kaiser: Germans with noble-sounding surnames more often work as managers than as employees. *Psychological Science* 24, no. 12: 2437–2444.
- Sprietsma, M. 2009. Discrimination in grading? Experimental evidence from primary school. Discussion Paper No.: 09–074. ZEW: Centre for European Economic Research.
- Sprietsma, M. 2013. Discrimination in grading: Experimental evidence from primary school teachers. *Empirical Economics* 45, no. 1: 523–538.
- Sprietsma, M. 2009. Discrimination in grading? Experimental evidence from primary school. Discussion Paper No.: 09–074. ZEW: Centre for European Economic Research.
- Steinpreis, R., K. Anders, and D. Ritzke. 1999. The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex Roles* 41, no. 7: 509–528.
- Takiff, H., D. Sanchez, and T. Stewart. 2001. What's in a name? The status implications of students' terms of address for male and female professors. *Psychology of Women Quarterly* 25, no. 2: 134–144.
- Tversky, A., and D. Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185: 1124–1131.
- West, S., and T. Shults. 1976. Liking for common and uncommon first names. *Personality and Social Psychology Bulletin* 2, no. 3: 299–302.
- Widner, D., and S. Chicoine. 2011. It's all in the name: Employment discrimination against Arab Americans. *Sociological Forum* 26, no. 4: 806–823.

Wilson, S. 1995. Not tension but intention: A response to wong's analysis of researcher/teacher. *Educational Researcher* 24, no. 8: 19–22.

Zick, A., B. Küpper, and A. Hövermann. 2011. *Intolerance, prejudice, and discrimination: A European report*. Berlin: Friedrich Ebert Stiftung.

## Notes on contributor

Iman Nick is a US American sociolinguist in Germany. She holds a BA in Germanic languages and literature; a BSc in clinical and social psychology; an MA in German; an MSc in forensic psychology; and a PhD and German “Habilitation” in linguistics. She is the current President of the American Name Society.

Email: mavi.yaz@web.de