

Are the results of questionnaires measuring non-cognitive characteristics during the selection procedure for medical school application biased by social desirability?

Abstract

Introduction: A stronger consideration of non-cognitive characteristics in Medical School application procedures is desirable. Psychometric tests could be used as an economic supplement to face-to-face interviews which are frequently conducted during university internal procedures for Medical School applications (AdH, Auswahlverfahren der Hochschulen). This study investigates whether the results of psychometric questionnaires measuring non-cognitive characteristics such as personality traits, empathy, and resilience towards stress are vulnerable to distortions of social desirability when used in the context of selection procedures at Medical Schools.

Methods: This study took place during the AdH of Lübeck University in August 2015. The following questionnaires have been included: NEO-FFI, SPF, and AVEM. In a 2x1 between-subject experiment we compared the answers from an alleged application condition and a control condition. In the alleged application condition we told applicants that these questionnaires were part of the application procedure. In the control condition applicants were informed about the study prior to completing the questionnaires.

Results: All included questionnaires showed differences which can be regarded as social-desirability effects. These differences did not affect the entire scales but, rather, single subscales.

Conclusion: These results challenge the informative value of these questionnaires when used for Medical School application procedures. Future studies may investigate the extent to which the differences influence the actual selection of applicants and what implications can be drawn from them for the use of psychometric questionnaires as part of study-place allocation procedures at Medical Schools.

Keywords: Medical Education, College Admission Test, Psychometrics, Personality Inventory, Social Desirability

Katrin U. Obst¹

Linda Brüheim²

Jürgen Westermann³

Alexander Katalinic¹

Thomas Kötter¹

1 University of Lübeck, Institute of Social Medicine and Epidemiology, Lübeck Germany

2 University of Lübeck, Department of Quality Management and Organizational Development, Lübeck Germany

3 University of Lübeck, Institute of Anatomy, Lübeck Germany

1. Introduction

For many years, there have been up to five graduates applying for each place available at Medical School [1]. So far, Abitur (final exam) grades were considered the main criterion for study-place allocation. Taking additional selection criteria into account, such as clinical or non-cognitive characteristics, universities allocate up to 60% of their study places via university internal procedures for Medical School applications (AdH, Auswahlverfahren der Hochschulen). However, even in the AdH, the Abitur grade is the most important criterion [2]. This state of affairs has been criticized by many stakeholders and a reform of these application procedures has been demanded [3], [4], [<http://www.aerzteblatt.de/nachrichten/sw/Studienplatzvergabe?nid=63689>].

Although the Abitur grade predicts study grades as well as whether the standard period of study can be main-

tained by individuals [5], [6], [7], using this as the main criterion is problematic because it does not tell us anything about the characteristics that define a “good physician”, including clinical skills, empathy, resilience to stress, and certain personality traits [8]. These characteristics are desirable for faculty and society. In fact, studies show that these characteristics predict study success on the one hand and, on the other hand, who may become a “good physician” [9], [10]. Voltmer et al. show that work-related behavior and experience patterns have an impact on the health of medical students and young physicians [11], [12], [13]. Health is a basic prerequisite for the professional and empathetic treatment of patients [14], [15]. Physicians’ empathy appears to enhance patient satisfaction and also influences the correctness of diagnoses and treatment success in a positive way [16]. Furthermore, certain personality traits, such as *openness to experience* or *agreeableness*, are associated with the

ability to empathize [17] and linked directly to patient satisfaction [18] and study success [19].

To identify non-cognitive characteristics at some German universities, including Lübeck University, face-to-face interviews are an integral part of the AdH (e.g., [20], [21], [22]). These guideline-based interviews are conducted by a commission. However, they are resource-consuming and doubtful in their validity [23]. For these reasons, the use of supplementary validated questionnaires is advised [24]. Experience with the use of questionnaires to measure non-cognitive characteristics during the AdH is rare. These questionnaires are, however, well validated in the contexts of research studies [25], [26], [27]. It remains open as to whether such questionnaires are influenced by social desirability when answered in the context of Medical School application procedures [28], [29]. In this case answers to the questionnaires would have little meaning and would not be of much help in deciding to whom a study place should be given. Therefore, before AdH procedures are supplemented by questionnaires to measure non-cognitive characteristics, these should be tested for distortions of social desirability. Low distortions of social desirability are a prerequisite for using such instruments during AdH procedures.

On the basis of the aforementioned studies discussing which characteristics predict who may become a “good physician”, this study concentrates on measures of empathy, personality traits, and work-related behavior and experience patterns. The main question of this study is: Are questionnaires which measure these non-cognitive characteristics prone to social desirability distortion?

We expect that applicants of the AdH show enhanced tendencies to present themselves in a socially-desirable way and in accordance with what they think characterizes a “good physician”.

2. Methods

2.1 Participants

This study investigated applicants for Medical School at Lübeck University who participated in the face-to-face interviews which took place as part of the AdH 2015. Overall, 240 applicants were invited to the interviews in 2015; 228 accepted the invitation. 226 of these were included in our study (see table 1) and two did not participate.

2.2 Employed instruments

The following questionnaires have been tested for the applicability in the AdH.

SPF

The SPF [25] is a German translation and revision of the Davis Interpersonal Reactivity Index [30]. Empathy is measured through four facets:

1. Perspective taking (ability to adopt another’s point of view),
2. fantasy (ability to empathize with fictional characters),
3. empathic concern (ability for other-oriented emotions, such as pity), and
4. personal distress (self-oriented emotions such as uneasiness, which may occur in close or problematic interpersonal interactions).

Being a self-oriented emotion, personal distress is not considered in the empathy score.

For the medical context, the Jefferson Scale of Physician Empathy (JSPE [31]) might be more common. However, its wording makes it easy to guess what it measures, which makes it particularly vulnerable to social-desirability effects. This is why we decided to use a less obvious measurement of empathy, even though this measure is not customized to the medical context.

NEO-FFI

The NEO-FFI by Costa & McCrae (German translation by Borkenau & Ostendorf [26]) is a personality questionnaire which measures personality according to the Big Five Model. Personality is described by five independent factors: neuroticism, extraversion, openness to experience, conscientiousness, and agreeableness.

Neuroticism comprises characteristics such as nervousness and uncertainty, but also inappropriate reactions to stress. *Extraversion* describes, among other characteristics, the extent of the sociability and optimism a person has. *Openness to experience* measures to what extent new experiences or changes are welcomed as well as creativity and a thirst for knowledge. *Agreeableness* includes social emotion and characteristics such as altruism, faith, cooperation, and softness. *Conscientiousness* comprises characteristics such as orderliness, reliability, and discipline.

Individuals are asked to indicate their agreement or disagreement with 60 items on a five-point Likert Scale. Means have been calculated for the items of each factor.

AVEM

The AVEM [27] is a multidimensional personality-diagnostics procedure which measures behavior and experiences with regard to job-related stress and its influences on health. Self-reports are gathered on a five-point Likert Scale according to the following 11 dimensions:

1. *subjective significance of work;*
2. *career ambition;*
3. *tendency to exert;*
4. *striving for perfection;*
5. *emotional distancing;*
6. *resignation tendencies;*
7. *offensive coping with problems;*
8. *balance and mental stability;*

9. *satisfaction with work*;
10. *satisfaction with life*;
11. *experience of social support*.

Based on the interplay of these 11 dimensions, four profiles, or patterns of behavior and experiences can be derived by means of cluster analysis:

- Pattern G (“health”): high but not excessive work commitment, high resilience and positive emotion, including a high satisfaction with life and experience of social support.
- Pattern S (“unambitious”): reduced work commitment combined with a high ability to distance from work and a high satisfaction with life.
- Risk pattern A (“overexertion”): high work commitment but low resilience, satisfaction with life is impaired.
- Risk pattern B (“burnout”): low work commitment, especially low significance of work and low career ambition, combined with low ability to distance from work, high tendencies to resign, low satisfaction with life.

For this study we used the short version of AVEM with 44 items. Wordings were adapted to the students’ situation. All data collected in this study constitute self-reports and not objectively-measured characteristics.

2.3 Study design

Applicants were randomly administered to one of two experimental conditions (one factorial, experimental design). At each stage, 12 applicants had their application interview simultaneously. We assigned these 12 simultaneously-interviewed applicants to one of the two experimental conditions in an alternating fashion. They were then asked to complete a questionnaire at the computer. The allocation of interview time slots for each applicant was realized by means of computer-generated random numbers and this was conducted by the interview organization team.

Applicants in the first group (*alleged application condition*) were persuaded by instruction that their answers to the questionnaire would be considered in their application for Medical School in the event that the interview result was equivocal (excerpt of instruction: In the case of equivocal judgments of the commission, the results of this questionnaire will help us to allocate the study places among the invited applicants in a fair way). Study information was given immediately after completion of the questionnaire.

Applicants in the control condition (*study condition*) were informed about the study and its aim prior to completing the questionnaire (excerpt of instruction: We would like to invite you to complete a set of questions which will help us to optimize study-place allocation procedures in the future).

Participants of this *study condition* had not been informed about the aim to test the questionnaires for social-desirability influences. However, it was explicitly pointed out that the answers to the questionnaire had no influence on interview results or their application to Medical School.

This experimental design allows comparison of the questionnaire answers under a condition of personnel selection, which decides the future life of the applicants with regard to the answers of the same questionnaires used in the common research condition. As such, we wanted to test whether, and to what extent, answers to our selected questionnaires might be changed by a personnel selection situation and, thus, whether our questionnaires were prone to social-desirability effects [28], [29].

By means of an additional scale for social desirability (SES-17, [32]), we tested whether the *alleged application condition* enhanced social-desirability tendency in general compared to the *study condition*. This scale functions as a control for the experimental manipulation.

At the end of the questionnaire, applicants had the opportunity to make comments about the study. These comments are included in the data analysis.

2.4 Analysis

Questionnaire data were imported into and analyzed with the statistic software IBM SPSS Statistics (version 22). All group comparisons were statistically controlled for age, gender, and Abitur grade. To control for alpha cumulation due to multiple testing, data were Bonferroni-corrected for each questionnaire.

2.5 Ethics

To maintain the deception, we could not inform applicants in the *alleged application condition* about the voluntariness of their participation in this study in advance. Immediately after completion of the questionnaire, however, applicants were informed, orally and in written form, about the study and its aims. Additionally, we sought their written consent. In the case of no written consent we withdrew the data set. However, all participants who completed the questionnaire also gave their written consent to the study. The procedure was discussed and authorized by the Ethical Committee of the University of Lübeck (file reference: 15-072).

3. Results

Firstly, we tested whether applicants in the *alleged application condition* showed higher scores for the social-desirability scale (SES-17) than applicants in the *study condition*. We could not show a statistically significant difference between the two experimental conditions (*study condition*: $M=13.11$, $SD=2.78$ vs. *alleged application condition*: $M=13.42$, $SD=2.42$, $F[1,220]=0.53$, n.s.). Nevertheless, there were differences between the two experimental conditions for all questionnaires that might be regarded as social-desirability effects (see Figure 1, Figure 2, Figure 3 and Table 2).

Accordingly, the empathy score differs significantly. Applicants of the *alleged application condition* seem to be

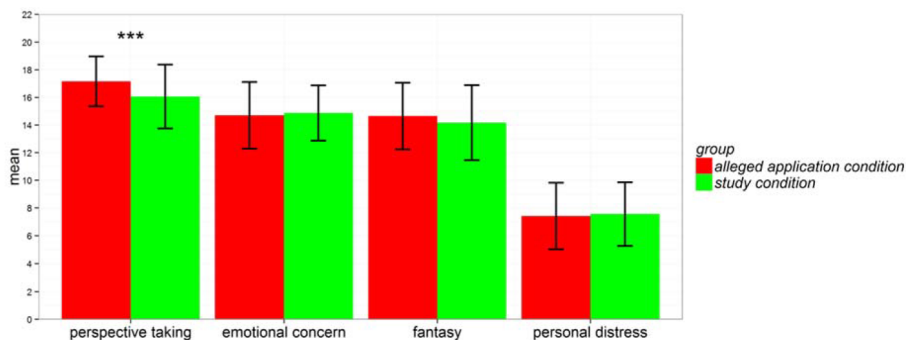
Table 1: Socio-demographic data and Abitur grade

	study condition	alleged application condition	total
sex (m [%]/w [%])	41 (36%)/73 (64%)	37 (33%)/77 (67%)	78 (34%)/150 (66%)
M age (SD)	20.93 (2.41)	20.91 (2.62)	20.92 (2.51)
M Abitur grade (SD)	1.54 (0.23)	1.52 (0.24)	1.53 (0.23)

Table 2: Statistical characteristics for significant differences in empathy, personality traits, and work-related behavior and experiences

	study condition	alleged application condition	F	df	p	dCohen
empathy (SPF)						
M empathy score (SD)	45.11 (5.19)	46.51 (4.70)	3.91	1,220	.05	.26
M perspective taking (SD)	16.07 (2.31)	17.16 (1.80)	14.94	1,220	<.01*	.26
personality (NEO-FFI)						
M neuroticism (SD)	2.34 (0.29)	2.43 (0.24)	6.14	1,220	.01*	.33
M agreeableness (SD)	2.12 (0.32)	2.21 (0.26)	4.49	1,220	.04	.28
dimensions of work related behavior and experience (AVEM)						
M subjective significance of work (SD)	14.61 (2.81)	15.29 (2.30)	4.57	1,219	.03	.29
M career ambition (SD)	16.09 (2.36)	16.69 (1.89)	4.54	1,217	.03	.29
M offensive coping with problems (SD)	16.25 (2.25)	17.44 (2.08)	17.66	1,217	<.01*	.57
M balance and mental stability (SD)	15.67 (2.59)	16.30 (2.33)	3.99	1,216	.05	.27

M=Mean; SD=standard deviation
 SPF and AVEM dimensions: sums; NEO-FFI: mean
 *significant with Bonferroni corrected level of significance: NEO-FFI $\alpha=0.01$; AVEM $\alpha=0.005$; SPF $\alpha=0.01$



***Bonferroni-corrected significant

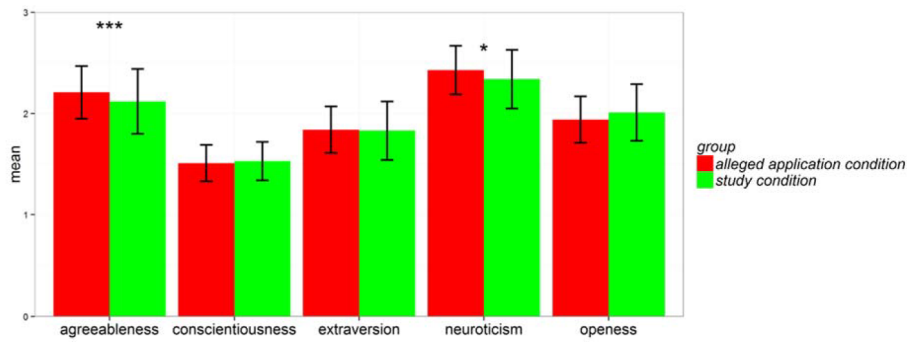
Figure 1: Mean comparison of the empathy scale SPF for the two experimental conditions: *alleged application condition* vs. *study condition*

significantly more empathic than applicants of the *study condition*. Examining the facets of the SPF more closely it becomes obvious that this difference relies mostly on a significant difference in the *perspective-taking* facet (see Figure 1).

Moreover, applicants of the *alleged application condition* show higher scores for the factors of neuroticism and agreeableness of the NEO-FFI compared to the *study*

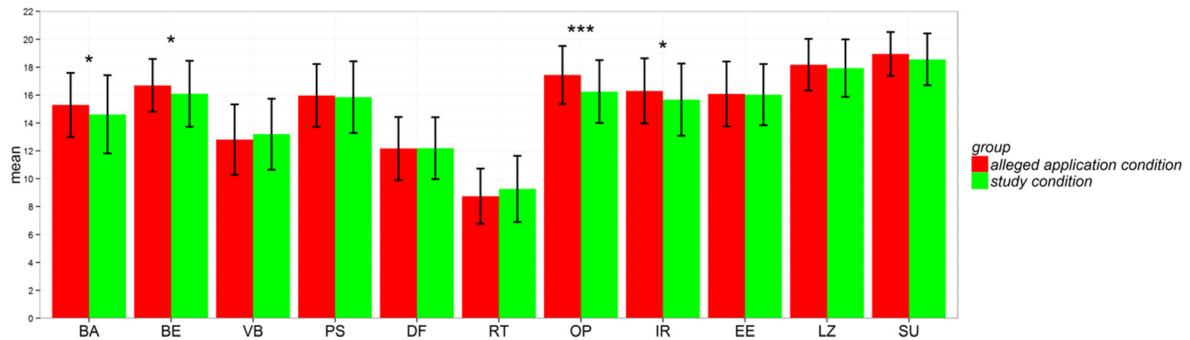
condition. The difference in the factor of agreeableness, however, did not withstand Bonferroni correction (see Figure 2).

The 11 dimensions of the AVEM revealed the following results (see Figure 3): applicants of the *alleged application condition* showed higher scores for *offensive coping with problems*. There were also higher scores for *subjective significance of work*, *career ambition*, and *balance*



*uncorrected significant; ***Bonferroni-corrected significant

Figure 2: Mean comparison of the NEO-FFI for the two experimental conditions: *alleged application condition* vs. *study condition*



* uncorrected significant; *** Bonferroni-corrected significant

BA=Subjective significance of work, BE=Career ambition, VB=Tendency to exert, PS=Striving for perfection, DF=Emotional distancing, RT=Resignation tendencies, OP=Offensive coping with problems, IR=Balance and mental stability, EE=Satisfaction with work, LZ=Satisfaction with life, SU=Experience of social support

Figure 3: Mean comparison of the AVEM for the two experimental conditions: *alleged application condition* vs. *study condition*

and mental stability. However, the last three dimensions reached significance only uncorrected. Remarkably, these differences in the AVEM-dimensions' results showed a three-times higher probability of reaching the AVEM pattern G for applicants in the *alleged application condition* compared to applicants in the *study condition* ($OR=3.15$, $WALD[1]=5.35$, $p=.03$).

Since the other AVEM patterns were relatively rare (pattern S: $n=4$; risk-pattern A: $n=19$; risk-pattern B: $n=2$), we did not analyze them.

Taken together, we can show differences between the experimental conditions for each scale. These differences have small-to-medium effect sizes (see Table 2).

Expectations about distortion effects are also evident in some of the comments applicants made during the evaluation of the interviews at the end of the day: in total, 20 comments referred to the study. From these, eight comments referred to general aspects, such as appreciation of such studies on the one hand or doubts about their informative values on the other. Four applicants from the *alleged application condition* did not believe that the questionnaire was part of the Medical School Application procedure. Six comments referred to aspects of social desirability. In fact, applicants in both experimental conditions pointed out that questionnaires might be distorted by social desirability when used in the context of study-place allocation (see Table 3). Even though these

six comments were not representative, they correspond neatly with the direction of differences that we reported in all questionnaires for our two experimental conditions.

4. Discussion

This study tested whether selected questionnaires for measuring personality traits, including empathy and coping with stress, are vulnerable to social-desirability effects when used in the framework of Medical School applications.

Our results show that the scores of all selected questionnaires are indeed distorted when used in the context of Medical School applications. Even after the Bonferroni correction for multiple testing, there were statistically-significant differences in the subscales of all implemented questionnaires. These differences had small-to-medium effect sizes [33]. Moreover, the AVEM differences appear to cumulate, which resulted in a three-times higher probability of reaching the AVEM pattern G for applicants in the *alleged application condition* compared to the *study condition*. This difference is remarkable in and of itself. The use of personality questionnaires in the framework of personnel selection is controversially discussed in the fields of Personality and Organizational Psychology. Studies in which participants are asked explicitly to present themselves in the best possible fashion (fake

Table 3: Comments from the evaluation of the interviews referring to the study

<i>study condition</i>
<p>„Das war eine ziemlich gute Erfahrung, ich bin aber der Meinung, solche Fragebögen sollten bei der Auswahl nicht berücksichtigt werden, da verfälschte Antworten gegeben werden.“ [It was a pretty good experience. However, I believe that such a questionnaire should not be taken into account in the selection procedures because answers are distorted]</p> <p>„Der Fragebogen lässt sich leicht durch idealisierte Angaben manipulieren und ist für die Auswahl der Teilnehmer unfair.“ [The questionnaire is easily manipulated by idealized answers and, thus, is not fair to the applicants]</p> <p>„Ich denke, dass 15 Minuten längere Auswahlgespräche mehr aussagen als die Studie. Wenn es um einen Medizinstudienplatz geht, glaube ich, ist jeder bereit zu schwindeln.“ [I think that 15 minutes more interview time says more than this study. Where it concerns a study place at Medical School, everybody is ready to deceive]</p>
<i>alleged application condition</i>
<p>„Ich denke, dass die Bewerber/ Studienteilnehmer intuitiv das ankreuzen, von dem sie ausgehen, was das "Gegenüber" hören will, vor allem, wenn es Teil des Auswahlverfahrens ist.“ [I think that applicants make theirs crosses intuitively as to where they think the other side wants them to be, especially when it comes to a place at Medical School]</p> <p>„Ich hätte den Fragebogen komplett anders ausgefüllt, wenn ich ihn für ein Aufnahmekriterium gehalten hätte.“ [I would have completed the questionnaire in a completely different way if I had thought it was one of the admission criteria]</p> <p>„Teilweise sind die Fragen doch sehr offensichtlich und die Antwortmöglichkeiten dafür zu pauschal. Evtl. etwas verschlüsselte Fragen einbringen.“ [Partly, the questions are very obvious and the answer options too general. Maybe more encrypted questions would be useful]</p>

good paradigms) show expectation- and stereotype-conforming behavior [34], [35], [36], [37], [38]. According to these studies, personality tests do indeed seem prone to social-desirability effects. However, other studies show only few effects of social desirability on personality tests [39], [40], [41]. Especially when participants are put in seemingly real personnel-selection procedures, the effects of social desirability seem to be rather small [40]. This might explain the small effect sizes reported in this study. Being a paradigm that does not involve faking goodness may also explain why applicants seem to represent themselves as more neurotic in the *alleged application condition* compared to the *study condition*. Applicants tried to represent themselves in accordance with what a “good” student and a “good” physician might be. This is why applicants probably tried to avoid representing themselves as too self-confident, especially considering the fact that self-reflection is part of the picture of an ideal physician [42].

Regarding empathy questionnaires or the AVEM, the authors are not aware of comparable studies investigating social desirability.

Since the number of cases (226) is rather small to reliably detect differences in this range of effect sizes, this study may rather underestimate the influence of social desirability on the selected questionnaires [43]. Thus, although the reported differences might be rather occasional and

small at first glance, based on our data we cannot assume that these questionnaires measure personality traits, empathy and stress resilience in a valid way.

However, there is one crucial limitation. Although we can show differences between the two experimental conditions for all selected questionnaires we cannot show a significant difference between the experimental conditions on the social desirability scale. This means that the distortions we could show for the other questionnaires might not be due to social desirability at all. Regarding our data, we could not rule out this limitation. The SES-17 [32] scale used in this study measures social desirability by very global everyday situations that have little to do with Medical School applications or personnel selection in general. Social desirability, on the other hand, describes a behavioral tendency to conform as best as possible within situation-specific norms and expectations [29]. Thus, it is possible that the Medical School application procedure triggers social desirability in a more context-specific way for which the SES-17 scale might not be sensitive enough.

Moreover, a stronger wording of the instruction text in the *alleged application condition* (see Methods) might have enhanced differences between the experimental conditions. Regarding their direction, the differences do indicate that applicants of the *alleged application condition* strived to represent themselves in terms of an ideal

student and ideal physician. This interpretation is also supported by the open-text comments of the evaluation of the Medical School application interviews. Taken together, this indicates that the SES-17 scale was probably not an appropriate control scale.

Overall, the results reported in this study show that psychometric questionnaires should be used with caution for Medical School application procedures, although their use might be highly desirable. We recommend further evaluating and probably re-editing psychometric questionnaires to reduce their vulnerability for social-desirability effects. Whether the differences shown here affect the actual selection procedure remains an open question. For example, it is conceivable that applicants with high tendencies of social desirability might reach higher ranks and thus have a higher probability in gaining the desired study place. Furthermore, it remains open as to what implications this could have regarding the matter of how to select those applicants who have the highest probability of becoming “good” physicians [44]. The results reported here make clear that research in this field is desperately needed regarding when psychometric questionnaires should be used in Medical School application procedures.

5. Conclusions

- Using psychometric questionnaires for Medical School application procedures is not problem-free: questionnaires tested in this study show distortions of social desirability. However, if answers are influenced by a desire to represent themselves in terms of the ideal student or physician, the answers lose their predictive value and, thus, are of no use for study-place allocation.
- Future research is required to evaluate the actual impact of those social-desirability effects and to reveal alternative procedures.

Acknowledgements

We thank Jessica Lückert, Sophia Marie Saftien, Karl Böse, Karen Sievers, and Josefin Wagner for their support in the preparation and implementation of this study.

Competing interests

The authors declare, that they have no competing interests.

As part of her job role at Lübeck University, Linda Brüheim was involved in the analyses of the Medical School application procedures (AdH).

The study was funded by Lübeck University, Medical School Lübeck and the Institute of Social Medicine and Epidemiology, Lübeck University.

References

1. Stiftung für Hochschulzulassung. Daten der bundesweit zulassungsbeschränkten Studiengänge an Hochschulen. Dortmund: Stiftung für Hochschulzulassung; 2014. Zugänglich unter/available from: <http://bit.ly/12Vo1xR>
2. Stiftung für Hochschulzulassung. Verordnung über die zentrale Vergabe von Studienplätzen durch die Stiftung für Hochschulzulassung. Dortmund: Stiftung für Hochschulzulassung; 2014. Zugänglich unter/available from: <http://bitly.com/1wStw9w>
3. Bundesvertretung der Medizinstudierenden in Deutschland e.V. Positionspapier Zulassungsverfahren zum Medizinstudium in Deutschland. Bonn: Bundesvertretung der Medizinstudierenden in Deutschland; 2015. Zugänglich unter/available from: <http://bit.ly/1lvhLgh>
4. Bundesärztekammer. Stellungnahme der Bundesärztekammer zu den Themen für den "Masterplan Medizinstudium 2020", insbesondere hinsichtlich möglicher Maßnahmenvorschläge. Berlin: Bundesärztekammer; 2015. Zugänglich unter/available from: <http://bit.ly/2dlRE0u>
5. Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Health Sci Educ Theory Pract.* 2001;6(2):159-175. DOI: 10.1023/A:1011489618208
6. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ.* 2002;324(7343):952-957. DOI: 10.1136/bmj.324.7343.952
7. Maslov Kruzicevic S, Barisic KJ, Banozic A, Esteban CD, Sapunar D, Puljak L. Predictors of attrition and academic success of medical students: a 30-year retrospective study. *PLoS ONE.* 2012;7(6):e39144. DOI: 10.1371/journal.pone.0039144
8. Shen H, Comrey AL. Predicting medical students' academic performances by their cognitive abilities and personality characteristics. *Acad Med.* 1997;72(9):781-786. DOI: 10.1097/00001888-199709000-00013
9. Hojat M, Robeson M, Damjanov I, Veloski JJ, Glaser K, Gonnella JS. Students' psychosocial characteristics as predictors of academic performance in medical school. *Acad Med.* 1993;68(8):635-637. DOI: 10.1097/00001888-199308000-00015
10. Adam J, Bore M, McKendree J, Munro D, Powis D. Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Med Educ.* 2012;12:69. DOI: 10.1186/1472-6920-12-69
11. Voltmer E, Kötter T, Spahn C. Perceived Medical School Stress and the Development of Behavior and Experience Patterns in German Medical Students. *Med Teach.* 2012;34(10):840-847. DOI: 10.3109/0142159X.2012.706339
12. Voltmer E, Kieschke U, Spahn C. Arbeitsbezogenes Verhalten und Erleben bei Ärzten im dritten bis achten Berufsjahr. *Z Psychosom Med Psychother.* 2007;53(3):244-257. DOI: 10.13109/zptm.2007.53.3.244
13. Kötter T, Tautphäus Y, Obst KU, Voltmer E, Scherer M. Health-promoting factors in the freshman year of medical school: a longitudinal study. *Med Educ.* 2016;50(6):646-656. DOI: 10.1111/medu.12987
14. Kjeldstadli K, Tyssen R, Finset A, Hem E, Gude T, Gronvold NT, Ekeberg O, Vaglum P. Life satisfaction and resilience in medical school - a six-year longitudinal, nationwide and comparative study. *BMC Med Educ.* 2006;6:48. DOI: 10.1186/1472-6920-6-48

15. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet*. 2009;374(9702):1714–1721. DOI: 10.1016/S0140-6736(09)61424-0
16. Neumann M, Scheffer C, Tauschel D, Lutz G, Wirtz M, Edelhäuser F. Physician empathy: definition, outcome-relevance and its measurement in patient care and medical education. *GMS Z Med Ausbild*. 2012;29(1):Doc11. DOI: 10.3205/zma000781
17. Costa P, Alves R, Neto I, Marvão P, Portela M, Costa MJ. Associations between medical student empathy and personality: a multi-institutional study. *PLoS ONE*. 2014;9(3):e89254. DOI: 10.1371/journal.pone.0089254
18. Duberstein P, Meldrum S, Fiscella K, Shields CG, Epstein RM. Influences on patients' ratings of physicians: Physicians demographics and personality. *Patient Educ Couns*. 2007;65(2):270–274. DOI: 10.1016/j.pec.2006.09.007
19. Lievens F, Coetsier P, De Fruyt F, De Maesseneer J. Medical students' personality characteristics and academic performance: a five-factor model perspective. *Med Educ*. 2002;36(11):1050–1056. DOI: 10.1046/j.1365-2923.2002.01328.x
20. Hallfahrt T, Reinke S, Westermann J. Exzellenz und Engagement gesucht. *Dtsch Arztebl*. 2009;106:A1296–A1298.
21. Waydhas C, Heue M, Nast-Kolb D. Strukturierte Interviews zur Studienplatzvergabe in der Medizin: Erste Erfahrungen. *GMS Z Med Ausbild*. 2007;24(4):Doc186. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2007-24/zma000480.shtml>
22. Hampe W, Klusmann D, Buhk H, Münch-Harrach D, Harendza S. Reduzierbarkeit der Abbrecherquote im Humanmedizinstudium durch das Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftsteil (HAM-Nat). *GMS Z Med Ausbild*. 2008;25(2):Doc82. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000566.shtml>
23. Kötter T, Obst K, Brüheim L, Eisemann N, Voltmer E, Katalinic A. Können psychometrische Tests den Erfolg im Auswahlgespräch zum Medizinstudium vorhersagen? Eine Querschnittsstudie an einer deutschen Hochschule. *Gesundheitswes*. 2015., DOI: 10.1055/s-0035-1554705
24. Wissenschaftsrat. Empfehlungen zur Reform des Hochschulzugangs. Berlin: Wissenschaftsrat; 2004. Zugänglich unter/available from: <http://bit.ly/1Z20sry>
25. Paulus C. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. Saarbrücken: Universität des Saarlandes; 2009. Zugänglich unter/available from: <http://bit.ly/1kHRCmV>
26. Borkenau P, Ostendorf F. NEO-Fünf-Faktoren-Inventar. 2., neu normierte und vollständig überarbeitete Auflage. Göttingen: Hogrefe; 2008.
27. Schaarschmidt U, Fischer AW. Arbeitsbezogenes Verhaltens- und Erlebensmuster. 2. Auflage. Frankfurt am Main: Swets Test Services; 2003.
28. Edwards AL. The social desirability variable in personality assessment and research. Oak Brook: The Dryden Press; 1957.
29. Mummendey HD. Methoden und Probleme der Kontrolle sozialer Erwünschtheit (Social Desirability). *Z Different Diagn Psychol*. 1981;2:199-218. Zugänglich unter/available from: <http://bit.ly/1TJtnn3>
30. Davis MH. A multidimensional approach to individual differences in empathy. Austin: University of Texas; 1980. Zugänglich unter/available from: <http://bit.ly/22IFq7w>
31. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. *Am J Psychiatry*. 2002;159(9):1563–1569. DOI: 10.1176/appi.ajp.159.9.1563
32. Stöber J. Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität. *Diagnostica*. 1999;45:173-177. DOI: 10.1026//0012-1924.45.4.173
33. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2. Auflage. Hillsdale, N.J.: Lawrence Erlbaum Assoc Inc; 1988.
34. Viswesvaran C, Ones DS. Meta-analyses of fakability estimates: Implications for personality measurement. *Educ Psychol Meas*. 1999;59:197–210. DOI: 10.1177/00131649921969802
35. Alliger GM, Dwight SA. A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educ Psychol Meas*. 2000;60:59–72. DOI: 10.1177/00131640021970367
36. Krahé B, Herrmann J. Verfälschungstendenzen im NEO-FFI: Eine experimentelle Überprüfung. *Z Different Diagn Psychol*. 2003;24:105–117. DOI: 10.1024//0170-1789.24.2.105
37. Hagemann D, Amelang M, Bartussek D, Stemmler G. *Differentielle Psychologie und Persönlichkeitsforschung*. 6. vollst. überarb. Aufl. Stuttgart: Kohlhammer; 2006.
38. Galic Z, Jerneic Ž, Kovacic MP. Do applicants fake their personality questionnaire responses and how successful are their attempts? A case of military pilot cadet selection. *Int J Select Assess*. 2012;20:229–241. DOI: 10.1111/j.1468-2389.2012.00595.x
39. Ones DS, Viswesvaran C. The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection. *Human Perform*. 1998;11:245–269. DOI: 10.1080/08959285.1998.9668033
40. Hossiep R, Paschen M, Mühlhaus O. *Persönlichkeitstests im Personalmanagement: Grundlagen, Instrumente und Anwendungen*. Göttingen: Hogrefe Verlag; 2000.
41. Amelang M, Schäfer A, Yousfi S. Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of fake good instructions. *Psychol Beitr*. 2002;44:24–41.
42. Hibbeler B. Zwischen Samaritertum und Ökonomie: Was ist ein "guter Arzt"? *Dtsch Arztebl*. 2011;108:A2758.
43. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175–191. DOI: 10.3758/BF03193146
44. Dwight SA, Donovan JJ. *Warning: Proceed with caution when warning applicants not to dissimulate*. Dallas, TX: Society of Industrial and Organizational Psychology; 1998.
45. Schmitt N, Oswald FL, Kim BH, Gillespie MA, Ramsay LJ, Yoo TY. Impact of elaboration on socially desirable responding and the validity of biodata measures. *J Appl Psychol*. 2003;88(6):979–988. DOI: 10.1037/0021-9010.88.6.979
46. Marcus B. *Persönlichkeitstests in der Personalauswahl*. *Z Psychol*. 2003;211:138–148. DOI: 10.1026//0044-3409.211.3.138

Corresponding author:

Dr. med. Thomas Kötter, MPH
 University of Lübeck, Institute of Social Medicine and
 Epidemiology, Ratzeburger Allee 160, D-23562 Lübeck
 Germany, Phone: +49 (0)451/500-51220
 thomas.koetter@uksh.de

Please cite as

Obst KU, Brüheim L, Westermann J, Katalinic A, Kötter T. Are the results of questionnaires measuring non-cognitive characteristics during the selection procedure for medical school application biased by social desirability? *GMS J Med Educ.* 2016;33(5):Doc75.
DOI: 10.3205/zma001074, URN: urn:nbn:de:0183-zma0010749

This article is freely available from

<http://www.egms.de/en/journals/zma/2016-33/zma001074.shtml>

Received: 2016-01-06

Revised: 2016-08-04

Accepted: 2016-08-12

Published: 2016-11-15

Copyright

©2016 Obst et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Werden die Ergebnisse von Fragebögen zur Messung nicht-kognitiver Eigenschaften im Rahmen des Auswahlverfahrens für Medizinstudienplätze durch soziale Erwünschtheit verzerrt?

Zusammenfassung

Zielsetzung: Eine stärkere Berücksichtigung nicht-kognitiver Eigenschaften in der Studienplatzvergabe im Fach Medizin wäre prinzipiell wünschenswert. Der Einsatz psychometrischer Fragebögen könnte eine ökonomische Ergänzung zum Auswahlgespräch, das vielerorts Teil des Auswahlverfahrens der Hochschulen (AdH) ist, darstellen. Diese Studie untersucht, inwiefern Ergebnisse psychometrischer Fragebögen zur Erfassung nicht-kognitiver Eigenschaften wie Persönlichkeit, Empathie und Stress-Resilienz anfällig für Verzerrungen durch sozial erwünschtes Antwortverhalten sind, wenn Sie im Kontext der Studienplatzvergabe im Fach Medizin eingesetzt werden.

Methodik: Die Studie fand im Rahmen des AdH der Universität zu Lübeck im August 2015 statt. Untersucht wurden folgende Fragebögen: NEO-FFI, SPF und AVEM. In einem experimentellen 2x1-Zwischensubjekt-Design wurde das Antwortverhalten zwischen einer vermeintlichen Auswahlbedingung und einer Kontrollbedingung verglichen. In der Auswahlbedingung wurde den Bewerberinnen und Bewerbern suggeriert, dass die Fragebögen Teil des Auswahlprozesses seien. In der Kontrollbedingung wurde im Vorhinein mitgeteilt, dass die Fragebögen Teil einer wissenschaftlichen Begleitstudie sind.

Ergebnisse: Für alle untersuchten Fragebögen weisen die Ergebnisse Unterschiede auf, die auf soziale Erwünschtheit hindeuten. Dabei sind nicht die gesamten Fragebögen betroffen, sondern einzelne Subskalen.

Schlussfolgerung: Die Aussagekraft der untersuchten Fragebögen für den Einsatz in der Auswahl-situation muss in Frage gestellt werden. Zukünftige Studien müssen untersuchen, inwiefern sich die hier gefundenen Unterschiede tatsächlich auf den Auswahlprozess auswirken und welche Implikationen sich daraus für deren Einsatz bei der Studienplatzvergabe im Fach Medizin ergeben.

Schlüsselwörter: Medizinische Ausbildung, Auswahlverfahren, psychometrische Testung, Persönlichkeitsinventar, Soziale Erwünschtheit

Katrin U. Obst¹

Linda Brüheim²

Jürgen Westermann³

Alexander Katalinic¹

Thomas Kötter¹

1 Universität zu Lübeck, Institut für Sozialmedizin und Epidemiologie, Lübeck, Deutschland

2 Universität zu Lübeck, Dezernat Qualitätsmanagement und Organisationsentwicklung, Lübeck, Deutschland

3 Universität zu Lübeck, Institut für Anatomie, Lübeck, Deutschland

1. Einleitung

Seit Jahren bewerben sich konstant etwa 5 Abiturientinnen und Abiturienten auf einen Medizinstudienplatz [1]. Bisher ist die Abiturnote bei der Studienplatzvergabe das Hauptkriterium. Um - in engen gesetzlichen Grenzen - auch Auswahlkriterien zur Anwendung kommen zu lassen, die stärker auf klinisch-praktische bzw. nicht kognitive Eigenschaften abzielen, werden an vielen Hochschulen bis zu 60% der Medizinstudienplätze im Rahmen des Auswahlverfahrens der Hochschulen (AdH) vergeben. Auch im AdH ist die Abiturnote allerdings das wichtigste Kriterium [2]. Dies wird von verschiedenen Interessen-

gruppen kritisiert, eine Reform des Zulassungsverfahrens gefordert [3], [4], [<http://www.aerzteblatt.de/nachrichten/sw/Studienplatzvergabe?nid=63689>].

Problematisch an der Abiturnote als Hauptauswahlkriterium ist, dass sie zwar die Einhaltung der Regelstudienzeit und gute Noten vorhersagt [5], [6], [7]. Eigenschaften, die einen guten Arzt/eine gute Ärztin ausmachen, wie etwa klinisch-praktische Fähigkeiten, Empathie, Stress-Resilienz und bestimmte Persönlichkeitseigenschaften werden durch die Abiturnote eher nicht prädiert [8]. Dabei zeigen Studien, dass auch solche für Fakultäten und Gesellschaft wünschenswerte nicht-kognitive Eigenschaften eine Vorhersagekraft für ein erfolgreiches Medizinstudium und für die Frage, wer ein guter Arzt/eine gute Ärztin wird, besitzen [9], [10]. So können Voltmer et

al. belegen, dass arbeitsbezogene Verhaltens- und Erlebensmuster einen bedeutenden Einfluss auf die Gesundheit von Medizinstudierenden und jungen Ärztinnen und Ärzten haben [11], [12], [13]. Diese ist wiederum Grundvoraussetzung für einen kompetenten und einfühlsamen Umgang mit Patienten [14], [15]. Ärztliche Empathie scheint nicht nur die Patientenzufriedenheit, sondern auch die Richtigkeit von Diagnosen und den Behandlungserfolg positiv zu beeinflussen [16]. Bestimmte Persönlichkeitseigenschaften, z.B. Offenheit für Erfahrungen und Verträglichkeit, sind wiederum mit Empathiefähigkeit [17] und auch direkt mit Patientenzufriedenheit [18] verknüpft, nicht zuletzt aber auch mit Studienerfolg [19].

An einigen deutschen Hochschulen, unter anderem in Lübeck, sind daher Auswahlgespräche zur Erfassung nicht-kognitiver Eigenschaften Teil des AdH (z.B. [20], [21], [22]). Diese meist anhand eines Leitfadens von einer Kommission durchgeführten Gespräche sind ressourcenintensiv und von unklarer Validität [23]. Unter anderem deswegen wird der ergänzende Einsatz validierter Fragebogeninstrumente empfohlen [24]. Die Erfahrungen mit einer fragebogengestützten Erfassung nicht-kognitiver Eigenschaften im Rahmen des AdH sind allerdings bisher äußerst begrenzt. Zwar gibt es Fragebögen, die sich im wissenschaftlichen Kontext bewährt haben [25], [26], [27]. Unklar ist jedoch, ob die mit diesen Fragebögen erzielten Ergebnisse in der Auswahl-situation durch sozial erwünschte Antworten beeinflusst werden [28], [29]. In diesem Fall wären die Antworten in den Fragebögen wenig aussagekräftig und kaum hilfreich in der Entscheidung um die Studienplatzvergabe. Bevor das AdH also durch fragebogengestützte Messungen nicht-kognitiver Eigenschaften ergänzt werden kann, muss deren Anfälligkeit für soziale Erwünschtheit überprüft werden. Eine geringe Anfälligkeit wäre die Voraussetzung dafür, dass diese als aussagekräftige Messinstrumente in das AdH aufgenommen werden könnten.

Aufbauend auf den zuvor genannten Studien zur Frage, welche Eigenschaften vorhersagen, wer am wahrscheinlichsten ein guter Arzt/eine gute Ärztin werden könnte, konzentrierte sich diese Studie daher zum einen auf Empathie und Persönlichkeitseigenschaften, zum anderen auf bestimmte Verhaltens- und Erlebensweisen im Zusammenhang mit Studium und Arbeit.

Dabei lautete die Fragestellung dieser Studie: Sind ausgewählte Fragebögen zur Erfassung der genannten nicht-kognitiven Eigenschaften im Rahmen des AdH anfällig für Verzerrungen durch soziale Erwünschtheit?

Es wurde erwartet, dass Bewerberinnen und Bewerber im Kontext des AdH eine erhöhte Tendenz aufweisen, sich im Sinne einer Tendenz zu sozial erwünschtem Antworten möglichst positiv im Sinne einer/eines guten Studierenden und angehenden Ärztin/Arztes darzustellen.

2. Methoden

2.1 Stichprobe

Untersucht werden sollten Medizinstudienplatzbewerberinnen und -bewerber, die 2015 am Auswahlgespräch im Rahmen des AdH der Universität zu Lübeck teilnahmen. Insgesamt wurden 2015 240 Bewerberinnen und Bewerber zu einem Auswahlgespräch eingeladen, 228 folgten dieser Einladung. Hiervon konnten 226 Bewerberinnen und Bewerber in die Studie eingeschlossen werden (siehe Tabelle 1), 2 nahmen nicht teil.

2.2 Verwendete Instrumente

Es wurden die folgenden Fragebögen auf ihre Einsatzfähigkeit im AdH überprüft:

SPF

Der SPF [25] ist eine Übersetzung und Überarbeitung des Interpersonal Reactivity Index von Davis [30]. Empathie wird hierbei durch vier Facetten konzeptualisiert:

1. *Perspective taking* (die Fähigkeit, die Sichtweise des Gegenübers einzunehmen),
2. *Fantasy* (die Fähigkeit, sich spontan in fiktive Figuren hineinzusetzen),
3. *Empathic concern* (die Fähigkeit zu fremdorientierten Emotionen wie z.B. Mitleid) und
4. *Personal distress* (eigenfokussierte Emotionen wie etwa Unruhe, die in engen oder problematischen interpersonellen Interaktionen auftreten können).

Als eigenfokussierte Emotion geht *Personal distress* nicht in die Berechnung des Gesamt-Empathiewertes mit ein. Zur Erfassung von Empathie im medizinischen Kontext ist die Jefferson Scale for Physician Empathy [31] möglicherweise geläufiger. Aus ihrer Itemformulierung ist ihr Zweck jedoch sehr leicht ableitbar, was sie stark anfällig für Tendenzen sozialer Erwünschtheit macht. Aus diesem Grund haben wir uns für eine weniger offensichtliche Messung von Empathie entschieden, auch wenn diese nicht spezifisch auf den medizinischen Kontext zugeschnitten ist.

NEO-FFI

Das NEO-FFI nach Costa & McCrae (deutsche Übersetzung von Borkenau & Ostendorf [26]) ist ein Persönlichkeitsfragebogen, der Persönlichkeit auf der Basis des Big-Five-Modells misst und Persönlichkeit anhand folgender fünf, voneinander unabhängiger Faktoren beschreibt: Neurotizismus, Extraversion, Offenheit für Erfahrungen, Gewissenhaftigkeit, Verträglichkeit.

Neurotizismus umfasst dabei Eigenschaften wie Nervosität und Unsicherheit, aber auch z.B. unangemessene Reaktionen auf Stress. *Extraversion* umschreibt u.a. das Ausmaß an Geselligkeit oder Optimismus. *Offenheit für Erfahrungen* erfragt inwiefern neue Erfahrungen oder

Tabelle 1: Soziodemographische Daten und Abiturnote

	Bedingung: Wissenschaftliche Studie	Bedingung: Vermeintliche Auswahlsituation	Gesamt
Geschlecht (m [%]/w [%])	41 (36%)/73 (64%)	37 (33%)/77 (67%)	78 (34%)/150 (66%)
M Alter (SD)	20.93 (2.41)	20.91 (2.62)	20.92 (2.51)
M Abiturnote (SD)	1.54 (0.23)	1.52 (0.24)	1.53 (0.23)

Abwechslung wertgeschätzt werden sowie das Ausmaß an Kreativität und Wissbegierde. *Verträglichkeit* erfasst soziale Emotionen und Eigenschaften wie Altruismus, Vertrauen, Kooperativität und Nachgiebigkeit. *Gewissenhaftigkeit* umschreibt Eigenschaften wie Ordnungsliebe, Zuverlässigkeit und Disziplin.

Auf 60 Items wird dazu die Zustimmung bzw. Ablehnung mittels einer 5 stufigen Likert-Skala erfragt. Über die Items eines Faktors wird ein Mittelwert berechnet.

AVEM

Der AVEM [27] ist ein mehrdimensionales, persönlichkeitsdiagnostisches Verfahren, mit dem Verhalten und Erleben im Hinblick auf berufliche Arbeitsbelastungen und ihr Einfluss auf Gesundheit erfasst werden. Dazu werden mittels einer 5-stufigen Likertskala Selbsteinschätzungen zu den folgenden 11 Dimensionen erfasst:

1. *Subjektive Bedeutsamkeit der Arbeit,*
2. *Beruflicher Ehrgeiz,*
3. *Verausgabungsbereitschaft,*
4. *Perfektionsstreben,*
5. *Distanzierungsfähigkeit,*
6. *Resignationstendenz bei Misserfolgen,*
7. *Offensive Problembewältigung,*
8. *Innere Ruhe und Ausgeglichenheit,*
9. *Erfolgs erleben im Beruf,*
10. *Lebenszufriedenheit und*
11. *Erleben sozialer Unterstützung.*

Aus dem Zusammenspiel der 11 Dimensionen können mittels Clusteranalyse vier verschiedene Profile bzw. Verhaltens- und Erlebensmuster abgeleitet werden:

- **Muster G („Gesundheit“):** hohes, aber nicht exzessives Arbeitsengagement bei gleichzeitig hoher Widerstandskraft und positiven Emotionen wie etwa Lebenszufriedenheit oder dem Erleben sozialer Unterstützung.
- **Muster S („Schutz / Schonung“):** vermindertes Arbeitsengagement bei gleichzeitig hoher Distanzierungsfähigkeit und Lebenszufriedenheit.
- **Muster A („Selbstüberforderung“):** hohes Arbeitsengagement bei niedriger Widerstandsfähigkeit und nur eingeschränkter Lebenszufriedenheit.
- **Muster B („Burnout“):** niedriges Arbeitsengagement, vor allem niedrige Bedeutsamkeit der Arbeit und niedriger beruflicher Ehrgeiz mit jedoch verminderter

Distanzierungsfähigkeit, einer hohen Resignationstendenz und einer niedrigen Lebenszufriedenheit.

Im Rahmen dieser Studie kam die auf die Situation von Studierenden angepasste Version mit 44 Items zum Einsatz.

Bei den mittels SPF, NEO-FFI und AVEM erhobenen Angaben handelt es sich um Selbstauskünfte, nicht um objektiv gemessene Eigenschaften.

2.3 Studiendesign

Die Studie umfasst zwei Versuchsbedingungen, denen die Bewerberinnen und Bewerber zufällig zugeordnet wurden (einfaktorielles experimentelles Design). Dabei wurden jeweils die 12 zeitgleich für das Auswahlgespräch eingeteilten Bewerberinnen und Bewerber abwechselnd einer der beiden im Folgenden beschriebenen Bedingungen zugeordnet und computergestützt direkt im Anschluss an das Auswahlgespräch befragt. Die Zuteilung der Zeiten war dabei im Vorfeld mittels computergenerierter Zufallszahlen durch das Organisationsteam der Auswahlgespräche erfolgt.

Der ersten Gruppe (Bedingung *vermeintliche Auswahlsituation*) wurde mittels des Instruktionstextes (Auszug: „Sollte das Urteil der Auswahlkommission nicht eindeutig sein, werden die Ergebnisse dieses Fragebogens uns dabei helfen, die Studienplätze möglichst gerecht auf die eingeladenen Bewerber(innen) zu verteilen.“) suggeriert, dass die Fragebögen Teil des Auswahlverfahrens seien und deren Ergebnisse in die Entscheidung über die Zulassung zum Medizinstudium einfließen würden. Eine Studienaufklärung über die Studie und ihre Ziele erfolgte durch die Versuchsleiterin direkt im Anschluss.

Die zweite Gruppe (Kontrollbedingung *wissenschaftliche Studie*) wurde vorab darüber informiert, dass es sich um eine wissenschaftliche Studie handelt (Auszug aus dem Instruktionstext: „Wir möchten Sie einladen, eine Reihe von Fragen zu beantworten und damit zur Verbesserung des Auswahlverfahrens in der Zukunft beizutragen.“), jedoch nicht darüber, dass soziale Erwünschtheit Gegenstand dieser Studie war. Dabei wurde explizit darauf hingewiesen, dass die Beantwortung der Fragen keinerlei Einfluss auf das Auswahlgespräch hat.

Mittels dieser Versuchsanordnung konnte die Beantwortung der Fragebögen unter der Bedingung einer für die weitere berufliche Biographie der Bewerberinnen und Bewerber entscheidende Auswahlsituation mit der Beant-

wortung derselben Fragebögen unter der üblichen Bedingung einer wissenschaftlichen Befragung verglichen werden. So sollte geprüft werden, ob und wenn ja, wie sich die Beantwortung der ausgewählten Fragebögen durch eine vermeintliche Auswahl-situation verändert, also ob sie anfällig für Tendenzen der sozialen Erwünschtheit ist [28], [29].

Mittels einer zusätzlich erhobenen Skala zur sozialen Erwünschtheit (SES-17 [32]) sollte getestet werden, inwiefern das Auswahlverfahren überhaupt die Tendenz zur sozialen Erwünschtheit gegenüber einer wissenschaftlichen Studie erhöht. Diese Skala dient als Kontrolle für die experimentelle Manipulation.

Am Ende des Fragebogens wurde den Bewerberinnen und Bewerbern die Gelegenheit gegeben, in einem Freitextfeld Kommentare zur Befragung abzugeben, die in die Auswertung mit einbezogen wurden.

2.4 Auswertung

Die Fragebogendaten wurden in die Statistiksoftware IBM SPSS Statistics (Version 22) importiert und mit dieser ausgewertet.

Bei allen Gruppenvergleichen wurde statistisch für Alter, Geschlecht und Abiturnote kontrolliert. Zur Kontrolle der α -Kumulation bei mehrfachem Testen wurde fragebogenweise eine Bonferroni-Korrektur vorgenommen.

2.5 Ethik

Um die Täuschung aufrecht zu erhalten, konnte in der Bedingung *vermeintliche Auswahl-situation* nicht vorab auf die Freiwilligkeit der Teilnahme hingewiesen werden. Direkt im Anschluss an die Befragung erfolgte daher die Aufklärung mündlich durch die Versuchsleiterin sowie zusätzlich in schriftlicher Form. Es wurde eine nachträgliche schriftliche Einverständniserklärung eingeholt. Bei fehlendem Einverständnis wären die Daten gelöscht worden. Alle Teilnehmerinnen und Teilnehmer, die einen Fragebogen ausgefüllt hatten, erklärten jedoch im Nachhinein ihr Einverständnis. Das Vorgehen war zuvor mit der Ethikkommission der Universität zu Lübeck beraten und positiv von dieser begutachtet worden (Aktenzeichen 15-072).

3. Ergebnisse

Es wurde zunächst geprüft, ob die Bewerberinnen und Bewerber der Bedingung *vermeintliche Auswahl-situation* höhere Werte auf der Skala zur sozialen Erwünschtheit (SES-17) aufweisen als die Bewerberinnen und Bewerber der Bedingung wissenschaftliche Studie. Ein statistischer Unterschied konnte nicht gezeigt werden (*wissenschaftliche Studie*: $M=13.11$, $SD=2.78$ vs. *vermeintliche Auswahl-situation*: $M=13.42$, $SD=2.42$, $F[1,220]=0.53$, n.s.). Dennoch zeigen sich einzelne Unterschiede in allen verwendeten Fragebögen zwischen den beiden Versuchsbedingungen, die eine Anfälligkeit für Verzerrungen nahele-

gen (siehe Abbildung 1, Abbildung 2, Abbildung 3 und Tabelle 2).

So weist der Gesamtwert der Empathieskala einen signifikanten Unterschied auf. Bewerberinnen und Bewerber in der Bedingung *vermeintliche Auswahl-situation* scheinen signifikant empathischer zu sein als in der Kontrollbedingung. Betrachtet man die einzelnen Facetten des SPFs, zeigt sich, dass dieser Unterschied primär auf einen signifikanten Unterschied in der Facette *Perspective taking* zurückgeführt werden kann (siehe Abbildung 1).

Weiterhin haben Bewerberinnen und Bewerber in der Bedingung *vermeintliche Auswahl-situation* höhere Werte für die Faktoren Neurotizismus und *Verträglichkeit* des Persönlichkeitsfragebogens NEO-FFI als in der Kontrollbedingung. Der Unterschied im Faktor *Verträglichkeit* ist jedoch nach Bonferroni-Korrektur für multiples Testen nicht mehr signifikant (siehe Abbildung 2).

Für die 11 Dimensionen des AVEM ergibt sich folgendes Bild (siehe Abbildung 3): Bewerberinnen und Bewerber in der Bedingung *vermeintliche Auswahl-situation* haben höhere Scores für *Offensive Problembewältigung*. Weiterhin erreichen sie höhere Scores für die Dimensionen *Subjektive Bedeutsamkeit der Arbeit*, *Beruflicher Ehrgeiz* und *Innere Ruhe/Ausgeglichenheit*. Die letzten drei Dimensionen sind jedoch nur unkorrigiert signifikant. Bemerkenswerterweise führen die Unterschiede in den Dimensionen des AVEM dazu, dass Bewerberinnen und Bewerber der Bedingung *vermeintliches Auswahlgespräch* eine etwa drei Mal höhere Wahrscheinlichkeit haben, dem Muster G des AVEM zugeordnet zu werden als Bewerberinnen und Bewerber der Bedingung wissenschaftliche Studie ($OR=3.15$, $WALD[1]=5.35$, $p=.03$). Da die übrigen Muster in dieser Stichprobe unabhängig von der experimentellen Manipulation relativ selten auftraten (Muster S: $n=4$; Risikomuster A: $n=19$; Risikomuster B: $n=2$), wurde auf ihre Auswertung verzichtet.

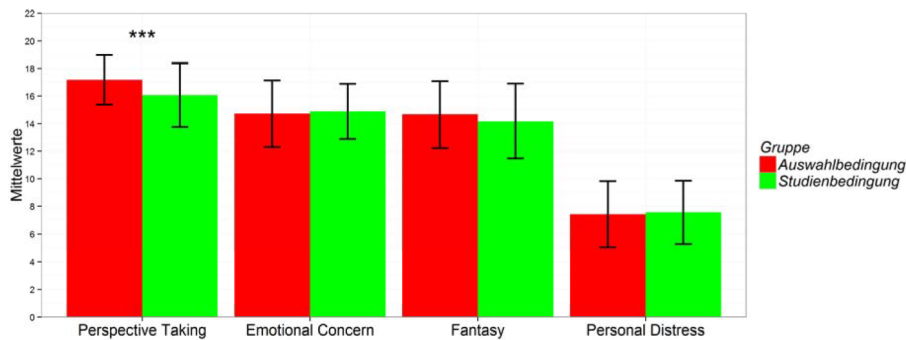
Es können also für jede der hier ausgewählten Skalen signifikante Unterschiede zwischen den beiden Versuchsbedingungen nachgewiesen werden. Diese Unterschiede weisen kleine bis mittlere Effektstärken auf (siehe Tabelle 2).

Die Erwartung verzerrter Antworten in den Fragebögen zeigt sich auch in einigen Kommentaren, die die Befragten im Rahmen der Evaluation des Auswahlgesprächs am Ende des Tages gemacht haben: Von insgesamt 20 Kommentaren zur Studie bezogen sich insgesamt acht Kommentare eher auf allgemeine Einschätzungen, wie etwa ein Begrüßen derartiger wissenschaftlicher Studien oder Zweifel an deren Aussagekraft. Vier Teilnehmer der Bedingung *vermeintliche Auswahl-situation* hatten Zweifel, ob diese Fragen wirklich Teil des Auswahlverfahrens seien. Insgesamt sechs Kommentare bezogen sich auf Aspekte der sozialen Erwünschtheit. Tatsächlich äußerten Befragte beider Versuchsbedingungen Bedenken hinsichtlich verzerrter Antworten, wenn Fragebögen Teil des Auswahlverfahrens sein würden (siehe Tabelle 3). Auch wenn diese sechs Kommentare kein repräsentatives Meinungsbild darstellen, sie decken sich doch gut mit

Tabelle 2: Statistische Kennwerte für die signifikant unterschiedlichen Dimensionen von Empathie, Persönlichkeit sowie arbeitsbezogenem Erleben und Verhalten

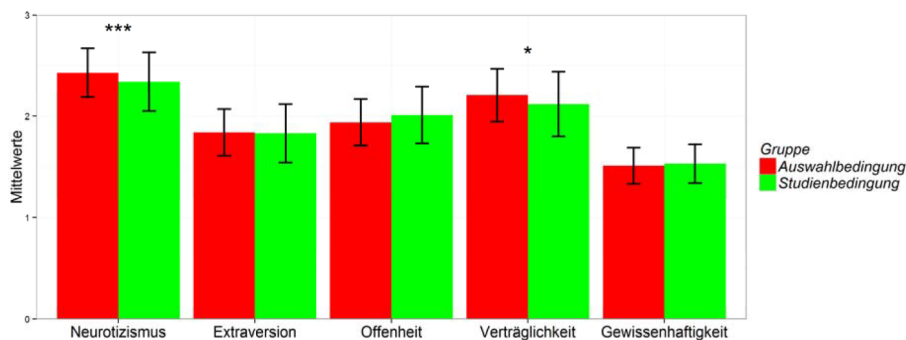
	Bedingung: Wissenschaftliche Studie	Bedingung: Vermeintliche Auswahlsituation	F	df	p	d _{Cohen}
Empathie (SPF)						
M Empathie-Gesamtwert (SD)	45.11 (5.19)	46.51 (4.70)	3.91	1,220	.05	.26
M Perspective Taking (SD)	16.07 (2.31)	17.16 (1.80)	14.94	1,220	<.01*	.26
Persönlichkeit (NEO-FFI)						
M Neurotizismus (SD)	2.34 (0.29)	2.43 (0.24)	6.14	1,220	.01*	.33
M Verträglichkeit (SD)	2.12 (0.32)	2.21 (0.26)	4.49	1,220	.04	.28
Dimensionen arbeitsbezogenen Verhaltens und Erlebens (AVEM)						
M Subjektive Bedeutsamkeit der Arbeit (SD)	14.61 (2.81)	15.29 (2.30)	4.57	1,219	.03	.29
M Beruflicher Ehrgeiz (SD)	16.09 (2.36)	16.69 (1.89)	4.54	1,217	.03	.29
M Offensive Problembewältigung (SD)	16.25 (2.25)	17.44 (2.08)	17.66	1,217	<.01*	.57
M Innere Ruhe / Ausgeglichenheit (SD)	15.67 (2.59)	16.30 (2.33)	3.99	1,216	.05	.27

M=Mittelwert; SD=Standardabweichung
 SPF und AVEM-Dimensionen: Summenwert; NEO-FFI: Mittelwert
 *signifikant auf Bonferroni-korrigiertem Signifikanzniveau: NEO-FFI $\alpha=0.01$; AVEM $\alpha=0.005$; SPF $\alpha=0.01$



*** Bonferroni-korrigiert signifikant

Abbildung 1: Mittelwertvergleiche der Empathieskala SPF für die beiden Versuchsbedingungen vermeintliche Auswahlsituation vs. wissenschaftliche Studie



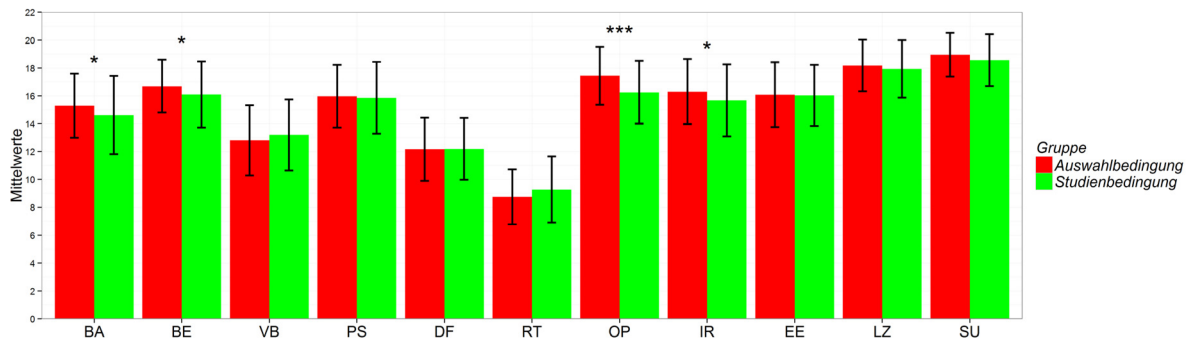
* = unkorrigiert signifikant; *** Bonferroni-korrigiert signifikant

Abbildung 2: Mittelwertvergleiche des NEO-FFI für die beiden Versuchsbedingungen vermeintliche Auswahlsituation vs. wissenschaftliche Studie

der Richtung der Unterschiede, die wir in den Fragebögen zwischen den Versuchsbedingungen zeigen konnten.

4. Diskussion

Diese Studie prüft die Frage, inwiefern bestimmte etablierte Fragebögen, die Persönlichkeit, Empathie und den



* = unkorrigiert signifikant; *** Bonferroni-korrigiert signifikant
 BA = subjektive Bedeutsamkeit der Arbeit, BE = beruflicher Ehrgeiz, VB = Verausgabungsbereitschaft, PS = Perfektionsstreben,
 DF = Distanzierungsfähigkeit, RT = Resignationstendenz, OP = offensive Problembewältigung, IR = Innere Ruhe/ Ausgeglichenheit,
 EE = Erfolgserleben im Beruf, LZ = Lebenszufriedenheit, SU = soziale Unterstützung

Abbildung 3: Mittelwertvergleiche des AVEM für die beiden Versuchsbedingungen vermeintliche Auswahl-situation vs. wissenschaftliche Studie

Tabelle 3: Kommentare aus der Evaluation des Auswahlgespräches, die sich auf Verzerrungstendenzen des Fragebogens beziehen

Bedingung: <i>vermeintliche Auswahl-situation</i>
„Das war eine ziemlich gute Erfahrung, ich bin aber der Meinung, solche Fragebögen sollten bei der Auswahl nicht berücksichtigt werden, da verfälschte Antworten gegeben werden.“
„Der Fragebogen lässt sich leicht durch idealisierte Angaben manipulieren und ist für die Auswahl der Teilnehmer unfair.“
„Ich denke, dass 15 Minuten längere Auswahlgespräche mehr aussagen als die Studie. Wenn es um einen Medizinstudienplatz geht, glaube ich, ist jeder bereit zu schwindeln.“
Bedingung: <i>wissenschaftliche Studie</i>
„Ich denke, dass die Bewerber/ Studienteilnehmer intuitiv das ankreuzen, von dem sie ausgehen, was das "Gegenüber" hören will, vor allem, wenn es Teil des Auswahlverfahrens ist.“
„Ich hätte den Fragebogen komplett anders ausgefüllt, wenn ich ihn für ein Aufnahmekriterium gehalten hätte.“
„Teilweise sind die Fragen doch sehr offensichtlich und die Antwortmöglichkeiten dafür zu pauschal. Evtl. etwas verschlüsseltere Fragen einbringen.“

Insgesamt 20 Kommentare, die sich auf Aspekte der Studie bezogen, davon 13 in der Bedingung *wissenschaftliche Studie* und 7 in der Bedingung *vermeintliche Auswahl-situation*.

Umgang mit Stress messen, im Rahmen des AdH anfällig für sozial erwünschtes Antwortverhalten sind. Die Ergebnisse zeigen, dass die mit allen ausgewählten Fragebögen erzielten Scores durch den Einsatz in einer Auswahl-situation beeinflusst werden. Selbst nach Bonferroni-Korrektur finden sich statistisch signifikante Unterschiede in Subskalen aller verwendeten Instrumente. Die Unterschiede zeigen dabei kleine bis mittlere Effektstärken [33]. Zudem scheinen sich einzelne Unterschiede im AVEM aufzukumulieren, was sich in einer mehr als dreifach so großen Wahrscheinlichkeit für ein Muster G in der *vermeintlichen Auswahlbedingung im Vergleich zu Kontrollbedingung wissenschaftliche Studie* äußert. Allein dieser deutliche Unterschied ist bemerkenswert. Der Einsatz von Persönlichkeitsfragebögen in Personalauswahlverfahren wird vor allem in der Persönlichkeits- sowie der Arbeits- und Organisationspsychologie kontrovers diskutiert. Studien, bei denen Probandinnen und

Probanden explizit aufgefordert wurden, sich möglichst gut darzustellen (fake-good-Paradigma), zeigen durchaus erwartungs- und stereotypenkonformes Antwortverhalten [34], [35], [36], [37], [38]. Persönlichkeitstests in Auswahl-situationen scheinen demnach tatsächlich anfällig für soziale Erwünschtheit zu sein. Andere Studien hingegen belegen kaum Einflüsse sozialer Erwünschtheit auf die Beantwortung von Persönlichkeitstests [39], [40], [41]. Insbesondere, wenn Probandinnen und Probanden in eine vermeintlich reale Auswahl-situation gebracht wurden, scheinen die Effekte sozialer Erwünschtheit kleiner zu werden [40], was auch die relativ geringen Effektstärken dieser Studie erklären könnte. Auch könnte der zunächst überraschende Befund, dass sich Bewerberinnen und Bewerber in der vermeintlichen Auswahl-situation neurotischer dargestellt haben als in der Kontrollbedingung, dadurch erklärt sein, dass es sich um kein reines fake-good-Paradigma handelte. Die Bewerberinnen und

Bewerber waren vielmehr gefordert, sich möglichst positiv im Sinne einer / eines guten Studierenden und angehenden Ärztin/Arzt darzustellen. Möglicherweise haben sie daher bewusst vermieden, sich allzu selbstsicher darzustellen, gehört doch ständige Selbstreflexion zum Bild eines „guten Arztes“/einer „guten Ärztin“ [42].

Vergleichbare Studien zur Anfälligkeit von Empathie-Fragebögen oder des AVEM für sozial erwünschtes Antworten gibt es nach aktuellem Kenntnisstand nicht.

Da die Fallzahl von 226 eher zu gering war, um Unterschiede in diesem Effektstärkenbereich zuverlässig zu detektieren, werden die Auswirkungen sozialer Erwünschtheit in dieser Studie eher unterschätzt [43]. Auch wenn die gefundenen Unterschiede auf den ersten Blick eher vereinzelt oder gering erscheinen mögen, kann also aufgrund der vorliegenden Daten nicht davon ausgegangen werden, dass die Fragebögen in einer Auswahl-situation tatsächlich valide Persönlichkeit, Empathie und Stressresilienz messen.

Einschränkend muss angemerkt werden, dass zwar Unterschiede in den ausgewählten Fragebögen zwischen den Versuchsbedingungen nachgewiesen werden können, ein signifikanter Unterschied in der Skala für soziale Erwünschtheit zwischen den Bedingungen jedoch nicht. Daraus könnte man schließen, dass diese Verzerrungen möglicherweise andere Ursachen als eine erhöhte Tendenz zur sozialen Erwünschtheit haben. Dies kann auf der Basis dieser Ergebnisse nicht ausgeschlossen werden. Die hier verwendete Skala SES-17 [32] erfasst soziale Erwünschtheit mit sehr globalen Alltagssituationen, die wenig mit der Studienplatzvergabe oder einem Auswahlprozess allgemein zu tun haben. Soziale Erwünschtheit hingegen umschreibt eine Verhaltenstendenz, bei der eine Person bemüht ist, sich möglichst konform mit situationsspezifischen Erwartungen und Normen zu verhalten [29]. Es wäre daher denkbar, dass das Auswahlverfahren soziale Erwünschtheit deutlich kontext-spezifischer erhöht und die SES-17 somit nicht empfindlich genug ist, um experimentell induzierte Unterschiede zu messen. Möglicherweise hätte eine stärkere Formulierung im Instruktionstext der Bedingung *vermeintliche Auswahl-situation* (siehe Methoden) zu einem stärkeren Unterschied zwischen den Bedingungen geführt. Betrachtet man jedoch die Richtung der Unterschiede in den Fragebögen, deuten diese durchaus darauf hin, dass die Bewerberinnen und Bewerber in der Bedingung *vermeintliche Auswahl-situation* bemüht waren, sich möglichst positiv im Sinne einer/eines guten Studierenden und angehenden Ärztin/Arzt darzustellen. Dies wird auch durch die Freitextkommentare der Evaluation des Auswahlgesprächs unterstützt. Dies spricht dafür, dass die Skala SES-17 als Kontrollskala möglicherweise nicht gut geeignet war.

Insgesamt zeigen die Ergebnisse dieser Studie, dass ein Einsatz etablierter Fragebögen zur Messung nicht-kognitiver Eigenschaften in der Vergabe der Medizinstudiensplätze zwar wünschenswert ist, jedoch mit Vorsicht geschehen sollte. Ratsam wäre eine weitere Evaluation von Fragebögen entsprechend der oben diskutierten Fragen

und ggf. eine Überarbeitung dieser, um sie weniger anfällig für Verzerrungen zu machen. Inwiefern sich die gefundenen Einschränkungen auf den tatsächlichen Auswahlprozess auswirken, etwa, weil Bewerber mit einer hohen Tendenz zur sozialen Erwünschtheit höhere Rangplätze erhalten und daher eher in den Genuss eines Studienplatzes kommen oder inwiefern dies indirekt Auswirkungen auf die Frage hat, ob auf diese Weise wirklich diejenigen ausgewählt werden, die mit höherer Wahrscheinlichkeit gute Mediziner werden, bleibt Gegenstand zukünftiger Forschung [44]. Die hier vorgestellten Ergebnisse belegen jedoch, dass derartige Forschung dringend nötig ist, wenn psychometrische Verfahren zukünftig ergänzend im Auswahlverfahren der Hochschulen zum Einsatz kommen sollen.

5. Schlussfolgerungen

- Der Einsatz etablierter Fragebögen zur Erfassung nicht-kognitiver Eigenschaften im Rahmen des AdH scheint nicht unproblematisch zu sein: Die hier überprüften Fragebögen weisen Einflüsse sozialer Erwünschtheit auf. Wenn die Ergebnisse jedoch durch den Wunsch, sich möglichst positiv im Sinne eines idealen Studierenden oder angehenden Arztes darzustellen, beeinflusst werden, sind ihr prädiktiver Wert und damit ihr Nutzen für die Studienplatzvergabe unklar.
- Weitere Forschungsarbeiten zu den tatsächlichen Auswirkungen sozial erwünschter Antworten sowie möglichen alternativen Verfahren sind daher nötig.

Danksagungen

Wir bedanken uns bei Jessica Lückert, Sophia Marie Saftien, Karl Böse, Karen Sievers und Josefin Wagner für die Unterstützung bei der Vorbereitung und Durchführung der Befragungen.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Linda Brüheim war im Rahmen ihres Beschäftigungsverhältnisses an der Universität zu Lübeck mit der Auswertung des AdH betraut.

Die Studie wurde aus Eigenmitteln der Universität zu Lübeck, der Sektion Medizin der Universität zu Lübeck und dem Institut für Sozialmedizin und Epidemiologie der Universität Lübeck finanziert.

Literatur

1. Stiftung für Hochschulzulassung. Daten der bundesweit zulassungsbeschränkten Studiengänge an Hochschulen. Dortmund: Stiftung für Hochschulzulassung; 2014. Zugänglich unter/available from: <http://bit.ly/12Vo1xR>

2. Stiftung für Hochschulzulassung. Verordnung über die zentrale Vergabe von Studienplätzen durch die Stiftung für Hochschulzulassung. Dortmund: Stiftung für Hochschulzulassung; 2014. Zugänglich unter/available from: <http://bitly.com/1wSTw9w>
3. Bundesvertretung der Medizinstudierenden in Deutschland e.V. Positionspapier Zulassungsverfahren zum Medizinstudium in Deutschland. Bonn: Bundesvertretung der Medizinstudierenden in Deutschland; 2015. Zugänglich unter/available from: <http://bit.ly/1lvhLgh>
4. Bundesärztekammer. Stellungnahme der Bundesärztekammer zu den Themen für den "Masterplan Medizinstudium 2020", insbesondere hinsichtlich möglicher Maßnahmenvorschläge. Berlin: Bundesärztekammer; 2015. Zugänglich unter/available from: <http://bit.ly/2dlRE0u>
5. Salvatori P. Reliability and validity of admissions tools used to select students for the health professions. *Adv Health Sci Educ Theory Pract.* 2001;6(2):159-175. DOI: 10.1023/A:1011489618208
6. Ferguson E, James D, Madeley L. Factors associated with success in medical school: systematic review of the literature. *BMJ.* 2002;324(7343):952-957. DOI: 10.1136/bmj.324.7343.952
7. Maslov Kruzicevic S, Barisic KJ, Banozic A, Esteban CD, Sapunar D, Puljak L. Predictors of attrition and academic success of medical students: a 30-year retrospective study. *PLoS ONE.* 2012;7(6):e39144. DOI: 10.1371/journal.pone.0039144
8. Shen H, Comrey AL. Predicting medical students' academic performances by their cognitive abilities and personality characteristics. *Acad Med.* 1997;72(9):781-786. DOI: 10.1097/00001888-199709000-00013
9. Hojat M, Robeson M, Damjanov I, Veloski JJ, Glaser K, Gonnella JS. Students' psychosocial characteristics as predictors of academic performance in medical school. *Acad Med.* 1993;68(8):635-637. DOI: 10.1097/00001888-199308000-00015
10. Adam J, Bore M, McKendree J, Munro D, Powis D. Can personal qualities of medical students predict in-course examination success and professional behaviour? An exploratory prospective cohort study. *BMC Med Educ.* 2012;12:69. DOI: 10.1186/1472-6920-12-69
11. Voltmer E, Kötter T, Spahn C. Perceived Medical School Stress and the Development of Behavior and Experience Patterns in German Medical Students. *Med Teach.* 2012;34(10):840-847. DOI: 10.3109/0142159X.2012.706339
12. Voltmer E, Kieschke U, Spahn C. Arbeitsbezogenes Verhalten und Erleben bei Ärzten im dritten bis achten Berufsjahr. *Z Psychosom Med Psychother.* 2007;53(3):244-257. DOI: 10.13109/zptm.2007.53.3.244
13. Kötter T, Tautphäus Y, Obst KU, Voltmer E, Scherer M. Health-promoting factors in the freshman year of medical school: a longitudinal study. *Med Educ.* 2016;50(6):646-656. DOI: 10.1111/medu.12987
14. Kjeldstadli K, Tyssen R, Finset A, Hem E, Gude T, Gronvold NT, Ekeberg O, Vaglum P. Life satisfaction and resilience in medical school - a six-year longitudinal, nationwide and comparative study. *BMC Med Educ.* 2006;6:48. DOI: 10.1186/1472-6920-6-48
15. Wallace JE, Lemaire JB, Ghali WA. Physician wellness: a missing quality indicator. *Lancet.* 2009;374(9702):1714-1721. DOI: 10.1016/S0140-6736(09)61424-0
16. Neumann M, Scheffer C, Tauschel D, Lutz G, Wirtz M, Edelhäuser F. Physician empathy: definition, outcome-relevance and its measurement in patient care and medical education. *GMS Z Med Ausbild.* 2012;29(1):Doc11. DOI: 10.3205/zma000781
17. Costa P, Alves R, Neto I, Marvão P, Portela M, Costa MJ. Associations between medical student empathy and personality: a multi-institutional study. *PLoS ONE.* 2014;9(3):e89254. DOI: 10.1371/journal.pone.0089254
18. Duberstein P, Meldrum S, Fiscella K, Shields CG, Epstein RM. Influences on patients' ratings of physicians: Physicians demographics and personality. *Patient Educ Couns.* 2007;65(2):270-274. DOI: 10.1016/j.pec.2006.09.007
19. Lievens F, Coetsier P, De Fruyt F, De Maeseneer J. Medical students' personality characteristics and academic performance: a five-factor model perspective. *Med Educ.* 2002;36(11):1050-1056. DOI: 10.1046/j.1365-2923.2002.01328.x
20. Hallfahrt T, Reinke S, Westermann J. Exzellenz und Engagement gesucht. *Dtsch Arztebl.* 2009;106:A1296-A1298.
21. Waydhas C, Heue M, Nast-Kolb D. Strukturierte Interviews zur Studienplatzvergabe in der Medizin: Erste Erfahrungen. *GMS Z Med Ausbild.* 2007;24(4):Doc186. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2007-24/zma000480.shtml>
22. Hampe W, Klusmann D, Buhk H, Münch-Harrach D, Harendza S. Reduzierbarkeit der Abbrecherquote im Humanmedizinstudium durch das Hamburger Auswahlverfahren für Medizinische Studiengänge - Naturwissenschaftsteil (HAM-Nat). *GMS Z Med Ausbild.* 2008;25(2):Doc82. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2008-25/zma000566.shtml>
23. Kötter T, Obst K, Brühem L, Eisemann N, Voltmer E, Katalinic A. Können psychometrische Tests den Erfolg im Auswahlgespräch zum Medizinstudium vorhersagen? Eine Querschnittsstudie an einer deutschen Hochschule. *Gesundheitswes.* 2015., DOI: 10.1055/s-0035-1554705
24. Wissenschaftsrat. Empfehlungen zur Reform des Hochschulzugangs. Berlin: Wissenschaftsrat; 2004. Zugänglich unter/available from: <http://bit.ly/1Z20sry>
25. Paulus C. Der Saarbrücker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie: Psychometrische Evaluation der deutschen Version des Interpersonal Reactivity Index. Saarbrücken: Universität des Saarlandes; 2009. Zugänglich unter/available from: <http://bit.ly/1kHRCmV>
26. Borkenau P, Ostendorf F. NEO-Fünf-Faktoren-Inventar. 2., neu normierte und vollständig überarbeitete Auflage. Göttingen: Hogrefe; 2008.
27. Schaarschmidt U, Fischer AW. Arbeitsbezogenes Verhaltens- und Erlebensmuster. 2. Auflage. Frankfurt am Main: Swets Test Services; 2003.
28. Edwards AL. The social desirability variable in personality assessment and research. Oak Brook: The Dryden Press; 1957.
29. Mummendey HD. Methoden und Probleme der Kontrolle sozialer Erwünschtheit (Social Desirability). *Z Different Diagn Psychol.* 1981;2:199-218. Zugänglich unter/available from: <http://bit.ly/1TJtnn3>
30. Davis MH. A multidimensional approach to individual differences in empathy. Austin: University of Texas; 1980. Zugänglich unter/available from: <http://bit.ly/22IFq7w>
31. Hojat M, Gonnella JS, Nasca TJ, Mangione S, Vergare M, Magee M. Physician empathy: definition, components, measurement, and relationship to gender and specialty. *Am J Psychiatry.* 2002;159(9):1563-1569. DOI: 10.1176/appi.ajp.159.9.1563
32. Stöber J. Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität. *Diagnostica.* 1999;45:173-177. DOI: 10.1026//0012-1924.45.4.173

33. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2. Auflage. Hillsdale, N.J.: Lawrence Erlbaum Assoc Inc; 1988.
34. Viswesvaran C, Ones DS. Meta-analyses of fakability estimates: Implications for personality measurement. *Educ Psychol Meas.* 1999;59:197–210. DOI: 10.1177/00131649921969802
35. Alliger GM, Dwight SA. A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educ Psychol Meas.* 2000;60:59–72. DOI: 10.1177/00131640021970367
36. Krahé B, Herrmann J. Verfälschungstendenzen im NEO-FFI: Eine experimentelle Überprüfung. *Z Different Diagn Psychol.* 2003;24:105–117. DOI: 10.1024//0170-1789.24.2.105
37. Hagemann D, Amelang M, Bartussek D, Stemmler G. *Differentielle Psychologie und Persönlichkeitsforschung.* 6. vollst. überarb. Aufl. Stuttgart: Kohlhammer; 2006.
38. Galic Z, Jerneic Ž, Kovacic MP. Do applicants fake their personality questionnaire responses and how successful are their attempts? A case of military pilot cadet selection. *Int J Select Assess.* 2012;20:229–241. DOI: 10.1111/j.1468-2389.2012.00595.x
39. Ones DS, Viswesvaran C. The Effects of Social Desirability and Faking on Personality and Integrity Assessment for Personnel Selection. *Human Perform.* 1998;11:245–269. DOI: 10.1080/08959285.1998.9668033
40. Hossiep R, Paschen M, Mühlhaus O. *Persönlichkeitstests im Personalmanagement: Grundlagen, Instrumente und Anwendungen.* Göttingen: Hogrefe Verlag; 2000.
41. Amelang M, Schäfer A, Yousfi S. Comparing verbal and non-verbal personality scales: Investigating the reliability and validity, the influence of social desirability, and the effects of fake good instructions. *Psychol Beitr.* 2002;44:24–41.
42. Hibbeler B. Zwischen Samaritertum und Ökonomie: Was ist ein "guter Arzt"? *Dtsch Arztebl.* 2011;108:A2758.
43. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39(2):175–191. DOI: 10.3758/BF03193146
44. Dwight SA, Donovan JJ. *Warning: Proceed with caution when warning applicants not to dissimulate.* Dallas, TX: Society of Industrial and Organizational Psychology; 1998.
45. Schmitt N, Oswald FL, Kim BH, Gillespie MA, Ramsay LJ, Yoo TY. Impact of elaboration on socially desirable responding and the validity of biodata measures. *J Appl Psychol.* 2003;88(6):979–988. DOI: 10.1037/0021-9010.88.6.979
46. Marcus B. Persönlichkeitstests in der Personalauswahl. *Z Psychol.* 2003;211:138–148. DOI: 10.1026//0044-3409.211.3.138

Korrespondenzadresse:

Dr. med. Thomas Kötter, MPH
 Universität zu Lübeck, Institut für Sozialmedizin und
 Epidemiologie, Ratzeburger Allee 160, 23562 Lübeck,
 Deutschland, Tel.: +49 (0)451/500-51220
 thomas.koetter@uksh.de

Bitte zitieren als

Obst KU, Brüheim L, Westermann J, Katalinic A, Kötter T. Are the results of questionnaires measuring non-cognitive characteristics during the selection procedure for medical school application biased by social desirability? *GMS J Med Educ.* 2016;33(5):Doc 75. DOI: 10.3205/zma001074, URN: urn:nbn:de:0183-zma0010749

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2016-33/zma001074.shtml>

Eingereicht: 06.01.2016

Überarbeitet: 04.08.2016

Angenommen: 12.08.2016

Veröffentlicht: 15.11.2016

Copyright

©2016 Obst et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.