ePREREGISTRATION

Project Title: Validation of a Construct-Based Situational Judgement Test To Assess Medical School Applicants

This is an identical (but blinded) version of the (non-blinded) preregistration registered on March 23, 2020 at OSF Registries. The link to the actual (non-blinded) preregistration will be provided upon publication.

A. BACKGROUND

Theoretical Approach

Besides professional competencies future physicians are likewise expected to possess social skills that allow for an appropriate handling of patients, colleagues, or other stakeholders in the medical context. Admission decisions for medical schools should, therefore, already consider applicants' social skills in addition to other criteria like cognitive ability [publication blinded]. In this context, situational judgement tests (SJTs) had been proposed as a suitable measure to efficiently assess social skills in a high number of medical school applicants without preselection based on cognitive criteria (Lievens, & Coetsier, 2002). An SJT is a low-fidelity simulation that provides participants with a critical situation and asks to rate or describe reactions to this situation. Although this method has advantages, prior SJTs that were derived from everyday situations in medical studies could not show a meaningful factor structure and construct validity (Christian, Edwards, & Bradley, 2010).

To solve this problem, current research started to concentrate more on the underlying structure of SJTs. Instead of developing SJTs that reflect actual everyday situations as accurate as possible, SJTs are developed with a focus on an intended construct and the theory behind it (e.g., Guenole, Chernyshenko, & Weekly, 2017). Thus, the test development of SJTs is shifting from a top-down event-perspective that aims at describing broad social skills to a bottom-up theory-guided process that assesses distinct personality traits. These construct-based SJTs function like personality tests as the presented reactions reflect different levels of the intended trait (Mussel, Gatzka, & Hewig, 2016). Following this line of research, some researchers

developed SJTs that measure narrow traits and show good psychometrical properties (e.g., Mussel et al., 2016; Olaru, Burrus, MacCann, Zaromb, Wilhelm, & Roberts, 2019). Those SJTs assess personality in general, but target no special sample or practical purpose and are formulated with a generic wording. However, there are construct-based SJTs available that are designed for medical school selection and that are formulated with wording of the medical context (De Leng, Stegers-Jager, Born, &, Themmen, 2018; Husbands, Rodgerson, Dowell, & Patterson, 2015). Nevertheless, both SJTs assess integrity as only one aspect of personality and do not assess a more comprehensive set of traits that is crucial for medical school applicants.

Prior work from research on high-fidelity simulations identified assertiveness, warmth, and emotional stability to be crucial traits for medical school applicants (Breil, Wigger, Finke, Dinter, & Back, 2019). Assertiveness and warmth can be understood as narrow definitions of agency and communion which are generally acknowledged as two general factors that are able to describe behavior in social interactions (Wiggins, 2991). Emotional stability was identified to be an additional factor when overt interpersonal behavior should be comprehensively described (Leising & Bleidorn, 2011). As there is currently no SJT available that assesses these three traits with a focus on the medical context, our work group together with researchers from other medical faculties developed the [blinded]-SJT in accordance with the guidelines suggested by Guenole et al., 2017 (see Appendix A for a detailed description of the development process and Appendix B for example tasks; [the herein presented SJT name was blinded as it refers to an identifiable group of researchers]). A pretest and first pilot study indicated promising relations to intended constructs for assertiveness and warmth (see Appendix D for a summary of results from the pretest and pilot study). However, the expected factor structure could not be confirmed and emotional stability showed no discriminant validity. In a revision process, we decided to concentrate only on assertiveness and warmth, defined narrower traits, and revised the items accordingly. A revised version of the initial

[blinded]-SJT resulted, named [blinded]-SJT-r (see Appendix A for details of the revision process and Appendix C for example tasks).

Study Goals and Hypotheses

The aim of the present project is to validate the [blinded]-SJT-r that intends to measure assertiveness and warmth in medical school applicants. In doing so, we investigate the construct validity, nomological network, the face validity, and an ad hoc criterion.

We expect the [blinded]-SJT-r to measure two distinct factors with one factor reflecting assertiveness and one factor reflecting warmth. To validate the constructs, we first investigate relations to a self-report of the Big Five and their facets. Assertiveness in the [blinded]-SJT-r is similar defined to the assertiveness facet of extraversion and warmth is similar defined to the compassion facet of agreeableness. Although the [blinded]-SJT-r and the Big Five self-report are both completed by the same individual, we do not expect the relations between the [blinded]-SJT-r scales and the Big Five variables to show high relations as the two methods still are different. While the [blinded]-SJT-r only refers to hypothetical behavior in the medical context, the Big Five self-report measures general, actual behavior from past experiences. Regarding the convergent validity we expect the following:

Hypothesis 1a: Assertiveness is moderately positively related to extraversion and especially to the assertiveness facet.

Hypothesis 1b: Warmth is moderately positively related to agreeableness and especially to the compassion facet.

When completed as a self-report, the assertiveness and compassion facet show moderate relations to other facets of the Big Five. But as the results from the [blinded]-SJT-r should not highly correlate to the results of the Big Five self-report, we expect the following for the discriminant validity:

Hypothesis 2a: Assertiveness is not or only weakly related to the other Big Five traits and their facets.

Hypothesis 2b: Warmth is not or only weakly related to the other Big Five traits and their facets.

As we assume that assertiveness and warmth represent narrow definitions of agency and communion, we investigate in the next step, how both construct are related to self-reported interpersonal style. As agency and communion are conceptualize within a circumplex (Wiggins, 2003), both variables can be measured distinct scales and in each possible combination (Wiggins, Trapnell, & Phillips, 1988). For relations to the [blinded]-SJT-r scales, we expect the following relations:

Hypothesis 3a: Assertiveness is moderately positive related to the assured-dominant scale and moderately negatively related to the unassured-submissive scale.

Hypothesis 3b: Warmth is moderately related to the warm-agreeable scale and moderately negatively related to the coldhearted scale.

The other scales of the circumplex represent combinations of agency and communion and we do not formulate additional hypotheses for relations between the [blinded]-SJT-r scales and these additional scales.

In a third step to validate the construct validity, we are interested in how the [blinded]-SJT-r scales relate to broader social skills. We therefore include a self-report of an inventory that assesses social skills with primary and secondary factors ([publication blinded], 2009). The inventory includes secondary factors that correspond to our definition of assertiveness and warmth and we therefore expect the following relations:

Hypothesis 4a: Assertiveness is moderately positively related to agency as a primary factor and especially to assertiveness as a secondary factor of agency.

Hypothesis 4b: Warmth is moderately positively related to social orientation as a primary factor and especially to prosociality as a secondary factor of social orientation.

To explore the nomological network of the [blinded]-SJT-r scales, we examine their relations to subclinical grandiose narcissism. Narcissism is a construct that is visible in interpersonal situations and has already been examined in its relations to the Big Five and agency and communion (Bradlee, & Emmons, 1992). Present frameworks conceptualize narcissism as a bi- or multidimensional construct. Here, we adapt the Narcissistic Admiration and Rivalry framework, an approach that distinguishes between admiration and rivalry as two dimensions of narcissism (Back, Küfner, Dufner, Gerlach, Rauthmann, & Denissen, 2013). Admiration is characterized by an agentic strategy to promote a grandiose self and shows positive relations to agentic behavior (Back et al., 2013). Rivalry, on the other hand, is characterized by an antagonistic strategy and shows negative relations to communal behavior (Back et al., 2013). Accordingly, we expect the following relations to the [blinded]-SJT-r scales:

Hypothesis 5a: Assertiveness is moderately positively related to narcissistic admiration.

Hypothesis 5b: Warmth is moderately negatively related to narcissistic rivalry.

The rationale behind an SJT is that people behave in the SJT as they will behave in real situations, i.e., they show behavioral consistency (Lievens & De Soete, 2012). If our [blinded]-SJT-r is able to measure assertiveness and warmth, we expect the scales to be related to actual assertive and warm behavior. As we are currently not able to obtain participants actual behavior, we use participants past behavior as an ad hoc criterion to validate the [blinded]-SJT-r (for an identical approach see Mussel et al., 2016; Olaru et al., 2019). Indicators of past behavior are more general and not restricted to the medical context like the [blinded]-SJT-r and we expect relations to be moderate. The according hypotheses read as following:

Hypothesis 6a: Assertiveness is moderately positively related to agentic behavioral indicators.

Hypothesis 6b: Warmth is moderately positively related to communal behavioral indicators.

Within an open research question, we want to examine the face validity of the [blinded]-SJT-r. Face validity can be an important factor for completing a test as it can guide participants' motivation (Derous, & Born, 2005). If participants have an idea what the test is used for and can relate to the content, they might complete the test with a higher motivation. On the other hand, if the face validity is too high, a test might become too easy and participant might adapt their answer to get a more favorable result (i.e., they 'fake'; Guenole et al., 2017). As, up to our knowledge, face validity of a construct-based SJT has not been investigate under high-stakes circumstances yet, we formulate this as an open research question without direction:

Research Question 1: How is face validity related to performance in the [blinded]-SJT-r?

Within the scope of a second research question, we want to investigate whether the kind of material has an influence on how high the face validity of the [blinded]-SJT-r scales is. Prior research showed that situations do not always matter in SJTs (Krumm, Lievens, Hüffmeier, Lipnevich, Bendels, & Hertel, 2015) and participants can score equally when their answers are only based on presented reactions without the corresponding situations.

Research Question 2: How is the type of material in the [blinded]-SJT-r related to its face validity?

B. METHOD

Data Collection

Data will be collected within the admission process for medical studies [blinded] in March 2020.

Sample. The sample consists of participants of the admission process who completed the SJT as part of the admission tests and agreed to be contacted for further studies after the tests. A total number of N = 4237 registered for the admission tests and about half of them

agreed to be contacted for additional studies. In a similar study one year ago, about 250 participants completed an additional online survey.

Procedure. The admission tests take place on the 11th, 12th and 15th of March 2020 at three medical faculties in [blinded]. Almost all registered persons will complete the SJT as part of the admission tests as this part is needed for the application at the medical faculty in [blinded]. The SJT consists of two components and in the current study we are focusing on the second component (i.e., [blinded]-SJT-r). Participants have 25 minutes to complete the [blinded]-SJT-r. Some days after they sat the admission tests, participants will be invited via e-mail to complete an additional online survey that includes personality questionnaires, behavioral indicators, and procedural questions. Every participant receives an individual link that allows connecting the data from the online survey to results from the admission tests. The online survey is conducted in LimeSurvey version 2.62.2+170203 and will take about 40 minutes. All variables presented in the measure section are mandatory. Participants who completed the whole online survey get a feedback on their results of the included questionnaires and have the chance to win one of 100 Amazon gift cards, each of which is worth 50€.

Measures.

[blinded]-SJT-r. We developed the written [language blinded] [blinded]-SJT-r that aims to assess two traits: warmth and assertiveness (see Appendix A). An SJT task presents the participants with a critical situation that could happen to them when they study medicine and with behavioral reactions to this situation. Each task comes with three behavioral reactions that reflect different levels of the respective construct (coded as 1 = low level, 2 = medium level, and 3 = high level). All tasks are written from the 1st person perspective and address the participants directly. The [blinded]-SJT-r consists of 30 tasks with 15 tasks per construct (see Appendix C for example tasks). Participants are asked to choose one behavioral reaction that corresponds most to their reaction in this situation (i.e., would-do instruction).

Big Five. A 60-item version of the [language blinded] adaption of the BFI-2 was included to measure the Big Five and their facets [publication blinded]. Participants answer how much they agree to each item on a scale from 1 (*do not agree at all*) to 5 (*agree totally*).

Interpersonal Adjective List. A list of 32 adjectives reflecting agentic and communal interpersonal behavior was included [publication blinded]. Participants answer how much each item applies to them on a scale from 1 (*extremely inaccurate*) to 8 (*extremely accurate*).

Inventory of Social Competencies. 83 items of the long version of the [language blinded] Inventory of Social Competencies [publication blinded] were included to measure the three main factors social orientation, agency, and self-control and their facets. Participants answer how much each item applies to them on a scale from 1 (*does totally not apply*) to 4 (*does totally apply*).

Narcissism. A short 6-item version of the Narcissistic Admiration and Rivalry Questionnaire (NARQ-S; Leckelt et al., 2018) was included to measure narcissistic admiration and rivalry with 3 items each. Participants answer how much each item applies to them on a scale from 1 (*does not apply at all*) to 6 (*applies completely*).

Behavioral Indicators. To measure an ad hoc criterion, we included 10 items that aim at participants' behavior in the last six months (see Buchanan, & Bardi, 2015). 5 items were designed to assess agentic behaviors and an example item reads "stand up for your own needs". 5 items were designed to assess communal behaviors and an example item reads "show understanding for others' problems". The instruction asks participants to indicate how often, relative to the chance they had to show this behavior, they showed this behavior. They answer on a scale from 0 (*never*) to 4 (*always*) with an extra option that reads "I did not have the chance".

Face validity. To identify whether participants are able to identify the targeted construct in an [blinded]-SJT-r task, we show participants parts of the [blinded]-SJT-r again and ask them

to indicate which social competency is targeted (we use ‘social competency’ instead of ‘trait’ here, because we assume that this wording is easier for participants and that they do not distinguish between social competencies and traits). Participants are randomly assigned to one of six conditions that are combinations of different [blinded]-SJT-r material (either a) only the situation, b) only the behavioral reactions, or c) situation and behavioral reactions, i.e. the whole task) and different answering format (either a) open answer or b) single selection of a construct). If participants have the single selection answering format, they are provided with six facets of the Big Five (anxiety, assertiveness, energy, politeness, warmth, depression) and corresponding short definitions. Each participant sees [blinded]-SJT-r material of six tasks in total with three tasks per construct. For this study, we only analyze data from the single selection conditions.

C. ANALYSIS PLAN

Data preparation

No data has been prepared or analyzed until the upload of this preregistration. We first recode the original answers from the [blinded]-SJT-r to the trait level that they reflect (coded as 1 = low level, 2 = medium level, and 3 = high level). For each participant two SJT variables will then be computed, one variable representing the assertiveness scale and one representing the warmth scale. Both scales are computed as the sum of the 15 items that belong to the respective scale. Raw answers from the personality and social competency questionnaires (Big Five, interpersonal adjective list, inventory of social competency, narcissism) will be prepared as described in their references. Items from the behavioral indicator questions will be aggregated into one variable for agency and one for communion. For a variable that reflects their ability to identify the intended trait, participants will receive one point for each construct that they identified correctly from the material they are provided with. Accordingly, this variable can range from 0 to 6.

Analyses of Hypotheses 1-6

First, we will compute factor analysis to identify how many factors the [blinded]-SJT-r represents. If our assumption of two factors holds true, we start to check our hypothesis. For each hypothesis, we compute a zero-order correlation between the assertiveness or warmth score of the [blinded]-SJT-r and the other variable. If we state that the [blinded]-SJT-r scales relate to a special facet or secondary factor, we compute correlations for all facets or secondary factors of the primary factor, use the Fisher-*r*-to-*z* formula to transform them, compute one-sided 95% confidence intervals, and compare the intervals. If the correlation to the expected facet or secondary factor is significantly higher (i.e., the confidence intervals do not overlap), we assume the hypothesis to be confirmed. To investigate Research Question 1, we compute a two-sided zero-order correlation between each SJT variable and the ability-to-identify-criterion variable. For Research Question 2, we differentiate the participants according to the material they saw and (if homogeneity of variance exists) compute an analysis of variance (ANOVA). If differences between the groups exist, we will compute Tukey's test for post-hoc analysis. If no homogeneity of variance exists, we compute Welch's ANOVA and Games-Howell's test for post-hoc analysis.

Additional Analyses We will further check, whether the [blinded]-SJT-r scales can be predicted by the variables from Hypotheses 1-6. Therefore, we compute regression analyses with the assertiveness or warmth scale as the dependent variable and the respective other variable as the independent variable. We will also analyze possible subgroup differences in the [blinded]-SJT-r scales for gender, age, and cultural background. If differences exist, we will include these variables as control variables in a second regression analysis. We further look at relations between the [blinded]-SJT-r scales and a cognitive variable (i.e., GPA). If a relation exists, we will enter the cognitive variable as a control variable into the regression analysis.

Inference Criteria

We refer to a significance level of $p < .05$ as being significant. Significance tests for hypotheses will be computed one-tailed according to the assumed direction. Significance tests for the research questions will be computed two-tailed. We will provide 95% confidence intervals for all effects. Interpretation of the effect size will be guided by the suggestions for personality research by Gignac & Szodorai (2016). Our term of a moderate relation corresponds to their term of a medium relation.

Data Exclusion

We expect that almost all participants will complete all items of the [blinded]-SJT-r as they get the instruction that missing [blinded]-SJT-r items will contribute negatively to their results. However, participants who completed less than 80% (i.e., less than 12 items) of a single [blinded]-SJT-r scale will be excluded for analyses regarding this scale. If a participant provides no data or less than 80% of data for another measure, this participant will be excluded for analyses regarding this specific measure.

D. REFERENCES

Back, M. D., Küfner, A. C., Dufner, M., Gerlach, T. M., Rauthmann, J. F., & Denissen, J. J. (2013). Narcissistic admiration and rivalry: Disentangling the bright and dark sides of narcissism. *Journal of Personality and Social Psychology*, *105*(6), 1013-1037.

Bradlee, P. M., & Emmons, R. A. (1992). Locating narcissism within the interpersonal circumplex and the five-factor model. *Personality and Individual Differences, 13*, 821–830.

[publication blinded]

Buchanan, K., & Bardi, A. (2015). The roles of values, behavior, and value-behavior fit in the relation of agency and communion to well-being. *Journal of Personality*, *83*(3), 320-333.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63,* 83–117.

[publication blinded]

De Leng, W. E., Stegers-Jager, K. M., Born, M. P., & Themmen, A. P. (2018). Integrity situational judgement test for medical school selection: judging 'what to do'versus 'what not to do'. *Medical education*, *52*(4), 427-437.

Derous, E., & Born, M. P. (2005). Impact of face validity and information about the assessment process on test motivation and performance. *Le travail humain*, *68*(4), 317-336.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78.

Guenole, N., Chernyshenko, O. S., & Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, *17*(3), 234-252.

Husbands, A., Rodgerson, M.J., Dowell, J. & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC Med Educ*, *15***,** 144.

[publication blinded]

[publication blinded]

Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests? *Journal of Applied Psychology*, *100*, 399-416.

[publication blinded]

Leckelt, M., Wetzel, E., Gerlach, T. M., Ackerman, R. A., Miller, J. D., Chopik, W. J., ... & Richter, D. (2018). Validation of the Narcissistic Admiration and Rivalry Questionnaire Short Scale (NARQ-S) in convenience and representative samples. *Psychological assessment*, *30*(1), 86-96.

Leising, D., & Bleidorn, W. (2011). Which are the basic meaning dimensions of observable interpersonal behavior?. *Personality and Individual Differences*, *51*(8), 986-990.

Lievens, F., & Coetsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment*, *10*(4), 245-257.

Lievens, F., & De Soete, B. (2012). Simulations. *In* N. Schmitt (Ed.), *The Oxford Handbook of Personnel Assessment and Selection* (pp. 383‑410). Oxford: Oxford University Press.

Mussel, P., Gatzka, T., & Hewig, J. (2018). Situational judgment tests as an alternative measure for personality assessment. *European Journal of Psychological Assessment, 34*(5), 328–335.

Olaru, G., Burrus, J., MacCann, C., Zaromb, F. M., Wilhelm, O., & Roberts, R. D. (2019). Situational Judgment Tests as a method for measuring personality: Development and validity evidence for a test of Dependability. PloS one, 14(2).

Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behavior. In W. Grove & D. Cicchetti (Eds.), Thinking clearly about psychology: Essays in honor of Paul E. Meehl (Vol. 2., pp. 89–113). Minneapolis, MN: University of Minnesota Press.

Wiggins, J. S. (2003). Paradigms of personality assessment. New York: Guilford.

Wiggins, J. S., Trapnell, P. & Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research, 23*, 517–530.

E. Appendices

Appendix A: Development of the [blinded]-SJT and [blinded]-SJT-r

The development of the [blinded]-SJT started with a workshop with an expert in SJT development and researchers from several medical faculties in [blinded]. Within the workshop, a summary of prior SJT tasks based on a classic approach was discussed and the development of construct-based SJT tasks was trained. Based on these insights and material, researchers from the same medical faculties developed SJT situations (N =229) and related reaction (four to nine per situation) that aimed at measuring assertiveness, warmth, or emotional stability. After another critical review, a total of 171 SJT tasks were selected and rated by two other independent construct experts. The experts rated the SJT tasks according to the intended construct and the reactions according to the level of the construct. If the experts agreed on the intended construct and on the construct level of at least three reactions (every reaction measuring another construct level), this SJT task was selected for a pretest study. A total of 60 SJT situations with three reactions that reflected three levels of the construct (i.e., low, medium, high) resulted. The 60 SJT tasks consisted of 20 task per construct.

The pretest was an online study that was completed by undergraduate psychology students and included the 60 SJT tasks and a Big Five questionnaire (BFI; [publication blinded]). The results showed promising reliability and validity (see Appendix D).

For the following pilot study, 15 SJT tasks per construct were selected based on their discriminatory power. The pilot study was done as a voluntary study after a science test that was mandatory within the selection process for medical studies at the medical faculty in [blinded] (the [blinded]-SJT was also completed at several other medical faculties, but we concentrate on the [blinded] sample here). About 77% of the participants of the science test completed the [blinded]-SJT (N = 845; we marked a completion when at least 90% of the items were answered). Those participants who agreed to be contacted for further studies after

the selection process received an invitation to an additional online study two weeks later. This online study included different personality questionnaires (see Appendix D) and was completed by 245 participants. The results of the pilot study were different from the pretest study with lower reliabilities and lower validity. Especially emotional stability showed low reliability and no concurrent validity.

Based on these results, we decided to concentrate on assertiveness and warmth and started an intense review process. First, situations of the remaining 30 SJT tasks were classified with regard to the facet of assertiveness or warmth they aim at. If the content did not match the definition of the facet, the content was edited accordingly. Second, it was checked whether each situation included a dilemma that prevents participants from easily choosing the highest level of the construct, i.e., to prevent faking. Third, four to seven reactions per situation were developed according to their fit to the situation and intended facet. Two construct experts that were not involved in the development process rated only the situations with regard to the intended construct and commented insecurities. The situations were edited accordingly. Three other construct experts then got the SJT situations and reactions and rated the construct level of the reactions. We selected three reactions per situation that were equally rated by the experts and represented all three levels of the construct. The resulting [blinded]-SJT-r includes 15 situations per assertiveness and warmth with three reactions that display either a low, medium, or high level of the construct (see Appendix B for example tasks).

Appendix B: Example Tasks From [blinded]-SJT

An example task for assertiveness in the [blinded]-SJT reads the following:

[blinded as the original language is visible]

An example task for warmth in the [blinded]-SJT reads the following:

[blinded as the original language is visible]

An example task for emotional stability in the [blinded]-SJT reads the following:

[blinded as the original language is visible]

Appendix C: Example Tasks From [blinded]-SJT-r

An example task for assertiveness in the [blinded]-SJT-r reads the following:

[blinded as the original language is visible]

An example task for warmth in the [blinded]-SJT-r reads the following:

[blinded as the original language is visible]

Appendix D: Summary of the [blinded]-SJT From the Pretest and Pilot Study

	Study					
	Pretest			Pilot		
Sample						
Type	Undergraduate psychology students			Applicants for medical studies		
Number	54			247		
Gender	91% female			69% female		
Age	range: 18-48 (M = 23.02, SD = 5.66)			range: 17-35 (M = 19.99, SD = 1.96)		
	Scale					
	Assertiveness	Warmth	Emotional Stability	Assertiveness	Warmth	Emotional Stability
Reliability	α = .66	α = .77	α = .71	α = .48	α = .63	α = .30
Big Five						
Extraversion	.34*	.25	.38**	.15*	.05	.30**
Agreeableness	-.06	.46**	.14	-.02	.32**	.23**
Neuroticism	-.12	-.30*	-.45**	-02	-.10	-.25**
Conscientiousness	.19	.28*	.31*	.05	.11	.27**
Openness	-.07	.05	.30*	.17**	.08	.12*
Interpersonal Adjective List						
Assured-dominant				.14*	-.05	.19**
Unassured-submissive				-.14*	-.02	-.25**
Warm-agreeable				.01	.24**	.18**
Coldhearted				-.04	-.16*	-.23**
Inventory Social Skills						
Social orientation				-.02	.25**	.28**
Agency				.12	.00	.27**
Self-control				.04	.07	.28**
Reflection				.04	.08	-.04

Number in the pilot study refers to all participants who completed at least 90% of the SJT; Big Five were measured with BFI-2 [publication blinded], Interpersonal Adjective List shows four of eight scales [publication blinded], Inventory of Social Skills was the ISK [publication blinded]; * $p < .05$, ** $p < .01$.