# OSF Material 2: Detailed Description of the Assessment of All Variables of Interest in the Five Studies

Please note that whenever possible, we used measurements of the respective outcome variables that were assessed later in time than the ability measures. We always used the first assessments of self-rated and objective ability, and we aggregated the outcomes across as many points in time as possible to obtain more reliable measures. For descriptive statistics and internal consistencies, see OSF Material 3.

## Sample A: Study 2 by Dufner et al. (2012; see also Selfhout, Denissen, Branje, and Meeus, 2009)

In this study, a cohort of 489 first-year psychology students attending Utrecht University, The Netherlands, were invited to participate. They were divided into 20 work groups by the department. A total of 378 students agreed to take part. The initial sample (238 students) then consisted of the 10 work groups in which more than 80% of the students agreed to participate. Of these, we assessed all relevant variables in 188 first-year participants (157 female) between the ages of 17 and 31 ($M$ = 18.89, $SD$ = 1.67). The study consisted of six waves, of which only three contained all of the variables that were relevant for our purposes (T1 in the second week of the first semester, T2 4 months after T1, T3 8 months after T1). In each wave, participants completed an online questionnaire in which different indicators of intrapersonal adjustment were assessed, and participants rated themselves and all other members of their group in a round robin design on several indicators of social adjustment and personality traits. All outcome variables were aggregated across T2 and T3. No experimental manipulations were implemented.

**Self-rated ability.** We assessed self-rated reasoning ability during the execution of the round robin ratings of the first online survey with the item "Some people are dull

and not so intelligent, whereas other people are very intelligent and clever. Please rate yourself on this dimension". The scale ranged from 1 (*not intelligent, dull*) to 7 (*very intelligent, clever*). Retest reliability was .62.

**Objective ability.** An objective score for reasoning ability was assessed in the first online survey with a shortened 15-item version (see Denissen, Schönbrodt, van Zalk, Meeus, & van Aken, 2011) of Raven's advanced progressive matrices (Raven, Raven, & Court, 1962; $\alpha = .71$).

**Outcome measures.** For the category *global self-evaluation*, we assessed self-esteem with a German version (von Collani & Herzberg, 2003) of the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1965). The RSES comprises 10 items rated on a scale from 1 (*strongly agree*) to 4 (*strongly disagree*). Sample items are "I take a positive attitude toward myself," and "At times I think I am no good at all" (reversed scored). Self-liking was measured during the round robin ratings with the item "You don't like some people at all while you like other people very much. Please rate yourself on this dimension" on a scale ranging from 1 (*do not like him or her at all*) to 7 (*like him or her very much*).

For the category *well-being*, positive affect and negative affect were assessed with a German version (Krohne, Egloff, Kohlmann, & Tausch, 1996) of the expanded form of the Positive and Negative Affect Schedule (PANAS-X; Watson, Clark, & Tellegen, 1988; Watson & Clark, 1994) with the subscales *positive affect* (interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, active) and *negative affect* (distressed, angry, guilty, scared, hostile, irritable, ashamed, nervous, confused, afraid). Participants rated the extent to which they felt in accordance with each item during "the past week, including today" on a scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). We assessed depression with a modified version of the Beck Depression

Inventory (BDI; Beck, Steer, & Brown, 1996). The BDI comprises 21 symptoms. In the modified version, participants indicated for 21 symptoms of depression how much it applied to them on a scale from 1 (*completely true*) to 5 (*not at all true*). Example items are "I feel my future is hopeless and will only get worse," and "I blame myself for everything bad that happens."

For the category *self-rated agentic outcomes*, subjects' perception of their group influence was assessed with the item "Please indicate how much influence you had on your work group during the past week". A response scale between 1 (*no influence*) and 7 (*large influence*) was used.

For the category *self-rated communal outcomes*, we did not assess any variables in this study.

For the category *peer-rated agentic outcomes*, we assessed peer-rated reasoning ability during the round robin ratings with the item "Some people are dull and not so intelligent, whereas other people are very intelligent and clever. Please rate your group members on this dimension" on a scale ranging from 1 (*not intelligent, dull*) to 7 (*very intelligent, clever*). Furthermore, we assessed peer-rated group influence during the round robin ratings with the same item as used for the self-rating.

For the category *peer-rated communal outcomes*, we assessed the extent to which a person was liked by others in the round robin ratings, with the item "You don’t like some people at all while you like other people very much. Please rate your group members on this dimension," rated on the same scale as self-liking. During the round robin ratings, we also assessed peer-rated friendship quality with an item to indicate the degree to which participants were friends with each of their group members on a continuous scale, ranging from 1 (*distant acquaintance*) to 7 (*best friend*). Peer-rated emotional support was measured with the items "My friend really likes me," and "When

I am sad, my friend cheers me up," rated on a scale ranging from 1 (*does not apply at all*) to 7 (*fully applies*).

**Sample B: The PILS Study (Personality and Interaction Laboratory Study; see Geukes et al., 2017; see also osf.io/q5zwp)**

The participants who provided all relevant variables consisted of 295 students (162 female) between the ages of 18 and 39 (M = 23.8, SD = 3.95) at the Johannes Gutenberg University in Mainz, Germany. The initial sample size (311 before excluding subjects with missing data on relevant variables) was determined with a time-based procedure: A period of 5 months was fixed before data collection began, and we assessed as many participants as possible. No analyses were run until the data collection was completed. In the first phase of the study, participants completed online-questionnaires concerning demographic and trait measures. They were then assigned to 54 groups of four to six persons each and attended three laboratory sessions of 1-2 hr each spaced exactly 1 week apart. The first session consisted of three tasks, and the second and third sessions of two tasks each. After each task and at the beginning of each session, we implemented round robin ratings concerning several relationship indicators and interpersonal perceptions, and all group members reported their state-affect. All variables from the round robin ratings were centered at the respective group means. We aggregated all outcome variables that were assessed during the laboratory sessions across the six assessments in Weeks 2 and 3 (i.e., before and after each task).

**Self-rated ability.** Self-rated reasoning ability and vocabulary knowledge were assessed in the online survey through a compared-to-average item each, within an extended German version of the Self Attributes Questionnaire (SAQ; Pelham & Swann, 1989). Participants rated their own ability concerning *reasoning ability* and concerning *large vocabulary* compared to their fellow students on scales comprised of the following:

*bottom 5%* over *lower 10%*, *lower 20%*, *lower 30%*, *lower 50%*, *upper 50%*, *upper 30%*, *upper 20%*, *upper 10%*, and *top 5%*.

**Objective ability.** In the first laboratory session, an objective score for vocabulary knowledge was assessed with a standard German test for the assessment of vocabulary knowledge (MWTB [Mehrfachwahl Wortschatz Test], version B; Lehrl, 2005; α = .63). The MWTB test consists of 36 sets of five alternative letter combinations, of which only one is an actually existing word. For the assessment of reasoning ability, participants conducted the shortened version of Raven's advanced progressive matrices test (see Sample A; α = .71).

**Outcome measures.** For the category *global self-evaluation*, we assessed satisfaction with oneself in the state affect ratings during the sessions by asking participants to rate the item "I am satisfied with myself". All state affect ratings were assessed on a scale from 1 (*does not apply at all*) to 6 (*applies perfectly*). During the round robin ratings, we assessed self-liking with the item "I like myself". All items in the round robin design were rated on a scale ranging from 1 (*does not apply at all*) to 6 (*applies perfectly*).

For the category *well-being*, when participants filled out the state affect ratings during the sessions, we asked them how much they felt "determined," "active," and "optimistic" at the moment. Furthermore, we assessed the valence of participants' affect using the affect grid (Russell, Weiss, & Mendelsohn, 1989). This measure requests individuals to rate their affective valence and arousal on a 9 x 9 coordinate system. We used only the valence score that was assessed on a scale ranging from 1 (*negative emotion*) to 9 (*positive emotion*).

We used a number of different measures to assess *self-rated agentic outcomes*. Participants rated the item "I am trusting my abilities" in the state affect ratings, on a

scale ranging from 1 (*does not apply at all*) to 6 (*applies perfectly*). In addition, we used the item "I can imagine myself as a good leader" during the round robin ratings (see above).

For the category *self-rated communal outcomes*, we used the two items "I can imagine myself as a good friend," and "I am trustworthy" in the round robin ratings (see above).

For the category *peer-rated agentic outcomes*, we used the items "I can imagine this person as a good leader," and "This person is intelligent" during the round robin ratings (see above).

For the category *peer-rated communal outcomes*, we used the items "I can imagine this person as a good friend," "I like this person," and "This person is trustworthy" in the round robin ratings (see above).

**Sample C: The Connect Study (see Geukes et al., 2017; see also osf.io/2pmcr)**

A full cohort of psychology freshman (140 students, starting in October 2012) at the University of Münster, Germany, was invited to take part in this study; 131 students agreed to participate. A total of N = 91 subjects (74 female) provided all relevant variables. Their ages ranged from 18 to 42 ($M$ = 20.6, $SD$ = 3.38).

The study startet with an introductory session including a zero aquaintance experiment. Participants filled out four online surveys and one follow-up survey (Survey 1 at the very beginning, surveys after 2 [S2], 8 [S3], and 19 [S4] months, and the follow-up assessment [S5] after approx. 32 months) that contained questions about demographic information and trait measures. During the study, participants filled out a total of 23 online diaries (time-based assessment) to measure relationship indicators and interpersonal perceptions, out of three different versions (three diaries per week during the first 3 weeks, one diary per week until the end of the first semester, and two

follow-up diaries along with the last two online surveys at S4 and S5, respectively). There were also three phases of event-based assessment (T1: the first three weeks of the first semester, T2: one week approx. 2 months after T1, T3: one week approx. 3 months after T1), where participants filled out a questionnaire via a smartphone app after each interaction with another subject, and rated several behavioral and interpersonal perceptions, as well as their state affect during the interactions. Approximately 1 year (13 months) after the beginning of the study, most subjects participated in a laboratory session, where intellectual ability data was assessed.

All outcomes from online surveys were aggregated across S2, S3, S4, and S5. The outcomes from the time-based diary assessment were aggregated across all diary data from the fourth week on. All outcome measures from the event based assessment were aggregated across the periods T2 and T3.

**Self-rated ability.** Self-rated reasoning ability and vocabulary knowledge were assessed in the first online survey as two compared-to-average items in an extension of the SAQ (see Sample B; retest reliabilities were .65 for the item “reasoning ability” and .77 for “large vocabulary”).

**Objective ability.** Vocabulary knowledge was assessed in the laboratory approximately 1 year after the beginning of the study with the MWTB (see Sample B; $\alpha = .19$). At the same occasion, participants’ reasoning ability was measured with the shortened version of Raven's advanced progressive matrices test (see Sample A; $\alpha = .56$).

**Outcome measures.** For the category *global self-evaluation*, we assessed self-esteem with the RSES (see Sample A) and satisfaction with oneself exactly as in Sample B, during the online survey. Furthermore, satisfaction with the self was assessed during the event-based assessments. Participants rated the item "When interacting I was...

*dissatisfied with myself* vs. *satisfied with myself*" on a bipolar scale ranging from 1 to 7. During the time-based assessments (diaries), participants rated their own likability with the item "I find this person... *unlikeable/annoying* (-5, transformed to 1 in data) vs. *nice/likeable* (5, transformed to 11 in data)" which was presented together with a photo of themselves.

For the category *well-being*, we participants rated in the online survey how much they felt "determined," "active," "optimistic," and in a "good mood" on a scale from 1 (*does not apply at all*) to 6 (*applies perfectly*). During the event-based assessments, "bad mood" was assessed with the item "When interacting I was... *in a good mood* vs. *in a bad mood*" on a bipolar scale ranging from 1 to 7.

For the category *self-rated agentic outcomes*, we used the items "leadership ability," "assertive," "independent," and "ambitious" from the SAQ (see Sample B) and participants rated the item "I trust in my abilities" on a scale from 1 (*does not apply at all*) to 6 (*applies perfectly*). Furthermore, participants filled out a German version (Schütz, Marcus, & Sellin, 2004) of the Narcissistic Personality Inventory (NPI; Raskin & Hall, 1979; Raskin & Hall, 1981) in the online survey, for which we computed the leadership/authority subscale (see Ackerman et al., 2011) consisting of eight items. Each NPI item consists of a forced choice between two contradictory statements (e.g., "I am not sure if I would make a good leader", coded as 1, vs. "I see myself as a good leader", coded as 2). All these measures were assessed during the online survey.

For the category *self-rated communal outcomes*, we assessed trustworthiness with the SAQ (see Sample B). Furthermore, participants judged their own behavior and their interaction partners' behavior during the event-based assessments on bipolar scales ranging from 1 to 7, including the item "*friendly* (1) vs. *unfriendly* (7)". During the time-based assessments (diaries), participants rated their own warm-heartedness with the

item "This person is... *critical/combative* (-5, transformed to 1 in data) vs. *understanding/warmhearted* (5, transformed to 11 in data)," which was presented together with a photo of themselves. In an analogous way, they rated their cold-heartedness (*affectionate/trustworthy* vs. *cold-hearted, manipulative*).

For the category *peer-rated agentic outcomes*, we used several variables that were measured in the diaries. Participants rated how well they knew their peers with the item "I knew this person before today..." on a scale ranging from 1 (*not at all*) to 6 (*good friends*). Also, they rated their peers' leadership ability with the item "This person is a good leader ..." (checkbox), which was only presented for targets who were rated as "met at least once" on the knowing variable. For all targets who were additionally chosen as "interaction partner" (photo selection for "With whom did you interact last week? Interaction means, every encounter with one or more persons that lasted longer than 5 minutes and offered the opportunity to react to the behavior of the other person(s)"), subjects also rated the target's "unintelligence" with the item "This person is... *intelligent* (-5, transformed to 1 in data) vs. *unintelligent* (5, transformed to 11 in data)."

For the category *peer-rated communal outcomes*, there were several measures in the diaries. Participants rated the likability of persons they had met at least once with the same item used for self-rated likability. They also rated their friendship with these persons with the item "This person is a friend of mine ..." (checkbox). For all targets with whom the subject had interacted, we additionally assessed relationship satisfaction ("I am ... with the relation to that person"; 0 = *very unsatisfied*; 10 = *very satisfied*), relationship importance ("The relation to that person is ... to me"; 0 = *not important*, 10 = *very important*), conflict potential ("I have ... with that person"; 0 = *never conflicts*, 10 = *conflicts at every occasion*), emotional support ("In case of emotional troubles I can ..."; 0

= *never turn to this person*, 10 = *always turn to this person*), and acceptance ("I feel ... by this person"; 0 = *not accepted at all*, 10 = *entirely accepted*). Interaction partners also rated warm-heartedness and cold-heartedness on the same scales as for the self-rated communal outcomes. Furthermore, we used peer ratings of interaction negativity ("For me, the interaction with her/him was...;" 1 = *positive*, 7 = *negative*) and unfriendly behavior (1 = *friendly*, 7 = *unfriendly*), which was measured in the event-based assessments (smartphone app).

**Sample D: Study 1 by Dufner et al. (2012)**

This study was an online investigation with German-speaking internet users. We did not define an aspired sample size before the start of the online assessments. When the number of participants reached 2,306, we considered the sample size large enough and ended data collection. No analyses were run until the data collection was completed. Out of this initial sample, we used data from $N$ = 2,047 (259 participants excluded because they aborted the questionnaire, did not follow instructions, or had missing data; and one 12-year old participant was also excluded) subjects (1,431 female) aged between 17 and 76 ($M$ = 27.61, $SD$ = 8.78). Participants first completed an online survey where they reported self-perceptions and psychological adjustment and fulfilled a vocabulary test. They then invited a friend to provide peer ratings of psychological adjustment and personality. No experimental manipulations were implemented in this study.

**Self-rated ability.** Self-rated vocabulary knowledge was assessed as the aggregate of the six items “I have a large vocabulary,” “I am unfamiliar with many foreign words that other people know” (reversed scored), “I am intelligent,” “I know more things than other people,” “I am not very knowledgeable” (reversed scored), and “I consider myself erudite”, rated on a scale from 1 (*does not apply to me*) to 7 (*fully applies to me*; $\alpha = .93$).

**Objective ability.** We assessed an objective score for vocabulary knowledge in the online survey with the MWTB (see Sample B; α = .68).

**Outcome measures.** For the category *global self-evaluation*, we assessed self-esteem with the RSES (see Sample A).

For the category *well-being*, we assessed optimism and pessimism as subscales of a German version (Glaesmer, Hoyer, Klotsche, & Herzberg, 2008) of the Life Orientation Test (LOT-R; Scheier, Carver, & Bridges, 1994). Each subscale consists of three items, which are rated on a scale ranging from 0 (*applies perfectly*) to 5 (*does not apply at all*). Sample items are "I always see my future optimistically" for optimism and "Things develop almost never as I would like them to" for pessimism. Positive affect and negative affect were assessed with the German version of the PANAS-X (see Sample A). Depression was assessed with the German short version of the Center for Epidemiological Studies Depression Scale [Allgemeine Depressionsskala] (ADS-K; Hautzinger & Bailer, 1993). For 15 items, participants rated how often they had specific feelings during the last week. Sample items are "During the last week, it was hard for me to concentrate" and "During the last week, I was sad" with response options from "less than 1 day" (1), over "1-2 days", "3-4 days", to "5-7 days" (4). Life satisfaction was assessed using a German version (Schumacher, Klaiberg, & Brähler, 2003) of the Satisfaction with Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) consisting of five items. Sample items are "In most ways my current life is close to my ideal" and "I am satisfied with my life", rated on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

For the category *self-rated agentic outcomes*, we used the leadership/authority subscale of the NPI (see Sample C), where we modified the response format such that participants indicated their agreement with narcissistic statements of the original

forced-choice items on a 2-point scale (1 = *disagree*, 2 = *agree*). We also assessed an "agentic compared-to-average score", where participants answered the item "Please estimate how much the following characteristics are present in you, compared to a mean person of the same age, gender and origin" for ten agentic adjectives (e.g., assertive, achievement-oriented), on a scale ranging from 0 (*much less present than for a mean person of the same age, gender, and origin*) to 9 (*much more present than for a mean person of the same age, gender, and origin*).

For the category *self-rated communal outcomes*, we used the same principle of compared-to-average ratings as for self-rated agentic outcomes, but with ten communal adjectives (e.g., helpful, honest).

For the category *peer-rated agentic outcomes*, we used peer-reports of the same compared-to-average ratings as in self-rated agentic outcomes.

For the category *peer-rated communal outcomes*, we used peer-reports of the same compared-to-average ratings as in self-rated communal outcomes. Additionally, emotional support was assessed in the peer-report with the items "My friend really likes me", and "When I am sad, my friend cheers me up", rated on a scale ranging from 1 (*does not apply at all*) to 7 (*fully applies*). We also assessed how much a person was liked by others with the items "I like my friend", "My friend is likeable", "I get along with my friend very well", and "I can talk with my friend", rated on a scale ranging from 1 (*not at all*) to 6 (*very much*).

**Sample E: The Self-Insight (SI) Study by Dufner, Arslan, Hagemeyer, Schönbrodt, and Denissen (2015)**

This study took place in Berlin, Germany, and consisted of two waves of data collection (approx. 14 months apart). The investigation aimed for an overall N of > 200 (because this sample size is sufficient for detecting an average effect size in social

psychology, Richard, Bond, & Stokes-Zoota, 2003). The initial sample consisted of 209 participants, the relevant variables were assessed in a total of *N* = 202 participants (134 female). Their age ranged from 22 to 36 years (*M* = 27.41, *SD* = 2.96) and they were all students, but none of them studied psychology. At each of the two waves, participants filled out online questionnaires, participated in a laboratory session, provided daily-diary assessments during a period of 14 days and invited at least three informants who knew them well to provide ratings about them. For our data analysis, we used all outcome variables measured at the second wave. For the variables assessed via diary at the second wave, we aggregated all variables across the 14-days diary assessment. The study involved an experimental manipulation that is independent of the current research question.[1]

**Self-rated ability.** We assessed self-rated vocabulary knowledge in the first online survey with an own scale, consisting of the six items "I do not know many foreign words that are known by other people" (reversed scored), "My vocabulary is not really extensive" (reversed scored), "I know the meaning of most German words", "I also know words that are used rarely", "I have a large vocabulary", and "When I do a crossword, it will rarely happen that there is a word that I don't know". Participants rated these items on a scale ranging from 0 (*does not apply at all*) to 7 (*applies perfectly*; α = .82).

**Objective ability.** We assessed an objective score for vocabulary knowledge in the first laboratory session with the MWTB (see Sample B; α = .72).

**Outcome measures.** For the category *global self-evaluation* we assessed self-esteem with the RSES (see Sample A) during the online survey. Additionally, we assessed self-esteem in the diary with four adapted items of the RSES (see Sample A), namely the items "Today I thought I was a looser" (reversed scored), "Today, I had a

positive attitude towards myself", "Today I was satisfied with myself" and "Today, I felt really useless" (reversed scored).

For the category *well-being*, we assessed positive affect and negative affect with the PANAS-X (see Sample A), depression with a modified version of the BDI (see Sample A), and life satisfaction with the SWLS (see Sample D), all in the online survey. In addition, we assessed optimism with the LOT-R (see Sample D).

For the category *self-rated agentic outcomes*, we assessed leadership ability in the online survey with the leadership/authority subscale of the NPI (see Sample C). We also assessed compared-to-average ratings in the online survey, for ten agentic adjectives (see Sample D). The only difference to Sample C was that participants were asked to compare themselves to an abstract other person and additionally to a concrete peer-group. We thus got scores for self-rated "agentic outcomes compared to average abstract other" and "agentic outcomes compared to average concrete other".

For the category *self-rated communal outcomes*, we used the same principle of compared-to-average ratings compared to abstract and concrete others, but with communal adjectives (see Sample D).

For the category *peer-rated agentic outcomes*, we assessed peer-rated social influence with an own scale, consisting of the five items "When group decisions have to be made, my peer often has big influence", "Others listen to my peer", "My peer exudes authority in a natural manner", "It often happens that my peer is ignored by others" (reversed scored), and "If my peer was the leader of a working group, other members would follow his instructions", rated on a scale ranging from 0 (*does not apply at all*) to 7 (*applies perfectly*).

For the category *peer-rated communal outcomes*, we assessed peer-rated liking by the items "I like this person" and "This person is likeable". They were rated on a scale

ranging from 0 (*not at all*) to 6 (*very much*). On the same scale, we assessed peer-rated quality of interactions with the items “I get along with this person”, “I am friends with this person”, “I know this person well”, “I can talk with this person”, and “I go out with this person”. We also assessed peer-rated conflict potential with the four items “My peer and I argue frequently”, “I would describe the relationship to my peer as very harmonic” (reversed scored), “There are conflicts between my peer and myself”, and “Even when we are together for a long time, my peer and I rarely argue with each other” (reversed scored). These items were rated on a scale ranging from 0 (*does not apply at all*) to 7 (*applies perfectly*).

References

Ackerman, R. A., Witt, E. A., Donnellan, M. B., Trzesniewski, K. H., Robins, R. W., & Kashy, D. A. (2011). What does the narcissistic personality inventory really measure? *Assessment, 18*(1), 67–87. doi:10.1177/1073191110382845

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory*. San Antonio, TX: Psychological Corporation.

Denissen, J. J. A., Schönbrodt, F. D., van Zalk, M., Meeus, W. H., & van Aken, M. A. (2011). Antecedents and consequences of peer-rated intelligence. *European Journal of Personality*, *25*, 108-119. doi:10.1002/per.799

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 71–75. doi:10.1207/s15327752jpa4901_13

Dufner, M., Arslan, R. C., Hagemeyer, B., Schönbrodt, F. D., & Denissen, J. J. A. (2015). Affective contingencies in the affiliative domain: Physiological assessment, associations with the affiliation motive, and prediction of behavior. *Journal of Personality and Social Psychology, 109*(4), 662–676. doi:10.1037/pspp0000025

Dufner, M., Denissen, J. J. A., Van Zalk, M., Matthes, B., Meeus, W. H. J., van Aken, M. A. G., & Sedikides, C. (2012). Positive intelligence illusions: On the relation between intellectual self-enhancement and psychological adjustment. *Journal of Personality, 80*, 537-571. doi:10.1111/j.1467-6494.2011.00742.x

Geukes, K., Breil, S. M., Hutteman, R., Küfner, A. C. P., Nestler, S., & Back, M. D. (2017). *Explaining the longitudinal interplay of personality and social relationships in the laboratory and in the field: The PILS and CONNECT study*. Manuscript in preparation.

Glaesmer, H., Hoyer, J., Klotsche, J., & Herzberg, P. Y. (2008). Die deutsche Version des Life-Orientation-Tests (LOT-R) zum dispositionellen Optimismus und Pessimismus [The German version of the Life-Orientation-Test (LOT-R) for dispositional optimism and pessimism]. *Zeitschrift Für Gesundheitspsychologie, 16*(4), 26–31. doi:10.1026/0943-8149.16.1.26

Hautzinger, M., & Bailer, M. (1993). *Allgemeine Depressions Skala (ADS).* Weinheim, Germany: Beltz.

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der 'Positive and Negative Affect Schedule' (PANAS) [Investigations with a German version of the Positive and Negative Affect Schedule (PANAS)]. *Diagnostica, 42*(2), 139–156.

Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B* (5th ed.). Balingen: Spitta Verlag.

Pelham, B. W., & Swann, W. B. (1989). From self-conceptions to self-worth: On the sources and structure of global self-esteem. *Journal of Personality and Social Psychology, 57*(4), 672–680. doi:10.1037/0022-3514.57.4.672

Raskin, R. N. & Hall, C. S. (1979). A narcissistic personality inventory. *Psychological Reports, 45*(2), 590. doi:10.2466/pr0.1979.45.2.590

Raskin, R. N. & Hall, C. S. (1981). The Narcissistic Personality Inventory: Alternative form reliability and further evidence of construct validity. *Journal of Personality Assessment, 45*(2), 159–162. doi:10.1207/s15327752jpa4502 10

Raven, J. C., Raven, J., & Court, J. H. (1962). *Advanced progressive matrices set II*. Oxford, England: Oxford Psychologists Press.

Richard, F. D., Bond Jr., C. F., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology, 7*, 331-363.

Rosenberg, M. (1965). *Society and the adolsecent self-image*. Princeton, NJ: Princeton University Press.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*(3), 493–502.

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67*, 1063–1078.

Schumacher, J., Klaiberg, A., & Brähler, E. (2003). *Diagnostische Verfahren zu Lebensqualität und Wohlbefinden* [Diagnostic techniques for the assessment of well-being and quality of life]. Göttingen, Germany: Hogrefe.

Schütz, A., Marcus, B., & Sellin, I. (2004). Die Messung von Narzissmus als Persönlichkeitskonstrukt: Psychometrische Eigenschaften einer Lang- und einer Kurzform des Deutschen NPI (Narcissistic Personality Inventory) [Measuring narcissism as a personality construct: Psychometric properties of a long and a short version of the German Narcissistic Personality Inventory]. *Diagnostica, 50*, 202–218. doi:10.1026/0012-1924.50.4.202

Selfhout, M., Denissen, J. J. A., Branje, S., & Meeus, W. H. J. (2009). In the eye of the beholder: Perceived, actual, and peer-rated similarity in personality, communication, and friendship intensity during the acquaintanceship process. *Journal of Personality and Social Psychology, 96*(6), 1152–1165. doi:10.1037/a0014468

von Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg [A revised version of the German adaptation of Rosenberg's self-esteem scale]. *Zeitschrift für Differentielle und Diagnostische Psychologie, 24*(1), 3–7. doi:10.1024//0170-1789.24.1.3

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. doi:10.1037/0022-3514.54.6.1063

Watson, D. & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - expanded form*. Iowa City: University of Iowa. doi:10.1037/0022-3514.65.4.836

## Footnotes

[1]About half of the participants (*N* = 110) were randomly assigned to an intervention, which took place after the completion of Wave 1. Participants in the intervention group received either a letter or an e-mail containing a large package of information about numerous personality test scores and among these was the Picture Story Exercise (PSE). A small amount of PSE scores were accidentally matched to the wrong persons when the feedback was provided. After the completion of the Wave 2 assessments, we corrected this mistake and provided feedback on the actual PSE scores (averaged across the two waves).