

## Antecedents and consequences of peer-rated intelligence

Jaap J. A. Denissen

Humboldt-University Berlin

Felix D. Schönbrodt

LMU München

Maarten van Zalk

Örebro University

Wim H. J. Meeus & Marcel A. G. van Aken

Utrecht University

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

Publisher link: <http://onlinelibrary.wiley.com/doi/10.1002/per.799/abstract>

### Abstract

The current study investigated the antecedents and consequences of peer-rated intelligence in a longitudinal round robin design following previously unacquainted members of small student work groups. Social relations analyses were used to calculate target effects as an index of peer-reputations. In addition, self-ratings, university grades, and objective indicators of intelligence were assessed. Results indicated that peer-reputations were reliable and stable and also accurate to some degree, being correlated with objective intelligence. On the other hand, they were also shown to be biased in terms of correlating with interpersonal liking (increasing across time) and idiosyncratic rating tendencies (decreasing across time). Agreement between self-ratings and peer-reputations increased, but this was likely due to “constructivist” identity negotiation processes, as peer-reputations did not become more accurate over time. Finally, it was shown that peer-rated intelligence reputations have important real-life consequences, being correlated with changes in academic achievement across two 4-month periods and with dropout from university after 8 months. Overall, the pattern of results demonstrates the utility of a socioanalytic perspective in analyzing personality and social processes.

*Keywords:* Academic achievement, intelligence, social relations analysis, person perception, longitudinal study, interpersonal processes

Much of personality psychology still uses participants' subjective reports on their own and others' personality, even though such reports have been criticized as biased (Baumeister, Vohs, & Funder, 2007; Schwarz, 1999). In the case of intelligence, the situation is strikingly different: Since the days of Binet and Simon, researchers have mostly relied on achievement tests, the outcomes of which hardly depend on participants' self-views. Because intelligence tests have a long and successful history in terms of predictive validity and objectivity (Kaufman & Lichtenberger, 2005), it seems understandable to rely on them instead of subjective perceptions of intelligence (Paulhus, Lysy, & Yik, 1998). However, there are both psychometric and substantive reasons to focus on the antecedents and correlates of peer-rated intelligence, which we discuss in the following

From a psychometric perspective, peer-ratings of personality have been treated more favorably than self-ratings. According to Hofstee (1994), such aggregated ratings are more reliable and valid than self-ratings and should be used as the gold standard of personality assessment (at least when it comes to traits such as the Big Five). In addition, such aggregated peer-ratings can be seen as personality *reputations* with important interpersonal and group outcomes (Hogan, 1996). In the case of intelligence, such outcomes may manifest themselves in the following contexts. For example subjective perceptions of other people's ability may lead to important consequences when they are formed in contexts such as unstandardized job interviews, because future employers will tend to hire the applicant they see as most competent (Jones, 1989). Moreover, in educational settings, a classic finding is also that teachers' perceptions of pupils' ability (even if these are unfounded) are related to differential treatment and academic performance, with pupils who are seen as brighter being treated more favorably and obtaining higher grades (Rosenthal & Jacobson, 1968). Finally, being perceived as intelligent is seen as an important criterion in mate choice and may

therefore influence biological fitness (Miller, 2001; Prokosch, Coss, Scheib, & Blozis, 2009).

### **Antecedents of Intelligence Perceptions**

Given the practical and theoretical relevance of peer-rated intelligence, it becomes important to understand the origins of such perceptions. One overall framework to investigate this question is by applying a modified Brunswikian lens model analysis (Brunswik, 1956, see Figure 1) to the phenomenon (see Back 2011, this issue). According to this framework, latent personality traits give rise to valid behavioral cues (cue validity) that may be picked up by observers (cue utilization). On the other hand, observers's ratings may also be distorted by a number heuristics that focus attention to invalid indicators. Before we review potentially valid and invalid indicators, however, it is necessary to establish first whether peer-rated intelligence is related to measured intelligence in the first place (observer accuracy).

**Observer accuracy.** Regarding the accuracy of ability ratings, most research has focused on self-rated instead of peer-rated intelligence. An early meta-analysis by Mabe and West (1982) summarized effect sizes from 55 studies and reported a mean correlation of .29 with measured performance. Comparable estimates have been found in more recent studies (Furnham & Chamorro-Premuzic, 2004a; Paulhus et al., 1998). By comparison, less is known about the validity of peer-ratings. An important exception is the work conducted by Borkenau and his colleagues. For example, Borkenau and Liebler (1993) measured the intelligence of 50 target dyads (mostly romantic couples) with a standardized test battery and let them rate each other's intelligence using a single-item measure. Results indicated that partner ratings correlated .31 with measured intelligence (a correlation of .29 was found between self-ratings and measured intelligence), demonstrating a modest degree of validity.

In a study that is somewhat comparable to the current one, Paulhus and Morgan

(1997, Study 2) followed 103 third-year university students belonging to small (4-5 members) working groups, who met weekly for 20 minutes across a period of 7 weeks. Members were discouraged from interacting with each other outside of these group sessions. All participants rated their own and each other's intelligence twice, after 2 and 7 weeks of acquaintanceship. Thus, this study allows for an investigation of timely patterns of accuracy. Results indicated that after two weeks, the correlation between the results of the intelligence test (the short Wonderlic Personnel Test) and participants' subjective ratings was .20 and .12 for self- and peer-ratings, respectively. After seven weeks, this correlation had increased to .35 in the case of the self-ratings and .37 for the peer-ratings, thus being slightly higher than the level of accuracy that is usually found in the literature. It should be noted, however, that this may have been due to the fact that participants were explicitly requested to engage in discussions regarding their own personality (i.e., problem solving ability).

**Possible valid indicators of intelligence.** To choose plausible candidates as indicators of intelligence, it is useful to consider the core processes underlying intelligence. According to theoretical and laypeople's account, both general reasoning and verbal ability constitute such proximal skills (Sternberg, Conway, Ketron, & Bernstein, 1981). Such latent abilities are involved in solving new and complex problems and understanding and communicating difficult thoughts and ideas, respectively. Accordingly, an individual's demonstrated achievement in educational settings, especially when this setting involves verbal discussions, seems ideal as a cue to estimate this person's level of ability. Consistent with this, there exist a broad range of studies that provide evidence that high academic achievement is linked to subsequent increases in self-rated ability. Most of this research has focused on children in educational settings and found that grades predict changes in academic self-concept (Marsh & Craven, 2006; Marsh & O'Mara, 2008; Marsh, Trautwein, Ludtke, Koller,

& Baumert, 2005), which is closely related to self-rated intelligence.

Much less is known about the determinants of peer-rated intelligence, however, though it is plausible that peers would also rely on behaviors related to verbal fluency and academic achievement as cues. Regarding the latter, it could be argued that peer-ratings should be even more highly correlated with achievement than self-ratings, since peers typically have less information regarding factors that may attenuate the diagnostic value of this cue (e.g., if less intelligent people get good grades by investing a lot of effort). Accordingly, intelligence ratings by peers should be at least as strongly associated with grades than is the case for self-ratings. Relevant for this prediction, Gest, Rulison, Davidson, and Welsh (2008) assessed 427 elementary school pupils (i.e., between 6 and 9 years old) 5 times across a period of 3 years. The grade point average (GPA) was drawn from information on participants' report cards, whereas an aggregated measure of peer-rated intelligence was obtained via sociometric nominations. Results indicated that grades were strongly concurrently correlated with peer-rated intelligence (average  $r = .60$ ) and also predicted longitudinal changes in peer-rated intelligence (average cross-lagged path = .17).

The importance of verbal fluency cues for peer-ratings of intelligence has been demonstrated in a series of impressive studies by Borkenau and his colleagues. For example, Borkenau and Liebler (1995) recorded participants' behavior during a 90-second episode consisting of various short tasks and had 6 objective judges code the videos with regard to 48 behavioral attributes. Consistent with lay and expert theories of intelligence, measured intelligence correlated most strongly with indicators of verbal fluency, such as the use of easily understandable ( $r = .44$ ) and non-halting ( $r = .53$ ) speech. Interestingly, these cues were also correlated with unacquainted raters' judgments after seeing the video clips (with sound), with correlations of .22 and .32, respectively. However, partner-rated intelligence was not

significantly correlated with these valid cues. In another study (Borkenau, Mauer, Riemann, Spinath, & Angleitner, 2004), intelligence ratings of observers who watched participants read newspaper headlines aloud were correlated .53 with a mixed battery of intelligence subtests (including verbal ones) and .35 with the Raven progressive matrices (which is a measure of general intelligence). This suggests that peers can correctly infer the intelligence of others by relying on cues of verbal fluency.

**Socially constructed peer-rated intelligence reputations.** As seen above, peer-rated intelligence can be accurate to a certain degree, informed by valid cues such as verbal fluency and academic achievement. On the other hand, having a reputation of being intelligent in the eyes of one's peers may be a phenomenon that is also somewhat independent from actual ability. In that regard, peer-reputations may partly represent a social construction that is informed by cues that are not linked to more objective, psychometric “reality” and may therefore be termed invalid. For example, academic achievement is an imperfect cue for intelligence since it is also dependent on personality factors like conscientiousness (e.g., Wagerman & Funder, 2007). In addition, in the study by Borkenau and Liebler (1995), partner ratings of intelligence were correlated with invalid indicators (i.e., not associated with measured intelligence) such as physical slenderness ( $r = .22$ ) and having a well-proportioned body ( $r = .26$ ). These invalid cues were also used by unacquainted raters, suggesting that intelligence ratings partly depend on a “what is beautiful is good” stereotype ascribing intelligence to good-looking individuals (Eagly, Ashmore, Makhijani, & Longo, 1991).

Another way in which peer-reputations may represent a social construction that is independent of psychometric measures of intelligence is that research has showed that individuals may engage in a process of identity negotiation when other people regard them in a different way than they view themselves (Swann, 1987). In the case of intelligence, which is

a highly positively valued characteristic, they may thus be especially determined to convince discrepant peers of the validity of their self-views, resulting in the successful modification of peer-ratings. We know of no naturalistic, longitudinal study that has investigated the ability of self-ratings of intelligence to predict corresponding peer-ratings (and vice versa), but it is certainly an interesting theoretical possibility.

### **Outcomes of Intelligence Perceptions**

As stated above, intelligence perceptions are theoretically and empirically related to a number of important outcomes. In the current section, we review empirical evidence to back up these predictions. Because only a very limited number of studies have investigated the outcomes of peer-rated ability, we also focus on the outcomes of self- and teacher-rated intelligence in this section. Regarding the former, research on academic self-concept has established that self-perceived ability is associated with changes in academic performance (Marsh & Craven, 2006; Marsh & O'Mara, 2008; Marsh et al., 2005). Theoretically, this causal ordering is termed “self-enhancement” because it is posed that perceptions of competence lead to subsequent improvement in corresponding ability, though it is not well-known what mediates this association.

In the area of teacher expectations, several investigators have found that perceptions of ability can be associated with changes in achievement. In the most prominent study in this regard, Rosenthal and Jacobson (1968) randomly provided teachers with bogus feedback regarding the ability of their pupils and found that positive teacher expectations were associated with superior academic achievement over time. Since then, a number of studies have tried to investigate the mediating behaviors that account for this effect of interpersonal expectations. According to a meta-analysis by Harris and Rosenthal (1985), four clusters of behaviors can be identified as such: Teachers with high expectations 1) display a

warmer interpersonal style, 2) express more positive feedback, and provide students with 3) richer input in terms of the quantity and difficulty of material and 4) more opportunities to display their skills (e.g., by interacting more frequently with students). Accordingly, it seems that teachers communicate expectations in subtle ways to their pupils that affect subsequent outcomes.

Finally, Gest and his colleagues recently provided the first empirical evidence that peer-rated ability is associated with subsequent changes in both self-concept (Gest, Domitrovich, & Welsh, 2005) and achievement (Gest et al., 2008; see also Hughes, Dyer, Luo, & Kwok, 2009). Unfortunately, however, it is not well-understood what processes mediate this association, as both changes in self-concept and effort have been empirically discounted as possible mediators in the study by Gest et al. (2008). Instead, these authors speculated that the effects of peer-rated intelligence may be mediated by influencing 1) achievement-related discourse between peers (e.g., by providing praise for academic achievements), 2) opportunities for informal peer tutoring that promote academic skills, and 3) selective peer affiliations (e.g., peers with a reputation for high intelligence may mutually support each other in academic situations leading to improved skills). Thus, as in the case of teach-rated ability, it may be that peer perceptions act as self-fulfilling prophecies by influencing communication patterns or by selectively reinforcing achievement-related behaviors, even if those expectations are not explicitly communicated and fail to leave a mark on self-ratings of intelligence.

### **The Current Study**

On the basis of the literature reviewed above, the current study set out to test the following hypotheses using a longitudinal round robin design. First, it was expected that peer-reputations of intelligence are reliable in terms of both interrater agreement (consensus) and

temporal stability (Hypothesis 1), and show a substantial degree of convergence with self-ratings (Hypothesis 2). Furthermore, it was expected that these peer-reputations show a substantial degree of accuracy as indexed by a significant correlation with measured intelligence (Hypothesis 3) and are correlated with verbal fluency and academic grades, which have been shown to be valid indicators in previous studies (Hypothesis 4). On the other hand, we expected these correlations to be less than perfect, since peer-rated intelligence reputations should also show significant correlations with indicators such as liking and people's own self-ratings (assumed similarity), which can be seen as resulting from social construction processes (Hypothesis 5). Furthermore, we expected the pattern of correlations as specified in Hypotheses 1-5 to change over time, with increasing validity and less dependence on liking as participants get to know each other better (Hypothesis 6). Finally, we expected peer-rated intelligence reputations to be significantly correlated with important longitudinal outcomes such as changes in academic achievement and dropout from university (Hypothesis 7).

The hypotheses listed above were tested using a powerful design, tracking previously unacquainted participants across three different measurements spanning a period of 8 months (i.e., almost an entire academic year). We assessed both self- and peer-reports and analyzed the latter using the social relations model (SRM) to disentangle different sources of variance underlying these ratings (D. A. Kenny & La Voie, 1984). Finally, we measured psychometric intelligence using a well-validated objective test and obtained information on academic achievement from university records. With these tools, we could go beyond previous research in a number of important ways. First, the current study goes beyond research in artificial laboratory settings by studying peer-ratings in a naturalistic context (i.e., university setting), in which both valid and invalid indicators of intelligence were available.

We also expanded experimental research by investigating important real-life outcomes of peer-rated intelligence. Second, existing naturalistic research by Gest and colleagues (2005; 2008) was expanded by focusing on young adults (who have been found to be more resistant to social influence; Steinberg & Monahan, 2007) and on university dropout as an outcome in addition to academic achievement per se. Third, we included possible indicators of intelligence to investigate the possible precursors to peer-reputations as well as measured intelligence to disentangle genuine social-constructivist processes from a possible kernel of (psychometric) truth could possibly underlie them.

## **Method**

### **Participants**

Participants were psychology freshmen attending Utrecht University in the Netherlands. For educational purposes, freshmen in the Netherlands are randomly placed in groups in which they work together during the entire academic year to complete a substantial part of the psychology curriculum. The students were assigned to one of the 20 introduction groups of around 25 people each ( $N = 489$ ). E-mails, flyers, posters, and an announcement during the first university lecture generated attention for the current study. A total of 378 participants (77% of all first-year students) stemming from 18 groups signed up for the study via a website. Of these, the 10 groups in which more than 80% of the participants registered for the current study were selected for participation. In these 10 groups of 238 individuals, 221 individuals registered for the current study (93% enrollment rate). The mean age of these individuals was 18.9 ( $SD = 1.6$ ) and 181 (82%) were females. The majority of participants (92%) were of Dutch origin. The participants that were selected ( $n=221$ ) and those that were not ( $n= 157$ ) did not differ significantly ( $p > .10$ ) in gender and age.

## **Procedure**

Starting the second week of their University freshmen year, participants filled out the questionnaires online by accessing a website using a personal password (Wave 1). After 4 months, a second wave of data collection took place at the end of the first semester (Wave 2). After an additional 4 months (one month before the final examinations of the second semester), participants completed a final wave of data collection (Wave 3). At Wave 1, all participants rated their own and their fellow group members' intelligence and liking using a round robin design. Single items were used because letting participants rate all other participants on multi-item scales would place too high demands on their motivational as well as cognitive abilities. Rating scales appeared in randomized order to avoid response sets. At Wave 3, participants completed a shortened version of the Raven Advanced Progressive Matrices (Raven, Raven, & Court, 1962). Confidentiality of all answers was explicitly guaranteed. The questionnaires used in the current study were part of a larger battery that took a median of 40 minutes to complete. Participants received 20€ (around \$25), 2 hours of course credit, and personality profile feedback at the end of the study.

## **Instruments**

*Perceived intelligence* was measured with the following item: “Some people are dull and not so intelligent, whereas other people are very intelligent and clever. Please rate yourself and your group members on this dimension.” A response scale between 1 (“not intelligent, dull”) and 7 (“very intelligent, clever”) was used.

*Objective intelligence* was assessed by using a 15-item (shortened) version of the Raven Advanced Progressive Matrices. This version was created by selecting all items with item-total correlations with the full Raven version higher than .35 according to the published manual). Note that 12 of these items were also included in the short version by Bors and

Stokes (1998). To boost compliance, participants were required to declare (by clicking a checkbox) that they would take the test in a quiet environment in which they could concentrate without being distracted by others. They were told that the test would take about 15-20 minutes, but that no time limits would be enforced (on average, participants took 20 minutes to complete the test,  $SD = 9.83$ ). Items consisted of a matrix of 3 x 3, with two rows of 3 and one row of 2 abstract figures that following logically on each other according to an abstract rule. In the lower right corner, an empty cell was presented. Participants were asked to pick (by selecting a radio button) one out of 8 alternative figures listed at the bottom of the webpage that would fit this empty cell. Before starting the test, two practice items with low difficulty were presented, which needed to be answered correctly to ensure that participants understood the format of the test. It was also stated that the items would be presented in order of increasing difficulty. Item difficulties ranged between .94 (the first item) and .43 (the last item). Items were aggregated to form a composite scale of cognitive ability, with a reliability of .71. The mean score on this scale was 10.72 ( $SD = 2.80$ ), while the median was 11. The distribution had a skew of -.51 ( $SE = .18$ ) and a kurtosis of -.20 ( $SE = .35$ ) and ranged between 3 and 15. Inspection of the distribution indicated that the most frequent score was 12.

*Liking* was assessed with the following item: "You don't like some people at all while you like other people very much. Please rate yourself and your group members on this dimension" (i.e., self-liking was also assessed but not used in the current study). Participants rated the likeability of each person on a Likert scale ranging from 1 (do not like him or her at all) to 7 (like him or her very much). The reliability of the aggregated liking score (across raters) ranged between .59 and .63 (calculated with the HLM software package estimating the reliability of the intercept of a model with target as Level 2 clustering variable).

*Clarity of speech* was assessed in Wave 3 only, using the following item: "Some

people are easy to follow and express themselves clearly, whereas other people are difficult to follow and express themselves in a confusing manner. Please rate yourself and your group members on this dimension". A rating scale between 1 (easy to follow, expresses him-/herself clearly) and 7 (difficult to follow, expresses him-/herself in a confusing manner) was used. Values were recoded so that high values represent high clarity of speech.

*Academic achievement during high school* was assessed at Wave 1 using a self-report item ranging from 1 to 10 (corresponding to the standard grading scale in the Netherlands). *In contrast, academic achievement in university* was directly obtained from university records for 190 participants (out of 194 remaining active in the current study) who had given us written permission to access this information. The curriculum consisted of 4 examination blocks, of which 3 were used in the current study (see Figure 1). In Block 1, which took place approximately 2 months after Wave 1 and 2 months before Wave 2, participants took the exams of Introduction to Psychology (History) and Methods and Statistics 1. Block 2 took place one week after Wave 2 and three months before Wave 3 and consisted of exams in Biological Psychology and Methods and Statistics 2. Finally, Block 3 took place 3 weeks after Wave 3 and consisted of exams in Personality Psychology and Social Psychology; the average of the corresponding grades was used as outcome variable for Wave 3.

### **Statistical Procedure**

Applying the SRM framework by Kenny (1994), proportions of variance within the SRM and individual perceiver and target effects were calculated. For this purpose, we used the package TripleR

(Schmukle, Schönbrodt, & Back, 2009) in the R Environment for Statistical Computing (R Development Core Team, 2008). As multiple groups were investigated, the

significance of variance components can be estimated by treating groups as the unit of analyses: Separate social relations estimates were derived for each group and weighted one-sample *t*-tests were employed to determine whether the mean of these estimates significantly differed from zero. As all variables were assessed using single items, relationship variance cannot be partitioned from error variance. In SRM analyses, group differences are removed from the target and perceiver effects, which makes them independent of group membership and appropriate for conventional least squares procedures that assume independence. However, other indices that were not derived from the SRM (like the self-rating of intelligence) were not corrected for group membership. If such indices are to be correlated with target or perceiver effects, partial correlations have to be employed, where the group effect is partialled out from the non-SRM variables by  $g-1$  dummy variables representing  $g$  groups (D. A. Kenny, Kashy, & Cook, 2006). In the remainder of the paper, all reported correlations and regressions are controlled for group differences in such a manner<sup>1</sup>.

## Results

### Social Relations Analyses and Variance Components

In a first step, IQ ratings were decomposed into relative perceiver, target, and relationship components (see Table 1). All variance components were highly significant, which allowed for investigating the individual perceiver or target effects in further analyses

---

1. Both predictors and criteria are controlled for group differences by a dummy coded variable indicating group membership. Group membership therefore is treated as a fixed effect. As group membership was randomly assigned, it could be argued that a mixed effects model with groups as a random effect would be more suitable, which, however, requires a sufficient number of groups. In the current study, additional analyses employing mixed effect models revealed that results are equivalent to fixed effects models.

(whereas in Table 1 relative variances are reported for ease of interpretation, the significance tests were performed on the absolute variances). The reliabilities for the perceiver effect were .93, .94, and .91 in the respective wave. The reliabilities of peer-reputations, which are indexed by the target effects, were .83, .89, and .86. Thus, peer-reputations of intelligence were reliable in terms of interrater agreement (consensus; Hypothesis 1).

The largest proportion of variance in all waves was the relationship component (which also included error variance). Besides that, the perceiver variance was approximately twice as large as the target variance. The amount of target variance (i.e., consensus) in IQ ratings (13-18%) is comparable to 15% of variance in interpersonal attraction (D. A. Kenny, 1994), it is somewhat smaller than the 30% that have been reported by Paulhus and Reynolds (1995) for both short- and longer-term round robin ratings of the Big Five.

(e.g., about 10-15% of variance in interpersonal attraction is attributable to the target; D. A. Kenny, 1994)

**Self-Other Agreement.** To investigate the self-other-agreement in intelligence ratings, the self-rating was correlated with the target effect within each wave. The resulting correlations demonstrate a significant and considerable amount of self-other-agreement (Hypothesis 2), which even rose with increasing acquaintance:  $r_s = .22, .44, \text{ and } .43$  ( $p_s < .01$ ), in the respective wave (see also Table 2). The correlation coefficient in Wave 1 significantly differed from the coefficients in Waves 2 and 3.

**Assumed similarity.** Assumed similarity, also known as the false consensus bias (Ross, Greene, & House, 1977), refers to the tendency to assume that other people think or feel as oneself, or, in the terms of the SRM, the tendency to perceive others in the same way as one perceives oneself. It is calculated by correlating the self-rating with the perceiver effect. In our study, both variables show a significant correlation in each wave:  $r_s = .33, .45,$

and .51 ( $ps < .001$ ; the correlation coefficients significantly differed between Waves 1 and 3).

This finding is consistent with Hypothesis 5 and indicates that persons who rate themselves as more intelligent also falsely tend to assume that others are also more intelligent.

### **Stability and concurrent validity of self- and peer reputations of intelligence.**

For a first examination of the interrelations of grades, IQ, self-, and peer-reputations (target effects), bivariate correlations between all measures were calculated (Table 2). As predicted in Hypothesis 1, peer-reputations have a high stability ( $r = .67$  to  $.87$ ), which is even larger than the stability of self-ratings ( $r = .50$  to  $.62$ ) and grades ( $r = .58$  to  $.62$ ).

Regarding concurrent validity, measured intelligence (Raven) is only related to grades in Wave 2, peer-reputations are associated with both measured intelligence and grades.

### **Usage and Validity of Indicators**

How accurate were peer-reputations of intelligence when compared to intelligence as measured by the Raven IQ test? Bivariate correlations between the target effect of intelligence ratings and the Raven score were 0.25, 0.27, and 0.22 in Waves 1 to 3 ( $ps < 0.001$ ), which was consistent with Hypothesis 3.

What informs peer-reputations of intelligence after participants get to know each other? Either objective markers (e.g., grades) or relatively unbiased reputations, like clarity of speech, could be used as indicators. In this regard, two concepts have to be distinguished: usage and validity. While some possible indicators are valid, but are not used for impression formation, some other indicators might be used, that are invalid. To assess both dimensions, we ran multiple regressions within Waves with exam grades (only in Waves 2 and 3), the intelligence self-rating, liking, and clarity of speech target effect (only Wave 3) as predictors. In one set of regressions, the peer-rating served as dependent variable; significant regression weights would suggest that peers use the indicator in question (or some other indicator that is

directly correlated with it), especially in cases in which peers are not able to influence the indicator (such as when independent observers code the indicators, which was not the case in the current study). In the other set of regressions, the Raven score served as dependent variable; significant regression weights would suggest that the indicator in question is a valid predictor of intelligence (see Table 3).

Grades emerged as invalid predictors of intelligence, indicating that they are not driven by differences in measured intelligence in the first year of the psychology curriculum. Notwithstanding, they were strongly used by participants to inform intelligence ratings of their peers (Hypothesis 5). Only intelligence self-ratings of the target emerged as both valid and used correlates (we are hesitant to call them cues as they cannot be directly observed by outside observers) of intelligence ratings. Although verbal fluency has been found as a marker of general intelligence (Borkenau et al., 2004), clarity of speech turned out to be an invalid predictor of intelligence, though it was robustly associated with peer-reputations (which is consistent with Hypothesis 4)(and verbal fluency has also been found as a marker of general intelligence; Borkenau et al., 2004)(and verbal fluency has also been found as a marker of general intelligence; Borkenau et al., 2004). Liking was an invalid indicator that is used after a short acquaintance (Hypothesis 5), but dropped after longer acquaintance (Hypothesis 6). Concerning the overall variance explained by the regression models, the models for indicator usage were much stronger ( $R^2$  up to .51) than the models for indicator validity ( $R^2 = .04$ ).

### **Predicting Academic Success**

**Cross-lagged effects between self-ratings, peer-reputations, and grades.** As has already been shown in the paragraph about self-other-agreement, the intercorrelation of self and peer-reputations increases over time. To analyze whether this increasing convergence is driven by an assimilation of self-ratings to peer-reputations or vice versa, a cross-lagged

path analysis was conducted. To further extend the model, the cross-lagged effect of grades was included: how do self and peer-reputations at Waves 1 to 3 predict the grades assessed after these ratings? Do grades (which are certainly known to one-self and maybe known to peers) in term predict self-ratings and peer-reputations of intelligence assessed after the exams?

Following models were fit using the OpenMx package (Boker et al., 2010): a) a baseline model which incorporates only the temporal stability coefficients for self and peer-reputations and for grades ( $M_{\text{stability}}$ ), b) a model with cross-lagged paths from self-ratings to grades and grades to self-ratings ( $M_{\text{self}}$ ), c) a model with cross-lagged paths from peer-reputations to grades and grades to peer-reputations ( $M_{\text{peer}}$ ), d) a full cross-lagged model with cross-lagged paths between self and peer-reputations, as well as between self and peer-reputations to grades and back ( $M_{\text{full}}$ ), e) a reduced model, where all insignificant paths from the full model were removed ( $M_{\text{reduced}}$ ).

The final model ( $M_{\text{reduced}}$ ) showed the best fit concerning several indices ( $\chi^2 = 67.9$ ,  $p < .001$ ; RMSEA = .100, 95% CI [.068, .133]; CFI = .953). If grades were controlled for measured intelligence (i.e., the Raven score), results did not change (all paths from Raven to grades are insignificant). Path coefficients (see Figure 1) indicated low but significant cross-correlations between self and peer-reputations between Wave 1 and Wave 2. More importantly, there were significant paths from peer-reputation to grades, even after controlling for the stability in grades (Hypothesis 7). Except a weak path from grades (Block 1) to self-ratings (Wave 2), self-ratings were not connected to grades or vice versa. While the proposed structural model suggests a causal influence from peer-reputations to grades, a possible alternative explanation could include a third variable influencing both peer-reputations and grades. Possible variables could be the time spent learning or the time spent at the university,

which would influence both peer-reputations and grades. Both variables have been assessed in the study at each wave, and supplementary analyses which included these two variables into the model (with paths from the new variables to peer-reputations and subsequent grades, as well as stability paths between the new variables) showed that the paths from peer-reputations to grades did not diminish. Furthermore, model fit (as measured by the CFI) got worse after inclusion of the new variables. Hence, although the existence of other third variables never can be ruled out, the current results so far support the interpretation of a significant path from peer-reputations to subsequent grades.

**Student dropout.** Students that quit the university could be identified by the fact that they were no longer rated by their peers at Wave 3. In the current sample, 25 participants dropped out from university (11.3% of the sample). To examine whether intelligence reputations could predict future dropout, we ran a logistic regression (0=continuing, 1=dropout) with intelligence peer-reputation and intelligence self-rating from Wave 1 as linear predictors. As the Raven test was applied at Wave 3, no scores are available for dropout participants. Likewise, as many students already dropped out before Wave 2, these analyses could only be run for variables assessed at Wave 1 (at this time, students had not received any grades yet, so we could not control for actual academic achievement).

Logit estimates were  $-0.97$  ( $p = .04$ ) for peer-reputations, but not significant for self-ratings ( $p = .30$ ), with one-tailed tests of significance. This finding indicates that future drop-out students were evaluated lower on their intelligence at the beginning of the semester (even after mere two weeks of acquaintanceship) than their continuing peers. Noteworthy, the intelligence self-rating was not a significant predictor of student dropout.

### **Discussion**

The current study set out to investigate the antecedents and consequences of peer-

rated intelligence using a round robin design that tracked ratings of previously unacquainted individuals across a period of eight months. Furthermore, we also assessed self-ratings of intelligence as well as objective intelligence and documented academic achievement, thus relying on multiple independent sources. We addressed and confirmed the hypotheses, that peer-reputations about intelligence are consistent, temporally stable, and correlated to self-ratings. Furthermore we could confirm the hypothesis that peer-reputations are partly based on valid indicators (self-rating) and partly on invalid indicators (liking, clarity of speech, and grades), even though the validity of individual indicators did not always turn out as expected (e.g., we expected grades to be a valid indicator of intelligence). Finally, we could show that peer-reputations predict grades (which were assessed after the assessment of the peer-reputations), and the future dropout of students could be predicted by peer-reputations of intelligence after only two weeks of acquaintance. All in all, our results give rise to the following five broad conclusions.

First of all, peer-rated intelligence was found to be reliable in terms of inter-rater agreement and stability. Because they are agreed-upon by group members, they can be seen as veritable *reputations* of intelligence. This also means that such ratings can be meaningfully aggregated to form more reliable indices of personality, as suggested by Hofstee (1994). The fact that intelligence ratings were also temporally stable suggests that people to some degree are “stuck” with being known for high or low intelligence. In fact, even after a mere two weeks of interacting with each other, group members already had reputations that correlated highly with peer-reputations eight months later, suggesting that intelligence impressions arise fairly quickly (Ambady & Rosenthal, 1992) and may be difficult (but not impossible; see below) to change.

Second, these peer-rated intelligence reputations "matter" as they predicted

important real-life outcomes in our study. Specifically, these reputations were correlated with changes in academic achievement between Waves 1 and 2, and between Waves 2 and 3 (i.e., across two four-month-periods), even after controlling for initial levels of achievements. In other words, if a participant was initially seen as intelligent by their fellow group members, this was associated with subsequent increases in academic achievement, thus living up to their reputations. This finding is consistent with results by Gest et al. (2008) and Hughes et al. (2009), though we are the first ones to demonstrate this effect in an older and more selected sample (see limitations section below). We are also the first authors to control for objective intelligence in demonstrating associations with academic achievement. By doing so, we were able to show that the consequences of reputations survived a statistical control for measured ability, which is inconsistent with the possibility put forward by Gest et al. (2008) that peer-reputations are more accurate. After all, the association between peer-reputations and intelligence would have been driven by their shared association with measured ability, which would have rendered the statistical effect of peer-reputations statistically insignificant. The fact that this did not happen suggests that peer-reputations may exist as a socially constructed reality independent of psychometric reality.

In addition to uncovering an association between peer-rated intelligence and changes in academic achievement, we are the first to demonstrate that such reputations are also associated with participants quitting their studies. Accordingly, it can be said that not only did students accommodate to social expectations that were held in their social setting (which may be rephrased as a kind of social conformity to “fit in with the group”), these expectations may have played a role in their decision to leave this setting altogether. This was likely a decision not made easily, since it has important financial and job-related consequences to quit or change one's studies. It is all the more remarkable that such

consequences were predicted by impressions that were made after only 2 weeks of acquaintanceship, with no real opportunity for peers to observe each other's academic ability in the form of grades.

Of course, our results are consistent with the classic finding from Rosenthal and Jacobson (1968), who found that pupils of teachers who had been manipulated into (incorrectly) believing that they were more intelligent than others subsequently achieved better academically. Like in the current study, these initial interpersonal expectations acted as a self-fulfilling prophecy that has been shown to depend on a warmer interpersonal style, the expression of more positive feedback, and the provision of students with richer input and opportunities to display their skills. Although we unfortunately did not assess such possible mediators, our results are consistent with the existence of such processes in the case of peer-rated intelligence as well. Gest et al. (2008) raised the possibility that the effect of peer intelligence reputations are driven by similar processes such as the provision of more positive feedback to peers who are deemed intelligent and the fact that such individuals are sought out for peer tutoring, which has been shown to stimulate academic achievement. More research is needed to empirically test these possible mediators as observed in everyday peer interactions.

Third, and consistent with our hypotheses, the current study found that peer-reputations of intelligence were based on grades and verbal fluency, but these indicators unexpectedly turned out to be invalid. Specifically, these indicators did not correlate significantly with objective intelligence as measured with the Raven test, which goes against lay and scientific theories of intelligence (Sternberg et al., 1981). What could be driving this result? Of course, part of the explanation may lie in the fact that the Raven test is a marker of fluid intelligence, whereas verbal fluency may be considered more a marker of crystallized intelligence (though it should be noted that verbal behaviors correlated with the Raven in the

study by Borkenau et al., 2004). Apart from this psychometric explanation, it needs to be acknowledged that whereas objective intelligence has been consistently found to predict academic achievement in high school students (e.g., Laidra, Pullmann, & Allik, 2007), a more mixed picture emerges regarding associations with achievement in university students. Specifically, some authors reported significant correlations, (Farsides & Woodfield, 2003; Furnham & Chamorro-Premuzic, 2004b; Lounsbury, Sundstrom, Loveland, & Gibson, 2003), but these are usually lower than the correlations found between personality factors (typically: conscientiousness) and achievement, and other authors failed to replicate them altogether (Furnham, Chamorro-Premuzic, & McDougall, 2003).

Whereas these cited investigations studied the association between intelligence and achievement in British and American students, one group of researchers did so using Dutch psychology students and is thus the most relevant benchmark for the current study. Specially, Busato, Prins, Elshout, and Hamaker (2000) correlated measured intelligence with academic achievement and found only a small and inconsistent association ( $r$  between .03 and .15). Instead, correlations with the personality factor conscientiousness were more consistent and larger ( $r$  between .16 and .21), leading them to conclude that “the lesson for students from the present research [...] might be that it actually does not matter if you use elaborate processing strategies [i.e., contingent on measured intelligence], as long as you just work hard and conscientiously enough” (Busato et al., 2000, p. 1064). Consistent with this, a correlation of .31 ( $p < .01$ ) was found in the current study between self-reports of conscientiousness and average achievement. Accordingly, at least in the Netherlands, it seems that it is falsely assumed that success in university is dependent to a large degree on objective intelligence, whereas it seems actually more dependent on effort and motivation and thus more in line with Dweck's “incremental” notion of academic achievement (e.g., see

Blackwell, Trzesniewski, & Dweck, 2007). Of course, stronger correlations may have been found with samples with less range restrictions, although it should also be noted that the intelligence distribution did not suggest a strong ceiling effect.

Fourth, the current results indicated that agreement between self- and peer-reports of intelligence increased across time. As such, they are consistent with past research suggesting that there is a positive association between the length of acquaintance between targets and judges and their level of agreement with each other (e.g., Funder & Colvin, 1988; Funder, Kolar, & Blackman, 1995). In the past, such increasing agreement has been ascribed to corresponding accuracy levels, as for example suggested by (Funder & Colvin, 1988, p. 149): “This result implies that agreement among acquaintances' judgments must derive at least partly from experience with and observation of the person who is judged.” In contrast, our results suggests that increasing agreement across acquaintance is not rooted in increased validity of peer-reputations, as there was no corresponding increase in correlation with objective intelligence. Rather, the results of the current cross-lagged analyses provide direct support for the existence of identity negotiation mechanisms (Swann, 1987), as there were reciprocal associations between self- and peer-reports of intelligence (controlling for objective intelligence) between Wave 1 and 2. Between Wave 2 and 3, by comparison, there was only a significant association between initial peer-reputations and changes in self-rated intelligence, suggesting that individuals may conform to group impressions after some degree of interaction.

On first sight, the current findings may thus be most consistent with a social constructivist view of reputation negotiation, with personality impressions being most dependent on social persuasion processes. However, it should also be noted that both self- and peer-rated intelligence were significantly correlated with objective intelligence in all

measurement waves, suggesting that socially constructed intelligence perceptions also contain a certain degree of accuracy. In fact, a lens model analysis (Brunswik, 1956) suggested that self-rated intelligence is the only measured indicator that was both used by peers and valid in terms of its correlation with measured intelligence. Accordingly, individuals demonstrated some degree of self-insight into their objective level of intelligence, and peers managed to pick up on this “kernel of truth” to form their own impressions of these individuals, regardless of any additional sources of bias these ratings were also found to contain in the current study.

Fifth, and finally, the overall conclusion of the current study is that peer-rated intelligence reputations can be seen as both accurate and invalid. Specifically, whereas they correlated with objective intelligence, they also correlated (and more strongly so) with sources of bias such as liking and assumed similarity. Interestingly, these sources of bias showed a striking pattern of longitudinal development, with liking being most strongly associated with intelligence reputations during the initial phase of acquaintanceship, after which this association decreased in strength until the time of the second assessment and disappeared altogether by the time of the third assessment. Accordingly, individuals seemed to rely less on generalized positivity impressions as they become acquainted. In contrast, assumed similarity effects increased across time and were strongest at the third assessment. This effect is calculated by correlating participants' self-rating with their perceiver effect and thus represents a more idiosyncratic, person-centered source of bias. For example, a person who considers him- or herself to be low on intelligence would “project” this low rating into other people in general, regardless of whether they like this person (although we cannot rule out the possibility that the effect is due to response tendencies affecting both self- and other-ratings).

In addition, an unexpected source of bias was detected in the current study, which also grew in terms in strength as participants became more acquainted. Specifically,

intelligence reputations became increasingly dependent on participants university grades, which were (as stated above) not valid indicators of intellectual ability because they were not correlated with objective intelligence. In a way, however, we believe it is not too much of a stretch to suggest another form of “accuracy” for peer-reputations of intelligence. That is, such reputations at least created a "social reality" that was more powerful than psychometric, “objective” reality in the form of the IQ test, as the former predicted changes in achievement and even dropout from university, whereas the latter did not have any corresponding effects. In a way, this “social reality” was therefore much more of a “press” for behavior than either objective intelligence or self-ratings of intelligence, thus reaffirming the utility of a socio-analytic approach to personality research (Hogan, 1996).

### **Limitations**

Of course, the current study also suffers from some limitations, which we are not afraid to admit. One cluster of limitations concerns our rather specific sample, consisting of Dutch first-year university students of psychology, who were predominantly female. For example, it may be that academic achievement is more related to objective intelligence in other places around the world or even in other academic fields within the Netherlands. An obvious concern is also the issue of restriction of range, as we focused only on university students, who usually have above-average intelligence. However, objective intelligence still correlated significantly with intelligence perceptions, suggesting that we had enough variation to produce effects; in fact, the existence of some range restriction makes it all the more impressive that we were able to predict important outcomes like changes in grades and dropout from university. A more valid concern in this context is that the work groups that were studied were of course not necessarily representative of “normal” social groups, as they were especially focused on academic discussions, in the context of which members received

performance feedback. In more informal and emotion-focused groups without such achievement feedback, reputations may be based on entirely different cues.

The second cluster of limitations concerns the fact that our measurements focused on somewhat narrow operationalizations of key constructs. For example, we used only a single item to let people rate each other's intelligence, so we could not operationalize intelligence perceptions in a multifaceted or hierarchical way. That said, single items are usually less reliable, making it less likely to find effects in general and more impressive that we found some nevertheless. It should also be noted that employing a round robin design in larger groups put lots of demands on the time and concentration of participants, so it is often necessary to rely on very short scales (Denissen, Geenen, Selfhout, & van Aken, 2008). Furthermore, we assessed intelligence using only a short test of fluid ability, using an online assessment procedure. Even though this produced a reliable intelligence score that correlated significantly with theoretically related constructs (such as self-ratings and peer-reputations) and did not seem to be affected that severely by ceiling effects, different results may have been found if a more diverse battery of cognitive tests under more controlled circumstances is used.

Finally, a more fine-grained picture of the antecedents and consequences of peer-rated intelligence may have emerged if we had extended the scope of our investigation to cover all principles of the PERSOC model (Back et al, 2011, this issue). For example, we only managed to time the first assessment after participants had already interacted for two weeks, which prohibited the investigation of cues at zero-acquaintance. Future investigations would ideally have participants rate each other's intelligence using only photos as cues, and perhaps a second time in a getting-acquainted setting without previous interaction. Furthermore, it would have been preferable to code more objective cues using video- or audio-taped

interactions using neutral observers in order to find out what behaviors are associated with impressions of intelligence (Principle 3), how these behaviors are taken up by the participants of social interactions (Principle 4) and lead to a change in relationship dispositions (e.g., admiring another person because of a supposedly high intelligence level (Principle 1) that may in turn even affect the way in which member of the social group are treated (Principle 2). In this context, it is telling that Rosenthal, who was the first one to discover the effect of interpersonal expectations of competence in teachers (Rosenthal & Jacobson, 1968), went on to establish the paradigm of “thin slices” of behavior to study the proximal behavioral cues related to such expectations (Ambady & Rosenthal, 1992; Harris & Rosenthal, 1985). Such efforts are needed to uncover the proximal processes that drive the effects of interpersonal ability expectations by peers as well.

### References

- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256-274.
- Baumeister, R., Vohs, K., & Funder, D. (2007). Psychology as the science of self-reports and finger movements. *Perspectives on Psychological Science, 2*, 396-403.
- Blackwell, L., Trzesniewski, K., & Dweck, C. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*, 246-263.
- Boker, S., Neale, M., Maes, H., Metah, P., Kenny, S., Bates, T., Estabrook, R., et al. (2010). OpenMx: The OpenMx Statistical Modeling Package. R package version 0.2.5-1050. Retrieved from <http://openmx.psyc.virginia.edu>
- Borkenau, P., & Liebler, A. (1993). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*, 546.
- Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*, 1.
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599-614.
- Bors, D., & Stokes, T. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement, 58*, 382.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments (2d ed.)*. Berkeley: University of California Press.

- Busato, V., Prins, F., Elshout, J., & Hamaker, C. (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. *Personality and Individual Differences, 29*, 1057-1068.
- Denissen, J., Geenen, R., Selfhout, M., & van Aken, M. (2008). Single-item big five ratings in a social network design. *European Journal of Personality, 22*, 37.
- Eagly, A., Ashmore, R., Makhijani, M., & Longo, L. (1991). What is beautiful is good, but...: A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110*, 109-128.
- Farsides, T., & Woodfield, R. (2003). Individual differences and undergraduate academic success: The roles of personality, intelligence, and application. *Personality and Individual Differences, 34*, 1225-1243.
- Funder, D., & Colvin, C. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*, 149-158.
- Funder, D., Kolar, D., & Blackman, M. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of Personality and Social Psychology, 69*, 656.
- Furnham, A., & Chamorro-Premuzic, T. (2004a). Estimating ones own personality and intelligence scores. *British Journal of Psychology, 95*, 149-160.
- Furnham, A., & Chamorro-Premuzic, T. (2004b). Personality and intelligence as predictors of statistics examination grades. *Personality and Individual Differences, 37*, 943-955.
- Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences, 14*, 49-66.

- Gest, S. D., Domitrovich, C. E., & Welsh, J. A. (2005). Peer academic reputation in elementary school: Associations with changes in self-concept and academic skills. *Journal of Educational Psychology, 97*, 337-346.
- Gest, S. D., Rulison, K. L., Davidson, A. J., & Welsh, J. A. (2008). A reputation for success (or failure): the association of peer academic reputations with academic self-concept, effort, and performance across the upper elementary grades. *Developmental Psychology, 44*, 625-635.
- Harris, M., & Rosenthal, R. (1985). Mediation of interpersonal expectancy effects: 31 meta-analyses. *Psychological bulletin, 97*, 363-386.
- Hofstee, W. (1994). Who should own the definition of personality? *European Journal of Personality, 8*, 149-162.
- Hogan, R. (1996). A socioanalytic perspective on the five-factor model. In J. S. Wiggins (Ed.), *The five factor model of personality: Theoretical perspectives* (pp. 163-179). New York: Guilford.
- Hughes, J. N., Dyer, N., Luo, W., & Kwok, O. M. (2009). Effects of peer academic reputation on achievement in academically at-risk elementary students. *Journal of applied developmental psychology, 30*, 182.
- Jones, E. (1989). The framing of competence. *Personality and Social Psychology Bulletin, 15*, 477.
- Kaufman, A., & Lichtenberger, E. (2005). *Assessing adolescent and adult intelligence*. Hoboken: John Wiley & Sons.
- Kenny, D. A. (1994). *Interpersonal perceptions: A Social Relations analysis*. New York: Guilford.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York:

Guilford.

Kenny, D. A., & La Voie, L. (1984). The Social Relations Model. *Advances in Experimental Social Psychology, 18*, 142-182.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences, 42*, 441-451.

Lounsbury, J., Sundstrom, E., Loveland, J., & Gibson, L. (2003). Intelligence, "Big Five" personality traits, and work drive as predictors of course grade. *Personality and Individual Differences, 35*, 1231-1239.

Mabe, P., & West, S. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*, 280-296.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science, 1*, 133-163.

Marsh, H. W., & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin, 34*, 542.

Marsh, H. W., Trautwein, U., Ludtke, O., Koller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development, 76*, 397-416.

Miller, G. (2001). *The mating mind*. New York: Anchor.

Paulhus, D., Lysy, D., & Yik, M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality, 66*, 525-554.

- Paulhus, D., & Morgan, K. (1997). Perceptions of intelligence in leaderless groups: The dynamic effects of shyness and acquaintance. *Journal of Personality and Social Psychology, 72*, 581-591.
- Paulhus, D., & Reynolds, S. (1995). Enhancing target variance in personality impressions: Highlighting the person in person perception. *Journal of Personality and Social Psychology, 69*, 1233-1242.
- Prokosch, M., Coss, R., Scheib, J., & Blozis, S. (2009). Intelligence and mate choice: Intelligent men are always appealing. *Evolution and Human Behavior, 30*, 11-20.
- R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: Basic Books.
- Raven, J. C., Raven, J., & Court, J. H. (1962). *Advanced progressive matrices set II*. Oxford: Oxford Psychologists Press.
- Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3*, 16-20.
- Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279-301.
- Schmukle, S. C., Schönbrodt, F. D., & Back, M. D. (2009). TripleR: A package for round robin analyses using R (version 0.3). Retrieved from <http://www.persoc.net/Toolbox/TripleR>
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.
- Steinberg, L., & Monahan, K. (2007). Age differences in resistance to peer influence. *Developmental psychology, 43*, 1531-1543.

Sternberg, R., Conway, B., Ketron, J., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology*, *41*, 37-55.

Swann, W. (1987). Identity negotiation: Where two roads meet. *Journal of Personality and Social Psychology*, *53*, 1038-1051.

Wagerman, S., & Funder, D. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality*, *41*, 221-229.

Table 1. *Variance components and descriptives for round robin intelligence ratings*

	Wave 1	Wave 2	Wave 3
Sample size	220	202	194
Mean (SD) perceiver effect	4.81 (0.67)	4.84 (0.74)	4.86 (0.71)
Mean (SD) target effect	4.80 (0.46)	4.84 (0.56)	4.86 (0.59)
Mean (SD) self-rating	5.13 (0.89)	5.20 (0.89)	5.19 (0.9)
Perceiver variance (relative)	0.34	0.36	0.30
Target variance (relative)	0.13	0.18	0.18
Relationship + error variance (relative)	0.54	0.46	0.52
Total absolute variance	1.16	1.42	1.37

*Note:* All variance components significantly differ from zero with  $p < .001$ .

Table 2: *Bivariate correlations between grades, Raven score, self-ratings, and peer-reputations in Waves 1 to 3.*

	1	2	3	4	5	6	7	8	9
1. Peer-reputation, W1	-								
2. Peer-reputation, W2	0.67	-							
3. Peer-reputation, W3	0.66	0.87	-						
4. Self-rating, W1	0.22	0.28	0.27	-					
5. Self-rating, W2	0.31	0.44	0.39	0.50	-				
6. Self-rating, W3	0.29	0.43	0.43	0.48	0.62	-			
7. Grade W1	0.38	0.64	0.61	0.11	0.24	0.25	-		
8. Grade W2	0.34	0.53	0.52	0.12	0.22	0.21	0.58	-	
9. Grade W3	0.30	0.52	0.53	0.14	0.14	0.26	0.51	0.62	-
10. Raven	0.25	0.27	0.22	0.18	0.21	0.23	0.08	0.16	0.07

*Note:* All correlations  $> .15$  are significant ( $p < .05$ ). W1-3 = Wave 1-3.

Table 3: *Standardized regression weights of multiple regressions. Validity was operationalized by the Raven score; usage was operationalized by the intelligence target effect*

Indicator	Validity			Usage		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Grades	-	0.05	0.06	-	0.53***	0.48***
Intelligence self-rating	0.19**	0.19*	0.21*	0.26***	0.31***	0.21***
Liking	0.12	0.10	0.01	0.19**	0.15*	0.00
Clarity of speech	-	-	-0.02	-	-	0.29***

*Note.* Adjusted R<sup>2</sup> for the model predicting the Raven score (validity) was 0.04 in each of

Waves 1 to 3, and 0.10, 0.51, and 0.51 for the model predicting the target effect (usage).

\*  $p < .05$ . \*\*  $p < .01$ . \*\*\*  $p < .001$ .

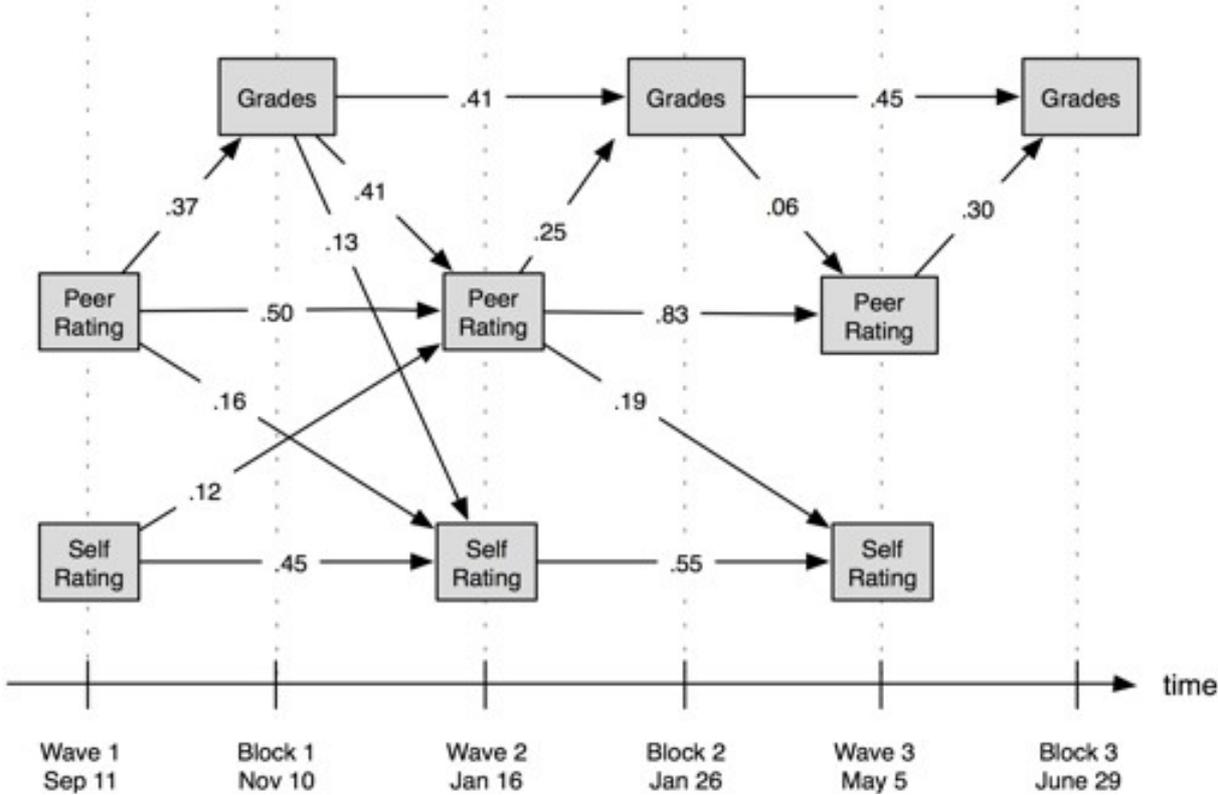
Table 4. *Goodness-of-fit indices and chi-square difference tests of nested structural models of grades and intelligence ratings.*

<b>Model</b>	<b>AIC</b>	<b>BIC</b>	<b><math>\chi^2</math></b>	<b><i>df</i></b>	<b>Comparison</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>
$M_{\text{stability}}$	185.5	41.8	245.5	30			
$M_{\text{self}}$	171.2	43.1	221.2	25	$M_{\text{stability}}$ vs. $M_{\text{self}}$	24.2***	5
$M_{\text{peer}}$	110.4	12.7	160.4	25	$M_{\text{stability}}$ vs. $M_{\text{peer}}$	85.1***	5
$M_{\text{reduced}}$	25.9	-22.7	67.9	21	$M_{\text{peer}}$ vs. $M_{\text{reduced}}$	92.5***	4
$M_{\text{full}}$	34.4	-10.0	66.4	16	$M_{\text{full}}$ vs. $M_{\text{reduced}}$	1.5	5
					$M_{\text{self}}$ vs. $M_{\text{reduced}}$	153.3***	4

Note. Models are ordered along *dfs*.

Note: \*\*\*  $p < .001$ .

Figure 1: Reduced model (M3r).



Note. All paths coefficients are significant with  $p < .05$ .