

## Fake review

Vazire, S., Mehl, M. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of Self-Ratings and Other-Ratings of daily behavior

Linus Neumann

The article addresses the old, but yet unsatisfactory ascertained question whether self- or observer-judgements about a person's personality or behavior (here: behavior) are more accurate than the other in certain domains. The authors introduce to personality psychology a method to obtain (a) objective (i. e. not favoring target's or raters' perspectives) and (b) ecologically valid (i. e. obtained in a naturalistic setting) behavioral data: The electronically activated recorder (EAR), whose results are correlated with self- and close-informants-ratings on the ACT. Results indicate wide areas of overlap in accuracy as well as distinct advantages of either perspective in specific domains. Namely, self-ratings were more accurate for arguing and watching TV, whereas informant-ratings were better predictors of talking one-on-one, attending a lecture and spending time with others.

This creative study has several mentionable strong features. The authors developed a behavioral measure that is molecular, yet vastly callous towards manipulation and applicable in a naturalistic setting. Considering the amount of information that is gathered about naturalistic behavior, the method is quite unobtrusive and has the interesting feature that participants never know whether their behavior is recorded or not. This could be argued to minimize influences on participants' behavior. The results are presented in a very comprehensible and illustrative manner using the Johari-window. It is worth positive mentioning that the claimed effects are illustrated using different statistical methods. Furthermore, the authors attempt to illustrate and outrule possible aggregation effects. All in all, this indicates a very care- and thoughtful handling of the data.

The focus of this study is of eminent significance to the field of social and personality psychology and is clearly addressed in a very promising manner.

The authors remain critical about their work and devise a fine limitations-section that doesn't just end in criticism, but leads over to recommendations for future research. I believe that the chosen approach will prove very fruitful for future research.

To establish this fruitfulness, some aspects of the study could be addressed and discussed more thoroughly to add even more to the undoubted strength of the paper. Room for this could be gained by leaving out the trivial and redundant Study I, which merely replicates a stable and well-established finding and does not add to the results at this early stage yet, due to limited sample size.

The body of literature that is cited by the authors mostly refers to personality psychology. However, the ACT-questionnaire is not theory-driven and was chosen and adapted to suit the measurement device EAR. Undoubtedly, prediction of behavior is a core aspect of personality perception, but it can be discussed how the items even refer to personality: items like "eating" or redundant reverse coded item-pairs like "talking to the same sex" / "talking to the opposite sex" and "outdoors" / "indoors" do not add to the strength of the paper, as the criteria of several significant findings overlap. So far, the measures appear to be objective, but arbitrary. If at all personality-related, the items apparently mostly refer to the scale of extraversion. Also, it is noticeable that only positive aspects of the ACT seem to be detectable and the behavior that is more accurately predicted by peers is the behavior that they are involved in. To sum up this critique, one can refer to the introductory quote of a rabbit in a field, which does not quite suit the ACT-use in the study: Participants know if they are on the phone or not, and the ACT is the estimate of this exact behavior – and not, how or why they were on the phone.

Considering the existing literature, the interesting question is the one addressing unique insights.

The biases and problems that are summed up in the introduction refer to personality assessment and do not apply to the ACT-captured behavior. But since the authors chose to investigate behavior, one could argue that a more adequate approach would be asking for the exact time the person performs a behavior in a given period of time, and then calculate the participants' means, rather than referring to a virtual norm, that is nonexistent, hard to grasp for participants, and therefore a source of error.

The issues of (a) reactivity to the EAR and (b) the short time between the self-ratings and the EAR-assessment remain almost undiscussed. Participants' ratings of obtrusiveness are reported, but are not sufficiently convincing to me to completely drop this subject. This is especially striking, as the assessed behaviors are molecular and small enough in number to memorize over a 4-day period.

In addition to this, self-ratings were assessed before, while informants' ratings were assessed after the EAR-phase. To avoid confounding EAR- and informant-reports-data, the possible influence of behavior in the crucial 4 days on the informants' ratings should have been out-ruled by assessing all ratings at the same time. Since the informants' ratings were economically assessed via the Internet, the different points in time of completing the questionnaire add a source of error to the informants' ratings that could have been avoided.

In addition to these main chances of improvement, some minor criticism could be raised:

- The influence of social desirability on the informants' ratings in terms of “valuing me and mine” is not discussed, although only positive behaviors could be predicted
- I have already addressed to redundancy of Study I at this early stage of research, due to insufficient sample size. However, to be able to refer the results of both studies to each other, questions should have been phrased in the same manner („...compared to the average person!“) and anchor points of the Likert-scale should have been defined in both ACT-assessments.
- At a closer look, some of the findings appear trivial to some extent:
  - prediction: gains in delta-R<sup>2</sup>s are due to the aggregation of self- and other-reports. The interesting part of this paper are the independent contributions.
- The SECSI rating tool is introduced and its scales are explained, yet not addressed anymore throughout the paper, although scales like activity, mood and interaction might be even closer to current concepts of personality than the ACT.
- It is a common claim in papers that address personality perception, that self-ratings are used, because researchers believe they are the most valid measure. Often times however, they are just chosen for plain convenience.