Reviewer's report

Title: Reliability and Validity of the Structured Interview of Personality Organization (STIPO)

Version: 1 Date: 8 March 2013

Reviewer: eve Caligor

Reviewer's report:

See comments below, numbered 1-4, about ways in which the authors could improve the quality and interest of the paper. All recommendations are for discretionary revisions.

This paper describes the design and findings of a study evaluating the reliability and validity of the German version of the STIPO. The STIPO is a semi-structured interview for assessment of 7 core dimensions of personality functioning relevant to personality disorder within the framework developed by Otto Kernberg. The STIPO can be rated on a 5-point scale, reflecting the overall subjective rating of the interviewer, or on a scale derived from summation of scores of the 0-2 individual item scores. This study reports on the former. 125 patients were recruited and 122 were included in the sample. Assessment instruments included the STIPO, SCID I and II and 8 additional self-report questionnaires.

Inter-rater reliability:
The authors report that inter-rater reliability of the instrument was high, with ICC of .885-1.0.

1. Comments: The authors report ratings using the STIPO 5-point scoring system. It would be interesting to know how inter-rater reliability compares when ratings are scored using mean summation scores, as reported by Stern et al. It seems potentially problematic to have different groups reporting findings that use different scoring systems for the same instrument.

Internal consistency:
The internal consistency for most of the scales was in the “good” range ” with the exception of the reality testing scale, which the authors comment on in the discussion.

Comments: None

Validity:
The authors used a variety of instruments to generate hypotheses about the correlations between these questionnaires and particular, pre-selected, STIPO domains. The authors report that all correlations were in the predicted direction.

2. Comments: Many of the questionnaires are unfamiliar to English-speaking
readers. Perhaps the authors could say a bit more about the content domains and psychometric properties of these instruments.

3. Comments: The presentation of data in this section is the weakest part of the paper. The authors report correlations between the particular STIPO scale identified in their hypothesis and the questionnaire; based on the report of Stern et al, as well as the nature of the underlying theoretical model, one would anticipate that other scales might correlate as well. Can the authors provide additional data or address this matter in the text? Also, can the authors comment in the discussion on some of the individual findings in Table 3? This would be especially helpful for readers unfamiliar with the questionnaires.

Discriminant validity:

The authors report on correlations between the various STIPO domains and patients with PD vs no PD diagnosis and then between those with Cluster B and Cluster C diagnoses. Overall STIPO scores discriminated between patients with PD and those without, and even more interesting, between Cluster B and Cluster C.

4. Comment: These are the most interesting findings reported in the paper (Tables 4 and 5). The authors might expand their discussion to comment further on these data and their implications for the value of the instrument.

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.