Do Rater Certification Procedures Identify Poor Raters?
Daniel DG¹ and Kott A²
¹Bracket, McLean VA; ²Bracket, Prague, CZ

INTRODUCTION
Prior to rating in a CNS clinical trial, site investigators are required to demonstrate proficiency in rating key outcome measures. A common test utilized to evaluate rater competence is successful rating of a videotaped interview either at the investigators’ meeting or online. The relationship between performance on this test procedure and performance rating patients at the site is poorly defined.

METHODS
Prior to study initiation, raters were trained to rate the PANSS at investigators’ meetings and/or online by highly interactive procedures, including slide presentations and rating of videotaped patient interviews and interview skills training. In order to be approved to rate in these global schizophrenia trials, investigators were required to successfully rate a videotaped patient interview by scoring at least 80% of the 30 PANSS items within the acceptable range set by an expert panel. Raters who failed the first attempt were given remediation and retested. If successful on either the first or second attempt investigators were approved to rate patients at their sites.

After study initiation, a proprietary audio/video recording system was utilized to record PANSS rating procedures at the site in ten schizophrenia clinical trials conducted in North America, Europe, South America, and Asia. Sites uploaded video or audio recorded ratings assessments for review by calibrated external reviewers. External reviewers provided feedback on ratings quality on an ongoing basis to the site and Bracket.

RESULTS
SUMMARY OF RESULTS I
When rating their first patient at the site, raters who successfully scored the videotaped test interview on the first attempt (N=189) had a higher rate of matches with consensus scores established with external reviewers (76.3%) than raters who required two attempts (n=43) to successfully score the videotaped test interview (67.3%) (t=3.39; df=230; p < 0.05).

SUMMARY OF RESULTS II
The number of deviations from the expert panel in scoring the recorded interview was modestly but statistically significantly correlated with the number of deviations from the external expert in scoring their first patient at the site (r=0.24, n=232, p<.001).

DISCUSSION
Investigators who failed their first attempt to rate the certification videotape to predetermined standards made more errors rating patients at their site than investigators who rated the certification video successfully on the first attempt.

Performance rating a videotaped PANSS interview prior to study participation modestly, but statistically significantly correlated with subsequent performance rating patients at the site.

The results suggest that performance on pre-study screening procedures is a useful factor in identifying raters at risk and thus may be a useful component of risk based monitoring procedures.

Future analyses will address contributions of geography, educational level, clinical and scale experience to ratings performance.