

**[S20.002] Poor Inter-Rater Reliability between Local Faculty and ABPN Examiners in Evaluating the Neurology Clinical Skills Exam: Preliminary Results**

*Lori Schuh, Detroit, MI, Zachary London, Ann Arbor, MI, Robert Neel, Cincinnati, OH, Charles Brock, Tampa, FL, Brett Kissela, Cincinnati, OH, Lonni Schultz, Detroit, MI, Douglas Gelb, Ann Arbor, MI*

**OBJECTIVE:** To study inter-rater reliability between local faculty examiners (LFE) and American Board of Psychiatry and Neurology (ABPN) examiners in evaluating the Neurology Clinical Skills Examination (NEX). **BACKGROUND:** The ABPN has recently replaced the traditional, centralized oral examination with the locally administered NEX. The ABPN postulated the experience with the NEX would be similar to the Mini-Clinical Evaluation Exercise, a reliable and valid assessment tool. The reliability and validity of the NEX has not been established, and concerns exist that bias may allow previously unacceptable candidates to be ABPN certified. **DESIGN/METHODS:** NEX encounters were videotaped at four neurology programs. The encounters were de-identified as much as possible. LFE and ABPN examiners graded the encounters using two evaluation forms: an ABPN sanctioned form and a modified form with a contracted 3 point rating scale. Standardized instructions were sent to all evaluators. Demographic data were obtained on all residents, faculty and ABPN examiners. Statistics used were Cohen's kappa and Intraclass correlation coefficients (ICC). **RESULTS:** 85 videotaped NEX encounters of 32 residents were evaluated by 20 LFE and 18 ABPN examiners. The inter-rater reliability for a determination of pass versus fail for each encounter was poor (kappa 0.33;0.11,0.55 95% CI), with ABPN examiners assigning a failing grade to 18% of encounters judged by LFE as passing, and a passing grade to 8% of encounters judged by LFE as failing. ICC between LFE and ABPN examiners for each performance rating on the sanctioned NEX form was poor with a range of 0.15-0.48, and did not improve with the contracted rating form (ICC range 0.12-0.39). **CONCLUSIONS/RELEVANCE:** There is poor inter-rater reliability between LFE and ABPN examiners. If the ABPN examiners' evaluations are considered to be a "gold standard", these results suggest LFE may be biased in assessing resident NEX performance. Future studies should address the inter-rater reliability of ABPN examiners. Supported by: This research was supported by an AAN Education Research Grant