# Online calibration for P-MCAT: A neural network based approach

Wednesday, 26th July - 15:10: Machine Learning (Benjamin Banneker) - Oral

*Dr. Lu Yuan (Collaborative Innovation Center of Assessment for Basic Education Quality,Beijing Normal University), Ms. Yingshi Huang (Collaborative Innovation Center of Assessment for Basic Education Quality,Beijing Normal University), Prof. Ping Chen (Collaborative Innovation Center of Assessment for Basic Education Quality,Beijing Normal University)*

Online calibration is a key technology for calibrating new items in computerized adaptive testing (CAT). As the multidimensional polytomous data become popular, the marginal maximum likelihood estimation with an expectation maximization (MMLE/EM)-based online calibration methods applicable to multidimensional CAT with polytomously scored items (P-MCAT) have been proposed (Yuan et al., 2022). However, the existing methods are mainly based on the MMLE/EM algorithm, which suffer from convergence problem when faced with high dimensionality and inaccessible proper initial values. To conquer these challenges, a neural network (NN)-based online calibration framework was put forward in this study. The new method differs profoundly from the traditional ones in that the parameter estimates of new items are obtained by learning the patterns between input and output data instead of finding solutions to the log-marginal likelihood function. This change in perspective gains critical benefits: (1) sidestepping the high-dimensional integration and high-dimensional optimization procedures on the ability posterior calculation and item parameter estimation, respectively; (2) reducing the impact of initial values by approaching item parameters indirectly via the optimization of network parameters. In this study, full-scale simulation studies were conducted to evaluate the performance of the NN-based method and MMLE/EM-based methods under varying scenarios. Furthermore, an alternative solution was presented for traditional methods to obtain appropriate initial values. Results showed that both the NN-based method and the alternative solution found their strengths in recovering the item parameters of new items while the MMLE/EM methods struggled to converge when more than three dimensions were involved in the test.