Title of Project:
Investigating the Combined Effects of Rater Expertise, Working Memory Capacity, and Cognitive Functionality on the Scoring of Second Language Speaking Performance

Researcher:
Qie Han
Teachers College, Columbia University
qie.han@tc.columbia.edu

Research Supervisor:
Prof. James Purpura
Teachers College, Columbia University


Qie Han

TIRF Research Topic Investigated:
Language Assessment

Project Summary:
In L2 performance assessment, raters can affect test validity due to variance in scores caused by differences in their own attributes rather than test takers' ability. To improve scoring validity, we must investigate what rater characteristics are likely to contribute to this construct-irrelevant variance. A rater experiential characteristic, rater expertise, has been found to shape scoring performance by L2 researchers. However, to harness expertise, raters must also rely on two cognitive characteristics, i.e., working memory capacity (WMC) and cognitive functionality (defined in terms of strategies related to information processing), to utilize knowledge, information, and skills in the scoring process. Seeing the importance of all three rater characteristics to scoring, this dissertation will be the first to investigate their combined influence on raters' scoring performance.

To this end, 90 raters with varied rating-related knowledge and experience will complete a rater background survey designed to measure rater expertise, score 27 spoken responses from the Aptis test, and finish one verbal WM task. Then, six raters will be randomly selected from the 90 to verbally report their thought processes during scoring. Quantitative and qualitative analysis procedures including hierarchical regression and verbal protocol analysis will be conducted on the raters' data to explore: (1) the joint effects of rater expertise and WMC on scoring performance, in terms of their relative contributions to scoring performance and their potential interaction, and (2) the cognitive functionality of raters (as reflected by strategies identified from raters' verbal reports) a nd how this functionality shapes raters' scoring performance.