measures), performance phase (metacognitive monitoring measure) and self-reflection phase (causal attributions and adaptive inferences measures) were asked before, during and after completing the task, respectively. Verbal responses were recorded verbatim and afterwards coded by two independent assessors. kappa coefficient of 0.89-0.98 was attained by two coders. Information was processed by descriptive and inferential statistics.
Results: Descriptive statistics showed that most participants (88.2–43.4%) reported task-specific processes for SRL measures. Multiple logistic regression analyses revealed that Students who exhibited higher self-efficacy (odds ratio [OR] 1.48, 95% confidence interval [CI] 1.03–2.12) and reported task-specific processes for metacognitive monitoring (OR 9.04, 95% CI 1.37–59.64) and causal attributions (OR 9.8, 95% CI 1.96–39.34) measures were more likely to be high previous performers. Only the causal attributions measure (OR 23, 95% CI 4.57–115.76) was associated with the learning task performance. Univariate analyses demonstrated that low previous female performers had significantly lower self-efficacy beliefs than low previous male performers ($p = 0.035$). There were no sex differences on any of the other microanalytic measures ($p > 0.05$). Phi coefficient revealed significant correlations between several SRL measures within and across the three phases.
Discussion: We identified important associations between SRL microanalytic measures and previous biomedical science performance and subsequent performance on a biomedical science learning task. We also recognized high levels of inter-rater reliability and significant relationships between the SRL measures of three phases of self-regulation. Comparing our findings with wider context of using SRL microanalytic approaches in science learning demonstrated that a SRL microanalytic protocol can detect differences in self-regulatory processes throughout the three phases of cyclical model of SRL among participants at different levels of prior science achievement (2).
Conclusion: These findings have implications to offer the educators a framework for providing feedback on SRL processes which is essential for effective feedback and improving performance. There are also potential applications for directing the content of formal remediation interventions in the early years of medical school.
References: 1. Lucieer SM, Jonker L, Visscher C, Rikers RM, Themmen AP. Self-regulated learning and academic performance in medical education. Med teach 2015; 1-9.
2. DiBenedetto MK, Zimmerman BJ. Differences in Self-Regulatory Processes among Students Studying Science: A Microanalytic Investigation. Int J Educ Psych Assess 2010;5 (1):2–24.

#9F3 (125960)
Developing an optimal model for sequential OSCE using CTT and IRT based psychometric properties

Sara Mortaz Hejri*, Tehran University of Medical Sciences, Tehran, Iran

Mohammad Jalili (Tehran University of Medical Sciences, Tehran, Iran)

Introduction: In objective structured clinical examination (OSCE), an increase in the number of stations is associated with an increase in reliability, but is more resource-consuming. Hence, sequential testing has been proposed in which all students first participate in a short test. Students who fail this test will participate in the supplementary OSCE. The results of the screening test should be able to predict the students' performance in the main test with a reasonable accuracy. The purpose of this dissertation was to introduce an optimal screening test based on the following factors: number of screening test stations, method of selecting screening test stations, and the cut-scores of the screening test
Methods: We used two datasets from a 10-station OSCE. Psychometric properties of stations were determined according to classical test theory (CTT) and item response theory (IRT). Then, several hypothetical screening tests were designed with different numbers of stations (three modes), different psychometric properties (5 modes in CTT and 2 modes in IRT), and different cut-scores (two levels: normal and stringent). Each hypothetical screening was compared to the main test.A desirable composite outcome was defined for each screening test, comprising of: passing percentage of more than 50%, positive predective value (PPV) equal to one, and negative predective value (NPV) greater than 0.25
Results: The OSCEs failure rates were 5.7% (n=6) and 10.9% (n=29) in 2011 and 2013, respectively. According to CTT-based screening tests, 20 out of 60 hypothetical OSCEs yielded the desirable outcome. Fourteen out of these 20 tests had stringent pass levels. The number of stations were as follows: 9 tests had five stations, 5 tests had four stations and 6 tests had three stations. According to IRT-based screening tests, 2 out of 6 OSCEs had the desirable outcome. Both these tests were composed of stations with high discrimination value and had stringent cut-score. One test had 5 stations, and the other had 4 stations.
Discussion: We proposed and evaluated an optimal model for the sequential design of an OSCE. While several studies have investigated different aspects of sequential OSCE, we believe this disssertation adds to the existing literature in several ways: For designing the screening OSCEs, we simultaneously considered three factors: number of stations, criteria for selection, and the cut-off score. None of the previous studies have taken all three parameters into account in their model. We also defined a composite outcome for evaluating the accuracy of the screening tests, while previous studies have considered various separate