Overview
The National Board of Medical Examiners (NBME) is an independent, not-for-profit organization that serves the public by developing, administering, and conducting research on high-quality assessments for healthcare professionals.

NBME programs include the United States Medical Licensing Examination®; an extensive offering of achievement tests for courses offered by medical schools; and numerous client examinations in medicine and other health professions. The variety of assessment programs creates a wealth of data and opportunities for applied and theoretical research that can impact practice.

The NBME employs approximately 25 doctoral level psychometricians and assessment scientists, as well as several MDs specializing in medical education. Staff is recognized internationally for its expertise in statistical analysis, psychometrics, and test development.

Interns will interact with other graduate students and NBME staff, and will present completed projects or work-in-progress to NBME staff. Internships typically result in conference presentations (e.g., NCME) and sometimes lead to publication or dissertation topics.

Requirements
- Active enrollment in doctoral program in measurement, statistics, cognitive science, medical education, or related field.
- Experience or coursework in one or more of the following: test development, IRT, CTT, statistics, research design, and cognitive science. Advanced knowledge of topics such as equating, generalizability theory, or Bayesian methodology is helpful. Skill in writing and presenting research. Working knowledge of statistical software (e.g., Winsteps, BILOG; SPSS, SAS, or R).
- Interns will be assigned to one or more mentors, but must be able to work independently.
- Must be authorized to work in the US for any employer. If selected, F-1 holders will need to apply for Curricular Practical Training authorization through their school’s international student office, and have a social security number for payroll purposes.

Compensation
Total compensation for the two months is approximately $9800, and is intended to cover all major expenses (food, housing, travel).

Research Projects
Interns will help define a research problem; review related studies; conduct data analyses (real and/or simulated data); and write a summary report suitable for presentation. Projects are summarized below. Applicants should identify 1 to 3 projects they prefer to work on and, if desired, to indicate any other research they would like to conduct.

1. **Alternative Scoring Models for Assessments Consisting of Several Performance Tasks.** This project explores scoring models for case-based performance assessments in which examinees complete several cases and provide numerous responses to the same case (e.g., examinee responses clustered within cases). Interns will work with USMLE Step 3 Computer-Based Case Simulations and/or USMLE Step 2 Clinical Skills exam.

2. **Extending Psychometric Speed-Accuracy Models to Data from Experimental Pacing Studies:** This project integrates hierarchical speed-accuracy models (van der Linden, 2007) with models for
response propensities and non-ignorable missing data. Specifically, the speed–accuracy–omission model proposed by Ulitzsch, Pohl & von Davier (2017) will be used to examine respondent behavior in recent experimental studies that varied the workload per item block.

3. **Exploring the Impact of Examination Timing.** High-stakes testing programs strive to provide the right amount of testing time — too much is wasteful and costly, too little impacts score validity. Possible projects will investigate exam timing, using real or simulated data, focusing on questions such as how test takers use time during the test, the differences between needing versus wanting more time, and metrics for monitoring speededness.

4. **Score Report Design.** This project involves the design of meaningful and useful score reports at the individual and aggregate level (e.g., school level). The project will evaluate data from focus groups, interviews, and surveys regarding the uses and interpretations of score feedback sent to medical schools about student performance. The project also will establish what feedback should be given to test takers based on current test design and scoring methods.

5. **Multivariate Generalizability and Score Profiles for Individuals and Institutions.** This project involves evaluating the quality of subscores using multivariate G-theory (Brennan, 2001). There is particular interest in assessing the reliability of score profiles at the level of the school, classroom, and institution by integrating Brennan’s work on score profile reliability (Brennan, 2001, p. 323) with the work of Kane & Brennan (1977) on the generalizability of class means.

6. **Test Security Evaluation and Data Analysis.** The intern will pursue research related to NBME test security priorities. Potential projects include development and operationalization of flagging rules for exposed MCQ material; new ideas and input from the literature on evaluating security risks in performance assessments; and data analysis of item response latencies to develop forensic methods to identify potential security risks from specific test centers or schools.

7. **Innovative Item Generation.** Provide input into and assist with implementing a research plan for evaluating items developed in innovative ways, including items developed via technology-assisted item development (e.g., automated item generation). This plan can include gathering qualitative data from expert review, as well as quantitative data such as item analysis and calibration results.

8. **Test Construction.** Potential projects include methods for evaluating and balancing multiple layers of content constraints; a literature review and plan for improving management of large item pools; assessing and developing best practices for different modes of testing; or developing evidence-based plans for managing item development processes to support different testing modalities (e.g., continuous vs. episodic testing; fixed-form vs. adaptive tests).

9. **Comparing Tests of Similar Constructs.** This project has two phases: (a) Conduct a quantitative and qualitative synthesis of the literature on the comparability of scores from two different physician licensure examinations. (b) Analyze data from primary sources to investigate the relationships among test scores from different licensure and certification tests (e.g., correlations; logistic regressions, concordance tables).

10. **Analysis of Ultimate Classification Error Rates for Test Batteries.** This project develops the concept of optimal cut score for pass/fail tests (Grabovsky & Wainer, 2017). When several tests have to be passed to obtain certification, the ultimate (overall) classification error rate depends on cut-scores set for individual tests and on the correlational structure of the tests in the battery. The project compares theoretically derived optimal cut-scores with those set by standard setting committees.

**Application**

Send cover letter outlining experience and project interests by number, along with a current resume to Joanne Ver Ploeg (JVerPloeg@nbme.org). Application deadline is **February 12, 2018**.

All applicants will be notified of selection decisions by **March 30, 2018**.