



THE NATIONAL CENTER FOR THE IMPROVEMENT OF EDUCATIONAL ASSESSMENT

2020 Summer Internship Program in Educational Assessment and Accountability

The National Center for the Improvement of Educational Assessment, Inc. (the Center) is a small non-profit organization that occupies a unique and influential niche at the intersection of educational measurement and educational assessment policy. The Center is pleased to offer **up to five (5) summer internships** for advanced doctoral students in educational measurement and/or assessment/accountability policy who want the opportunity to work with the Center's professionals on projects with direct implications for state and national educational policy.

The Center for Assessment

The Center was formed in 1998 as a not-for-profit corporation with a mission to increase student learning through improved assessment and accountability practices. The Center is located in Dover, NH (10 miles from the seacoast town of Portsmouth, NH and about an hour north of Boston, MA). The Center's thirteen professional staff members have earned doctorates in psychometrics, curriculum, or statistics and most have worked at high levels in state departments of education (e.g., assessment directors) or in testing companies. The combination of technical expertise and practical experience allows Center professionals to contribute effectively to cutting edge applications in educational measurement and policy.

The Center works directly with states (current contracts include more than 30 states or entities) and has working relationships with several national research and advocacy organizations such as the Council of Chief State School Officers (CCSSO), Achieve, and KnowledgeWorks. Some sample current projects of the Center include:

- ✓ Serving as technical leaders in the design and implementation of Innovative Assessment Demonstration Authority (IADA) projects with states pursuing this flexibility under the federal Every Student Succeeds Act (ESSA),
- ✓ Helping states devise student longitudinal growth systems for school accountability, and analyze the factors affecting the validity and reliability of such systems,
- ✓ Designing innovative, interactive assessment and accountability reporting systems designed to yield meaningful interpretations of student and school scores,
- ✓ Working with multi-state assessment consortia on a variety of issues ranging from assessment design and development to structuring systems for assisting the consortia in receiving relevant and timely technical advice, and
- ✓ Assisting states in developing comprehensive and coherent systems of assessment that serve summative and formative purposes. For example, the Center has been a national leader in designing systems to support competency-based and personalized learning models.



The Summer Internship Program

Each intern will work on **one major project** throughout the summer (to be negotiated between the intern and the Center mentor) and may participate with Center staff on other ongoing projects. The intern will have the opportunity to attend meetings and interact with state assessment personnel. Interns will be expected to **produce a written report and a proposal for a research conference (e.g., NCME, AERA)**, as evidence of successful completion of their project. One of the Center's senior staff will serve as the intern's primary mentor, but the interns will interact regularly with many of the Center's staff. **Potential** intern projects for 2020 may include the following¹:

- 1. Mitigating the impact of rater inaccuracies on test score scales:** As large scale assessments continue to include greater numbers of cognitively complex assessment tasks that ask students to respond in writing, there are both benefits and challenges. Such tasks offer the opportunity to gather rich and instructionally useful feedback for educators and students, but year-to-year score scale stability becomes harder to achieve due the potential influence of rater inconsistencies. Among the challenges associated with tests in which there is a greater emphasis placed on written responses is that the impact on score scales requires a heightened scrutiny due to its potential to threaten the comparability of scores within and across items and over time. This internship provides the opportunity for an advanced doctoral student interested in examining the psychometric impact of rater error on score scales and possible mitigation procedures.
- 2. Improving the interpretability of test score reports.** The aesthetics and quality of information presented on test score reports has improved over the last decade, but a survey of state individual student reports conducted by a 2019 Center summer intern (Tanaka, 2019) revealed that error associated with test scores was rarely reported. The *Standards for Educational and Psychological Testing* explicitly call for error or uncertainty of test scores to be included anytime scores are reported. When asked why error is not being reported, many test contractors and state assessment leaders reported that users did not understand how to interpret error and were frustrated trying to make sense of these reports. This internship combines assessment literacy and report design to better understand how we might produce more accurate and useful score reports. This project will involve reviewing assessment literacy research on how best to communicate measurement error, designing report mock-ups, and conducting cognitive laboratories with potential stakeholders to evaluate and refine draft designs.
- 3. Innovative Assessment System Additional Validity Evidence Collection and Analysis:** New Hampshire received a first-in-the-nation waiver from federal requirements related to state annual achievement testing in the 2014-15 school year. The Performance Assessment of Competency Education (PACE) innovative assessment system is now in its sixth year of

¹ More details about the Center for Assessment can be found at www.nciea.org. Please also navigate to the Internship page for additional details about potential projects.

implementation, operating under the Innovative Assessment Demonstration Authority (IADA) under the Every Student Succeeds Act, and the collection and analysis of validity evidence continues. This project has two major component to support a special validity study in one participating district: 1) analyze collected student work on performance tasks and bodies of work to examine the extent to which student achievement is accurately reflected in PACE annual determinations, teacher judgment survey ratings, and NH SAS (the state standardized test) results in relation to the achievement level descriptors; and 2) use cognitive labs to compare two designs for collecting teacher judgments of student achievement. The purpose of this project is to analyze the evidence to make an argument about ways to improve the accuracy of teacher judgments about student achievement and add additional information to the validity argument about how well PACE standards represent the depth and breadth of student achievement.

- 4. Evaluating assessment accommodations:** Guidance from the United States Department of Education (ED) for peer review of state assessment systems specifies that states must ensure that accommodated administrations of assessments are appropriate, effective, and allow for meaningful interpretation and comparison of results for all students (peer review element 5.3). These are challenging criteria to meet. The purpose of this project will be to help identify and document the range of practices and sources of evidence to help developers better address these criteria. It is anticipated this project will involve a literature review, a survey of state assessment practices, and potentially the development of guidance to help assessment leaders understand and document the impact of accommodated conditions on assessment outcomes.

Requirements to work on this project: Familiarity with special populations including students with disabilities and/or language learners. Good understanding of professional practices in large-scale, standardized assessment that bolster inclusiveness and accessibility.

- 5. Test Fairness: Exploring current practices and future directions:** Test developers often describe three primary goals for large-scale, high-stakes, assessments: validity, reliability, and fairness. We know comparatively less about the latter of these three. However, there is emerging scholarship that suggests our understanding of fairness should be broader and more tightly coupled with validity. The chief goal of this project is to better understand the ‘state of the states’ with respect to practices and, chiefly, documentation in support of fairness. Project activities will likely include:
 - A review of the literature to better understand the dimensions of fairness associated with large scale high stakes state tests.
 - An exploration of the leading development and design practices employed by states to support fairness.
 - A study of the prominent sources evidence that states have collected to document the extent that fairness has been addressed (e.g. technical manuals, research reports, peer review submissions).

One outcome of the project may be to identify ways for states to bolster their practices and documentation in support of fairness.

- 6. Evaluating the reliability and precision of school accountability performance scores:** This project involves analyzing the reliability/precision of states' school accountability performance scores, and evaluating states' accountability performance criteria in terms of the reliability/precision characteristics. School performance scores can be produced a variety of ways and are usually aggregates of several other scores, including status and growth performance on tests of English language arts and mathematics, high school graduation rates, and other measures. However produced, they can be viewed as similar to assessment scores, with technical properties including validity, reliability/precision, and fairness. A key contribution of this project will be to implement advanced modeling methods to determine empirical reliability/precision estimates of school performance scores.

Requirements to work on this project: Ability to do empirical modeling of complex scores using large data sets in SAS, R, or other programmable statistical software. Knowledge of states' school accountability systems and federal school accountability requirements (i.e., ESSA) is a plus.

- 7. Evaluating interim assessments against intended use and interpretation:** This project involves evaluating the validity and usefulness of select commercial and state-developed interim assessments through constructing and analyzing theories of action and interpretive validity arguments, and associated evidence. Interim assessments may be meaningfully differentiated from each other and from other summative or formative assessments by specifying in detail both a theory of action and an interpretive argument (cf., Bennett, Kane, & Bridgeman, 2011; Gong & Dadey, 2019). Differentiated claims and interpretations are inherently represented in the marketing literature and assessment reports (e.g., item, student, class, school, and district) made available by publishers of commercial and state-sponsored interim assessments. This project will involve evaluating several interim assessments against their intended uses and interpretations, using the theory of action/interpretive validity argument approach. Possible interim assessments to analyze include prominent commercial interim assessments, as well as state-sponsored interim assessments.

Requirements to work on this project: Deep familiarity with wide variety of instructional uses of interim assessment information and ability to extract and/or construct a theory of action regarding those instructional uses, and ability to create interpretive and evaluative validity arguments based on typical published assessment documentation (e.g., score reports, test blueprints, alignment studies); background in ELA or mathematics, including familiarity with the Common Core State Standards.

- 8. Analyzing solutions to complex assessment design problems: The example of NGSS assessments:** This project involves analyzing the strengths and limitations of multiple real-world solutions to a complex assessment design problem: how to assess the *Next Generation*



Science Standards. The complex structure and lack of specifications of the NGSS, as well as varying state values and constraints have led states to develop a number of very different assessment designs for their science assessments. Close analysis of testing programs' documentation will be used to depict what the solutions are to creating an NGSS assessment, how the designs are similar and different, and what are the strengths/limitations of each design. Particular care will be taken to differentiate different designs to accomplish the same thing, and different designs to accomplish different things. The documentation will include for each testing program theory of action, claims, score reports and supporting materials, test blueprints, standard setting materials, alignment study materials, and other relevant documents to the extent possible. The document analysis will be supplemented by interviews of key design architects of the testing programs.

Requirements to work on this project: Interest in promoting multiple solution approaches to assessment design problems; understanding of the NGSS's structure and content sufficient to understand forced assessment design trade-offs; ability to read assessment technical documentation (e.g., test blueprints, standard setting reports, alignment studies) and see implications for an interpretive validity argument.

Application Information

General Qualifications

The intern must have completed at least two years of doctoral course work in educational measurement, curriculum studies, statistics, research methods, or a related field. Interns with documented previous research experience are preferred. Further, interns must document their ability to work independently to complete a long-term project. We have found that successful interns possess **most** of the following skills and knowledge (the importance of the level of skills and knowledge in each of the areas described below is **dependent** on the specific project):

- ✓ Ability to work on a team under a rapid development model
- ✓ A deep understanding of educational assessment and its uses including policy and practice
- ✓ Content knowledge in a relevant discipline (e.g. science, mathematics, language arts)
- ✓ Depending on the project, working knowledge of statistical analysis through multivariate analyses as well as fluency with one or more statistical packages, e.g., SAS, SPSS, R
- ✓ A solid understanding of research design
- ✓ Psychometrics (both classical and IRT) with demonstrated understanding of the principles of reliability and validity
- ✓ An interest in applying technical skills and understanding major policy and practical issues
- ✓ Excellent written and competent spoken English skills

Logistics

The internship duration is 8 weeks and is **located at our offices in Dover, NH for the full 8 weeks.** The internship will start in early June 2020; the specific date will be determined by the intern and the mentor.



Support

The Center will provide a stipend of **\$6000** as well as a housing allowance and reasonable relocation expenses.

Application

To apply for the internship program, candidates should submit the following materials **electronically**:

- ✓ A letter of interest explaining why the candidate would be a good fit with the Center, what the candidate hopes to gain from the experience, and which project(s) the candidate's preferred project. Further, the letter should explain both what the candidate could contribute to the preferred project(s) and why the project(s) fits with the candidate's interests.
- ✓ Curriculum vita, and
- ✓ Two letters of recommendations (one must be from the candidate's academic advisor).

Of approximately 20-30 applicants, six to eight are identified for a telephone interview. Those interviewed by phone may be asked to submit one recent sole (preferred) or first-authored academic paper. Please do not submit the paper until it is requested.

**Materials must be submitted electronically (including letters of recommendations) to:
Sandi Chaplin at schaplin@nciea.org and received by February 14, 2020.**

Applicants selected for interviews will be notified by March 6, 2020 regarding their candidacy.

To learn more about the Center, please visit www.nciea.org.