The National Center for the Improvement of Educational Assessment, Inc.

2026 Scott Marion[1] Summer Internship Program in Educational Assessment and Accountability

The National Center for the Improvement of Educational Assessment (Center for Assessment) is pleased to offer **up to six (6) 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 National Center for the Improvement of Educational Assessment, Inc. (the Center) is a small non-profit organization occupying a unique and influential niche at the intersection of educational measurement and assessment policy. It was formed in 1998 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 16 professional staff members have advanced degrees in psychometrics, curriculum, or statistics. Most have worked at high levels in state education departments (e.g., assessment directors) or testing companies. The combination of technical expertise and practical experience allows Center professionals to contribute effectively to cutting-edge educational measurement and policy applications.

The Center works directly with more than 40 states and entities, as well as several national research and advocacy organizations, such as the Council of Chief State School Officers (CCSSO), the Aspen Institute, 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 evaluate the impact of the pandemic on student learning and the implications for ongoing support
- ✓ Working with states to develop student longitudinal growth systems for school accountability and evaluating 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,
- ✓ Serving as conceptual leaders as well as assisting states in developing balanced systems of assessment that serve summative and formative purposes. For example, several Center professionals are authors of several chapters in the 2024 National Academy of Education

[1] The ***Scott Marion Summer Internship Program*** in Educational Assessment and Accountability was named for the Center's former Executive Director, who has been instrumental in shaping the program. Dr. Marion is now the Principal Learning Associate at the Center.

volume, *Reimagining Balanced Assessment Systems,* and published a companion practical guidebook in 2025, *Implementing Balanced Assessment Systems*.

The Summer Internship Program

The Center's internship program is designed to support the learning and professional growth of advanced doctoral students. Each intern will work on **one major project** throughout the summer (to be negotiated between the intern and the Center mentor) and may participate in other Center projects. Most interns will have opportunities 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, NCSA)** as evidence of successful project completion. One of the Center's senior staff will serve as the intern's primary mentor, but the interns will interact with other Center staff members. **Potential** intern projects for 2026 may include the following.[2]

Potential Internship Projects

1. **Understanding Divergent Assessment Signals**
   State education leaders routinely confront multiple and sometimes conflicting indicators of student performance. State summative assessments, national measures such as NAEP, and commercial interim tests often present different pictures of achievement and growth, creating challenges for policy decisions and public communication.

   This project focuses on three central research questions.
   - First, what is the extent and nature of the problem? To what degree do state leaders encounter mixed messages about performance? Are these discrepancies increasing? Do they appear more frequently in specific grades, content areas, student groups, or states?
   - Second, what accounts for these mismatches? Divergence may stem from differences in constructs, timing, rigor, administration conditions, or a combination of these and other factors.
   - Finally, how do leaders interpret assessment information when signals diverge? What principles, tools, or promising practices can support more straightforward interpretation and use?

   The intern will engage in several research activities, including conducting a landscape analysis of common state and district assessments, reviewing and summarizing published information about these assessments to describe their intended interpretations and uses, analyzing data to describe state and group-level performance trends, and potentially interviewing state education leaders to understand how they navigate conflicting signals.

[2] *More details about the Center for Assessment can be found at www.nciea.org.*

2. **Scaling Decisions in Large-Scale Assessment**
Determining the features and characteristics of a reporting scale is an important but often underappreciated aspect of assessment design. Although producing a reporting scale can appear straightforward, typically involving a linear transformation of the underlying theta values, scaling practices vary considerably across programs. Key decisions include setting the scale range, defining the highest and lowest obtainable scores, determining whether to anchor specific values, and deciding how to handle rounding. For large-scale state assessments, these choices can meaningfully affect the reported scores of thousands of students. While some guidance exists in the academic literature, much of the most practical insight resides in program technical manuals and technical advisory committee documentation.

   The goal of this project is to review these sources to create a framework that describes common scaling approaches and their relative prevalence in the field. Building on this foundation, the intern will develop guidance outlining the advantages and limitations of different scaling decisions. This work will be enhanced by applying various scaling methods to a simulated dataset to demonstrate how different choices influence reported scores under different conditions.

3. **Modeling the Impact of Design and Growth on Summative Score Formation**
Through-year assessment programs have been introduced as an alternative to single end-of-year state summative assessments. A through-year assessment is one administered in multiple distinct sessions during a school year. The results of those administrations are used to produce summative determinations and support at least one additional use (Dadey et al., 2023). Several states have considered adopting a through-year assessment model, but have ultimately decided *not* to aggregate scores over time. One roadblock has been the lack of empirical research describing how aggregation choices interact with student learning trajectories to influence interpretations of student scores.

   This intern will design and implement a simulation study that explores how various score aggregation methods interact with different through-year designs and varying patterns of change to impact summative score interpretations. Specifically, the intern will:
   - Specify at least two through-year design structures (for example, fall/winter/spring interim windows or multi-testlet structures with varying content coverage).
   - Define student performance trajectories using parameterized models (such as random intercept–slope growth, nonlinear growth, or growth with forgetting), partially informed by a literature review of within-year growth trends.
   - Implement multiple aggregation approaches (such as window averaging, weighted schemes, model-based estimates, or alternative scoring rules).
   - Evaluate the performance of each approach using metrics such as RMSE, bias, and correlation with true ability.

This simulation study will help the field better understand how different student performance trajectories, through-year design options, and aggregation methods interact to support summative inferences. In doing so, this work will help frame future work and potential implementations of through-year assessment programs.

The ideal candidate should be familiar with the role that state assessment plays in K-12 public schools and have substantial experience working in R (other programming languages may be considered). Experience running simulation studies is preferable, but not required.

4. **Supporting Assessment Score Interpretation and Use by Leveraging Generative AI**
Most assessment programs provide general interpretive guides and, in some cases, more differentiated supports (e.g., professional learning, coaching, processes, and protocols). However, even these more differentiated supports rarely directly address the ways educators actually interpret and use scores. This internship project is meant to address these challenges. The project will develop supports that are specific, detailed, and scalable using Generative AI (GenAI). Key to doing so is aligning each of these supports to a "use case", which is a highly detailed use of assessment results.

   In addition, this project will also explore the *processes* by which these supports can be rapidly and efficiently created through GenAI. GenAI may help overcome the obstacles that have made it so challenging to provide specific, detailed support. GenAI could provide the means to develop, at scale, supports for a large number of use cases in a variety of formats across varied contexts.

   This work will be conducted in partnership with a large assessment consortium, providing a robust context and access to a working group of educators. The project will involve three sequential streams of work:
   - **Prototype Supports for a Single Use Case Using Generative AI.**
   Develop detailed, specific, and varied supports for a specific use case by using Generative AI. The key point is that existing interpretive materials will serve as context for the GenAI application and, potentially, for retrieval-augmented generation methods.
   - **Explore GenAI-Enabled Resource Retrieval and Revision.**
   Examine whether GenAI can assist in retrieving, adapting, or wholesale developing supports tailored to different educator contexts. For example, the project might develop and test a GenAI chatbot that retrieves, modifies, and contextualizes guidance or templates based on a teacher's needs. The project may also explore how well a GenAI tool can develop support for new, user-supplied use cases.
   - **Pilot and Evaluate Usefulness.**
   If feasible, test the clarity, relevance, and usability of the developed supports with a small group of educators. This work stream may include brief interviews or usability testing to identify which formats or approaches are most valuable. It may also include co-development of supports as well as GenAI approaches and processes.

The ideal candidate should have expertise in translating assessment results into concrete next steps for a variety of audiences and experience using GenAI tools. Experience working in a school, district, or state is also preferred.

5. **Understanding the Role of Accountability Indicators: Replicating a Statewide Analysis of Chronic Absenteeism and School Quality**
   State accountability systems rely on multiple academic and nonacademic indicators to communicate school quality, identify schools for support, and guide resource allocation. Yet most states rarely examine whether these indicators actually work together as intended, or how changes to indicator composition might alter school classifications. This internship project builds on a recently completed multi-year statewide analysis that used principal components analysis (PCA) and path analysis to evaluate the coherence, predictive relationships, and policy consequences of key accountability indicators. The study found that academic achievement, growth, and chronic absenteeism form a unified school-quality construct; that absenteeism predicts performance both directly and indirectly; and that removing absenteeism from an index-based model significantly reshapes school rankings and identification outcomes.

   The intern will replicate and extend this analytic framework across additional states by conducting structural and predictive modeling, simulating alternative indicator configurations, analyzing ranking stability, and translating findings for policy audiences. Through this work, interns will gain hands-on experience with accountability system design, reproducible research, and evidence-based policymaking, contributing to a national effort to strengthen the coherence, stability, and equity of state accountability systems.

   The ideal candidate is intellectually curious and interested in using data to understand how accountability systems function in practice. Candidates should be comfortable working with quantitative data and have experience with multivariate analysis techniques (i.e., factor analysis, path analysis, structural equation modeling), which will be applied to policy-relevant contexts. A willingness to engage deeply with systems, reflect on results, and establish communication insights for policy audiences is essential.

6. **Understanding the Challenges to Coherence in Balanced Assessment Systems: Extending Shepard's (1991) "Psychometricians' Beliefs About Learning."**
   Assessment systems are balanced if they coherently link the various assessments in the system, often through a shared model of learning and clear specification of the learning targets, comprehensively support multiple purposes and uses, and continuously document student progress over time. These properties of coherence, continuity, and comprehensiveness were originally put forth in *Knowing What Students Know* (NRC, 2001) and help create a powerful image of a high-quality system of assessments.

   This is why one of the foundational chapters in *Reimagining Balanced Assessment Systems* (2024) was focused on synthesizing modern theories of learning and development. Importantly,

not any theory of learning will do. We agree with Shepard, Penuel, and Davidson (2017) that the learning model underlying the assessment system must be based on the latest and most credible research. Unfortunately, balanced assessment systems have been challenging to implement for many reasons, not the least of which is a lack of clarity among practitioners and assessment specialists about how students learn and develop. Unfortunately, we do not have empirical evidence about the learning models used by those most responsible for assessment system development.

Lorrie Shepard published Psychometricians' Beliefs About Learning in 1991. In that study, she interviewed a systematic sample of 50 district assessment leaders. She documented that the majority held at least implicit views of learning aligned with behaviorist learning theories, which ran counter to the prevailing cognitive theories held by learning scientists. The intern selected for this project will use Shepard's study as a foundation to conduct surveys and interviews with state assessment and content (curriculum) leaders. This study will provide us with a more grounded understanding of how current state-level leaders think about learning theory, which can inform the design and development of balanced assessment systems. Candidates for this project must have a solid understanding of modern theories of learning. Experience conducting interviews, designing surveys, and analyzing qualitative data is preferred.

7. **Data-Rich and Information-Poor: Making Sense of Assessment Results to Support Decision-Making**

Evans and Marion (2024) wrote that educators often feel they are driving in a snowstorm when trying to make sense of the various sources of assessment information from state, district, and classroom assessments. Evans and Marion also wrote that educators are often asked to extract instructionally useful information from assessments that are not necessarily designed to yield instructional insights. Furthermore, even if there is instructionally useful information in some of the district and classroom assessments, educators and school leaders rarely possess the assessment and data literacy necessary to make sense of this snowstorm of information. And, for good reason. Extracting meaning from multiple and potentially conflicting sources of information is hard!

Therefore, the intern selected for this project will work with one of the Center's district clients to develop protocols and tools to help school leaders and teachers interpret available assessment results and use them productively. This project will include multiple phases:

- The intern will start with a landscape analysis of the multiple sources of information available to educators and leaders at select grade levels and content areas (e.g., 6th grade mathematics) in this district.
- The intern will review existing data analysis protocols and score interpretation guides already used by the district. Additionally, the intern will review the literature on assessment utility and evaluate similar protocols used elsewhere.

- The intern will use information from the first two phases to draft interpretation protocols for this specific district, again focusing on a limited number of grade/content-area combinations.
- Finally, the intern will conduct cognitive laboratories (think-aloud study) with a sample of district educators and leaders to pilot and then refine the protocol and tools.

Importantly, the intern will incorporate GenAI tools throughout this project, especially those developed to help educators and leaders sift through vast amounts of data to produce defensible interpretations and potential actions.

The intern selected for this project should have classroom teaching experience, experience using artificial intelligence tools, and a solid understanding of various types of assessments from the classroom to the state level.

Application Information

General Qualifications

The intern must have completed at least two years of doctoral coursework 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, including multivariate analyses, as well as fluency with one or more statistical packages (e.g., R, SAS, SPSS), and a solid understanding of research design
- ✓ Psychometrics (both classical and IRT) with a demonstrated knowledge 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 is for **8 weeks** at the **Center's Dover, New Hampshire office.** The internship will start in early June 2026; the intern and the mentor will determine the specific date.

Support

The Center will provide a stipend of **$8,000, a $3,000 housing allowance, reasonable relocation expenses, and a $1,000** travel stipend to support attendance at a conference in 2027 to present the results of the intern's project (contingent upon acceptance of a conference proposal).

Application

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

- ✓ A letter of interest explaining why you would be a good fit with the Center, and what you hope to gain from the experience. The letter must also identify your **preferred project** (no more than three) and explain what you could contribute to each project and why it fits your interests.
- ✓ Curriculum vitae.
- ✓ Two letters of recommendation (one must be from your academic advisor).

Approximately 10 to 15 applicants are identified for a virtual interview. Those interviewed virtually may be asked to submit one recent academic paper. Please do not submit the paper until it is requested.

Materials must be submitted electronically (including letters of recommendation) to: Sandi Chaplin at schaplin@nciea.org and received by February 6, 2026.

Applicants selected for interviews will be notified by March 2, 2026, regarding their candidacy.
To learn more about the Center, please visit www.nciea.org.