**Collaborative Research: Integrating Cognition and Measurement with Conceptual Knowledge: Establishing the Validity and Diagnostic Capacity of Concept Inventories**
**Project Summary**

This proposal is being submitted as proposal type Empirical Research to the Research and Evaluation on Education in Science and Engineering (REESE) program of the National Science Foundation, within Strand A, Emerging Research, Area of Inquiry 3: Measurement, Modeling and Methods for Research and Evaluation. The overarching goal of this project is to implement a comprehensive, multidisciplinary approach to the design and validation of concept inventories (CIs) that transforms how they can be used in STEM education and enhances their effectiveness and impact. To accomplish our goals we propose to focus our research on a specific concept inventory, the Thermal Transport Concept Inventory (TTCI). The TTCI provides an excellent test case for developing the larger research and validation methodology that we plan to extend to other concept inventories. We will "deconstruct" the TTCI in terms of the underlying conceptual knowledge and skills that it taps in critical areas of science and engineering, gather additional empirical data on student reasoning and performance, and analyze that data by applying modern psychometric analysis techniques focused on diagnostic modeling. Concrete outcomes expected from the proposed work include: (1) an improved TTCI that will be made widely available through the web and accompanied by interpretive tools for use by faculty; (2) an in-depth, multi-level analysis of the assessment validity and diagnostic capacity of the TTCI; (3) a comprehensive inventory of possible formative and summative uses of CIs in light of their interpretive structure, including what is needed by STEM practitioners to make good use of CIs; and (4) a research and development framework for designing, developing, evaluating, improving, and/or implementing CIs in other STEM areas.

**Intellectual Merit:** Concept inventories (CIs) have been developed and extensively studied in numerous STEM areas. These CIs are designed to identify common student misconceptions in the sciences and engineering using well-crafted multiple choice questions. Following the success of the Force Concept Inventory in Newtonian physics, the TTCI is designed to elicit student misconceptions in the thermal and transport engineering sciences such as fluid mechanics, heat transfer, and thermodynamics. The TTCI particularly focuses on concepts deemed important but difficult for undergraduate engineering students. Thus CIs represent a relatively unique form of assessment instrument with a range of possible uses from diagnostic and formative purposes that can guide instructional planning and decision-making, to summative purposes for evaluating overall learning and instructional effects a student, classroom and/or instructional program level. What differentiates CIs from typical assessment of student academic achievement is that they tend to be highly focused on a small set of key constructs and understandings within a limited disciplinary domain. A primary rationale for the proposed work is that despite the importance and usefulness of reliably measuring fundamental key concepts, as in the case of the TTCI and other CIs, the cognitive understandings and psychometric modeling in analyzing CIs has been either primitive (restricted to standardized unidimensional psychometric models or classical test theory), fragmentary or non-existent. The work proposed here would develop an explicit set of facets of understanding for diagnostic measurement and reporting, would gather multi-level empirical data of student performance on the TTCI (including student protocol and interview studies, and large scale data for psychometric modeling), would revise both the items and facets based upon these empirical data analyses, and would gather further data on the revised instrument.

**Broader Impacts:** In addition to work with the TTCI, we propose to examine the possibilities for expanded uses of the TTCI and other CIs for formative classroom purposes as well as for a variety of summative purposes. In addition, we will inventory other CIs for which there are available data that can be analyzed. A key broader outcome from the proposed work is a model for cross-disciplinary development, evaluation and improvement of CIs that can be adapted and applied to CIs other than the TTCI. A unifying perspective for the proposed work is that of the multiple aspects of assessment validity. Additional outcomes expected from this project are both an expansion of the psychometric modeling and methods for applying diagnostic models to CIs, and an expanded sense of assessment validity in the concrete and critically important area of measurement and assessment of key concepts in science and engineering.