1. Questions(s) or issue(s) for dialogue at Learning Network Conference session:

MSP projects continue to struggle with the development of evidence of the fundamental outcomes of the work they do with teachers. Improving teachers’ disciplinary content knowledge and student achievement in mathematics and science are two central goals of both the NSF MSP program and the projects therein. Bolyard and Moyer-Pakenham (2008) identified teacher subject matter knowledge as one of the seven key teacher quality characteristics that emerge from their review of the research and policy literature of the effect of teacher characteristics on student outcomes. Assessment of the improvement of teacher subject matter knowledge is a crucial outcome measure of the MSP disciplinary-based projects, especially given the current accountability pressures on professional development programs and NCLB/ARRA. The availability of valid instrumentation to measure improvements in teacher and student disciplinary knowledge is not adequate to address the wide range of content objectives in the MSP projects (Moyer-Packenham, Bolyard, Kitsantas, & Oh, 2008).

PIs and evaluators face multiple validity issues in the identification and/or development of appropriate instruments. They must balance construct validity issues with instruments that cover the content area more broadly than the professional development will address against the use of subsets of instruments that are too closely tied to project objectives. The development of project-specific instruments presents the PIs and evaluators with additional validity issues of being able to clearly identify the content objectives that the professional development entails, using adequate assessment instrument development processes, and developing evidence of the validity and reliability of instruments. This is especially cogent given the multiple objectives that most projects address that go beyond simply increasing teacher content knowledge and improving student achievement that require project staff and evaluator resources.

2. Context of the work within the STEM education literature and within your MSP project:

Within the BSSP project, the development of adequate teacher and student content knowledge has required significant evaluation effort to design assessment instruments that aligned with the science disciplinary content that was the focus of the professional development. Instrument construction processes have been developed that involve science educators working with the disciplinary faculty to identify and clarify the disciplinary objectives of the yearly project foci. Evaluators have searched multiple assessment sources to identify multiple choice items that are aligned with the content objectives. Those items are reviewed by project scientists and science educators for content accuracy and alignment with content objectives. The items that have accurate content and alignment are then compiled into a test with 1-3 items per objective. Item performance statistics are calculated and internal consistency of the test is determined to contribute evidence of test reliability. However, limited time and resources have confounded efforts to conduct any validity studies on the assessments.
This experience is parallel to that of the principal investigators and evaluators in the Penn Science Teacher Institute. There were no appropriate instruments for assessing the science content and process knowledge of their teacher participants, so these instruments were developed by faculty. They experienced a conflict between needing to maintain the exams as originally developed for statistical purposes, and wanting to edit the exam in order to better match assessment to instruction as the curriculum has evolved over the years.

The STEM education literature is replete with evidence of the difficulty of validly assessing the nature and extent of disciplinary content knowledge (Pelligrino, Chudowsky and Glaser, 2001). Issues of large scale science assessment are well also documented (Wilson and Bertenthal, 2005). However, relatively less research has focused on the development of tests of teacher disciplinary knowledge (Moyer-Packenham, Bolyard, Kitsantas, & Oh, 2008).

The RETA projects funded by NSF, most particularly MKT, LMT, LMT-ETS, MSP-MAP, MOSART and ATLAST (included in panel discussion) have developed research-based instruments whose psychometric properties have been well characterized. These projects identify particular domain knowledge that range in specificity from large domain topics such as astronomy and physical science to measuring motivation of students. The NSF Knowledge Management and Dissemination project has compiled citations for instruments to measure changes in teacher disciplinary knowledge as well as teacher pedagogical content knowledge. However, the availability of these instruments is problematic as few published reports in either the research or gray literature venues include complete instruments.

3. Claim(s) or hypothesis(es) examined in the work (anticipating that veteran projects will have claims, newer projects will have hypotheses):

Using well developed assessments developed by RETA projects requires the principal investigators to address the alignment between the grain science of RETA-developed assessments and the science and mathematics disciplinary knowledge of their own projects. Assessments like ATLAST tend to be too narrowly focused. Others, like DTAMS and MOSART, tend to be too broad. Rarely if ever do projects find an assessment that is "just right." Developing valid and reliable measures of teacher and student disciplinary knowledge is a labor-intensive and complex process. Such development is facilitated by involving disciplinary faculty in the development of learning objectives and providing opportunities for project staff to work together on the development of project-specific instruments. However, the technical quality of such instruments is still problematic, making the inferences drawn about the effectiveness of the professional development less robust than is wanted.

4. Evaluation and/or research design, data collection and analysis:

This presentation is not reporting on the results of any particular evaluation or research project conducted within the MSP project. Rather the project will present the process by which the BSSP and the Penn Science Teacher Institutes have addressed their teacher and student assessment requirements and determined the validity and reliability of their instruments. These development processes will be contrasted with those of the ATLAST and MOSART projects. Panelists will discuss the issues that have arisen as they have designed, implemented and evaluated the quality of the instruments they have used. Mechanisms to determine the validity of instruments will be highlighted, such as checking with teachers about the opportunity that they
5. Key insights (retrospective for veteran projects, prospective for newer projects) that have value for the Learning Network:

The major insights that have arisen in the BSSP project around the issue of instrument development is the requirement that discussion about the nature of the science content addressed in the professional development be raised early and often in the development of the yearly activities. Assessment instrument development has required considerable resources from the project and evaluation scopes of work. Assessment development is an iterative process. Off the shelf assessments are useful in some project contexts, but they are not useful in others. Project specific instruments provide useful formative and summative data for projects, but it is difficult to compare outcomes across projects when instruments of differing subject matter and psychometric quality are used.