REQUEST FOR PROPOSALS:
THE DEVELOPMENT AND IMPROVEMENT OF THE
MEASUREMENT OF CLASSROOM QUALITY
Date of Release: August 18, 2008

The Spencer Foundation and the William T. Grant Foundation, in collaboration, reopen our grants competition to support research on the development and improvement of the measurement of indicators of classroom quality in grades K through 12. Quality is defined as those features and processes of classrooms that are likely to cause improvements in important youth outcomes such as academic achievement and engagement or reducing disruptive and antisocial behavior. This collaboration was borne of the need to measure classroom features and processes in ways that support demonstrably valid inferences about classroom quality.

Classrooms, like families and youth-serving organizations, can provide young people with a forum in which to develop meaningful relationships with adults and peers. They provide learning activities, access to resources, and opportunities for cognitive, social, and emotional learning and identity development. Because of the amount of time youth spend in classrooms, these sites are opportune places for intervention. This RFP focuses on the common interest of both Foundations in the development and improvement of tools for measuring and understanding the quality of K–12 classrooms.

Classrooms are a critical context or "social setting" for the daily experiences of youth. Such settings vary in social processes, resources, and the arrangement of those resources. Examples of these features of settings include the content and structure of interactions between students, teachers, and instructional materials; social norms regarding behavior in classrooms; the availability of significant and meaningful roles for youth within classrooms; the student-teacher ratio; the ethnic, racial, or ability composition of the group of students; and the physical arrangements of the space.

The Foundations will support a small group of research projects, one to three years in duration, with award amounts ranging from $50,000 to $500,000, including all direct and indirect costs. Funding available for these projects totals $1 million per year. The first RFP was released on January 8, 2008, and we anticipate awarding those grants in September 2008. We plan to release this RFP a third time in August 2009. After the

third set of awards are made, we will evaluate the success of the work and decide whether to issue an additional RFP.

The Foundations will consider several types of proposals: *(a)* new, stand-alone measurement development studies; *(b)* add-on studies in which new measurement development work supplements an existing field study; and *(c)* further analysis of existing data to improve measurement. We anticipate that the third category, further analysis studies, will have budgets at the lower end of our announced award range, add-on studies will have budgets at the middle of the range, and new stand-alone studies will have budgets toward the upper end of the range.

Review will take place in two stages: the deadline for letters of inquiry is **November 3, 2008**, at 11:59 p.m. EST, and the deadline for invited full proposals is **February 16, 2009**.

Please note that this document contains hyperlinks. Windows users, press "Ctrl" on your keyboard, and then click on these links to access the Foundations' websites. Mac users, simply click on the links to access the sites. Additional information on the interests of the Spencer Foundation (www.spencer.org) and William T. Grant Foundation (www.wtgrantfoundation.org) can be obtained by visiting our websites.

Goals for the RFP
With this RFP, we plan to support a small number of grants to develop and improve the measurement of classroom indicators associated with favorable student outcomes in grades K through 12. We see these measures as "yardsticks" that are critical to better understanding classroom features and processes and to improve classroom quality. We believe this improved understanding will enhance theory development and the generalizability of our knowledge of classroom functioning and quality. We have three main goals, in order of priority, for this RFP. First and foremost, all studies should result in the development or improvement of the reliability and/or validity of one or more measures of classroom indicators linked to quality. Second, we encourage work that results in relatively low cost and scalable measures. Third, we encourage the development of tools that are both ultimately useful for improving practice and useful to practitioners.

Our long-term goal with this RFP is to build greater capacity for research on classrooms, schools, and other youth-serving organizations. We recognize that research focused on understanding and improving such settings is at a much earlier stage of development than research on understanding and improving individuals' well-being. Consequently, we are seeking to contribute to knowledge of classrooms and school settings and more broadly to social setting theory, setting-intervention theory, setting measurement, the implementation of field experiments, the enrichment of data for descriptive and longitudinal studies, and research design and analysis. The Foundations will expand our portfolios of

classroom measurement research by supporting the work of a small number of project teams, linking these teams together, and facilitating the teams' consultations with other experts on improving youth-serving organizations. In turn, we expect that grantees' work will contribute to the capacity of others in the field to do high-quality measurement and intervention research targeting classrooms and other social settings.

Background for the RFP
There is increasing recognition of the vital role of classrooms in fostering cognitive, social, and emotional development and the life-course outcomes to which they are theoretically linked. These outcomes include economic productivity, political participation, and health. Measurement, particularly of these individual-level outcomes, has become increasingly psychometrically sophisticated and prominent in the media and in public policy debates. "No Child Left Behind" and the nature of the measurement of individual-level achievement outcomes are regularly debated in daily newspapers. When we consider how to achieve positive outcomes by measuring and changing the key elements of classrooms, including the interactions between teachers and students, we find ourselves treading on a much more tenuous scientific foundation.

Schooling provides children access to academic/cognitive, social, and emotional skills mainly through experience in classrooms. Most of the 15,000 hours children spend in school during their lives are spent in classrooms. Formal instruction occurs primarily in classrooms. Sociologists have long argued that classrooms are places where children learn to "be in society." Schools help to prepare children to live in the world by teaching them how to effectively participate in classroom life.

While whole school reform has often been a focus of research and policy interest, the effects of these reforms must be felt at the classroom level in order for learning to flourish. There is also evidence of significant variation across classrooms in students' academic, social, and emotional development. For these reasons, our focus is on classrooms.

Classrooms as Dynamic Social Systems
The focus of the RFP is on measurement, but applicants should describe the constructs under study in relation to theory about classrooms. We are exploring the idea that classrooms function as dynamic social systems, consisting of certain features, defined as resources; the social, spatial, and temporal organization of these resources; and social processes. While all of these components of classrooms are important, our current thinking is that social processes are particularly important for influencing youth outcomes. *Social processes* refers to transactions between two or more groups of people in a setting and can include social interactions, teacher instructional and support practices, roles for youth and teachers vis-à-vis one another, classroom-wide norms and expectations, and distribution of decision-making power. *Setting resources* refers to the "inputs" into a setting, including financial (e.g., per pupil expenditure), material (e.g.,

facilities, curricular materials), temporal (e.g., how much time is available), and human (e.g., teachers' and youth's knowledge and skills). *Organization of resources* refers to how the resources are arranged and allocated in the setting (e.g., arranging students into whole- versus small-group instruction in classrooms). Applicants should describe how they conceptualize classrooms and will measure relevant setting-level constructs such as social processes, resources, and arrangement of resources.

We are ultimately interested in productive discussions of how classroom processes are interrelated (e.g., coordinated or dissonant) in their functioning. This can be considered a corollary to individual-level functioning: how are individuals' personalities, beliefs, skills, and behaviors interrelated? How do they change over time? For classrooms, we seek a similar understanding of how different aspects of settings are interrelated and function in a dynamic system. We recognize that theories on classrooms have not been integrated well across disciplines, and we encourage applicants to draw from various fields and disciplines (e.g., anthropology, education, organizational behavior, psychology, sociology). For more information on the William T. Grant Foundation's current thinking about social settings, read the *2007 Annual Report* essay "[Measuring Social Settings](#)" and the *2006 Annual Report* essay "[Social Setting Theory and Measurement](#)."

Applicants should provide a convincing rationale for the correlation between classroom setting measurement operations, the developmental level of the classrooms (e.g., elementary, middle, or high schools), and the student outcomes of interest (e.g., academic, behavioral, social). We seek measurement tools that have strong potential to be used for the benefit of practice.

Measurement of Classrooms
Investment in classroom measurement has been negligible compared to investment in the measurement of cognitive skill and assessment of the impact of policies on cognitive skill. A measurement agenda that neglects classroom quality seems incomplete. Well-studied policies such as increasing resources, using accountability to improve incentives, revamping governance and school organization, promoting competition, and expanding school choice can affect cognitive skills primarily because they affect the quality of interactions occurring in classrooms. A considerable body of evidence suggests significant variation in student outcomes across classrooms. However, attempts to identify the factors that explain why some classrooms yield better outcomes than others have had only limited success. One reason may be that many of the data sets with good measures of student outcomes (e.g., test score gains) have little or no information about characteristics of the classroom setting across different classrooms. Researchers are therefore compelled to focus on much more distal measures of teacher characteristics such as college attended, content-specific courses taken, certification status, degrees obtained, or SAT scores to explain differences in student outcomes.

Moreover, to the extent that schooling affects long-term economic productivity and health by enhancing social and emotional skills, it is essential to understand the classroom processes important for the development of those skills.

Teacher educators, educational psychologists, and ethnographers have long studied classroom life, and this work has significantly influenced conceptions of effective instructional practice, teacher training, and curriculum development. With a few important exceptions, this work has rarely been large scale and quantitatively rigorous. As a result, we know comparatively little about the important dimensions of classroom quality and very little about how interventions influence classroom quality.

Types of Methods and Sources
As stated previously, the overarching goal of this RFP is to develop and improve tools for measuring and understanding classroom indicators linked to favorable change in student outcomes. More specifically, we expect applicants' studies to focus on creating/improving the reliability and validity of classroom measures. Secondarily, we encourage studies that yield measures that are affordable, scalable, and useful for improving practice. The proposals we seek will build on decades of ethnographic and behavioral observation studies of classrooms. Assessing the features, that is, the resources (e.g., teacher experience) or the organization of resources (e.g., whole- vs. small-group instruction) in a classroom is a much less onerous and costly task than measuring classroom processes. Nevertheless, the measurement of resources and the arrangement of resources remain important. The measurement of classroom processes must be grounded in a setting-level theoretical framework and based on prior empirical data.

As we understand current practice, there are five tools that researchers are using to gather data regarding classrooms: behavioral observations, teacher logs or diaries, surveys or scales, interviews, and unobtrusive measures including administrative records. All are appropriate to consider for this RFP. The current state of each leaves open important questions for measurement development. While the varying methods and sources of the measurement of classroom quality raise unique questions, there are many common questions across methods and sources.

Some **_behavioral observation systems_** of teacher-student interaction or teacher practices focus on broadband dimensions that cut across classrooms of diverse content material, while others focus on instruction in specific subjects, such as reading or mathematics. In behavioral observation literature there is reasonable convergence on the judgment that three major types of social processes (which are called by several names) describe classroom settings: instruction, classroom management, and supportive climate. On the other hand, some content-specific observational systems examine the nature of classroom talk, subject

matter, teacher expectations, quality of instruction, and cognitive demand of task.

Behavioral observations raise a host of empirically important questions. What is the optimal level of abstraction in describing classroom behavior? How do very concrete observations versus higher-inference observations differ as a function of the particular outcome and its time frame of measurement? How can reliability of measurement and hence improved power of research design be best achieved? How many observations are necessary to achieve a reliable measure? How many raters are necessary over how many observations? Are multiple raters of the same events necessary? Assessing and improving the reliability of these measures generally requires a design in which rater effects can be separated from temporal instability using generalizability theory. Applicants examining behavioral observation systems should address as many of these questions as appropriate

Teacher logs have shown considerable promise. Recently, teacher logs of specific practices for specific content material have demonstrated important relationships to the development of cognitive skill. For teacher logs, both questions similar to and different from those for behavioral observations arise. For example, what is the optimal level of abstraction for teachers to use in logging their activities? How do very concrete teacher log reports versus more inferential log reports differ as a function of the particular outcome of interest and the time frame of measurement? How can we best achieve reliability of measurement and hence improve the power of research designed to test the impact of interventions on classroom quality? How many log reports are necessary? How does the reliability of the setting-level measure improve as the number of teacher-raters increases? As in the case of behavioral observations, generalizability theory can be helpful in assessing sources of error and improving reliability. Applicants examining teacher log reports should address as many of these questions as appropriate.

Self-report questionnaires and scales completed by students remain the most common methods of inferring classroom-level processes. These scales take a variety of forms—self-report of social climate, behavioral norms and expectations, and peer and teacher behavior. These self-report scales can vary in how subjective or behaviorally anchored responses are. The more subjective these reports are the less likely they are to reflect classroom processes. These sources of information can be invaluable and are often easier, less intrusive, and less costly to administer than behavioral observations and teacher logs. In addition, they reflect students' experience. On the other hand, this may represent more about the individual student than the classroom and correlate minimally with objective assessments of the setting. In assessing classroom-level processes, student reports are often simply aggregated and used as an index of classroom-level process. Can an aggregate self-report measure accurately tap classrooms as the unit of analysis? Far too little research

has been directed to issues of in-group agreement between student ratings within a class or to the reliability of student ratings. Finally, student reports of classroom quality are of limited utility in predicting changes in the outcomes of the same students. There is much to be learned in this area if educational researchers are to follow the advances made in the study of organizations as the unit of analysis. Thus, questions about the meaning and utility of in-group agreement (heterogeneity versus homogeneity) are of utmost importance in assessing classroom-level phenomena. Applicants using self-reports should address as many of these issues and questions (as well as other more basic questions of reliability, validity, cost-effectiveness, and usefulness) as possible.

Interviews with students or teachers have the advantage of being able to reveal more in-depth information than other methods. However, like surveys and scales, interviews can be extremely structured and close-ended in format. In this instance, interviews raise many of the same issues as surveys. Only when interviews are more open-ended with probes for deeper meaning do they have obvious advantages over self-report questionnaires and scales. Yet, at the same time, interviews are relatively costly and labor-intensive. Interviews of teachers raise issues of the reliability of a single respondent in inferring classroom-level processes, as do teacher logs and surveys. Applicants using interviews should address as many of these issues and questions as possible, as well as other more basic questions of reliability, validity, affordability, and usefulness.

When we turn to ***unobtrusive measures*** including administrative reports or artifacts (such as what is hanging on the walls of the classroom or the number and type of books on the shelves), we find very little systematic research on how well they might index classroom processes. If these methods prove to be valuable, they would be very important because of their ease of collection, low cost, and unobtrusiveness. Applicants are encouraged to explore issues of validity and reliability of these measures in conjunction with others. The design of novel measurement instruments to assess the quality of dynamic classrooms and the demonstration of their reliability, validity, affordability, and practical utility is strongly encouraged. Many questions arise when we consider the interrelationships among different types and sources of classroom measures. For examples of these questions, see the discussion below under the third type of question.

Types of Questions
Building on what we already know, how can we improve the tools for studying classroom quality?

1. How does the optimal strategy for classroom assessment vary as a function of the classroom-level construct and student outcome of interest? For example, some would argue that teacher reports through diaries, logs, or interviews can supply good information on how teachers organize instruction, how much time they spend on

various instructional tasks, what cognitive skills they emphasize, and how they think about subject matter. The analysis of student work may be an efficient way to gain clarity on the cognitive level of the work in various subject areas. On the other hand, direct observation of classroom processes may be needed to gain insight into the quality of interactions between teachers and students and among students. The quality of social interactions may be essential in promoting pro-social norms and social and emotional regulation as well as promoting academic skills. Each research strategy entails sources of error that must be minimized to obtain adequate reliability and validity. And each imposes cost constraints that may limit the sample size of classes under study.

2. How can we measure key dimensions of quality with high reliability and validity at a reasonable cost? In particular, how do sources of uncertainty such as rater and item inconsistency and temporal instability affect the reliability of specific classroom assessments? In light of these findings, how can researchers organize the training of raters, their allocation to classrooms, and the scheduling of their visits over time to achieve adequate reliability within reasonable budgetary constraints? For what aspects of classroom quality can self-report methods such as teacher diaries, logs, or interviews generate reasonably reliable and valid inferences about classroom quality, and at what cost?

3. Can different types (and sources) of measures (e.g., behavioral observations, teacher logs, surveys and scales, interviews, and unobtrusive measures) tap the same underlying construct(s) or is the type (and source) of measurement unique to the construct(s)? When good construct validity (via convergent and discriminant validity) is established among two or more instruments, how can this knowledge enable researchers or analysts to make better choices about the optimal instrument for the construct of interest on the grounds of affordability and practical utility?

Types of Studies
In addressing any of the preceding or related questions, we will consider supporting several different types of studies. In all of these types of studies, applicants should provide strong theoretical rationale that their classroom constructs reflect classroom features and processes linked to quality. In addition, applicants are expected to address many of the specific issues of reliability, validity, affordability, and practical utility as well as the conceptual issues discussed above.

1. **New, stand-alone** classroom measurement development studies. Here, studies may be initiated to develop and validate a new measure(s) of classroom quality based on a promising setting-level conceptual structure and pilot data.

2. **Add-on studies** of additional measurement development work to a study of classroom settings. Adding a new measure of classroom processes (e.g., behavioral observations) to an ongoing study that may only include teacher and student perceptions of classroom processes could be a good way to pursue several questions driving this RFP. Alternatively, and perhaps more commonly, studies may be initiated to improve existing measure(s) of classroom quality. Applicants proposing to add setting-level data collection and analysis to their ongoing studies should provide a strong rationale for the timing of setting-level data collection. It may be too late to examine changes in setting-level constructs over a span of time in studies that are too far along.

3. **Further analysis of existing data** to improve the measurement of classroom settings. For example, studies of existing data may be reconceptualized at a classroom-level and re-analyzed to improve the quality of measurement at that level.

While both Foundations are interested in all the questions noted in this RFP, we do not expect every project to address all these questions nor to address them equally well. Instead, applicants should explain which questions they will focus on and how those decisions guide their proposed research.

Requirements for Proposals
Projects target the development, refinement, and improvement of tools for measuring and understanding effective features and processes of K–12 classrooms. The classroom should be the unit of analysis in these studies. Applicants should clearly discuss their classroom constructs and measurement development strategies. Applicants should provide a sophisticated analysis plan for improving the measurement of these classroom constructs.

The criteria we will use in evaluating proposals are:
1. *Theoretical and Empirical Rationale.* The proposed work should focus on measurement, but projects should be theoretically grounded and supported by prior empirical evidence. We are looking for theory that identifies key classroom-level indicators and how they work in a dynamic classroom system. Applicants should provide a strong theoretical and empirical rationale linking their classroom-level constructs to variation in important youth-level outcomes. Youth outcomes can include, but are not limited to, academic performance, social and emotional competence, and identity development.
2. *Data Collection.* Applicants will need to collect or utilize data on their classroom constructs of interest. Applicants at a minimum should have preliminary evidence that the measure is promising in terms of reliability and validity. In new or add-on studies, applicants are encouraged to propose multi-method and multi-

source measurements of their major classroom constructs. All applicants, however, should consider the best methods for addressing their research questions and the limits of any single source. This RFP focuses on classrooms as the units of analysis, but the unit of data collection (i.e., source of data) may differ from the unit of analysis. Data collection plans should reflect consideration of the requisite time frame for detecting classroom-level changes and also indicate the extent to which data will be made available to other researchers. Applicants proposing to add onto existing studies should provide a compelling argument regarding the timing of classroom-level data collection.

3. *Design of the Measurement Study.* To improve the measurement of classroom features and processes, studies should be longitudinal. They may be descriptive or intervention studies. Applicants should describe the observation points and their rationale, as well as sample. The design should reflect serious consideration of the multiple sources of measurement error (e.g., rater differences, temporal instability, item inconsistency) that are most plausible given the research goals and data collection strategies.

4. *Data Analyses.* Projects should include an analytic plan for addressing each of the measurement development and improvement goals. The plan should be sufficiently sophisticated for the phenomena under study and reflect a clear understanding of the strengths and limits of various analytic techniques.

5. *Anticipated Products and Communication Plan.* The products should be measurement research tools, complete with technical and training manuals. Ideally, they should be affordable and easy to use. The plan should contain channels for ongoing communication with practitioners or policymakers as the research is conducted. Products such as journal articles, books, and reports should also be made available to the research community. Both the tools and the research data should be placed in the public domain at minimal cost.

6. *Staffing Plan.* Project teams will require expertise in classroom-level theory, measurement development, and data analysis. Applicants should create a project team and staffing plan with sufficient expertise in these areas. We encourage measurement development collaborations with interdisciplinary and mixed-method expertise. Ethnic diversity in staffing at all levels of project teams is a priority of both Foundations, and applicants should consider these issues in designing their project teams and work. Plans, for example, could involve career development and mentorship for junior scholars of color.

Awards
The Foundations estimate supporting a small group of projects with one-to three-year awards, ranging from $50,000 to $500,000, including all direct and indirect costs. We expect that smaller grant awards are suitable for studies that analyze existing data; mid-range awards are most suitable for add-on studies, and larger grants for newly initiated studies. The total funds available for these projects are $1 million per year. Awards will be made annually for at least three years.

Capacity-Building Support
In addition to supporting each project independently, the Foundations will provide separate support for capacity-building activities for our grantees. The details of these activities will depend on the cohort of grants that we ultimately fund.

We anticipate that these capacity-building activities will include:
- *Bringing project teams of researchers together on a regular basis.* Before and after initiation of the funded studies, we anticipate team meetings to plan and stimulate cross-fertilization of ideas. Applicants do not need to budget for travel for these meetings. The Foundations will support the travel through other funding.

- *Linking grantees with other relevant funding opportunities such as the announcements listed below.* Additional funding opportunities may be posted on our website as they become available.

  o Methodology and Measurement in the Behavioral and Social Sciences (R01) - PAR-08-212: http://grants.nih.gov/grants/guide/pa-files/PAR-08-212.html;
  o Methodology and Measurement in the Behavioral and Social Sciences (R03) - PAR-08-214: http://grants.nih.gov/grants/guide/pa-files/PAR-08-214.html; and Methodology and Measurement in the Behavioral and Social Sciences (R21) - PAR-08-213: http://grants.nih.gov/grants/guide/pa-files/PAR-08-213.html;
  o Social and Behavioral Context for Academic Learning specific 84.305A Education Research RFA: http://ies.ed.gov/ncer/funding/soc_beh/index.asp

- *Identifying and building resources in classroom-level theory and measurement.* As we continue to make progress in these areas, we will post resources on our website.

We expect that grantees will become part of our ongoing efforts to produce work that will strengthen the field's capacity to conduct intervention research on classroom settings and to explain observed variation in student outcomes of interest. As projects develop, we will explore opportunities to work with grantees on creating and disseminating these materials. Examples may include a monograph on measurement

tools for assessing classroom features and processes, a report on observational methods and coding for classroom mechanisms, or a paper on troubleshooting methodological issues in conceptualizing, measuring, and analyzing classroom-level processes.

Hypothetical Examples

1. To more efficiently administer, plan, and deploy limited resources in the conduct of experimental field trials, researchers need to know the differential costs, reliability, and validity of different methods of measurement. In this hypothetical scenario, a research team is conducting a randomized cluster trial of an innovative literacy curriculum in fourth and fifth grade urban public school classrooms. They are interested in whether student-level literacy improvement is associated with improvements in classroom processes, studied here as use of classroom time and quality of teacher-student interactions. Currently, their principal source of information on the interactions and the use of time are teacher logs, collected at the conclusion of each day. They seek funding to assess the reliability and validity of the teacher logs and examine the costs of this mode of data collection. The team proposes to augment their current study by videotaping a random sample of classrooms and subjecting these videotapes to observational coding of time use and quality of teacher-student interactions. (These classrooms will have video cameras installed that will automatically record classroom activity throughout the literacy instructional session of each day). Analyses in the *add-on study* will permit the researchers to determine the relative costs of the two methods (including training, data collection, and data reduction), as well as differences in the reliability and validity of the scores reflecting time use and teacher-student interaction. The analysis will quantify independent sources of measurement error, including temporal instability and rater effects.

2. The field lacks reliable and valid measures of norms in classrooms in which everyone believes they are capable of achieving success. In this hypothetical example, a new mathematics curriculum is being launched in the classrooms of 40 middle schools (i.e., randomly assigned to 20 experimental classrooms and 20 control classrooms) to determine whether it improves mathematics achievement. Student-level outcomes are being assessed by pre- and post-standardized tests and the experimental implementation contrast by daily teacher logs designed to measure the components and sequencing of the new mathematics curriculum. A colleague approaches the intervention research team with the idea that perhaps the implementation of the new mathematics curriculum alone isn't leading to improvement, but rather that its effects may be moderated by classrooms in which teachers have established a normative cultural belief that everyone succeeds. There is strong theory to support this general hypothesis and there is evidence that

school norms in other areas such as aggression affect academic and behavioral outcomes. Thus, the researchers agree to request funds for an *add-on study* to develop a student self-report questionnaire to measure the shared classroom belief that everyone can succeed in mathematics. With the development of a reliable and valid measure of a shared classroom belief that everyone can succeed in mathematics, future researchers can evaluate whether the impact of a mathematics curriculum is moderated by classroom-level shared belief that everyone can succeed in math. The analysis will quantify the degree of inter-subjective agreement among students and item consistency to estimate classroom-level reliability.

3. There is a strong need to explore various options for improving measurement reliability, particularly with regard to behavioral observations and teacher logs. In this hypothetical example, a national evaluation study assessed the impact of a novel high school English-social studies curriculum that employed team teaching. Teacher logs were used as the primary mode of assessment of teacher practices. One major goal was to relate these teacher practices to classroom-level performance and student-engagement outcomes. Given that this co-teaching intervention had, at times, two teachers completing logs on the same subject and classroom, it presented a rare opportunity to conduct an intensive study of the reliability of teacher logs. For example, does the subject matter, the time of day (or days of the week) when the classes are held, or the time of day when the teacher completes the log affect the reliability of the ratings? Do teachers vary in the inter-rater reliability of their ratings across the academic year? Is it necessary for teachers to complete logs daily to achieve acceptable levels of reliability or can a cost-effective sampling strategy work reasonably well? If so, what is the ideal sampling strategy? Employing this unique national data set, a team of psychometric scholars propose to conduct *further analyses* to examine these questions.

4. There is a strong need for a low-cost and easy–to-administer instrument that reflects the quality of teacher-pupil interactions. In this hypothetical scenario, a state-wide education authority systematically collects administrative data and makes them available for research purposes. The administrative data consist of teacher characteristics (e.g., race, years of experience, teacher exam scores, age, type of degree and institution awarding degrees, certification status), school characteristics (e.g., size of student population, age of building, building condition), student-level characteristics (e.g., race/ethnicity, gender, reduced/free lunch status), and student-level test scores. Researchers are interested in developing a data set that reflects the quality of classroom teaching practices, which can be collected easily. This will allow educators and researchers to ask more sophisticated questions about what

types of classroom practices mediate between resources (e.g., teacher, school, and student characteristics) and student outcomes. After an extensive review of the literature on the assessment of classroom practices for third to fifth grade classes, classroom management, instructional support, and socio-emotional support emerge as central constructs. The team decides to apply for a *new, stand-alone study* to develop a brief, easy-to-administer instrument that taps these three constructs with high reliability and validity. A well-established behavioral observation instrument will be used as the criterion measure against which to validate the two new measures: teacher daily log reports and a brief easy-to-administer global behavioral observation instrument. Researchers will address a variety of questions about each of these new instruments. For example, how frequently do either the teacher logs or global behavioral observation measures need to be administered in order to achieve adequate rater consistency and temporal stability within budget constraints? If observers are involved, how much training will they need, and how many observers will be required to obtain adequate reliability?

Application Procedures

All applications will be submitted using the William T. Grant Foundation's electronic submission process. The process will proceed in two stages: letter of inquiry (LOI) and invited full proposals. The deadline for receipt of LOIs is November 3, 2008, at 11:59 p.m. EST. A team of senior staff from the two Foundations will evaluate all LOIs and invite a small group of finalists to submit full proposals. Invited full proposals will be due by February 16, 2009. Full proposals will undergo a rigorous scientific peer-review process involving a commissioned panel of external expert reviewers and the joint Senior Program Team of the Foundations.

Letter of Inquiry (LOI) Procedures

1. Applications are accepted through the William T. Grant Foundation website at http://www.wtgrantfoundation.org. Click on the USER LOGIN button on the home page. If you have not submitted online with us before, you will need to register on the website to obtain a Login ID and Password.
2. Select "Click here to start a new application."
   - For grant program, choose "Major Grants."
   - Enter your project title, which must begin with "Spencer-WTGF RFP 2009."
   - Select your tax-exempt organization.
   - After saving this information, you will be brought back to your home page. Click on the "Major Grants Letter of Inquiry" task.
3. Fill in text boxes for the following information.
   - Contact Information for the Principal Investigator only.
   - Principal Investigators. Contact information for the Principal Investigator will appear on this page. Add contact information

for each Co-Principal Investigator. Note that you may only designate one Principal Investigator, but you may list multiple Co-Principal Investigators.

- Project Information, including:
  - Project title (maximum of 15 words, must begin with "Spencer-WTGF RFP 2009");
  - Start and end dates of the project;
  - Grant request amount (direct and indirect costs combined for the full grant period);
  - Brief description of the project (maximum of 100 words). The brief description should be written in language appropriate for an educated lay audience, not for other researchers. Begin by stating the major measurement development/improvement questions. Then, briefly summarize the project's rationale and background, methods, design, and data analysis plan. Lastly, state what type of study this request is for—new stand-alone, add-on, or further analysis.
  - Discipline of Principal Investigator's most advanced degree.
  - Project (sample) demographics. Enter the age range of study participants. For gender, race/ethnicity, income, and locale, check all that apply.

4. Upload two files, one file containing the narrative and one containing all one-page curricula vitae. All pages of the application should use a 12-point font. Margins should be at least 1 inch on all sides. Do not use headers or footers. Documents should be single-spaced.
   - Narrative (five pages). The narrative should describe details of the project, including:
     - Major measurement questions guiding this work;
     - Theoretical and empirical rationale and a discussion of how the proposed activities will develop or further improve the measurement of classroom quality.
     - Study methods, including sample, design, and data collection procedures; and
     - Data analysis plan for addressing each of the study's questions.
   - Curricula Vitae. One-page vitae for the Principal Investigator and each Co-Principal Investigator. (If there are multiple vitae, they should be compiled and uploaded as a single document.)

Applicants will receive a confirmation email after submission of the Letter of Inquiry.