Phase-I Study of the Effects of Professional Development and Long-Term Support on Program Implementation and Scaling Up: Final Report[1]

EXECUTIVE SUMMARY

Paul R. Brandon, Alice K. H. Taum, Carlos C. Ayala, Donald B. Young,
Mary E. Gray, Thomas T. Speitel, Thanh Truc T. Nguyen, and Francis M. Pottenger III

Curriculum Research & Development Group, College of Education, University of Hawai'i at Mānoa

May 2007

This is the executive summary of the final report for a National Science Foundation (NSF) project (Grant No. REC022818). The project was conducted by researchers and curriculum developers at Curriculum Research & Development Group (CRDG), College of Education, University of Hawai'i at Mānoa. It was funded by the Interagency Educational Research Initiative (IERI), a collaboration among NSF, the U. S. Department of Education, and the National Institutes of Health. The project was a two-component, preparatory phase for a randomized study of the effects of variations in professional development (PD) duration and long-term training, with multimedia support, on (a) the implementation and student outcomes of middle-school inquiry-based science (here called *inquiry science*) on the wide dissemination ("scaling-up") of inquiry science. The version of inquiry science that we examined was CRDG's Foundational Approaches in Science Teaching (FAST) program. The project's two components were (a) the development of an alternative version of FAST PD for the first year of FAST to examine in the second phase of the study and (b) the development and validation of data collection instruments to use in the second phase.[2]

In this summary, we describe the development of the alternative version of FAST PD, describe the development of instruments, and summarize the findings of studies of the validity of data collected with the instruments.

Development of the Alternative Version of FAST PD

The alternative version of FAST PD that we developed is called FASTPro. It was developed by a team from CRDG's Learning Technology Section, including the section head, an instructional designer, two videographers, a professional video editor, and a graphic designer. The instructional designer and graphic designer also served as multimedia programmers. FASTPro consists of a one-week face-to-face institute (FASTStart), an electronic resource in the form of a multimedia DVD-ROM for the trained teachers (FASTeR), and an on-line course (FASTForward). FASTStart addresses the essential skills and concepts that are best taught in face-to-face PD. The knowledge and skills that are traditionally covered in the second week of FAST PD institutes are addressed in FASTeR and FASTForward. FASTeR includes video of FAST institutes, including video of the FAST trainer and of teachers in the role of students during the institutes; video showing FAST being taught in the classroom, including segments on students and segments on teachers; photographs and animation (in the form of slide shows) of setting up student investigations; and photographs of laboratory materials and equipment. FASTeR has one or more of these PD aids for 19 FAST 1 student science investigations. A survey of teachers about the extent to which they would use FASTeR showed strongly positive opinions about the DVD-ROM. The development of FASTForward was partially completed during the project and as of May 2007 was ongoing.

Instrument Development

[1]Any opinions, findings, and conclusions or recommendations expressed in this report are those of the authors and do not necessarily reflect the views of the National Science Foundation. An electronic copy of the full report, given in one file for the body of the report and another for the appendixes, is available at http://www.hawaii. edu/crdg/programs/pre/scup.html.

[2]No Phase-II grants were funded in the year that we applied for our second, follow-up grant. New funding for all IERI grants has been discontinued; the program is being phased out.

The instruments that were developed for comparing FASTPro with traditional FAST PD included the Inquiry Science Observation Code Sheet (ISOCS), the Inquiry Science Teacher Questionnaire (ISTQ), the Inquiry Science Questioning Quality (ISQQ) method, and the Inquiry Science Student Assessment (ISSA). The ISOCS is a method for coding and analyzing videotaped observations of teachers in inquiry science classrooms, with a focus on the interaction between teachers and students that is initiated by teachers' questions of students. It was developed in close consultation with FAST developers in more than 40 review-and-revision cycles over a two-year period. The ISTQ is a self-report instrument for collecting data on (a) the implementation of inquiry science in the classroom and (b) the context within which teachers implement inquiry science, including teacher demographics; teacher perceptions, behaviors, attitudes, opinions, interests, and beliefs; some classroom variables; and the support the school provides teachers to implement inquiry science. Implementation was measured with the ISTQ's Inquiry Science Implementation Scale. The context scales on the instrument included the Collaboration Frequency Scale, the Collaboration Benefits Scale, the Teacher Participation in Science Activities Scale, and the School Support for Inquiry Science Scale. Coefficient alphas were all high, and a test-retest study of the Implementation Scale showed high reliability. The ISSQ uses the paired-comparison method, conducted by expert judges, for measuring the quality of program implementation. A group of five FAST trainers served as judges and tried out the method. The ISSA is our student outcome measure. It includes multiple-choice items, a performance assessment, and attitudinal items that can be examined not only as outcome measures but also as measures of the participant responsiveness aspect of program implementation. It was developed and validated by Carlos C. Ayala of Sonoma State University, who collected validity data from over 400 students in the classrooms of 10 Hawai'i teachers .

Validity Data Collection and Analyses

After the lengthy development period, data for conducting validity studies were collected. Data for studies of the ISOCS and the ISQQ were from videotapes of 107 classroom periods of 16 FAST teachers in Hawai'i. Data for studies of the ISTQ were from two rounds of instrument administration to two samples of FAST teachers nationwide (*N*s = 79 and 156). Validity data for examining the ISSA were collected in 10 FAST classrooms in Hawai'i. An overview of the extent to which the findings of the data collection and analyses support conclusions favoring the validity of the data is are shown in Table ES-1.

Inquiry Science Observation Code Sheet (ISOCS) Validity

Evidence is given in the full report for the content-related validity, concurrent validity, and criterion-related validity of data collected with the ISOCS. The care with which the ISOCS development was conducted and the thoroughness of the process are evidence of content-related validity; they show that the instrument is designed to collect data that are (a) relevant to the measurement task (i.e., the data reflect what is intended to be observed) and (b) representative of the content domain. Videotapes were analyzed for nine of the 16 teachers whose classrooms had been observed. Analyses of the consistency of ratings between two coders showed a Pearson correlation of .99 between the numbers of codes assigned to the coding categories and a correlation of .53 between the total number of codes assigned to the teachers. We believe that these correlations are evidence of reliability. The consensus findings were less favorable. Agreement on the percentage with which two coders assigned identical codes in the first, independent round of coding ranged from 5% to 50%. These percentages are not high. Clearly, the reconciliation step in the coding process is essential for collecting ISOCS data. However, we believe that ISOCS inter-coder agreement results cannot be compared with the desirable or typical results for observations in which events are recorded in time periods (say, one code for every five-minute period). Our method is more stringent than this method, because our coders had to agree on precisely on codes at each observed moment of a class.

For the ISOCS concurrent validity analyses, we correlated the ISOCS results for the nine teachers with the results on the ISQQ. The Spearman's rho correlation of the ISQQ quality ranks with the percentage of codes that had been assigned in the student-comment code category = .52, and the Spearman's rho correlation of ISQQ ranks with the percentage that the teachers used follow-up statements and probing questions = .45. (The unit of analysis was the teacher; each teacher's percentage for a given code = the percentage of all the codes assigned for the teacher.) We believe that these correlations show a relationship between the two sets of results, thus supporting the validity of the ISOCS data.

For the ISOCS criterion-related validity analyses, we correlated the ISOCS results (in the form of teacher-student question-response exchanges, coded as the percentage of the total number of observed behaviors) with mean achievement test scores for six teachers. The correlation = .96—not a conclusive result, because the number of teachers analyzed is small, but nevertheless strongly suggestive of the validity of the observation data.

Inquiry Science Teacher Questionnaire (ISTQ) Validity

We reported evidence for the content-related validity, concurrent validity, and criterion-related validity of data collected with the ISTQ. Analyses were conducted of the ISTQ scales that measure inquiry science implementation and the context within which inquiry science is implemented. Data for the validity analyses were collected from a sample, closely representative of K–12 teachers nationwide, of 79 FAST teachers. Test-retest data for the Implementation Scale were collected from 156 FAST teachers nationwide.

The procedures with which the instrument was developed provide evidence of content-related validity. The results of reliability analyses, including coefficient alphas, factor analyses, and a test-retest study also provide content-related validity evidence. The alphas and the results of the factor analyses show that the implementation scale and the context scales are reliable. Because of the small number of teachers, we supplemented the factor analyses with parallel analyses; the results of all of them supported the appropriateness of our decision to conduct factor analyses. The results of the test-retest analysis of the Implementation Scale (Pearson correlation between the two administrations of the test = .76, and no variance due to occasion, as shown in a generalizability theory analysis) strongly support the reliability of data collected with the scale.

The results of concurrent validity analyses of ISTQ data show that the pattern of correlations among three context scales was as expected. The Pearson correlation of the Implementation Scale results with data collected on an implementation teacher log, which we developed and administered on multiple occasions to 66 of the teachers who completed the ISTQ = .66—substantial evidence for concurrent validity. The Pearson correlation of results on Teacher-Student Interaction factor of the ISTQ Implementation Scale with the ISOCS teacher-student interaction results =.50 and with the ISQQ ranks = .39. Both of these correlations are evidence of concurrent validity, although the strength of the evidence in both cases is tempered by outliers.

The results of the criterion-related validity analyses show a Pearson correlation of .37 between the Teacher-Student Interaction Subscale of the implementation scale with the ISSA multiple-choice posttest and a correlation of .43 with the ISSA extended-response item posttest. These correlations provide some evidence of criterion-related validity, although they were somewhat affected by an outlier on the ISSA.

Inquiry Science Questioning Quality (ISQQ) Validity

We gathered evidence of the content-related validity and concurrent validity of data collected with the ISQQ. The procedures with which the ISQQ was developed provide some evidence of content-related validity, although the strength of the evidence is qualified somewhat by the feedback of the five ISQQ judges: Some of the judges reported that it was difficult to make holistic judgments about quality, and some tended to add quality criteria of their own to those specified in the ISQQ procedures. Reliability findings, generated by five analyses, each from a different measurement tradition, also were mixed. Further analysis of the results showed that the judgments of two of the five ISQQ judges were less reliable than the judgments of three of the others, suggesting that the criteria for judging quality need revision and that the training period needs to be lengthened.

The results of concurrent validity analyses that are reported above in the section for the ISOCS provide good evidence for the validity of data collected with the ISQQ.

Inquiry Science Student Assessment (ISSA) Validity

We reported evidence for the content-related and concurrent validity of the achievement test, performance assessment, and seven attitudinal scales on the ISSA. Content-related validity evidence is found in the careful development of the instruments and in the alpha coefficients for components of the instrument. Concurrent validity evidence is found in the correlations among performance assessment components, of attitudinal scale scores among each other, and of the attitudinal scale results with the achievement test results, which all showed the expected pattern.

Table ES-1

The Extent to Which the Findings of Validity Studies Conducted During the Phase-I Study of the Effects of Professional Development and Long-Term Support on Program Implementation and Scaling Up Suggest Validity

Instrument or scale	Content-related validity evidence	Concurrent validity evidence	Criterion-related validity evidence
Inquiry Science Observation Code Sheet	*Strongly suggestive of validity:* •Careful, thorough development procedures *Moderately suggestive of validity:* •Reasonably high correlations between raters *Slightly suggestive of validity:* •Consensus results show lower percentages than desirable (although results are not comparable to observation rating studies)	*Strongly suggestive of validity:* •Correlations of .52 and .45 with quality ranks from the ISQQ	*Strongly suggestive of validity:* •Correlation of .96 with ISSA scores
Inquiry Science Teacher Questionnaire (ISTQ) Implementation Scale	*Strongly suggestive of validity:* •Careful, thorough development procedures •High test-retest correlation; no variance due to occasion (G-theory analysis) •Factor analysis and coefficient alpha results strong for Factor 1 (Teacher-Student Interaction) •Coefficient alpha results strong for Factor 2 (Connecting to the Outside World) *Moderately suggestive of validity:* •Coefficient alpha results acceptable for Factor 3 (Introducing the Investigation) *Contrary evidence:* •Small eigenvalues for Factors 2 and 3	*Strongly suggestive of validity:* •Correlation of .66 with teacher log results *Moderately suggestive of validity:* •Correlation of .50 between the Teacher-Student Interaction factor and the ISOCS teacher-student interaction results (small *N*; results perhaps affected by outlier) •Correlation of .39 with teacher quality rank (ISQQ) (small *N*; results perhaps affected by outlier)	*Moderately suggestive of validity:* •Correlation of .43 between the Teacher-Student Interaction factor and the ISSA total extended-response score and of .37 between the factor and the ISSA multiple-choice posttest
ISTQ context scales	*Strongly suggestive of validity:* •Careful, thorough development procedures	*Strongly suggestive of validity:* •The pattern of correlations about the total scale scores was as expected.	—
Inquiry Science Questioning Quality (ISQQ) method	*Strongly suggestive of validity:* •Careful, thorough development procedures *Slightly suggestive:* •Reliability results were mixed	*Strongly suggestive of validity:* •Correlations of .52 and .45 with teacher-student interaction on the ISOCS	—
Inquiry Science Student Assessment	*Strongly suggestive of validity:* •Careful, thorough development procedures •High coefficient alphas	*Strongly suggestive of validity:* Positive correlations found among attitude scales.	—