



Measuring The Effectiveness of Professional Development in Early Literacy: Lessons Learned

by Ormond Hammond, PhD*

November 2005

How do you bring high quality training to a teacher in a school on a remote island on the fringes of a small atoll somewhere in the Pacific Ocean? How do you even know what professional development is needed or what it means in that context? And how do you know if the training is effective?

In a region characterized by vast distances and even vaster cultural diversity, this was the challenge taken on by Pacific Resources for Education and Learning (PREL) as a part of its Regional Education Laboratory (REL) work for the past 4–5 years. The provision of professional development to teachers in the Pacific was the central focus of the research program called Pacific Communities with High-performance In Literacy Development (Pacific CHILD). The overarching research question has been, “What are the components of an effective professional development model for early reading improvement?”

The focus of this Research Brief is the methodology used to measure professional development (PD) effectiveness. It is *not* to present the *outcomes* of the research. Outcomes will be the focus of a later report. Here we will look at the needs that generated the research, what PREL did to meet those needs, and what lessons have been learned up to this time. What worked well? What did not? In particular, we will look at the development of a new instrument that was designed to measure the *quality* of PD as it is being delivered.

The Challenge: What Were The Research Needs?

The academic achievement of children whose first language is not English has long been a major educational concern. Those who come from cultural and linguistic minority backgrounds have been shown to fall short in school achievement. Measured through grades, retention in grade level, teachers’ judgments of student ability, and standardized tests, the academic performance of limited English proficient students generally lags behind other elementary school students (Moss & Puma, 1995).

Other mitigating factors further contribute to the region’s literacy dilemma, including lack of sufficient pre- and in-service training for teachers, lack of materials in local and English languages, unclear reading content standards, and shifting orthographies in Pacific languages.

Many students in the region are not learning to read in part because their teachers lack adequate content and pedagogical knowledge and skills to teach effectively. Poor teacher preparation is a general education problem, but it is particularly relevant to those who teach reading.

The Response: What Did We Do?

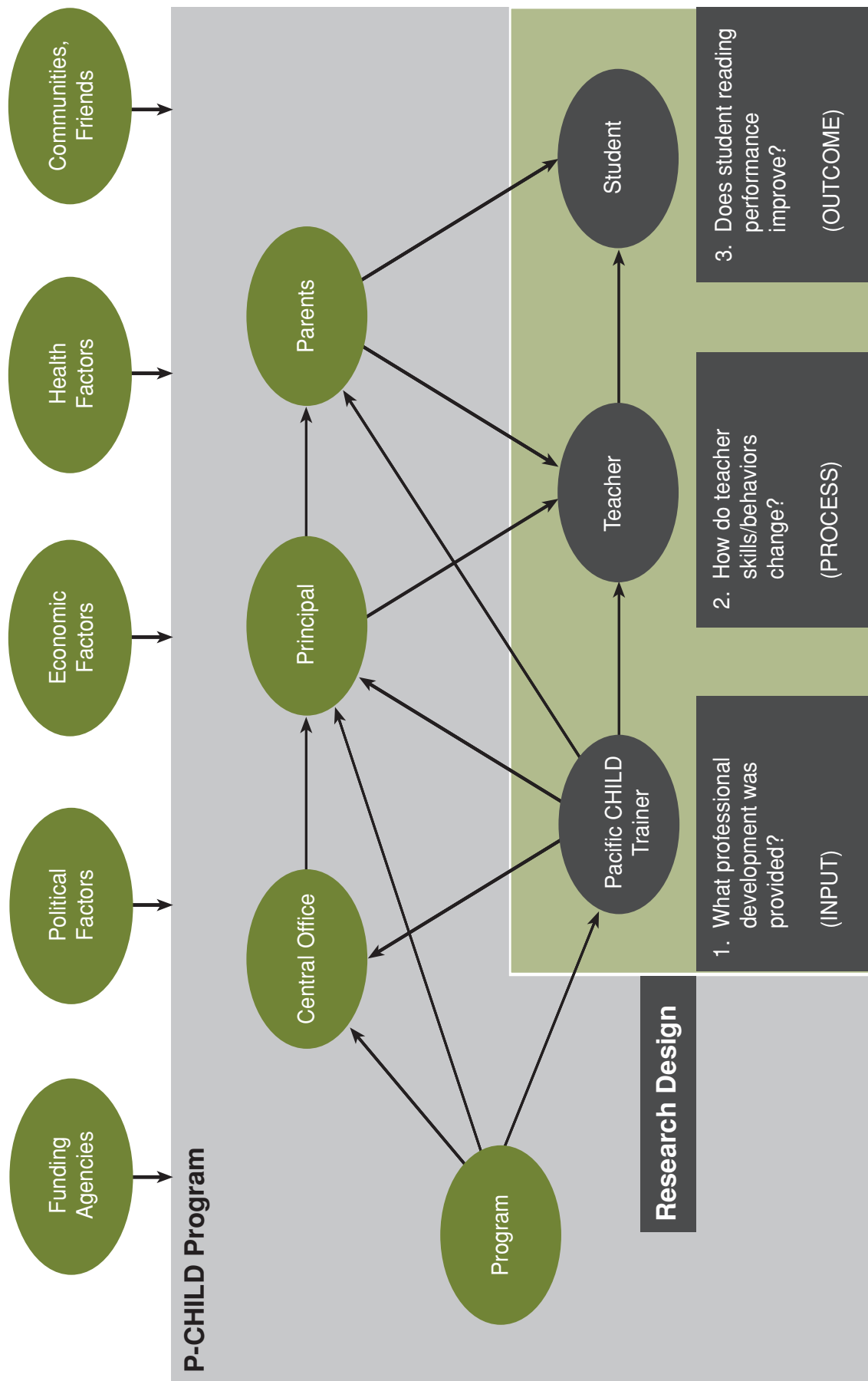
We developed a research model.

We began with a research design that isolated three of the components in a complex system: professional development, and teacher and student outcomes. The basic concept was to narrow the scope of the actual research to these three components, seen as stages in a logical chain: effective professional development logically leads to improvements in teacher knowledge and classroom skills, which, in turn, lead to an increase in student early reading proficiency (see figure 1).

The quest to identify the model’s components is based upon a multi-level, multi-method, grounded theory (Glaser and Strauss, 1967) approach. Data on professional development, and teacher, classroom, and student achievement have been collected to identify relationships between the professional development components and changes in teacher knowledge and instructional practices relative to student achievement data. To answer the research question, both quantitative and qualitative methodologies were used to provide the broadest range of analytical opportunities.

*Ormond Hammond, PhD is the Executive Director of PREL’s Center for Assessment, Research, and Evaluation.

Figure 1. A Conceptual Model for the Pacific CHILD Research Design in Context



We looked at “best principles” for evaluating professional development.

Much has been written about how to do professional development. Elmore (1995), for example, elaborates on what we know about the characteristics of successful professional development:

It focuses on concrete classroom applications of general ideas; it exposes teachers to actual practice rather than to descriptions of practice; it involves opportunities for observation, critique, and reflection; it involves opportunities for group support and collaboration; and it involves deliberate evaluation and feedback by skilled practitioners with expertise about good teaching. (p. 2)

However, the knowledge we have about how to organize successful professional development is still contextually limited. How do we organize successful PD practices for diverse teachers within diverse environments?

Michael Kamil summarized the National Reading Panel’s (NRP’s) review of scientifically-based studies on the effectiveness of teacher training, both pre-service and in-service. Only 11 studies of pre-service training met the NRP criteria for inclusion, and none of these looked at student outcomes or followed up on teachers after training. There were 21 studies reviewed on in-service training. Of these, 17 measured teacher outcomes and 15 showed improvement; 15 measured student outcomes and 13 showed improvement. Only in those cases where there were gains for teachers were there gains for students.

Guskey (1997; 2000) outlined several reasons why past efforts to identify elements of effective PD have not generated more definitive answers. For the Pacific CHILD research, we focused on the third reason: the neglect of quality issues. Efforts to identify effective PD elements focused more on the issues of quantity (i.e., documenting the presence or absence of particular elements). We decided to approach the task by establishing specific criteria to determine if a particular strategy was used appropriately, sensibly, and within the proper context. First we needed to find appropriate standards.

We developed a new instrument for determining the quality of professional development.

The Professional Development in Reading Observation (PDRO) Tool—We started with the idea that effective professional development is a specialized form of adult learning, and precepts of effective teaching should apply. Tharp and colleagues (Tharp, 1997; Tharp, Estrada, Dalton, & Yamauchi, 2000) have elaborated upon a set of pedagogical standards intended to inform instruction. There is some evidence that, when applied systematically, the Five Standards, as they have come to be known, result in affective, behavioral, and cognitive indicators of improved academic achievement, which is measured through self-report, as well as direct observation (Doherty, Hilberg, Pinal, & Tharp, 2003; Hilberg, Tharp, & DeGeest, 2000; Padron & Waxman, 1999).

In addition to the research on the Five Standards, the instrument draws upon research on adult learning and literature on effective professional development practices in reading (Brookfield, 1986; Center for Research on Education,

Diversity & Excellence [CREDE], 2002; Heimlich & Norland, 1994; Learning First Alliance, 2000; Rueda, 1998).

REL research staff developed an observation tool based on the Five Standards. A sixth standard was added to explicitly address activities inclusive of and responsive to diverse cultural practices. Table 1 includes the six standards with their accompanying descriptions.

Table 1. Standards for Teacher Professional Development

STANDARD 1	Facilitates learning and development through joint productive activity among professional development providers and participants.
STANDARD 2	Promotes learners’ expertise in professionally relevant discourse.
STANDARD 3	Contextualizes teaching, learning, and joint productive activity in experiences and skills of participants.
STANDARD 4	Challenges participants toward more complex solutions in addressing problems.
STANDARD 5	Engages participants through dialogue (instructional conversation).
STANDARD 6	Demonstrates techniques and activities that are inclusive of and responsive to diverse cultural practices.

The instrument enables observers to give a score to observed professional development on each of the six standards utilizing a 5-point, Likert-like scale with descriptors at the high, mid, and low points of the scale. More specifically, a score of 1 indicates the standard is not observed, 2 indicates the standard is emerging, 3 indicates the standard is developing, 4 indicates the standard is enacted, and 5 indicates the standard is integrated with two or more other standards.

The intent of the PDRO instrument is to be a data source that provides a rich description of PD events in research sites studied by PREL. PD event structures range from a single formal group (small or large) to one-on-one coaching or mentoring. These events are also characterized by individual delivery, as well as team delivery of PD.

We pilot tested the new instrument.

Observations using PDRO initially took place in spring 2003 with just a few observed PD events. Between February and May 2003, 13 observations were conducted using PDRO. The overall mean score of 2.2 reflected the category labeled as emerging on the 5-point rubric descriptions. While the research design does not track improvement of REL professional development, it is important to understand the gains made by our site-based staff over the course of this project.

A total of 51 PDRO records were collected and analyzed for the purposes of developing tests and measuring the larger research agenda of assessing the quality of PD as it is delivered. Table 2 illustrates the internal consistency of PDRO by item. The resulting standardized item alpha of .93 provides evidence of the instrument's reliability. While content, construct, and criterion validity were addressed, it is not possible to state that they have been established unequivocally. We did feel confident, however, that the instrument was a valuable addition to the set being used in the overall Pacific CHILD research. Table 3 shows the results of the PDRO over the course of the first year and a half.

Table 2. Professional Development in Reading Observation (PDRO) Instrument Item Reliability

	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5	Standard 6
Standard 1	1.000	.690	.739	.685	.589	.547
Standard 2	.690	1.000	.746	.794	.747	.661
Standard 3	.739	.746	1.000	.698	.621	.696
Standard 4	.685	.794	.698	1.000	.762	.704
Standard 5	.589	.747	.621	.762	1.000	.784
Standard 6	.547	.661	.696	.704	.784	1.000
Alpha =	.930					

Table 3. Professional Development in Reading Mean Scores

	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5	Standard 6
Mean Score	3.392	3.098	3.490	2.862	3.019	3.313
Total Mean Score	3.196					

Note. N = 51 observation records of professional development conducted at the five research sites (American Samoa, Kosrae, Marshall Islands, Palau, and Pohnpei).

For each standard, and overall, REL professional development providers were scoring on the average at the “developing” level according to the 5-point rubric descriptions. In other words, the PD providers were designing and carrying out activities that demonstrated a partial enactment of the standards. The total mean score of 3.196 demonstrated a marked improvement of REL PD from the inception of the PDRO during the spring of 2003. At this intermediate point, there was evidence supporting continued development of the instrument.

We also used other measures to describe the context of professional development.

These are summarized here simply to demonstrate the range of quantitative and qualitative measures used.

PD Log

The Weekly Professional Development Activity Log is designed to document (and monitor) the early literacy professional development activities conducted by REL staff. The log provides a greater depth of data on the PD experience at each site and allows us to more effectively factor out such components as content, duration, and format. The Web-based instrument provides information such as the number of PD activities that take place during a week, the location of the PD, the grouping format and strategies used, and grade levels served. There are also places for narrative information such as reflection and turning points that are noted by REL staff. The log also provides opportunity to code the motivation and need for the PD, the languages used during the PD event, and the PD content (early reading) and process (instructional strategies).



Teacher Interviews

The interviews tap teachers' experiences with Pacific CHILD PD. The interview asks for specific changes in teaching practices including when change happened, what caused it to occur, and why. The interview also seeks answers related to the professional development received throughout the year. Responses would provide a basis for looking at changes in teacher beliefs over time and changes in self-reported practice, and for a comparison of self-reported practice with classroom observations.

Staff Interviews

Our Advisory Panel (AP) suggested that the data should also include the background, perceptions, and experiences of REL staff, specifically those who provide direct PD to the schools. As a result, research staff developed a staff interview protocol. The interview taps REL staff's experiences as PD providers and as members of a larger research project. The interview asks for specific catalysts of change in their PD practices, including when change happened, what caused it to occur, and why. The interview also seeks answers related to the staff's perceptions of the research project and its impact on teacher and student practice. Responses provide a basis for looking at changes in PD practices over time, changes in self-reported practice, and a comparison of self-reported practice with observations using the PD Log and PDRO.

Where are we at this stage of the research process? We are in the final year of a 5-year project. We have collected data through the end of the 2004–2005 school year. These data will be fully analyzed in the next several months and a final report delivered.

The Findings: What have we learned?

Conducting research into what are effective PD practices in the Pacific region has given us many challenges, as well as provided a rich source of information and new research questions. Developing a new measure like the PDRO carries with it a whole stream of exciting directions for the future. The following paragraphs summarize what seem to be the major lessons learned to date.

In measuring the effectiveness of professional development in a complex multi-cultural context, it is necessary to use both quantitative and qualitative instruments and mixed methods.

There is no way that any single measure at any one of the three components of analysis would yield meaningful results. As intriguing as the initial findings of the PDRO on quality of PD are, they cannot give us the rich description of trainer–teacher–student interactions, nor can they tell us the intermediate or long-term outcomes of changes in teacher knowledge and skills and student achievement. Nonetheless, with the PDRO, we learned the value of systematically observing a PD event and quantifying the observations.

While it may be necessary to start with existing research methods or theories, it is sometimes beneficial to go beyond these in order to create new measures.

Development of the PDRO showed that it is possible to create a new measure that can provide useful data even while being pilot tested. There was a solid research base for the principles of effective teaching upon which the scales of the PDRO were based. But it was the cultural context of the Pacific that suggested adding a sixth scale to the instrument, extending the original theoretical concept.

Creating new measures means navigating a realm where establishing reliability and validity cannot be guaranteed.

Essentially, we learned the difficulties of trying to make solid research-based conclusions from new measures. Even with the best theoretical underpinnings, we simply have no way to solidify the construct validity of a measure like the PDRO. There is no comparative measure to relate it to. Once we complete the analysis of all three components, we will be better able to generalize about this, but more likely we will generate a plethora of new hypotheses. The process of validating any new measure takes a long time and many comparisons. So conclusions from this phase of the research will necessarily be quite tentative.

Measuring the effectiveness of PD in multicultural settings means constantly questioning one's assumptions.

Certainly we began this developmental research project with many assumptions about best PD practice, the components of an educational program, and what works in cultural context. Here are a few of the many assumptions we ended up discarding:

- We know what certain interventions are (e.g., “coaching”).
- Training staff will exhibit uniformly high standards.
- There is a discrete line between program (central office) staff and field staff.
- The provision and measurement of PD is something all cultures value.
- Staff can easily switch roles from service provider to research observer.

PD involves a complex system of learning.

We began with the assumption that there is a more or less linear progression from the provision of PD to changes in teacher knowledge and behavior to the logical outcome of improved student performance.

We learned that in PD learning progresses in a more systemic manner, including feedback loops in all directions. Measuring the quality of PD at just one point of that system can only capture a small percent of the causal soup. PD is a continuum and a system, not discrete between levels. Using the PDRO instrument led to more than just a numerical rating of the quality of PD. It led to knowledge about the interactive context.

At a more pragmatic level, we learned that the PDRO is a promising development, both in terms of advancing the con-

ceptualization of quality in PD and in helping to improve it in a formative sense. It was viewed as beneficial by those being observed in that it provided a context for mutual learning.

References

- Brookfield, S. D. (1986). *Understanding and facilitating adult learning*. San Francisco: Jossey-Bass.
- Center for Research on Education, Diversity, & Excellence. (2002). *The five standards for effective pedagogy*. Retrieved October 27, 2004, from www.crede.ucsc.edu/standards/standards.html
- Doherty, R. W., Hilberg, R. S., Pinal, A., & Tharp, R. G. (2003). Five standards and student achievement. *NABE Journal of Research and Practice, 1*(1), 1–24.
- Elmore, R. (1995). *Investing in teacher learning: Staff development and instructional improvement in Community School District #2*, New York City. New York: National Commission on Teaching and America's Future.
- Glaser, B., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine De Gruyter.
- Guskey, T. R. (1997, Spring). Research needs to link professional development and student learning [Electronic version]. *Journal of Staff Development, 18*(2), 36–40.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin.
- Heimlich, J. E., & Norland, E. (1994). *Developing teaching style in adult education*. San Francisco: Jossey-Bass.
- Hilberg, R. S., Tharp, R. G., & DeGeest, L. (2000). The efficacy of CREDE's standards-based instruction in American Indian mathematics classes. *Equity and Excellence in Education, 33*(2), 32–39.
- Kamil, M. (2005, June 8). *Research on Professional Development*. Presentation at A Focus on Professional Development in Early Reading. Honolulu, HI.
- Learning First Alliance. (2000). *Every child reading: A professional development guide*. Washington, DC: Author.
- Moss, M., & Puma, M. (1995). *Prospects: The congressionally mandated study of educational growth and opportunity. First year report on language minority and limited English proficient students*. Washington, DC: U.S. Department of Education.
- Pacific Resources for Education and Learning. (2003). *The professional development in reading observation (PDRO) instrument*. Honolulu, HI: Author.
- Padron, Y. N., & Waxman, H. C. (1999). Classroom observations of the five standards of effective teaching in urban classrooms with English language learners. *Teaching and Change, 7*(1), 79–100.
- Rueda, R. (1998). *Standards for professional development: A sociocultural perspective*. (Research Brief No. 2). Santa Cruz, CA: National Center for Research on Excellence, Diversity, and Education.
- Tharp, R. G. (1997). *From at-risk to excellence: Research, theory, and principles for practice*. (Research Report No. 1). Santa Cruz, CA and Washington DC: Center for Research on Education, Diversity and Excellence.
- Tharp, R. G., Estrada, P., Dalton, S. S., & Yamauchi, L. A. (2000). *Teaching transformed: Achieving excellence, fairness, inclusion, and harmony*. Boulder, CO: Westview.

Pacific Resources for Education and Learning (PREL) is an independent, nonprofit 501(c)(3) corporation that serves the education community in the U.S.-affiliated Pacific islands, the continental United States, and countries throughout the world. PREL bridges the gap between research, theory, and practice in education and works collaboratively with schools and school systems to provide services that range from curriculum development to assessment and evaluation.

Reprints available while supplies last. Please contact:



Pacific Resources for Education and Learning

900 Fort Street Mall ■ Suite 1300 ■ Honolulu, Hawai'i 96813

Phone: (808) 441-1300 ■ Fax: (808) 441-1385

U.S. Toll-free Phone: (800) 377-4773

U.S. Toll-free Fax: (888) 512-7599

Email: askprel@prel.org ■ Website: www.prel.org

Copies also available for download at <http://ppo.prel.org>.