Accounting for Every Child: The Evaluation of Reading Recovery®

Paper Presented at the 2006 Meeting of the American Educational Research Association
San Francisco, April 2006

Francisco X. Gomez-Bellengé
National Data Evaluation Center
College of Education
The Ohio State University

Inquiries may be directed to:
helpdesk @ndec.us
Or
614-688-8245

807 Kinnear Rd.
Columbus, Ohio 43212
http://www.ndec.us

Objectives

The purpose of this paper is to describe the national evaluation system for Reading Recovery (RR) in the United States. RR is an early literacy intervention serving over 120,000 first grade children every year. Through proprietary web-based technology and a relational database, the National Data Evaluation Center (NDEC) makes available evaluation reports to the RR network, allowing stakeholders to engage in process self-evaluations.

Description of Reading Recovery

RR was developed by New Zealand educator and researcher Dr. Marie M. Clay. In RR, individual students receive a half-hour lesson each school day for 12 to 20 weeks with a specially trained teacher. As soon as students can read within the average range of their class and demonstrate that they can continue to achieve, their lessons are discontinued, and new students begin individual instruction (Schmitt, Askew, Fountas, Lyons, & Pinnell, 2005).

The Trademark and the origins of the national evaluation

RR was implemented nationally in New Zealand in 1983 and a national data collection system was established the following year through the ministry of education (Schmitt et al., 2005). When RR was implemented in the United States, Marie Clay gave the trademark to The Ohio State University. The university in turn grants the trademark on a royalty-free basis to teacher training sites and university training centers, subject to annual renewal. Permission to use the trademark is contingent upon compliance with the *Standards and Guidelines of Reading Recovery in the United States*. One of these standards is to “Submit data on an annual basis to the NDEC using approved format, procedures, and materials” (Reading Recovery Council of North America, 2004).

How the evaluation works

In 2003-2004, the RR evaluation framework encompassed nearly 125,000 students, 15,000 teachers, 8,800 schools, 2,800 school systems, 495 teacher training sites and 22 universities. NDEC, located within the College of Education at The Ohio State University, has developed a web-based method of data collection, processing and dissemination. Following registration, teachers are provided with usernames and passwords. They then enter data about each child they serve (eight to ten), about themselves, and about their schools on the NDEC web site (http://www.ndec.us).

A three-tier computer server architecture (database server, application server and web server) provides the technological backbone of the system. It is capable of accommodating several thousand simultaneous on-line users. Previously, the center was staffed on an academic research project model, with a senior research associate, a research associate and temporary part-time students. The current setup is closer to a data processing center, with a director, three computer engineers and a customer service specialist. Two graduate students support the research function of the center but are not involved in day to day data processing.

The business layer of the web application verifies the accuracy and completeness of the data. Rules about data entry are enforced by the web site software. It prevents impossible values from being entered. It also forces the completion of required data fields, such as ‘date of first lesson’. When an improbable value is entered, a warning message is returned. Program data are stored on the NDEC computers. As soon as a teacher has entered some data about a student, this information is available on-line to the teacher leader and the university trainer.

Throughout the year, teacher leaders verify the data entered by the teachers, both for accuracy as well as to provide ongoing monitoring. At year-end, teacher leaders go through a check-out process, verifying the completeness of the information. Once done, they submit a command to the NDEC computers which then automatically start running a series of reports.

The NDEC Reports

As soon as the data are submitted, reports are run and stakeholders notified. For a given teacher training site, 30 or so reports adding up to thousands of tables are prepared. This happens in minutes. Stakeholders receive email notification that the reports are ready. They can then log on to the NDEC web site, download the reports and distribute and use them as appropriate.

Each level of the RR network receives one or more reports (for a more detailed discussion see (Gómez-Bellengé, 2004) The most detailed are the school-level reports, which contain nearly all the raw data sent by the teachers. The school report contains 26 tables while the short form of the district report contains 36 tables and charts. School districts, teacher training sites and university training center have available reports containing nearly three hundred tables and charts. Raw data in the form of Excel spreadsheets are also available.

Three basic types of data are collected; background demographic data, process data and outcome data. Background data on schools are combined with the Common Core dataset from the National Center for Education Statistics. Process data include length of interventions, other services received and various implementation and teacher

factors. Outcome data include pre- and post- literacy scores, status outcomes and performance relative to norms.

Utilizing the data

The RR evaluation is unusual in providing detailed data in user-friendly form to relevant stakeholders on a timely basis. The reports are made available in Microsoft Word format, which means they are editable documents. They are the basis for stakeholder-generated self-evaluations evaluations rather than externally-imposed evaluations, such as school or school district report cards published by many state departments of education.

Web-based computer technology allows the RR network to engage in evaluation best practices. Following is a breakdown of how the RR evaluation addresses or meets some of the *Program Evaluation Standards* (Joint Committee on Standards for Educational Evaluation, 1994):

- Utility Standard 1-Stakeholder identification: NDEC registers stakeholders annually, ensuring data will either be collected from them or be made available to them.
- Utility Standard 6-Report timeliness and dissemination: Reports are available within minutes of data submission to relevant stakeholders across the RR network.
- Feasibility Standard 1-Practical Procedures: Use of web-based technology and software-driven data cleaning minimize the time spent on data entry, verification and cleaning, while eliminating printing and mailing functions.

- Feasibility Standard 3-Cost effectiveness: The cost per student per year is under $4.
- Propriety Standard 1-Service Orientation: Every stakeholder and organization in the network receives relevant data on a timely basis. A Help Desk provides phone and email support.
- Accuracy Standard 1-Program Documentation: All relevant documentation is posted on the *Publications* page of the NDEC web site. Stakeholders can access current and archival data through protected web pages.

Clearly, the technology and the reports are but tools in a broader effort at evaluating RR at a local, regional, and national level. Evaluators are encouraged to seek and integrate sources of information that are not part of the national evaluation, such as local surveys, school district policies or state or district-based standardized test data.

Data-Driven Decision Making and the Impact of the Evaluation

The technological infrastructure provided by NDEC as well as in-services provided to teacher leaders allow these literacy specialists to engage in relatively sophisticated annual self-evaluations and in some cases use the raw data provided them to conduct action research.

The availability of data at all levels and on a timely basis has two main impacts. First, the wide dissemination of evaluation data is vital to continued support for RR at the school and school district level. The accountability that accompanies annual evaluation reports helps create collaborative structures and has become part of the culture of RR. Second, the availability of detailed data allows local decision makers to engage in data-

driven collaborative inquiry. For example, when federal and state mandates called for disaggregated data reporting, NDEC responded with these data. This allowed teacher leaders and school administrators to see how RR teachers fare with the different groups they serve. Because the reports emphasize data for factors that can be influenced at a given reporting level, such as productivity measures, they encourage decision-making.

The national evaluation

The Reading Recovery program evaluation uses a pretest-posttest two-group quasi-experimental research design. Given that this is an ongoing, annual internal evaluation, this is an exceptionally strong design (Whitehurst, 2002). The comparison group is a simple random sample of two first-grade children selected from each school served by RR. This random sample allows for both pre-post two-group comparisons as well as for setting national norms for the six tasks of the *Observation Survey*, the assessment instrument used by RR (Clay, 2002).

Because implementation decisions are local, the annual report serves as an outcome evaluation. National averages for various values, such as student absence rate or fall *Observation Survey* mean scores are a useful frame of reference for those evaluating local data. Part of the evaluation process consists of determining the proportion of students served who have reached an average reading range relative to the reference group. This is the main reason why data are collected on a comparison group.

What the data show about RR

The most notable aspect of RR evaluation data is the extent to which results replicate over time and space. A close examination of NDEC's national reports, as well as an examination of state reports, reveals remarkably consistent patterns in student

outcomes as measured by the six tasks of the *Observation Survey*, by success rates as measured by end of program status outcomes and by non-Reading Recovery based indicators such as placement rates in special education and grade retention rates.

Over the years, about 60% of all children served, even if for only one lesson, successfully discontinue their series of lessons, returning to regular classroom instruction. For those having the opportunity to receive a full series of lessons, about 75% successfully discontinue their series of lessons. As a group, children served by RR tend to enter first grade reading below RR text level 1 (20^{th} percentile), compared to level 4 for the overall first grade population, and, for those who discontinue, end first grade with average text levels around 19 (46^{th} percentile), compared to about 20 for the general population (Gómez-Bellengé & Thompson, 2005).

The percentage of children evaluated by RR teachers as having reached average reading levels (e.g., successfully discontinued) who are subsequently placed in special education services for Learning Disability consistently averages near zero percent, or about 154 for 72,000 successfully discontinued students on a national basis every year. This is a lower placement level than for the general population, even though these children were by definition at-risk at the beginning of first grade. Similarly, these children are extremely unlikely (159 out of 71,000) to be retained in grade because of reading difficulties.

Discussion

The availability of raw data allows school districts to conduct external evaluations of RR. Decision-makers are interested in student outcomes on state-mandated

standardized tests. Unfortunately, school districts rarely have the resources to conduct proper longitudinal studies.

One challenge of the RR internal evaluation is moving stakeholders from a top-down, formal outcome evaluation paradigm in which evaluation reports are written for someone else in a position of authority to a process, self-evaluation paradigm in which evaluation is ongoing and the emphasis is on using data to make decisions rather than writing reports. The scale of the network, with 15,000 teachers, 8,000 principals and the very small size of the NDEC staff (5) makes the dissemination of these concepts challenging. This is accomplished mostly through evaluation training sessions provided by 40 university faculty and the NDEC director to teacher leaders, who in turn work with teachers and administrators.

Another challenge is responding to the need for school districts to provide longitudinal follow-up data on state assessments for students formerly served by RR. Because RR is implemented in 52 federal entities and many states have more than one reading or writing assessment, the technical challenge of data entry and processing of such a large variety of data within the $4 per student budget is daunting.

In some ways, the evaluation of RR parallels that of many school district evaluation offices, where decision makers often want answers to questions for which either data are not available to provide scientifically acceptable answers or resources are not available to gather and analyze such data.

Reference List

Clay, M. M. (2002). *An observation survey of early literacy achievement*. (2nd ed.) Portsmouth, NH: Heinemann.

Gómez-Bellengé, F. X. (2004). Accounting for every child: The U.S. annual data evaluation. *Journal of Reading Recovery, 3*, 68-74.

Gómez-Bellengé, F. X. & Thompson, J. R. (2005). *U.S. norms for tasks of An Observation Survey of Early Literacy Achievement* (Rep. No. NDEC 2005-02). Columbus: The Ohio State University, National Data Evaluation Center.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of education programs*. (2nd ed.) Thousand Oaks, CA: Sage.

Reading Recovery Council of North America (2004). *Standards and guidelines of Reading Recovery in the United States* Columbus, OH: Author.

Schmitt, M. C., Askew, B. J., Fountas, I. C., Lyons, C. A., & Pinnell, G. S. (2005). *Changing Futures: The Influence of Reading Recovery in the United States*. Columbus, Ohio: Reading Recovery Council of North America.

Whitehurst, G. J. (2002). Evidence-based education (EBE). In *Student Achievement and School Accountability Conference*.