

# RESEARCH BRIEF

DEBRA HEATH

JANUARY 2006

---

## A COMPARISON OF METHODS FOR EVALUATING CHANGES IN TEACHER PRACTICE

---

### BACKGROUND

APS created the Reading Instruction Rubric in the spring of 2003 “to give the district, schools, administrators, and teachers data on the level of implementation of effective reading instruction within the district and at individual schools.” Developers also cast the rubric as a professional development tool school staff could use to foster high quality reading instruction in kindergarten through 8<sup>th</sup> grade. The rubric is divided into three components:

- *Classroom Context* concerns the classroom’s physical environment and materials.
- *Engaged Practice* addresses the effectiveness and comprehensiveness of a teacher’s instructional practice.
- *Assessment* studies the use of assessment in planning and implementing reading instruction.

The Assistant Superintendent’s Office used the Reading Instruction Rubric to assess teacher practice at selected elementary schools in the fall of 2003.<sup>1</sup> Third-party observers completed the *Engaged Practice* section. Teachers themselves completed the *Classroom Context* and *Assessment* sections.

In the spring of 2004, Research, Development and Accountability created an on-line survey based on the APS Reading Instruction Rubric. The survey was one component of a larger evaluation intended to describe APS’ Reading First program activities and their effects on school systems and teacher practices.<sup>2</sup> RDA administered the on-line survey to Reading First teachers in eight schools in April 2004.

The on-line Reading First survey included 27 items, verbatim, from the Reading Instruction Rubric.<sup>3</sup> Unlike the original rubric, however, the on-line survey asked teachers to rate their practice at two points in time: at the start of the school year and at its end. This “retrospective pre-test” method collected pretest and posttest data simultaneously. The intention was to capture changes in teacher practice between the start and end of the school year.

One benefit to the retrospective pretest method is that it economizes time and effort. Research also shows that it has the potential to yield more reliable and valid results because respondents

---

<sup>1</sup> The Reading Instruction Rubric was one of multiple instruments used by the Assistant Superintendent’s Office in a site review process conducted jointly with the NM State Department of Education.

<sup>2</sup> A PowerPoint presentation summarizing evaluation results may be obtained from RDA by calling 505-848-8710.

<sup>3</sup> Items were selected in consultation with the evaluation subcommittee of the Reading First Leadership Team.

use the same standards to rate themselves before the program and at the end of the program. In traditional pretests, participants often over-estimate their own knowledge and understanding. The traditional pre-posttest, therefore, may *under*-estimate knowledge gains. Retrospective pretests eliminate this “response shift bias” and may thereby produce more valid results.<sup>4</sup>

## RESEARCH QUESTIONS

***Research Question #1:*** *Were third-party observer ratings of teacher practice different from teachers’ self-assessment ratings?*

The Assistant Superintendent’s Office asked RDA to compare its real-time third-party ratings of *Engaged Practice* to the retrospective self-assessments teachers provided on the on-line survey. If results were similar, they would consider adopting a retrospective self-assessment approach rather than continuing the more costly third-party approach.

***Research Question #2:*** *Did retrospective pretest scores differ from traditional pretest scores?*

Another question of interest to RDA was whether teachers’ retrospective self-assessments on the on-line survey, completed in April 2004, would differ from their real-time self-assessments provided in the fall of 2003. Both surveys asked teachers to complete the *Classroom Context* and *Assessment* sections to reflect practice in the fall of 2003. This gave RDA an opportunity to compare traditional pretest and retrospective pretest results.

## METHODS

The Assistant Superintendent’s Office (ASO) collected data in the following manner:

- Trained observers visited teacher classrooms for a full class period and rated teacher practice on the *Engaged Practice* section of the Reading Instruction Rubric.
- Teachers themselves completed the rubric’s *Classroom Context* and *Assessment* sections.
- At the study school, 15 teachers participated in the ASO assessment.

RDA conducted the Reading First Retrospective Pretest On-Line Survey (RF) as follows:

- RDA collaborated with the Reading First Program Coordinator and each school’s Instructional Coach (IC) to gain the participation of all teachers of kindergarten through 3<sup>rd</sup> grade students. At the study school, this included one special education teacher.
- Teachers gathered in the school’s computer center to complete the on-line survey.
- RDA’s Reading First Technical Assistant set up the on-line survey on each teacher’s computer and answered questions as needed during survey completion.
- At the study school, 17 teachers participated in the on-line survey.

Each Reading Instruction Rubric item describes four levels of practice, displayed in progressive order from left to right, without labels. Figure 1 shows an item from the *Engaged Practice* section, with response options from the online survey. Teachers rated their practice “Now”

---

<sup>4</sup> Pratt, McGuigan & Katzev (2000). Measuring Program Outcomes: Using Retrospective Pretest Methodology. *American Journal of Evaluation.*, 21(3), 341-349.

(within the previous 2 weeks) and “Before” (during the first 9 weeks of the school year). The ASO version did not have the “Before” response options.

Figure 1. Example of a Reading Instruction Rubric item, displayed with response options from the Reading First retrospective online survey.

The lesson has no clear objectives.	The objectives are not directly tied to the components of a comprehensive reading curriculum or they do not match student need.	Clear objectives address the components of a comprehensive reading curriculum. They are matched to student need, but there are too few or too many objectives.	Clear lesson objectives address 1 to 3 components of a comprehensive reading curriculum and the objectives are matched to student needs.
BEFORE			
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
NOW			
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

RDA used SPSS statistical analysis software to run independent samples t-tests to test for equality of means between the ASO and RF assessment results.

- *Engaged Practice* section results were used to compare third-party observer ratings and retrospective self-assessment ratings (Research Question #1, n = 15).
- *Classroom Context* and *Assessment* results were used to compare teachers’ real-time pre-test ratings with their retrospective pre-test ratings (Research Question #2, n = 17).

## RESULTS

### *Research Question #1*

There were no significant differences between the ratings of third party observers and those of teachers themselves. Differences between the mean scores of the ASO observers and the mean scores of teachers responding to the on-line survey were insignificant at the  $p < .05$  level.

- This result suggests that the retrospective pre-test method may be a good alternative to the more costly third-party observer approach.

### *Research Question #2*

Teachers’ real-time pretest ratings tended to be higher than their retrospective pretest ratings. Mean scores on 6 out of 8 *Classroom Context* items and 3 out of 5 *Assessment* items were significantly higher in the real-time ASO survey than in the on-line retrospective survey.<sup>5</sup>

- This result aligns with research suggesting that traditional pretest self-reports tend to over-estimate pre-intervention knowledge, skills and practice.<sup>6</sup>

<sup>5</sup>  $p < 0.05$ . All real-time mean scores were higher than retrospective mean scores although differences were not statistically significant for 4 of the 13 measures.

<sup>6</sup> Umble, K., Upshaw, V., Orton, S., & Matthews, K. (2000). “Using the post-then method to assess learner change.” American Association for Higher Education Assessment Conference, Charlotte, NC.

Possible reasons for the differences between real-time and retrospective scores include:

- Teachers may have over-valued their classroom’s environment and materials and their own assessment practices before receiving Reading First (RF) professional development.
- The retrospective pretest may have given teachers an opportunity to assess pre-program practices with a common and clear set of instructional standards gained from RF training.

These interpretations are supported by findings from the larger Reading First program evaluation which suggest that RF interventions made teachers more aware of their own limitations.

Other explanations for higher real-time scores compared to retrospective pretest scores are possible. Two alternatives follow.

- The four-level continuum built into the Reading Instruction Rubric may have been too narrow for measuring change. If teachers felt they had improved but not enough to reach the highest level of instructional practice (Level 4), the retrospective pretest would have forced them to choose a Level 1 or Level 2 “Before” rating in order to give themselves a Level 2 or Level 3 “Now” rating.
- The desire to please others or paint the Reading First program in a positive light may have influenced teachers’ ratings. Teachers may have wanted to show that their practice had improved as a result of Reading First professional development. Consciously or unconsciously they may have selected levels that showed growth and/or reflected achievement of instructional standards, rather than choosing the levels that most closely reflected their actual practice.

## **LIMITATIONS**

The ASO third-party assessment data had many missing responses. Raters had been instructed to rate lack of practice as “Level 1” but some may have used blanks instead. Since it is not possible to interpret blanks with confidence, RDA coded the data as “missing.” This reduced the size of the sample on the *Engaged Practice* section in a possibly biased way, since many potential low-level ratings were eliminated.

The size of the study’s sample was small, with 17 teachers completing the online survey and 15 participating in the ASO assessment. Responses were anonymous so it is not possible to know exactly how many teachers participated in both assessments. More research is needed to explore the validity and generalizability of this study’s findings.

## **RECOMMENDATIONS**

RDA plans to incorporate the retrospective pretest method into other research and evaluation projects. This will amplify RDA’s understanding of the method’s validity and applicability to different types of questions and settings. It also will allow RDA to make a valuable contribution to the research and evaluation literature.