Schools frequently use Reading Recovery as an intervention to meet the needs of children who are most at risk of literacy failure or as a response to intervention (RTI) component (Clay, 2005; Dunn, 2007; Jones, Johnson, Schwartz, & Zalud, 2005). The specially designed individualized instruction, meticulous record keeping, and early delivery of Reading Recovery make it well-suited for both purposes.

After a rigorous review of research, the What Works Clearinghouse (WWC) report from the United States Department of Education’s Institute of Education Sciences confirmed the positive effects of Reading Recovery on student performance (What Works Clearinghouse, 2007). Research has also demonstrated that gains initially produced by Reading Recovery are sustained over time (Askew, Fountas, Lyons, Pinnell, & Schmitt, 1998; Briggs & Young, 2003; Brown, Denton, Kelly, & Neal, 1999; Forbes & Szymczuk, 2003; Homan, 2002; Jaggar & Simic, 1996; Lukas, 2001; Lyons, Pinnell, & DeFord, 1993; Ruhe & Moore, 2005; Schmitt & Gregory, 2005). Although studies demonstrate that students who successfully complete a series of Reading Recovery lessons and discontinue their need for intervention tend to remain within average performance of their peer groups in subsequent years, it must be noted that these effects depend on multiple factors beyond Reading Recovery. The expectation is that those who successfully completed their series of Reading Recovery lessons would be no more at risk than other average students at the end of first grade.

Reading Recovery in South Dakota
Schools have implemented Reading Recovery in South Dakota for more than a decade. South Dakota’s annual data consistently show positive gains for Reading Recovery students (Zalud, 2003, 2004, 2005, 2006, 2007). Stakeholders have expressed interest in two issues relating to progress of children after they have completed Reading Recovery: (a) students’ performance levels in years subsequent to Reading Recovery, and (b) how well the decision to discontinue the intervention predicts later success on the Dakota State Test of Educational Progress (Dakota STEP), South Dakota’s annual statewide assessment.

Purpose
This study was designed to (a) examine the Dakota STEP performance level categories (below basic or unable to perform grade-level content standards, basic or performing below grade-level expectations, proficient or meeting grade-level expectations, and advanced or exceeding grade-level expectations) attained by students who either successfully completed their series of lessons discontinuing their need for intervention or were recommended for further literacy support, and (b) analyze the relationship of Reading Recovery end of intervention status decisions to normal curve equivalency (NCE) scores attained on the third-, fourth-, and fifth-grade Dakota STEP. The researchers hypothesized that students who successfully completed a series of lessons and discontinued from Reading Recovery would likely fall within the proficient or advanced performance levels and attain NCE scores within or above the average of their grade-level peers. Additionally, we hypothesized that the students who were recommended for further literacy support would likely fall within the basic or below basic performance levels and achieve below average NCE scores compared to their grade-level peers.

Measures
Dakota STEP was used to assess students’ reading achievement. The foundation of the Dakota STEP is the Stanford Achievement Test Series, Tenth Edition, Abbreviated Battery, Form D (Stanford 10). The Stanford 10 was supplemented with items to assess the Dakota state content standards in both reading and
math, creating a standards-based test, which incorporated both norm-referenced and criterion-referenced elements. Augmenting the Stanford 10 to assess the state content standards allows South Dakota to satisfy the assessment requirements contained in the federal No Child Left Behind Act of 2001, and to obtain useful norm-referenced data that allow for comparison of its students’ achievement with their cohorts across the country. The Dakota STEP is administered to students in grades 3–8 and 11 in the spring of each year, and yields individual performance data in the form of mean scaled scores and mean national normal curve equivalents (Harcourt, 2005a). The mean scaled score and the mean national NCE were used to determine students’ later achievement for this study. The mean scaled scores were used to rank a student’s performance level as below basic, basic, proficient, or advanced. The performance level descriptors (also called proficiency levels) are written at a broad level and have the following definitions:

1. Advanced: A student performing at the advanced level exceeds expectations for that grade level. The student is able to perform the content standards for the grade at a high level of difficulty, complexity, or fluency.

2. Proficient: A student performing at the proficient level meets expectations for that grade level. The student is able to perform the content standards for the grade at the level of difficulty, complexity, or fluency specified by the standards.

3. Basic: A student performing at basic level performs below expectations for that grade level. The student is able to perform some of the content standards for below the level of difficulty, complexity, or fluency specified by the grade level standards.

4. Below Basic: A student performing at below basic is unable to perform the content standards for the grade. (Harcourt, 2005b, p. 24)

A normal curve equivalent is a standard score derived from the percentile rank that ranges from 1 to 99, with a mean of 50 and standard deviation of 21.06. The mean NCE is an indicator of performance of the typical student in the group in terms of percentile rank. Since percentile ranks cannot be averaged, all of the percentile ranks are converted to NCEs; the NCEs are averaged, and the mean NCE is converted to a percentile rank (Harcourt, 2005b).

Methods

The study focused on one Reading Recovery teacher training site which included several districts and schools. Schools that had consistently implemented Reading Recovery across a 3-year span were targeted for the study.

Data were collected on 176 students. The reading achievement of these students at the end of third-, fourth-, and fifth-grade was measured in two different ways: (a) the Dakota STEP mean scaled score which determined the performance categorical scores (basic, below basic, proficient and advanced), and (b) the NCE scores. Measuring students’ reading achievement using two different scores for the Dakota STEP allowed for two different types of analyses: Pearson chi-square and linear regression.

To complete the analysis, students’ available test scores were grouped into grade-level cohorts according to the year they participated in Reading Recovery: The students in Group A received Reading Recovery instruction in the 2000–01 school year; Group B received Reading Recovery instruction in the 2001–02 school year; and Group C in the 2002–03 school year. The number of students’ available scores is listed by grade level.

The study focused on one Reading Recovery teacher training site which included several districts and schools. Schools that had consistently implemented Reading Recovery across a 3-year span were targeted for the study.
and participation year in Table 1. Third-grade scores were collected for all three groups. Fourth-grade scores were collected for Groups A and B. Fifth-grade scores were collected for Group A only. Data revealed about 70% of the students in the study had successfully discontinued lessons, and 30% of the students were recommended for further literacy support at the end of their series of Reading Recovery lessons.

Results
Does the end of intervention status decision differentiate performance-level categories? The end of intervention status decision did differentiate categorical (below basic/basic and proficient/advanced) performance levels in third and fourth grades, but did not in fifth grade.

Third grade
The results of Pearson chi-square as reported in Table 2 revealed a significant difference in reading performance between the two end of intervention status groups in third grade \[\chi^2 (2, N = 142) = 6.879, p < .01\]. In third grade, slightly over half of those students who were recommended for further literacy support performed in below basic or basic levels, while the majority of students who successfully completed their series of lessons performed in proficient or advanced levels.

Fourth grade
The results of Pearson chi-square as reported in Table 3 did show a significant difference in reading performance between the two end of intervention status groups in fourth grade \[\chi^2 (2, N = 101) = 4.438, p < .01\]. Similar to the third-grade results, the fourth-grade results showed that many of the students who were recommended for further literacy support were mostly categorized in the proficient and advanced levels. This high level of performance by both groups showed that the end of intervention status decision did not differentiate later reading performance at this grade level.

Fifth grade
The results of Pearson chi-square as reported in Table 4 did not show a significant difference in reading performance between the two end of intervention status groups in fifth grade \[\chi^2 (2, N = 60) = 1.484, p > .05\]. By fifth grade, the students who had successfully completed their series of lessons discontinuing their need for Reading Recovery and the students who were recommended for further literacy support were mostly categorized in the proficient and advanced levels. This high level of performance by both groups showed that the end of intervention status decision did not differentiate later reading performance at this grade level.

Table 1. Available STEP Test Scores by Participation Year

<table>
<thead>
<tr>
<th>Year Served in RR</th>
<th>Number Tested in Grade 3</th>
<th>Number Tested in Grade 4</th>
<th>Number Tested in Grade 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A 2000–01</td>
<td>42</td>
<td>41</td>
<td>39</td>
</tr>
<tr>
<td>Group B 2001–02</td>
<td>44</td>
<td>38</td>
<td>17</td>
</tr>
<tr>
<td>Group C 2002–03</td>
<td>37</td>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

D = Discontinued; R = Recommended

Table 2. Frequencies of Third-Grade Dakota STEP Categorical Performance Level and End of Intervention Status Decision

<table>
<thead>
<tr>
<th>End of Status Outcome</th>
<th>Below Basic and Basic</th>
<th>Proficient and Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discontinued</td>
<td>28</td>
<td>71</td>
<td>99</td>
</tr>
<tr>
<td>Recommended</td>
<td>22</td>
<td>21</td>
<td>43</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>92</td>
<td>142</td>
</tr>
</tbody>
</table>

Test Statistic Value | df | p-value
Pearson Chi-Square   | 6.879 | 1.000 | 0.009*

Note: Casewise deletion missing data. Scores from one school did not provide performance categorical scores for 34 students for at least one grade level. Across the six districts, 60 students had missing NCE scores for at least one grade level. The missing data varied by grade. Analyses were run with available data resulting in different Ns for the two analyses.
Does the end of intervention status decision predict later reading achievement when using the NCE score, a continuous measure? Linear regression analysis revealed the end of intervention status decision to be a significant predictor of later achievement in third ($R = .255, p < .01$) and fourth grades ($R = .363, p < .01$) but not fifth ($R = .209, p > .05$). These data indicated that when compared using a national average norming curve, students in their third- and fourth-grade year who successfully completed their series of lessons and discontinued were performing at higher levels than students who were recommended for further literacy support. However, by the fifth-grade year, the end of intervention status decision showed no significant difference between the groups because both groups were performing within average levels or higher.

### Discussion

In this study, the end of intervention status decision was predictive of later achievement to an extent. In examining categorical performance data and the NCE data, students who successfully completed their series of lessons and discontinued were performing at higher achievement levels than the students who were recommended for further literacy support in third and fourth grades, despite the score used to measure achievement. In fact, about three-fourths of students who successfully discontinued the series of Reading Recovery lessons were still performing in the proficient or above performance level in third and fourth grade. However, by fifth grade, the end of intervention status decision was no longer predictive of students’ reading achievement. The lack of categorical performance level differentiation between the two end of intervention status groups in fifth grade resulted because the majority of students in both end of intervention status groups scored at the proficient or above performance levels on the state assessment meaning they read at or exceeded expectations for that grade level. By Grade 5 about 80% of all Reading Recovery students, whether successfully discontinued or recommended for further support, performed at or above grade-level expectations.

When using the NCE score as the measure for achievement, the results were similar to the categorical performance data. The Reading Recovery end of intervention status decision was a good predictor of later reading achievement in third and fourth grades but not fifth grade. This may

### Table 3. Frequencies of Fourth-Grade Dakota STEP Categorical Performance Level and End of Intervention Status Decision

<table>
<thead>
<tr>
<th>End of Status Outcome</th>
<th>Below Basic and Basic</th>
<th>Proficient and Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discontinued</td>
<td>15</td>
<td>51</td>
<td>66</td>
</tr>
<tr>
<td>Recommended</td>
<td>15</td>
<td>20</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>71</td>
<td>101</td>
</tr>
</tbody>
</table>

Test Statistic Value

**Pearson Chi-Square**

$df$ $p$-value

<table>
<thead>
<tr>
<th>Test Statistic Value</th>
<th>Pearson Chi-Square</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4.438$</td>
<td>1.000</td>
<td>0.0035*</td>
<td></td>
</tr>
</tbody>
</table>

* $p < .01$

Note: Casewise deletion missing data. Scores from one school did not provide performance categorical scores for 34 students for at least one grade level. Across the six districts, 60 students had missing NCE scores for at least one grade level. The missing data varied by grade. Analyses were run with available data resulting in different $N$ for the two analyses.

### Table 4. Frequencies of Fifth-Grade Dakota STEP Categorical Performance Level and End of Intervention Status Decision

<table>
<thead>
<tr>
<th>End of Status Outcome</th>
<th>Below Basic and Basic</th>
<th>Proficient and Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discontinued</td>
<td>6</td>
<td>33</td>
<td>39</td>
</tr>
<tr>
<td>Recommended</td>
<td>6</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>48</td>
<td>60</td>
</tr>
</tbody>
</table>

Test Statistic Value

**Pearson Chi-Square**

$df$ $p$-value

<table>
<thead>
<tr>
<th>Test Statistic Value</th>
<th>Pearson Chi-Square</th>
<th>$df$</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1.484$</td>
<td>1.000</td>
<td>0.223*</td>
<td></td>
</tr>
</tbody>
</table>

* $p > .05$

Note: Casewise deletion missing data. Scores from one school did not provide performance categorical scores for 34 students for at least one grade level. Across the six districts, 60 students had missing NCE scores for at least one grade level. The missing data varied by grade. Analyses were run with available data resulting in different $N$ for the two analyses.
be because most students, despite their end of intervention status, were performing within average ranges or higher in fifth grade. This supports the idea that providing one-to-one Reading Recovery instruction to students who are most at risk leads to the development of self-extending systems that enable students to continue to learn and profit from instruction well after the intervention has ended. However, it must also be considered that outside variables likely affected the results as more time passed.

Longitudinal studies have been found to be an effective tool to evaluate program effectiveness and can be used to identify specific benefits and disadvantages of program implementations (Tierney, 1991). Evaluating the effects of Reading Recovery by examining performance in subsequent years provides evidence that the program brings the lowest-achieving first-grade students into average ranges of reading achievement following the intervention. Examining how well the end of intervention status decision differentiated or predicted later reading achievement is an important component of program evaluation.

The results of this study suggest Reading Recovery professionals are making good decisions about discontinuing, and that children who receive a full series of lessons do well after the intervention has ended. However, decisions for the students who were recommended for further literacy support need to be investigated further. The results from the third and fourth grade suggest that the decision to recommend these students for further literacy support was justified because many of these students were performing at the basic level, meaning they were unable to meet grade-level expectations. In Grade 5, the results of the students who were recommended for further literacy support suggest that many of these students were meeting or exceeding grade-level expectations. This could be explained several ways. It could be that more of these students should have been considered as discontinued at the end of their series of lessons. Another possibility is that the further literacy support they received was appropriately matched to students’ needs. Yet another possibility could be that by Grade 5 the results may not be reflective of the decision, but the power of Reading Recovery regardless of the outcome status.

**Recommendations**

The results from these data suggest continued attention to the collaborative roles of teacher leader, Reading Recovery teacher, and classroom teachers. When these professionals work together to closely monitor a child’s success during the intervention, as well as following the interventions, sustained gains can result. More needs to be known about successful collaboration of these professionals.

In addition, more research is needed to examine classroom activities that support a student’s achievement in literacy when the child does not successfully complete the Reading Recovery lesson series. More in-depth study of how students recommended for further action perform in different interventions: Title I, special education, small literacy groups, after-school programs, etc. would be useful. Finally, more research is warranted that analyzes the acceleration rate by grade level and the acceleration curve between end of intervention status groups.

**References**


Susan Gapp is a reading professor in the Division of Curriculum and Instruction at The University of South Dakota. She formerly worked as a Reading Recovery teacher leader in South Dakota for 6 years.

Garreth Zalud is the chairman of the Division of Curriculum and Instruction at The University of South Dakota. He has been a Reading Recovery trainer and director of the USD Reading Recovery Training Center for 10 years. He serves on numerous Reading Recovery Council of North America and North American Trainers Group committees.

Dale Pietrzak is the assistant dean for the School of Education at The University of South Dakota. He teaches courses in school psychology, counseling, and educational research.


