Reading Psychology, 40:525–550, 2019

ISSN: 0270-2711 print / 1521-0685 online
DOI: 10.1080/02702711.2019.1629518

SELF-CORRECTIONS AND THE READING PROGRESS OF STRUGGLING BEGINNING READERS

JEROME V. D'AGOSTINO
Department of Educational Studies, The Ohio State University, Columbus, Ohio

ROBERT H. KELLY
EECE, Queens College CUNY, Flushing, New York

EMILY RODGERS
Teaching and Learning, The Ohio State University, Columbus, Ohio

*While there is consensus that self-corrections (SCs) ought to be coded as part of oral reading assessments, less agreement exists as to what, if any, role self-correcting plays in reading development. The purpose of this study was to address limitations of prior research and provide a more statistically accurate estimate of the role of SC in early reading progress. A dataset that included running records (*n = *3184) and baseline and posttest literacy achievement data for first-grade struggling readers (*n = *140) was used. Baseline achievement, errors, and total words served as covariates. To investigate the potential moderating effects of errors and words, we interacted SC with errors and with total words. We then conducted hierarchical linear modeling to examine the role of SC in beginning reading. Our findings indicate that SC significantly and positively predicted early reading progress for struggling readers. Further, SC was found to have additional predictive power for students who were at the earliest period of literacy development. Results suggest that early reading instruction ought to pay careful attention to self-correcting behavior during oral reading and that oral reading assessments may be enhanced with specific instructions about what to do with SCs.*

The study of self-corrections (SCs) following an error in oral reading has long been a cornerstone of literacy research. At least as far back as the 1930's, researchers detected a positive relation between reading proficiency and the predilection of the reader to self-correct errors (see e.g., Fairbanks, 1937). The

Address correspondence to Jerome V. D'Agostino, Department of Educational Studies, The Ohio State University, Columbus, OH 43214. E-mail: dagostino.22@osu.edu

impact of that early work on SCs can be found today in the plethora of oral reading assessments that contain coding conventions to mark SCs of errors (see e.g., Bader & Pearce, 2013; Beaver & Carter, 2009; Roe & Burns, 2011).

Yet despite its presence in numerous assessment coding schemes, SC data rarely are used for diagnostic or formative purposes, mainly because little is known about SCs role in the emergent literacy process. Although it now is a well-established finding that good readers tend to self-correct more than lower-achieving readers, there remains little, if any, affirmative evidence regarding the role of SC in literacy development. Few literacy experts likely would disagree that SCs serve as an indicator or sign of reading progression, but far fewer can assert with empirical support that SCs stimulate or propel readers to greater levels of reading proficiency.

Most SC research has been correlational in nature, and most, if not all prior studies, were not conducted with statistical control for initial reading levels, the number of errors made by the reader, and text-level difficulty. Advanced readers may self-correct more than less advanced readers simply because they have more opportunities to engage with more challenging text, make more errors, and consequently self-correct more. Prior SC researchers have claimed that SC ratio variables, such as dividing the SC count by the number of errors (Clay, 1969, 2000, 2013), or dividing SCs by the total number of words (Thompson, 1984), control for errors or text difficulty respectively, but in actuality the computation of such variables does not represent proper statistical control.

Without stronger empirical evidence, we will not adequately elucidate if correcting errors while reading "…contributes to the forward thrust of reading competency" (Clay, 2001, p. 195), or provides "…a window on the reading process" (Goodman & Goodman, 2004, p. 623), or merely "result from a failure to control for text difficulty level" (Share, 1990, p. 185). Prior research as a whole has yielded mixed conclusions about the role of SC in the early reading process, mainly due to the rather low quality of methodological rigor of those studies. The purpose of this study was to examine SC's potential contribution in predicting reading growth by employing better statistical control than found in past empirical investigations.

Literature Review

Theorizing the Role of SC in Early Reading Development

Self-correction in oral reading is defined here as the young reader's spontaneous correction of an error made while reading aloud (Clay, 2001, p. 185). An error is made when the reader reads a word that is not in the text; regardless of whether the change makes sense. Thus, errors should not be conflated with miscues because they are conceptually and theoretically different from each other (see Harmey and Kabuto (2018) for a discussion about the differences between errors and miscues). In the example that follows, the student makes an error by reading fish instead of shark, and then self-corrects:

> Text: This book has a shark swimming in the sea (Cowley, 2015).
>
> Student: This book has a fish swimming in the sea. Oh, not fish! That word is shark.

Self-correcting behavior is distinct from solving words correctly. For example, had the student decoded the word correctly by reading, "This book has a sh-, sh-, -ark, shark swimming in the sea." no error and thus no SC would be recorded. Note that by definition, a SC cannot occur unless an error has first occurred.

Conventional wisdom has it that learning and errors go hand in hand because errors provide opportunities to learn. Similar assertions are made about the role of errors and SCs in beginning reading development. Wood (1998, p. 202) argued that SCs are "self-instructive" while Allington (1983) noted that the presence of SCs strongly predicts reading progress. Clay (2001) seems to have gone further still, arguing that SC may even propel early reading development (e.g., Clay, 2001; Clay in Doyle, 2013).

Indeed, competing cognitive-constructivist views of the reading process converge on the view that problem-solving is vital for reading development, that the act of problem-solving itself while reading is the change agent for development to occur. Whether change occurs as a result of assimilation and accommodation as stage theorists argue (Chall, 1967) or as result of making adaptive choices, as overlapping waves theorists posit (Clay, 2001; Sharp, Sinatra, & Reynolds, 2008; Siegler, 1996) the opportunity to problem solve appears to be a central and critical feature to explain how young children progress from novice to skilled readers.

Self-correction may foster a beginning reader's solving of new words in that the reader's "executive system" for word identification is strengthened by corrective feedback (Vellutino & Scanlon, 2002, p. 590). In the example provided earlier, the assumption is that the student's substitution of fish for shark arguably made sense and fit English language syntax, but it neglected to take into account letter-sound information. The error provided an opportunity for the student to note the mismatch between the sounds in fish and the letters in shark and to self-correct, thus providing corrective feedback to the executive system. Had the student not noticed the mismatch the teacher could use the opportunity to teach the student to notice the letters in shark and how to use them to self-correct, thus expanding the executive system.

Extant Evidence about SC's Role is Weak

Despite the conventional wisdom that reading proficiency is correlated with self-correcting errors, at least among primary-grade students, much of the evidence that supports that assertion is derived from broader-based studies that focused on observing multiple factors involved in teacher–student interactions. Many studies relied on tallying differences among students across intact classroom reading groups in their reactions after making an error, and higher reading groups tended to spontaneously self-correct more than middle or lower reading groups (Chinn, Waggoner, Anderson, Schommer, & Wilkinson, 1993; Hoffman & Clements, 1984; Hoffman et al., 1984).

Note that examining the mean SC differences between intact groups does not address if SC was at least partially responsible for increased reading proficiency, because the prior proficiency level of the students before the measurement of SC remained unknown. As part of the analyses in their study, Hoffman and Clements (1984) regressed posttest achievement scores onto a number of potential predictors including pretest achievement, but they did not report the predictive capacity of SC. Harlin (1981) regressed spring Metropolitan Achievement Test scores of second-graders onto their SC levels, and found no effect for SC after including fall Concepts about Print scores, yet her goal was to address the combined predictive power of print awareness and SC rather than to use concepts about print as a prior achievement measure.

There has been some effort to examine the degree to which reading proficiency was related to change in SC rates over the school year. Cohen (1974) categorized first-grade children as either poor or good readers based on whether they were in the first or fourth quartiles of their monthly number of correct words read from either Open Court or Moonbeam tradebook reading passages. From November to June of the school year, she logged monthly SC rates, and found that good readers increased from virtually no SCs to about 17–19% of their errors, whereas poor readers increased from virtually no SCs to about 4% of their errors. Although the results seemed to imply that SC was beneficial to the progress of the good readers, the author did not statistically test for group differences while controlling for prior proficiency, nor did she attempt to link reading growth with SC rate changes.

Perhaps the most cited research pertaining to SC was conducted by Clay (1969) as a portion of her dissertation (Clay, 1966). She created four quartile groups based on students' scores on a combined word test when the children were six years of age. During the school year beginning at age 5, she recorded their SCs on a set of leveled classroom instructional books. Over their entire count of words read, she computed what she referred to as the self-correction ratio (SCR), which was the number of SCs divided by the number of total errors (including SCs and uncorrected errors). She found that the proportion of SCs to errors varied from 0.36, 0.26, 0.12, and 0.05 for the four quartile groups from highest to lowest, respectively.

Although she never addressed causation in her interpretation of the findings, one could (mis)construe that Clay (1969) was attempting to examine the causal nature of SCs because she referred to the classifying variable as "reading progress group." In actuality, students were not categorized based on the reading growth over the school year (they were categorized instead based on year-end word test achievement status), so she did not properly measure student growth as a function of SC activity while controlling for baseline proficiency.

Extant Approaches to Measuring SC are Flawed

Thompson (1981, 1984) focused on two aspects of Clay's (1969) SCR metric to support his claim that her results were too weak to draw causal conclusions. He maintained that Clay "confounded" SC with uncorrected errors because the proportion of SCs appears in both the denominator and numerator of her SCR index. Thus, according to Thompson, her measure simply detected that low-performing students had greater proportions of uncorrected errors compared to more proficient students, because as SCs decreased, uncorrected errors had to increase given her formula.

Besides claiming that Clay (1969) "confounded" SC and uncorrected errors, Thompson (1981, 1984) argued that her reliance on the SCR metric ignored text difficulty, in that more proficient students likely found the text easier to read than less proficient students. To address these two apparent problems with the SCR index, he suggested instead an index based on the division of SCs by total words read (SCTW), predicated on the assumption that total words represented a proxy for text difficulty.

In reanalyzing Clay's (1969) data, he found no differences between the four quartile groups on his new measure. From his findings, Thompson argued that group differences detected by Clay were likely due to text difficulty rather than the effects of SC, and he suggested that in order to properly isolate the causal role of SCs, one would need to compare low and high progress groups who made the same amount of total errors. Thompson's (1981, 1984) claims were convincing enough that subsequent SC researchers (Chinn et al., 1993; Hoffman & Clements, 1984; Hoffman et al., 1984) utilized his index (SCs divided by total words) instead of Clay's (SCs divided by total errors).

Share (1990) attempted to address Thompson's (1984) suggestion of comparing the SC rates of equally-proficient low- and high-progressing students. He matched six high progress Grade 1 children to six low progress Grade 2 children, and six high progress Grade 2 children to six low progress Grade 4 students on the Burt Word Reading Test. Students were asked to read a series of graded reading passages until their accuracy levels dropped below 95% (instructional level). After computing both SCR and SCTW indices for students on the read passages, he found no statistical difference between the four groups on either index, which he interpreted as evidence that, after controlling for proficiency level, reading growth (or progress) was not influenced by greater incidence of SC.

The validity of Share's (1990) claims was greatly diminished by limitations of his research design. Although it was assumed by the author that second graders who scored at the same level on the Burt Word Reading Test as first graders (and likewise fourth graders to second graders) must have progressed at a slower rate, one must assume that both groups entered school at the same proficiency level. Without proper baseline control, the claim is unwarranted. Furthermore, Share omitted the average number of graded passages read by each group (he provided the overall grand mean only), leaving the reader to assume that text difficulty was properly controlled and equivalent across the groups. Furthermore, based on his group averages and standard deviations, we computed effect sizes of about 0.52 and 0.30 on the Thompson Index (SCTW) and 0.30 and 0.17 on the Clay Index (SCR) for Grade 2 to Grade 1 students and Grade 4 to Grade 2 students, respectively. The rather substantial effect size differences between the grade levels may signify that his study was underpowered.

Remedying Problems with SC Measurement

Upon reviewing Clay's (1969) and Thompson's (1981, 1984) discussions of their respective metrics, it became apparent to us that both SCR and SCTW do not statistically control, in a covariate analysis sense, for the number of errors or words read, but instead address the potential moderating effects of those variables. Both SCR and SCTW are "per ratio" variables, one indicating the rate of SC per error (SCR) and the other per word (SCTW). Dividing one variable, x, by another variable, z, is the same as multiplying x by the reciprocal of z. This division process is identical to multiplying x by the reciprocal of z, $1/z$, which then results in the interaction of x and z in reciprocal terms. Thus, the SCR is the interaction between SC and the inverse of errors, and SCTW is the interaction between SC and the inverse of total words.

In past studies such as Clay (1966), Thompson (1984), and Hoffman et al. (1984) the predictive capacity of the SCR or SCTW interaction terms were examined in isolation, which likely yielded spurious estimates of the SC impact. Kronmal (1993) explained how using "per ratio" variables alone as predictors can lead to misleading results for at least three reasons. First, in regression analysis, to control for z, one enters it as a main effect covariate in a prediction model that includes x. If one wants to examine the degree to which z moderates the x–y relationship, one would include the main effects of x and z along with the x^*z interaction term. Excluding the main effects of x and z violates the principle of marginality (see Nelder, 1977).

Given that SCR and SCTW are interaction terms, neither represents the control of either errors or words. In order to account for variation in outcomes that was due to the number of errors made or words read, one would include those variables as covariates in statistical models. That is, interaction terms are not statistical controls in a covariation sense, but rather represent the moderating effect of the second variable.

"Per ratio" variables used in isolation result in fallacious regression models with spurious slopes and intercepts (Kronmal, 1993). If one were to divide SC by either error or words, one would need to divide all other variables in the statistical model, including the outcome variable, by the same divisor in order to produce an accurate set of regression coefficients. Doing so, however, creates variables and regression coefficients that are difficult to interpret because all variables have been converted to a "per ratio" metric. Instead of dividing SC by error or words, it seemed more sensible to include the main effects of SC, errors, and words in the analysis, and include non-reciprocal interaction terms, such as SC*errors or SC* words, if the goal is to examine the moderating effects of errors or words. We could not identify a single study that examined the role of SC on subsequent reading achievement in this manner.

Purpose and Research Questions

While there is strong theory to suggest that SC behavior plays a more important role in early reading development than merely providing evidence of processing, the existing empirical evidence is far too weak to draw any conclusions regarding the association of SC and growth in reading achievement. Prior correlations between SC and achievement status may have resulted because more proficient readers may have self-corrected more than less proficient readers, and thus, SC served as a proxy for pretest achievement levels. Prior research that relied on SCR or SCTW likely produced spurious results due to the statistical issues that result from using "per ratio" variables in isolation.

The purpose of this study, therefore, was to examine the role of SC in improving reading after considering prior reading achievement and to properly model the effects of SC, errors, and total words. The following questions guided our inquiry:

1. What is the relation between SC and the progress of beginning readers?
2. Do text difficulty (measured by total words) and number of errors moderate the relation of SC and reading progress?

We used running record data from first-grade students who were participating in an early literacy intervention, Reading Recovery (RR), along with their baseline and posttest *Observation Survey of Early Literacy Achievement* (OSELA; Clay, 2013) scores. A running record is an oral reading assessment tool developed by Clay (1967, 1972, 1979, 2000, 2013). Over each student's array of running records, we computed the student's average SC, errors, and total words. To investigate the potential moderating effects of errors and words, we interacted SC with each variable. We then conducted a series of hierarchical linear models (HLMs) to examine the role of SC in beginning reading.

We hypothesized that average number of SCs and average number of total words read would positively predict posttest scores, and average number of errors would negatively predict posttest scores, after controlling for OSELA pretest values. Thus, SCs were expected to significantly predict adjusted posttest scores after controlling for errors and words. Once those variables were included in the model, we did not expect errors or words to interact significantly with SCs.

Context for the Study

The RR context was an appropriate one to study SC behavior for several reasons. The design of the intervention requires that running records are administered in each lesson, and as such, the RR context provides a large number of data points on oral reading accuracy. Equally appealing about the RR context is that teachers receive training in how to administer the assessment in a standard way and indeed, interassessor agreement with calculating highest text level read with at least 90% accuracy has been found to be quite high (Denton, Cianco, & Fletcher, 2006, p. 29).

In addition, running records are widely used beyond the RR context and thus relevant to many elementary teachers and researchers (see e.g., Johnston & Afflerbach, 2015; Kragler, Martin, & Schreier, 2015); indeed the International Reading Association (IRA) noted in a position statement that excellent teachers use, among other assessment tools, running records to assess student progress (Santa et al., 2000). It is important to note, however, that we focused in this study on the contribution of SC in the emergent literacy process. RR was the context from which the data were gleaned, but no aspect of the intervention was examined in this study.

Method

Dataset

We conducted this research using a data set formed to study various characteristics of learning and instruction for young struggling readers. These data included pretest and posttest OSELA (Clay, 2013) scores, and daily running records for 140 students enrolled in 38 schools from a large urban Midwest school district who received RR while they were in first-grade during fall of 2013. The students were selected for RR because they had reading scores that placed them in the lowest 20% of first-grade students at their respective schools.

The average age of the students was 6.5 ($SD = 0.41$) at pretesting, and 82 (58.6%) of the 140 students were male. Less than 10% (12 of 140) had an identified disability, and 24 of 140 (17%) were English Language Learners. Over 95% of the students (134 of 140) were eligible for either free or reduced-price lunch, and in terms of ethnicity, about 59% (83 0f 140) were African-American, 21% were white (30 0f 140), 13% were Latino/a (18 of 140), 5% were bi-racial (7 of 140), and less than 1% (2 of 140) were Asian-American. Forty teachers in their RR training year offered the 140 students the intervention (3.5 students per teacher). Over 90% of the students received RR for more than 15 weeks, and none of the students were in the program for more than 21 weeks.

Measures

OSELA

The OSELA (Clay, 2013) contains six subtasks that are designed to measure achievement on key constructs of early literacy, including concepts about print (Concepts about Print [CAP]), oral reading (Text Reading Level [TRL]), letter knowledge (Letter Identification [LI]), reading vocabulary (Ohio Word Test [OWT]), writing vocabulary (Writing Vocabulary [WV]), and phonemic awareness (Hearing and Recording Sounds in Words [HRSW]).

We used the total score from the six tasks (D'Agostino, 2012), because it has been found to have distributional properties and other psychometric features that make it more advantageous for research purposes than the separate tasks scores (D'Agostino, Rodgers, & Mauck, 2018). The total score scale ranges from 0 to 800: 0 indicative of a student unable to identify any letters or words and reading at a pre-primer level, and 800 represents a student able to identify all upper and lower case letters, writing more than 100 words correctly and reading texts beyond third grade level. The alpha coefficient of the total score is above 0.90 (D'Agostino et al. 2018).

RUNNING RECORDS

Running records (Clay, 2000) are designed to record all the reading behaviors the student makes while reading aloud. Standard procedures are used to code accurate reading, substitutions, insertions, deletions, repetitions, SCs, and appeals for help. A running record contains the title and difficulty level of the book being read by the student, along with the student's total number of words read. The administration yields the number of errors, percent oral reading accuracy, and SC ratio.

Although we had the running records for each day of the week the student was present for RR lessons, we decided to analyze two running records per week for each student to make the data set more manageable. We selected the running records taken on Tuesdays and Thursdays of each week. Our rationale for selecting Tuesdays and Thursdays was that we expected there to be fewer missed lessons on those days than on Mondays or Fridays when school holidays frequently occur. In cases of missing data on a Tuesday or Thursday lesson, we selected the running record from the previous Monday lesson, and if that date also contained missing data, we then selected Wednesday's lesson for analysis. This selection protocol yielded 3814 running records across the 140 students.

For each student, we computed the student's average number of SCs across the running records. We also computed the average number of (1) errors, (2) total words read, and (3) book level for each student. We included SCs in the count of errors, following Clay's method (Clay, 2000). Our original intent was to use average book level in lieu of average number of words read because we reasoned that book level represented a more consummate measure of text difficulty. Though total words is a property considered in the book leveling process, other features such as words per T-unit and number of unique words also account for variation in text difficulty (Cunningham, Spadorcia, Erickson, Koppenhaver, Sturm, & Yoder (2005)). We discovered, however, that average total words and average book level correlated at 0.95, which supported Thompson's (1984) claim that total words is a sufficient proxy for text difficulty. Thus, and in accord with prior SC research, we decided to use average total words instead of average book level for further analyses.

Very little research has been conducted to document the interrater reliability of coded errors and SCs. As part of her dissertation research, Clay (1966) correlated one rater's coding of errors ($r = 0.98$) and SCs ($r = 0.68$) on four student running records taken two years apart. Given that only one rater and four students were included in the sample, and the long lag between time points (two years), the accuracy of those reliability estimates is questionable.

In a generalizability analysis of error coding in running records, Fawson, Ludlow, Reutzel, Sudweeks, and Smith (2006) found little variance between raters (no correlations were reported), but considerable variance across passages. Hence, running records taken across multiple books or passages were found to yield a sufficiently reliable estimate of a student's error rate.

In the current study, we relied on teachers' coding of errors and SCs taken across an array of about 27 running records on average per student on books that varied in terms of word count. We calculated the alpha coefficients for SCs, errors, and words across the panel of running records. The alpha coefficients were: 0.91 for SC, 0.80 for errors, and 0.89 for words read. We also analyzed the typical change rate on all three metrics, and found positive and significant increases for SCs, errors, and words. That is, across the 15- to 20-week interval, the typical student made gains in the amount of words in the books they read, the errors they made, and the number of errors corrected. Yet because the alpha coefficients were rather strong for all three indicators, there was little change in the rank order of the students on the measures from beginning to end of the intervention period.

Data Analysis

Each of the 40 teachers taught a small cluster of about three or four students, so we expected that at least some of the exit OSELA score variance would lie between teachers (hence, we anticipated a nonzero intraclass correlation). An HLM unconditional model verified our assumption – 15% of the variance in students exit scores was explained at the teacher level. To base the results on accurate standard errors and degrees of freedom, we conducted HLM analysis, with the key variables serving as Level-1 fixed effects and no Level-2 predictors. We first computed an unconditional model (no predictors), which yielded the 0.15 intraclass correlation, and the total student-level variance. Our second model included entry OSELA scores to determine the variance explained by pretest achievement levels. Our third model consisted of all three covariates, pretest OSELA, average error, and average total words.

We then entered average SC in our fourth model to examine the degree of additional variance explained by SC after considering the control variables of prior achievement, the typical error rate, and typical number of words read. The fifth model added two interaction terms, SC by errors and SC by total words, to investigate if errors or words moderated a possible SC effect. These interaction terms served as a more accurate way to examine the predictive capacity of SCR and SCTW.

Results

As can be seen in Table 1, students made, on average, about two SCs on each running record. Though the typical occurrence of SCs seems slight, note that the average number of SCs ranged from near zero to greater than five. Students made about seven and one quarter errors and read over 95 words on average across all books they read. Similar to SCs, the error and

TABLE 1 Variable descriptive statistics

Variable	*M* (SD)	Min	Max
Entry OSELA	343.12 (34.91)	221.00	428.00
Exit OSELA	460.34 (53.30)	312.00	562.00
Avg SC	1.95 (1.19)	0.16	5.06
Avg errors	7.24 (2.28)	1.79	15.48
Avg words	95.56 (29.56)	26.81	183.64

Avg errors: average number of self-corrections plus errors; Avg SC: average number of self-corrections; Avg words: average number of total words read; OSELA: *An Observation Survey of Early Literacy Achievement* (Clay, 2013).

total word variability is considerable across the 140 students. The typical SCR rate of 0.27 (1.95/7.24) is comparable to values found in Clay's research (1969).

Table 2 contains the correlations between the variables. It can be seen from the table that average SC, errors, and words all correlated in expected directions with both entry and exit OSELA scores. That is, students with lower entry and exit achievement levels tended to make more errors, read less words, and make fewer SCs than students who scored higher on the OSELA. Those with higher exit OSELA scores also tended to have higher entry achievement levels. Thompson (1981) maintained that students who read more words likely would make more SCs, and indeed the positive and significant correlation between SC and total words supported his claim. Though positive, the correlation between SCs and errors, however, was not significant—students who made more SCs were no more likely to make more errors than students who made fewer SCs. The sizable correlations between the outcome, exit OSELA, and entry OSELA, errors, and words warranted controlling for those variables while examining the contribution of SC in predicting posttest achievement.

The results of the HLM analyses are presented in Table 3. The unstandardized coefficients are provided in the cell entries, along with asterisks to indicate the degree of statistical significance of each coefficient (testing the null hypothesis that the coefficient value is zero). The Level-1, or student-level, variance of exit OSELA scores is provided for each model with the percentage of Level-1 variance explained for each model in comparison to the unconditional model (Model 1). Those

TABLE 2 Variable intercorrelations

Variable	2	3	4	5	6	7
1. Entry OSELA	0.59***	0.32***	−0.49***	0.61***	0.01	0.46***
2. Exit OSELA		0.41***	−0.72***	0.79***	0.09	0.54***
3. Avg SC (A)			0.13	0.39***	0.94***	0.93***
4. Avg errors (B)				−0.05	0.41***	0.10
5. Avg words (C)					0.33***	0.62***
6. A*B						0.88***
7. A*C						—

Avg errors: average number of self-corrections plus errors; Avg SC: average number of self-corrections; Avg words: average number of total words read; OSELA: *An Observation Survey of Early Literacy Achievement* (Clay, 2013).
** $p < 0.01$.
*** $p < 0.001$.

values were computed by subtracting the variance resulting from each conditional model from the unconditional variance (2375) and dividing the difference by 2375.

Model 1, the unconditional model, yielded an overall intercept of 460, which is the grand mean of Exit OSELA values presented in Table 1. After adding Entry OSELA as a predictor in Model 2, the student-level variance was reduced to 1624, which represented about a 32% reduction. The coefficient for Entry OSELA in Model 2 of 0.85 represents the unit change in exit OSELA values for every one-point change in the predictor. Thus, for each additional entry OSELA score, exit OSELA scores increased 0.85 points. For Model 3, we added average errors and average words, which increased the variance explained to nearly 65%. Note that average errors was not a significant predictor in the model, and once average errors and average words were added, entry OSELA no longer was significant. Thus, average words was the dominant covariate.

Model 4 included average SC along with the three covariates. After controlling for entry OSELA, average errors, and average words, the SC variable significantly and positively predicted posttest achievement. Average SC explained about an additional 1% of the student-level variance. An average SC increase of one point was associated with an exit OSELA increase of 6.79. In Model 5, the two interaction terms were added. The average SC–average error interaction was not significant, but the average SC–average

TABLE 3 Summary of hierarchical linear models predicting exit OSELA

	1	2	3	4	5	6
Intercept	460***	460***	460***	460***	460***	460***
Entry OSELA		0.85***	0.13	0.10	0.12	.14
Avg SC (A)				6.79**	42.73**	32.35***
Avg errors (B)			−2.69	−3.33*	−0.89	−3.08**
Avg words (C)			1.30***	1.22***	1.54***	1.58***
A*B					−1.16	
A*C					−0.22***	−0.25***
Student variance	2,375	1,624	844	816	728	745
Variance explained		31.62%	64.46%	65.64%	69.35%	68.63%

Cell values represent unstandardized beta coefficients. Avg errors: average number of self-corrections plus errors; Avg SC: average number of self-corrections; Avg words: average number of total words read; OSELA: *An Observation Survey of Early Literacy Achievement* (Clay, 2013). Student variance is the variance in exit OSELA at the student-level for each of the six HLM models. Variance explained is the percent of the total variance (2375) from the unconditional model (Model 1) explained by each successive conditional model.

*$p < 0.05$.

**$p < 0.01$.

***$p < 0.001$.

words interaction significantly and negatively predicted exit OSELA.

To examine if the average SC–average words interaction was significant due to a suppressor effect of the average SC–average error interaction, we removed the latter term to create a sixth model. As can be seen in Table 3, the average SC–average word interaction remained significant and in the same direction (in another model, we removed average SC–average word and retained average SC–average error, but the latter term was not significant). The average SC–average words interaction explained an additional 3% of the outcome variance (Model 6 compared to Model 4).

We conducted further analyses to examine why average SC–average words negatively predicted the outcome. A negative interaction could have resulted from favorable outcomes for students with either: (1) low word counts and high average SCs; or (2) from students with high word counts and lower average SC levels. To examine the pattern that led to the negative interaction, we created four groups of students based on median splits of the SC and word count variables. We conducted an ANCOVA controlling for entry OSELA and average errors with the two median split variables and the interaction term in the model (essentially duplicating Model 6).

As was the case in the HLM analysis, the interaction term was significant (results not shown). Figure 1 presents the means and 95% confidence intervals for the four groups on exit OSELA after controlling for average errors and entry OSELA scores. It can be seen in the figure that there was no significant mean difference between both SC groups, low and high, for those students who read higher-word count books on average. For students who had lower average word counts, students with greater SC levels had significantly higher adjusted OSELA scores than students who self-corrected less on average. (Figure 1).

Discussion

The purpose of this study was to first address limitations of prior SC research and then to provide a more statistically

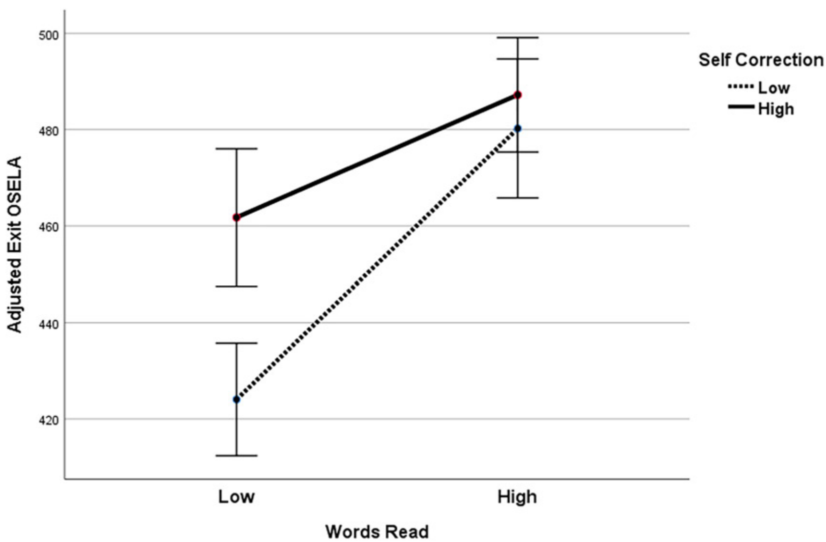

FIGURE 1 Adjusted exit OSELA mean differences for low and high average words and average self-correction. More frequent self-correction mattered for students who had lower average word read, but not for students who read more words. Error bars represent 95% confidence intervals.

accurate estimate of the association between SC and reading growth. We used a dataset from first-grade students who were participating in RR that included over 3000 running records and baseline and posttest *Observation Survey of Early Literacy Achievement* (OSELA; Clay, 2013) scores. Over each student's array of running records, we computed the student's average SC, errors, and total words. To investigate the potential moderating effects of errors and words, we interacted SC with each variable. We then conducted a series of HLM analyses to examine the role of SC in beginning reading.

Perhaps the primary reason that little is known about SC and learning to read is the lack of research that has properly examined the SC-achievement association. The great majority of prior research has not controlled for baseline achievement; moreover, attempts to control for student errors and text difficulty have resulted in metrics such as the SCR and SCTW that failed to serve as proper covariates. Instead, those "per ratio" metrics likely led to misleading statistical findings. Using a data set consisting of multiple running records of text reading over time, we attempted to rectify prior research limitations by properly controlling for prior achievement, errors, and text difficulty.

Self-correction Appears to Play a Role in Reading Development

We found that controlling for errors (Clay, 1969), text difficulty (Thompson, 1981), and prior achievement was necessary in order to more effectively isolate the association of SC and achievement growth. In our sample of 140 struggling readers in urban high-need schools, we found that pretest achievement accounted for over 30% of the posttest achievement variance. Adding average errors and average words read across the running records added about another third of the variance explained, but words read was the primary predictor among the three covariates. Average words appeared to serve as the best indicator of performance during the intervention period, and once considered in the model, seemed to render initial achievement status less critical. This finding seemed to support

Thompson's contentions that controlling for words read was vital to understand the role of SC.

With the three covariates considered, SCs per running record accounted for an additional 1% of the variance, which is sizable given that a large amount of the variance (nearly 65%) was explained before entering the predictor. A student one standard deviation (1.19) above the average SC mean (1.95) would have scored about eight points higher on the exit OSELA compared to a student at the mean average SC, which translates into a 0.15 standardized mean difference on the outcome. Although not a large effect size, it is not trivial either, and because SC explained a significant amount of variance apart from word count, appears to be more than a proxy for text difficulty.

SC Was More Associated with Achievement at Lower Reading Levels

Our finding that average total words and average book level correlated at 0.95 led us to conclude that, like Thompson (1984), total words is a sufficient proxy for text difficulty. Thus, our finding that an additional 3% of variance was explained by the SC–word interaction raises some interesting questions. Recall that the negative coefficient for the interaction term indicated a more pronounced SC effect for student who read fewer words on average compared to those students who read more words. The mean exit OSELA of students in the lower 50th percentile group on average words was significantly less (429) than students in the upper 50th percentile on word count (493), with an effect size difference of 1.20, revealing that SC appears to be more associated with reading proficiency for students in lower achievement levels.

Even though prior research has found that more proficient readers tend to SC more than less proficient readers (Chinn et al., 1993; Hoffman & Clements, 1984; Hoffman et al., 1984), which was also the case in this study (average SC correlated positively with entry and exit OSELA, see Table 2), it appears from our findings that self-correcting may benefit less proficient students to a greater extent. The finding lends support to Vellutino and Scanlon's (2002) argument that self-correcting provides important feedback to early systems of word identification that are just being constructed by the beginning reader, and perhaps the feedback system is more critical to students who are struggling the most to read.

Implications

The current perceived utility of self-correcting varies in practice. In some contexts like Reading Recovery or guided reading, educators analyze SCs to determine the sources of information used and neglected and this analysis guides subsequent instruction (Clay, 2013, p. 72; Fountas & Pinnell, 2012, p. 273). Indeed, Fountas and Pinnell, refer to self-correcting as one of 12 processing systems for reading, defining monitoring and correcting as a "Check on understanding and work to self-correct errors" (p. 273). Moreover, SCs that occur when a student is reading aloud during instruction are regarded as timely opportunities for the teacher to "… capitalise on the learner's shift in attending" (Clay, 2015, p. 193). In other contexts however where oral reading is analyzed, notably informal reading inventories, SCs are merely coded or counted and not analyzed to inform or guide instruction. Few if any other assessment devices besides running records (including derivatives of Clay's running records) provide specific instructions on what to do with the SC tally. Our findings challenge the wisdom of ignoring SCs in oral reading.

Limitations

The students in this study were those having the most difficulty learning to read, thus, their progress may not be indicative of typically progressing students. As such, we contend that our findings likely generalize to other similar struggling readers, but not necessarily to a broader population of students. Relatedly, the generalizability of this study is limited in terms of instructional context. RR, the intervention in this study, is based on a theoretical perspective that emphasizes accurate reading, problem solving, and self-correcting (Clay, 2005). As such, RR incorporates instructional components that provide opportunities for these types of reading behaviors. It is unknown, however, if one would find similar results in other learning contexts, ones that emphasize learning to read words or word parts in isolation, not embedded in text.

Besides the fact that we studied the effects of SC for struggling readers only, we had data on a relatively small pool of 140 students that were not drawn at random from the population, but instead represented all students who were selected for the intervention in the fall of the school year. Our sampling method, therefore, may have delimited the degree to which the sample represented struggling emergent readers, and though not considered an extremely small sample, 140 students may not have captured the full spectrum of characteristics found in that population.

Because the study design was based on the assessment of naturally-occurring SC actions, and thus, was non-experimental, our results do not provide a warrant to make causal claims. The findings, therefore, suggest that SC may propel reading proficiency, but until more rigorous experimental work is conducted, the answer to whether SC propels or reflects achievement will remain unanswered.

Perhaps one reason experiments designed to more properly isolate the SC effect have not been conducted is that self-correcting is, by nature, a self-initiating action that cannot be stimulated on demand by the researcher. It may possible, however, to implement a treatment situation that increases the likelihood that young readers will SC, by presenting them certain texts that offer more SC opportunities. But creating a condition that enhances the probability of a behavior is not the only barrier to studying SC experimentally. It is very unlikely that an SC signal will occur in a short duration over a few readings. In order to detect an effect, the researcher would need frequent observations over several weeks to accrue an adequate sample of SCs per student on an ample number of different books to yield a reliable overall estimate of each student's SC level. In our study, we extracted two running records per week for each student over 15–20 weeks. Thus is the tradeoff between experimentally-controlled conditions with strong internal validity, and a more naturally occurring situation that yields rich data over a substantial duration.

Although we found that higher levels of self-correcting were associated with greater levels of literacy achievement, we did not examine the degree to which particular types of SCs were related to the outcome. Future research may further delve into whether certain SCs reflect or propel reading growth to better understand the cognitive processes that occur while self-correcting, which may depend to some degree on the errors that are being self-corrected.

By the mid-1990s studies on SC became virtually dormant, likely due to research that suggested Clay's (1966) earlier research was flawed. Indeed, Clay's research on SC was limited in terms of drawing causal conclusions, but subsequent research was equally as flawed. Inappropriate statistical modeling and proper control of extraneous variables led to misinformation about the role of SCs in the reading process, which until now have remained unknown. The next step is to better understand how opportunities can be created to encourage the types of SCs that enhance and accelerate greater levels of reading proficiency.

ORCID

Robert H. Kelly http://orcid.org/0000-0001-9234-9143

References

Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. *The Elementary School Journal*, *83*(5), 548–559. doi:10.1086/461333

Beaver, J., & Carter, M. (2009). *Developmental reading assessment* (2nd ed.). Cranbury, NJ. Pearson Education, Inc.

Bader, L. A., & Pearce, D. L. (2013). *Bader reading and language inventory*. Cranbury, NJ. Pearson Education, Inc.

Chall, J. (1967). *Learning to read: The great debate*. New York, NY: McGraw-Hill.

Chinn, C. A., Waggoner, M. A., Anderson, R. C., Schommer, M., & Wilkinson, I. A. (1993). Situated actions during reading lessons: A microanalysis of oral reading error episodes. *American Educational Research Journal*, *30*(2), 361–392. doi:10.2307/1163240

Clay, M. M. (1966). Emergent reading behavior (Unpublished doctoral dissertation). University of Auckland, New Zealand.

Clay, M. M. (1967). The reading behavior of five year old children: A research report. *New Zealand Journal of Educational Studies*, *2*, 11–31.

Clay, M. M. (1969). Reading errors and self-correction behaviour. *British Journal of Educational Psychology*, *39*(1), 47–56. doi:10.1111/j.2044-8279.1969.tb02040.x

Clay, M. M. (1972). *A diagnostic survey and concepts about print test.* Auckland: Heinemann.

Clay, M. M. (1979). *Reading: The patterning of complex behavior.* Auckland: Heinemann.

Clay, M. M. (2000). *Running records for classroom teachers.* Portsmouth, NH. Heinemann.

Clay, M. M. (2001). *Change over time in children's literacy development.* Auckland: Heinemann.

Clay, M. M. (2005). *Literacy lessons designed for individuals part one: Why? When? and How?* Portsmouth, NH: Heinemann.

Clay, M. M. (2013). *An observation survey of early literacy achievement* (3rd ed.). Portsmouth, NH: Heinemann.

Clay, M.M. (2015). *Change over time in children's literacy development.* Portsmouth, NH: Heinemann.

Cohen, A. S. (1974). Oral reading errors of first grade children taught by a code emphasis approach. *Reading Research Quarterly*, *10*(4), 616–650. doi:10.2307/747504

Cowley, J. (2015). *What is a book?* Singapore: Hameray Publishing Group.

Cunningham, J. W., Spadorcia, S. A., Erickson, K. A., Koppenhaver, D. A., Sturm, J. M., & Yoder, D. E. (2005). Investigating the instructional supportiveness of leveled texts. *Reading Research Quarterly*, *40*(4), 410–427. doi:10.1598/RRQ.40.4.2

D'Agostino, J.V. (2012). *U.S. Norms and Correlations for An Observation Survey of Early Literacy Achievement (Report No. 2012–04).* Columbus OH: International Data Evaluation Center.

D'Agostino, J.V., & Rodgers, E. M., & Mauck, S.A. (2018). Addressing inadequacies of An Observation Survey of Early Literacy Achievement. *Reading Research Quarterly*, *53*, 51–69. doi:10.1002/rrq.181

Denton, C. A., Ciancio, D. J., & Fletcher, J. M. (2006). Validity, reliability, and utility of the observation survey of early literacy achievement. *Reading Research Quarterly*, *41*(1), 8–34. doi:10.1598/RRQ.41.1.1

Doyle, M. A. (2013). Marie M. Clay's theoretical perspective: A literacy processing theory. In D. E. Alvermann, N. J. Unrau, and R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 636–655). Newark, DE. International Reading Association.

Fairbanks. (1937). The relation between eye-movements and voice in the oral reading of good and poor silent readers. *Psychological Monographs*, *48*(3), 78–107. doi:10.1037/h0093394

Fountas, I. C., & Pinnell, G. S. (2012). Guided reading: The romance and the reality. *The Reading Teacher*, *66*(4), 268–284. doi:10.1002/TRTR.01123

Fawson, P. C., Ludlow, B. C., Reutzel, D. R., Sudweeks, R., & Smith, J. A. (2006). Examining the reliability of running records: Attaining generalizable results. *The Journal of Educational Research*, *100*(2), 113–126. doi:10.3200/JOER.100.2.113-126

Goodman, Y. M., & Goodman, K. S. (2004). To err is human: Learning about language processes by analyzing miscues. In R. B. Ruddell and. N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 620–639). Newark, DE. International Reading Association.

Harlin, R. (1981). Primary children's print awareness and self-correction behavior as predictors of reading achievement (Unpublished doctoral dissertation). University of Florida, Gainesville.

Harmey, S., & Kabuto, B. (2018). Metatheoretical differences between running records and miscue analysis: Implications for analysis of oral reading behaviors. *Research in the Teaching of English*, *53*(1), 11.

Hoffman, J. V., O'Neal, S. F., Kastler, L. A., Clements, R. O., Segel, K. W., & Nash, M. F. (1984). Guided oral reading and miscue focused verbal feedback in second-grade classrooms. *Reading Research Quarterly*, *19*(3), 367–384. doi:10.2307/747827

Hoffman, J. V., & Clements, R. O. (1984). Reading miscues and teacher verbal feedback. *The Elementary School Journal*, *84*(4), 423–439. doi:10.1086/461374

Johnston, P., & Afflerbach, P. (2015). Formative assessment used to its greatest advantage (Web log post). Retrieved from https://www.literacyworldwide.org/blog/literacy-daily/2015/09/17/formative-assessment-used-to-its-greatest-advantage.

Kragler, S., Martin, L., & Schreier, V. (2015). Investigating young children's use of reading strategies: A longitudinal study. *Reading Psychology*, *36*(5), 445–472. doi:10.1080/02702711.2014.884031

Kronmal, R. A. (1993). Spurious correlation and the fallacy of the ratio standard revisited. *Journal of the Royal Statistical Society*, *156*(3), 379–392. doi:10.2307/2983064

Nelder, J. A. (1977). A reformulation of linear models. *Journal of the Royal Statistical Society*, *140*(1), 48–77. doi:10.2307/2344517

Roe, B., & Burns, P. C. (2011). *Informal reading inventory: Preprimer to twelfth grade*. Belmont, CA: Wadsworth.

Santa, C. M., Williams, C. K., Ogle, D., Farstrup, A. E., Au, K. H., Baker, B. M., … Paratore, J. R. (2000). Excellent reading teachers: A position statement of the International Reading Association. *Journal of Adolescent & Adult Literacy*, *44*(2), 193–199.

Share, D. L. (1990). Self-correction rates in oral reading: Indices of efficient reading or artefact of text difficulty? *Educational Psychology*, *10*(2), 181–186. doi:10.1080/0144341900100205

Sharp, A. C., Sinatra, G. M., & Reynolds, R. E. (2008). The development of children's orthographic knowledge: A microgenetic perspective. *Reading Research Quarterly*, *43*(3), 206–226. doi:10.1598/RRQ.43.3.1

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York, NY: Oxford University Press.

Thompson, G. B. (1981). Individual differences attributed to self-correction in reading. *British Journal of Educational Psychology*, *51*(2), 228–229. doi:10.1111/j.2044-8279.1981.tb02476.x

Thompson, G. B. (1984). Self-corrections and the reading process: An evaluation of evidence. *Journal of Research in Reading*, *7*(1), 53–61. doi:10.1111/j.1467-9817.1984.tb00255.x

Vellutino, F. R., & Scanlon, D. M. (2002). The interactive strategies approach to reading intervention. *Contemporary Educational Psychology*, *27*(4), 573–635. doi:10.1016/S0361-476X(02)00002-4

Wood, D. (1998). *How children think and learn: The social contexts of cognitive development* (2nd ed.). Oxford. Blackwell.