Exploring Factors that Influence Quality Literature Circles

Chase Young & Kathleen A. J. Mohr

To cite this article: Chase Young & Kathleen A. J. Mohr (2017): Exploring Factors that Influence Quality Literature Circles, Literacy Research and Instruction, DOI: 10.1080/19388071.2017.1366606

To link to this article: http://dx.doi.org/10.1080/19388071.2017.1366606

Published online: 05 Sep 2017.
Exploring Factors that Influence Quality Literature Circles

Chase Young a and Kathleen A. J. Mohr b

aLanguage, Literacy and Special Populations, Sam Houston State University, Huntsville, Texas; bSchool of Teacher Education and Leadership, Utah State University, North Logan, Utah

ABSTRACT

Research indicates that literature circles are an authentic means for literacy development that students typically enjoy. To better understand the potential value and to add to the research base regarding literature circles, this study, involving 17 fourth graders, explores factors that may influence the quality of literature discussions, including reading ability, gender, personality types, and group size. The regression model generated a significant p-value of .019 and an adjusted R-squared of .66. Emotional stability was significant (p < .05), and group was significant at the p < .01 level indicating that groups comprised of three students engaged in higher-quality discussions. These findings are presented and limitations and implications of the study are discussed.

KEYWORDS

Children’s literature; instructional strategies and materials; literature-based instruction

Research in classroom-based literature circles has generally focused on implementation and design (Almasi, O’Flahavan, & Arya, 2001; Bond, 2001; Brabham & Villaume, 2000; Burns, 1998; Clark, 2009; Daniels, 2002; Peterson, 2016). Earlier research (Daniels, 1994) suggested roles (e.g., Director, Word Wizard) to help students initiate conversations, but more current recommendations promote open discussions with less restrictive preparatory methods and targeted mini-lessons (Evans, 2002). Research has also looked at individual student perceptions of literature circles, concluding that students enjoy literature circles (Daniels, 2002; Flowerday, Schraw, & Stevens, 2004; Peralta-Nash & Dutch, 2000). Then, McElvain (2010) studied the associated comprehension and indicated that students were in fact developing their understanding of the text through dialogue, a finding aligned with social constructivist theory (Vygotsky, 1978). In addition to understanding the teacher’s role in literature circles (e.g., mini-lesson, preparatory methods), research has explored individual student functions within small groups. Li et al. (2007) and Gnadinger (2008) found that students were capable of scaffolding and leading their peers. To add to the research base regarding literature circles, this study explored other factors that may influence the quality of literature discussions, including reading ability, gender, and personality types. In addition, a secondary regression model also considered the size of the group, a consideration not previously empirically studied in literature circles.
**Theoretical framework**

Scaffolding, which is a construct attributed to Bruner (Wood, Bruner, & Ross, 1976), relates to Vygotsky’s (1978) notion of the zone of proximal development (ZPD), and is a metaphor for providing needed support to students in an effort to maximize their success with challenging tasks and deeper cognition. Scaffolded instruction begins with the teacher’s curricular decisions in response to students’ needs. For example, if students are proficient in teacher-led classroom literature discussions, a teacher could decide that literature circles may be an appropriate next step.

However, implementing high-quality literature circles in elementary classrooms is a complex task, and in many cases requires a period of trial and error (Clarke & Holwadel, 2007). This period of facilitation might also be explained as making a difficult task achievable, an important process in scaffolding (Wood et al., 1976). The restructuring of literature circles through mini-lessons and student preparatory methods can help diminish the difficult transition to peer-led discussions (Burns, 1998; Clarke & Holwadel, 2007; Maloch, 2002).

The transition to internalization and independence begins with a knowledgeable other (Vygotsky, 1978) supporting a novice on a difficult task. The knowledgeable other is not necessarily smarter than the novice, but simply knows more about the topic. In addition to being more experienced, the knowledgeable other is often skilled at guiding the novice to a better understanding of a concept. The knowledgeable other does not necessarily dominate the instructional episode; rather he/she facilitates the learning by calibrating the difficulty and guiding a learner through a particular task. The teacher then slowly removes assistance when appropriate, which ultimately leads to student independence.

In the novice state, tasks are directed by inner speech (Berk, 1985), meaning that learners appropriate the language of the task from others. Tudge and Hogan (1997) suggest that students need a knowledgeable other to serve as an external mediator to help students internalize this inner speech. In other words, novices need access to the internal dialogue of an expert to talk through and master a given task. Inner speech during academic conversations is complicated, as discussants are constantly synthesizing what others are saying and determining appropriate responses. This inner speech may benefit students in literature circles where students are required to engage in a literature-based discourse to explore the meaning of a text. This particular discourse may not be automatic. The think-alouds demonstrated by the teacher should eventually become the students’ inner speech. Before the speech is internalized and accessed more automatically, students have to make conscious decisions while engaging in literate discussions. In theory, teachers who help students develop inner speech enable students to meet the demands of literate discussions (Smolucha & Smolucha, 1989).

From a practical standpoint, teachers verbalize their thinking so that students can learn the internal processes necessary for participating in analysis of texts and literature circle discussions. It is hypothesized that, once those processes are internalized by observing the teacher and from student practice, the tasks become automatic and the need for inner speech dissipates until the complexity of the task increases (Tudge & Hogan, 1997). In the case of literature circles, the structure of the activity rarely changes, but the content discussed is always shifting because of the texts. Students who are well rehearsed in the process of literate discussions are able to focus more on the content of the text, so
discussions are more characterized by functional and exploratory rather than reflectively talk, a type of discourse that focuses on group management (Maloch, 2002). Theoretically, then, the inner speech and self-regulation required in literature circle discussions, as well as other collaborative tasks, are internalized (Smolucha & Smolucha, 1989). Although this internalization may not become completely automatic, the tasks may require less attention.

Because of their significant role in the preliminary stages, it may be difficult for teachers to step back after the completion of their role in literature circles. Short, Kaufman, Kaser, Kahn, and Crawford (1999) juxtaposed the inclusion and exclusion of teachers in literature circle discussions. Teachers were included in four different groups with four different roles: teacher as facilitator, participant, mediator, or active listener. In another classroom, 9- to 11-year-olds engaged in peer-led discussions without the teacher. While each teacher role had a differing effect on the groups’ discussions, analysis of the discourse data suggests that there were only minor qualitative differences between teacher-led and peer-led groups. For example, student groups spent more time on each topic and talked about a smaller range of topics. Teachers encouraged discussion of a wider range of topics and deepened the discussion of relevant topics. Overall, both types of groups stayed on topic, and discussed relevant issues regarding the text. Although teacher support is sometimes needed, the absence of the teacher does not necessarily indicate unproductive discussions.

Short et al.’s study (1999) indicates that students are capable of engaging in literate discussions that are focused and relevant to the text. These findings are corroborated by another study involving 22 sixth grade students (Wiencek & O’Flahavan, 1994). The teacher minimized her control of the discussions and observed students engaged in literature circle discussions. The groups actively constructed meaning of text and were generally excited about the discussions. The differing opinions and interpretation of meaning of the text created a context for social interaction, a key component of learning from social learning perspectives (Vygotsky, 1978). However, it is possible that other factors beyond scaffolding, preparation, or group idiosyncrasies may also influence the quality of literature circles. For example, the gender configurations of group might influence the quality and direction of students’ discussions.

**Gender**

Tudge and Hogan (1997) theorized that gender makeup of a group changes the dynamics of collaborative task engagement. Although their claims were supported by theory, the authors did not provide empirical evidence to substantiate their argument. A study by Webb and Farivar (1994) investigated gender, comparing differently structured mix-gender groups completing a shared task. The researchers studied seventh and eighth graders in equal-gender (two boys and two girls), majority-boy (three boys and one girl), and majority-girl groups. The analysis revealed a small effect, indicating that gender can affect small-group discourse and task completion. However, when gender’s saliency was minimized by equally distributing males and females, the effect decreased. Interestingly, in both majority conditions, the girls were at a disadvantage. In majority-girl groups, the girls were observed seeking leadership from the boys. In the majority-boy groups, the girl was simply ignored altogether.
Related research on gender’s effect in small groups is often unclear because of less definitive findings (Lockheed, 1977, 1983). It seems that the effect gender has on small-group interaction might change as children enter adolescence. In Lockheed’s (1977) study of mixed-gender small groups, the 15- and 16-year-old boys were perceived as more adept leaders than the girls. In a later study (Lockheed, 1983) with primary school students, the boys were also perceived as leaders, but the analysis revealed that the perceptions did not transfer into the actual collaborative task as the boys and girls did not treat each other differently. In other words, the two studies suggest that males are perceived as leaders, but only older students’ perceptions match the observations by Lockheed during small group tasks. The researchers inferred that the young students were aware of gender-status differences, but the perceptions were not evident in actions until later in life. For example, younger students knew that boys and girls differ, but they did not seemingly act out the belief during discussions. However, the older students were observed treating males and females differently in groups, and thus acting on their beliefs that males were leaders. Both of these summarized studies were completed more than 30 years ago, so revisiting whether gender continues as a factor in the quality, productiveness, and equity of classroom interactions may be warranted.

**Personality**

Literature discussions are social learning situations and, thus, interpersonal in nature, but the nature of the interactions are dependent on the individuals involved. Intrapersonal factors, such as shyness or individual leadership qualities, influence student engagement (Li et al., 2007) and should be considered during social learning situations. The related literature rarely discusses the intrapersonal factors that likely impact the students’ social interaction, particularly during literature circles. Understanding more about personality factors that influence individual participation in literature circle discussions could provide information to educators on how to organize literature circles for more optimal interaction. Thus, understanding the influences of human factors, including the intrapersonal, personality aspects of the participants merits further investigation (Chan, 2010).

The five factor model that includes the “Big Five” personality traits is a common way to describe personalities (Anusic, Schimmack, Pinkus, & Lockwood, 2009) and has been used extensively with different age groups in psychological research. The Big Five traits are extroversion, agreeableness, conscientiousness, emotional stability, and openness. Extroverted students are described as enthusiastic and energetic. The model claims that agreeableness is demonstrated when students are compassionate and cooperative. Conscientious students are organized and efficient. Emotionally stable students are secure and confident. Finally, openness is a measure of a student’s curiosity or capacity to enjoy new experiences (Anusic et al., 2009; Barbaranelli, Caprara, Rabasca, & Pastorelli, 2003). All Big Five traits would seem to relate to the quality of interaction sought in literature circles. The Big Five are measured with a personality inventory, which is considered a reliable measure of personality yielding high coefficients (.90) across studies (Kaiser, Hunka, & Bianchini, 1971).

Personality traits might play a significant role in peer interaction during literature circles in addition to prior instruction and preparatory methods. One study (Young, 2014) investigated the Big Five personality traits, along with reading ability, as predictors of
quality verbal engagement in literature circle discussions. Six third grade literature discussions were video recorded and analyzed for quality verbal contributions. A quality of verbal engagement score was quantified by coding individual students’ discourse according to the three-story intellect model of posing increasingly higher-order questions of texts (Costa & Kallick, 2000). The verbal engagement score served as the dependent variable in a multiple linear regression. The independent variables were the Big Five personality factors along with reading ability. The Big Five were assessed using the Ten-Item Personality Inventory (TIPI; Gosling, Rentfrow, & Swann, 2003). Teacher ratings of students’ personality attributes agreed 100% with the students’ self-reported ratings. The regression indicated that higher reading ability, higher extroversion, and a lack of conscientiousness in student personalities positively predicted higher quality verbal engagement in literature circle discussions (Young, 2014). Thus, while intuitive, especially to seasoned classroom teachers, research indicates that personality and other student-specific characteristics likely contribute to the success of classroom literature circles.

Significance of the current study

The instructional design of literature circles has been researched extensively (Day & Ainley, 2008), but, beyond consideration of gender, there is a gap in the research when considering group configuration. Researchers (Clarke & Holwadel, 2007) report that changing pre-teaching methods, instilling positive conversational discourse, and providing a less structured preparation process to create a context for discussion enhanced literature circles; however, little research exists that considers the intentional placement of students in the groups. Perhaps text choice and reading ability should not be the only criteria that teachers use to determine the configuration of literature circles. The current study explores other factors that may influence the quality of literature circle discussion, and attempts to answer the following question:

To what extent do personality traits, reading ability, and gender predict the quality of verbal engagement (QVE) in fourth grade literature circles?

Method

Context and participants

This study was conducted in a school district in the U.S. southwest. The elementary school served 18% economically disadvantaged students. Student demographics in the school were 58% White (non-Hispanic), 19% Hispanic, 14% Black (non-Hispanic), 8% Asian/Pacific Islander, and <1% Native American.

The five literature discussion groups targeted for this analysis ranged from three to six students, and were selected as a convenience sample because the majority of the group members (N = 17) consented to participation allowing for examination of intact groups. The primary researcher filmed a total of 17 students in five groups in their classroom during the school day on two separate occasions. Students were often filmed during their Readers Theater performances throughout the year, so they were relatively comfortable. The discussions were concluded when the students felt they had sufficiently discussed the text; therefore the durations of the conversations varied (see Table 1 for more details).
The teacher was a departmentalized fourth grade language arts teacher who taught two different classes during the day. Her implementation of literature circles did not utilize roles and she encouraged open discussion. In preparation for literature circle discussions, students took notes while reading and wrote down questions that they wanted to ask the group. The teacher grouped her students by reading level and choice. In other words, she allowed students to choose their books, but she limited the number of choices based on individual students’ reading ability.

Although some of the students likely participated in various literature discussions in previous grade levels, the students had been participating in literature circles for 7 months in fourth grade before the study began. Also, because it was later in the school year, the students were quite familiar with one another and were likely more comfortable speaking with each other in groups. The students read their literature circle books every day and engaged in discussions on Mondays, Wednesdays, and Fridays. The teacher set no time limit on the literature circle discussion, and students were encouraged to discuss until there were no more discussion points or questions. The teacher did roam the classroom randomly listening to bits of the groups’ discussions, but rarely intervened unless the students had a question for her.

**Instrumentation**

**Ten item personality inventory**

The Big Five personality traits were measured with the TIPI (Gosling et al., 2003). The inventory was administered one-on-one with each participant at the beginning of the study. Thus, the primary researcher was available to answer any questions throughout the assessment. The profiles indicated levels of extroversion, agreeableness, conscientiousness, emotional stability, and openness. The convergence mean correlation across other measures of the Big Five was adequate (.65) as was the test and retest reliability (.72). Overall, the TIPI is considered a reliable assessment of the Big Five personality traits.

**Measure of academic progress**

Data from the spring administration of the Measure of Academic Progress (MAP; Northwest Evaluation Association, 2011) were used to determine students’ reading achievement. The Reading MAP test is a computer assessment that determines student reading achievement and progress based on grade-level norms. The MAP provides a
percentile score based on the normal performance of students at the same grade level. Thus, the average student’s score should be around the 50th percentile.

The MAP is an adaptive test based on item-response theory where the test reacts to student responses, thus becoming more difficult or easier as students answer items. The assessment produces a variety of reading measures including the percentile score based on the normal performance of students at the same grade level. Thus, the average student’s score should be around the 50th percentile. The current study employs the reported percentiles for the participating students. The reported MAP test and retest reliability range is .76 to .93. Ideally, reliability should not fall below .80, but the researchers explain that the reported range is due to the test question sets being different at each administration. The reported average Pearson correlation coefficient is .85, with a range of .69 to .80, statistically demonstrating the test’s acceptable reliability and validity (Northwest Evaluation Association, 2011). Although the assessment was also given in the Fall and Winter, the Spring administration was given 2 weeks before the study, and thus provided the most current results.

Analysis

The primary researcher collected a total of 136.7 minutes of footage from 10 sessions (two tapings of each group) in the spring semester. The groups averaged 27.34 minutes of discussion. The discussions ranged from 5 minutes and 14 seconds to 28 minutes and 27 seconds. The groups ranged from three to six students. (Not all students in each group studied were included in the analysis because of lack of parental consent.) Group information is summarized in Table 1.

In order to render a quality of verbal engagement score, the discussions were coded based on the quality of student contributions (see Table 2). The quality score assignment was based on the three-story intellect (Costa & Kallick, 2000). The three-story intellect was created to teach educators how to help their students think (Fogarty & McTighe, 1993). When the researcher observed students contributing at the varying levels, it was considered evidence that the students were thinking at differing levels. Although it had not been used in this manner previously, it made sense to analyze oral discussions using the three-story framework.

Statements and questions were awarded 1, 2, or 3 points (Figure 1) based on students’ contributions. The first level considers input characterized by recall of text information. Some examples of level-one contributions include: recall, describe, name, or identify. The next level, processing, requires higher-level thought from the reader. The responder is required to summarize, compare, sequence, infer, or analyze. The third level requires output, which include evaluating, speculating, predicting, generalizing, or judging. In this

<table>
<thead>
<tr>
<th>Score</th>
<th>Level</th>
<th>Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input</td>
<td>Name, Recall, Restate, Reread, Locate, Describe, State, Inform, Define, Identify, List</td>
</tr>
<tr>
<td>2</td>
<td>Process</td>
<td>Compare, Contrast, Classify, Distinguish, Explain (Why), Infer, Sequence, Analyze, Synthesize, Make Analogies, Reason</td>
</tr>
<tr>
<td>3</td>
<td>Output</td>
<td>Evaluate, Generalize, Imagine, Judge, Predict, Speculate, If/Then, Apply a Principle, Hypothesize, Forecast, Idealize</td>
</tr>
</tbody>
</table>
case, the researcher coded all utterances made during the recorded literature circle discussions, such as statements that evidence higher-level thinking. In addition, students’ questions were coded according to the elicited cognitive processes. For example, if a student asked, “What do you think the character will do next?,” then a score of 3 was assigned because the question elicited a prediction. Although the student was not making a prediction himself, he was using higher-level questioning to extend the discussion (Figure 1). Predicting is arguably a form of inferring; however, when students infer to predict, the students enter the hypothetical realm associated with the third level intellect. Such possible discrepancies were scored in favor of the student. Thus, a prediction (level 3) was also considered an inference (level 2); however, a score of 3 was assigned because the student hypothesized based on the inference. The scores were totaled into a QVE score.

The primary researcher and a graduate student independently coded 10% of the data to establish interrater reliability. The coders were initially in 79% agreement, and subsequently discussed all discrepancies until agreeing on 100% of the items. The most common discrepancy occurred when the graduate student coded statements, such as “yeah” as a level one, when the researcher had coded the same statement a zero because the contribution was merely a confirmation. Other common discrepancies were between level-two inferences that could also be described as level-three predictions (as noted above). According to the three-story intellect (Costa & Kallick, 2000), however, predictions were different because not only did the student process the text, but also the student’s verbalized output was hypothetical. For example, if the story says it is snowing, the reader can infer that it is winter. However, if a story says that a suspect has no alibi, the reader has to think beyond the text and predict a future event. In addition, as stated in the method, codes were scored in favor of the student.

The data, including reading percentile scores, gender, and personality inventories, were analyzed through multiple linear regression in R (R Development Core Team, 2010). The QVE score served as the dependent variable, and the independent variables were (a) MAP percentile score, (b) extroversion, (c) agreeableness, (d) conscientiousness, (e) emotional stability, (f) openness, and (g) gender.

Results

This research was guided by the question: to what extent do personality traits, reading ability, and gender predict the QVE in fourth grade literature circles? The statistical method of linear multiple regression was used to answer this question (Crawley, 2007).

Seventeen students were included in the analysis. Each student had a QVE score, reading percentile score, gender code, and ratings for each of the Big Five personality traits.

Matt: What is A.G.? [Infer = 2] This solicits an inference.
Julie: Well, it didn’t really say that—it just said it was on the suitcase. [Recall = 1]
Jeremy: Well, first, Stanley just thought it was probably a word. [Recall = 1]
Julie: Maybe it’s like initials. [Infer = 2]
Jeremy: He thought it was Adgy. [Recall = 1]
Matt: It’s probably initials. [Infer = 2]

Figure 1. Coded example from Holes (Sachar, 1998) transcription.
traits (Gosling et al., 2003). See Table 3 below for descriptive statistics. The QVE standard deviation is relatively large, which raised questions about whether the data represented a normal distribution. Therefore, subsequent analyses tested for outliers. Although the QVE scores ranged from a minimum of 4 to a maximum of 200, no outliers were detected by the Bonferonni test with a significance of $p < .05$. This large variation in QVE scores is an interesting finding in itself and seems to reflect the variability among students’ participation and contributions to literature discussions.

The linear model assumptions were also tested to ensure that the model was not misleading, biased, or inefficient. Global tests of model assumptions (global statistic, skewness, kurtosis, heteroscedasticity, and link function) were all met. The variance inflation factor was examined to test for multicollinearity and returned false. Models that have multicollinearity, with two or more highly correlated predictors, are not necessarily problematic when examining the model as a whole, but the highly correlated predictor variables cannot be analyzed individually because of the multicollinearity. However, in this model, the variables were not highly correlated, so the analysis of individual predictors was reliable.

The QVE score had a substantial standard deviation, thus indicating that students’ contributions varied greatly in quality. However, the mean reading score, on average, suggests students were reading above the expectation for typical fourth graders. Interestingly, the reported mean extroversion score indicates that the participating students were more extroverted than not; therefore, many of the students were likely comfortable sharing in groups. Student mean scores in agreeableness, conscientiousness, and openness were also in the medium to high range. The descriptive statistics reveal that emotional stability had the largest standard deviation and lowest mean of the personality factors. Therefore, students were more similar in the other personality factors than emotional stability, reflecting security and confidence. Because the subjects in this smaller sample were above average in many of the areas assessed, the researchers will not generalize these results, but make practical recommendations when applicable.

The results of the quantitative analysis suggest that the regression model was insignificant based on the overall $p$-value of .23 (Table 4). However, there was one significant factor, emotional stability ($p < .05$). Essentially, students who were more emotionally

### Table 3. Descriptive statistics.

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Median</th>
<th>$M$</th>
<th>$SD$</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>QVE</td>
<td>77.00</td>
<td>80.59</td>
<td>63.94</td>
<td>Max = 200, Min = 4</td>
</tr>
<tr>
<td>Reading ability</td>
<td>88.00</td>
<td>80.18</td>
<td>18.12</td>
<td>Max = 97, Min = 40</td>
</tr>
<tr>
<td>Extroversion</td>
<td>5.50</td>
<td>5.21</td>
<td>1.34</td>
<td>Max = 7, Min = 2.5</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>5.50</td>
<td>5.38</td>
<td>1.05</td>
<td>Max = 7, Min = 4</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>6.00</td>
<td>5.71</td>
<td>0.87</td>
<td>Max = 7, Min = 4</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>5.00</td>
<td>4.82</td>
<td>1.78</td>
<td>Max = 7, Min = 2</td>
</tr>
<tr>
<td>Openness</td>
<td>6.00</td>
<td>5.53</td>
<td>1.37</td>
<td>Max = 7, Min = 1.5</td>
</tr>
</tbody>
</table>
stable, described as secure and confident, provided more quality contributions within this context.

The regression model was circumspect because of the low degrees of freedom due to the limited number of participants in the study ($N = 17$). Even if the critical $p$-value were less than .05, the results cannot be generalized to larger populations because of the low power (.11).

After discussion of the results and the characteristics of the literature circles, it was determined that a second analysis should be conducted to investigate whether size of the group would help explain the variance in quality of verbal engagement (Table 5).

The secondary regression model including group size revealed a significant $p$-value of .019. The multiple $R$-squared was .83 with an adjusted $R$-squared of .66. The model results showed that 66% of the variance was captured by the 8 factors, 2 of which were significant. Emotional stability was still significant ($p < .05$). The additional factor, size, was significant at the $p < .01$ level indicating that the smaller the group size, the higher the quality of engagement.

The model shows that students in groups of three were correlated with higher verbal engagement in literature circle discussions. Also, students who were more emotionally stable (i.e., secure and confident) tended to engage in higher-quality discussions despite group size.

### Discussion

This study aimed to determine whether reading ability, gender, and personality traits influenced the quality of verbal engagement in fourth grade literature circle discussions.

#### Table 4. Regression model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE(B)</th>
<th>t</th>
<th>Sig. ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>−225.51</td>
<td>176.18</td>
<td>−1.28</td>
<td>0.23</td>
</tr>
<tr>
<td>Reading ability</td>
<td>1.24</td>
<td>0.28</td>
<td>0.78</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>65.26</td>
<td>40.05</td>
<td>1.63</td>
<td>0.14</td>
</tr>
<tr>
<td>Extroversion</td>
<td>−1.53</td>
<td>14.11</td>
<td>−0.11</td>
<td>0.92</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>13.64</td>
<td>13.01</td>
<td>0.91</td>
<td>0.39</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>−13.17</td>
<td>17.40</td>
<td>−0.76</td>
<td>0.47</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>24.51</td>
<td>10.32</td>
<td>2.37</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Openness</td>
<td>13.92</td>
<td>13.67</td>
<td>1.02</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Note. Multiple $R^2$: 0.56, Adjusted $R^2$: 0.23

#### Table 5. Secondary regression model.

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>SE(B)</th>
<th>$t$</th>
<th>Sig. ($p$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>71.09</td>
<td>143.79</td>
<td>0.49</td>
<td>0.63</td>
</tr>
<tr>
<td>Reading ability</td>
<td>−0.06</td>
<td>0.83</td>
<td>−0.07</td>
<td>0.95</td>
</tr>
<tr>
<td>Group size</td>
<td>−30.61</td>
<td>8.66</td>
<td>−3.53</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Gender</td>
<td>43.72</td>
<td>27.23</td>
<td>1.61</td>
<td>0.15</td>
</tr>
<tr>
<td>Extroversion</td>
<td>4.47</td>
<td>9.50</td>
<td>0.47</td>
<td>0.65</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>3.14</td>
<td>10.38</td>
<td>0.30</td>
<td>0.77</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>−9.09</td>
<td>11.59</td>
<td>−0.79</td>
<td>0.46</td>
</tr>
<tr>
<td>Emotional stability</td>
<td>17.08</td>
<td>7.16</td>
<td>2.39</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Openness</td>
<td>2.34</td>
<td>9.64</td>
<td>0.24</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Note. $R^2$: 0.83, Adjusted $R^2$: 0.66

All the previous factors were included with group size as an additional factor. The global test of assumptions was accepted, and no outliers were detected.
The limited number of subjects in this study makes it difficult to generalize the results, but they warrant a discussion about the findings because the analyses revealed interesting phenomena that are worthy of further research.

First, and not unexpectedly, the researchers noted the variation in student contributions. The high standard deviation of QVE scores indicate that not all students participated in peer-led discussions with the same quality. Literature circle research has not previously investigated individual factors that influence the discussion, and this study set out to explore that gap in the research. A previous study by the primary researcher (Young, 2014) revealed that third grade students with increased extroversion and lack of conscientiousness, combined with a higher reading ability, predicted higher quality verbal engagement in literature circles discussions. Whereas, the current study of fourth graders failed to corroborate those findings, results indicate that one personality trait, emotional stability, was significant in predicting the quality of engagement. Such findings suggest there might be some correlation between personality and discussion engagement, but the personality influences may differ with group dynamics or grade level. Because the previous study included third grade students, it is possible that reading ability wanes as a predictor as grade-level increases because students may be more proficient readers. Further research in various grade levels is needed to determine if predictors might change by age group.

In the first regression model used in this study, although only emotional stability was the statistically significant prediction, there was a positive trend toward gender. Its positive correlation suggests that the fourth grade girls were more likely to provide quality contributions in literature circle discussions, but further research could help clarify this tendency.

Another interesting factor in this study’s first model was extroversion. The previous study (Young, 2014) revealed that increased extroversion predicted higher quality of verbal engagement ($p < .05$). Yet in the current study, as a student’s extroversion decreased, the quality of verbal engagement increased, but not significantly. In fact, the extroversion factor was insignificant when predicting variance ($p = .91$) in this case. Because of the insignificance, the researchers decided to add an additional independent variable, group size, in a secondary analysis. Results from the second regression model suggests, within the context of the selected classroom, that smaller groups had significantly higher QVE means. Previous research on group size and productivity suggests that smaller groups are typically more productive than larger groups (Bass & Norton, 1951; Wheelan, 2009). Although the related research was conducted with adults, it was corroborated by this research in a fourth grade classroom and warrants further investigation of classroom discussions.

Research on group size indicates that larger groups often form sub-groups, each with varying productivity (Hare, 1981). According to Hare, groups of two and three are more unified in task completion. This phenomenon was present in this fourth grade classroom. According to the quantitative data in this study, groups of three were correlated with a higher QVE. It could be argued that smaller groups placed more responsibility on each member to participate. It may have been simply more difficult to avoid participating in the discussion with fewer group members. Conversely, larger groups may have made it easier for some students to sit back quietly while others worked toward the achievement of the task. The smaller groups might have allowed for more risk-taking and greater opportunity to hold each other accountable. Research articles (Almasi, 1995), books (Daniels, 2002), and literature circle websites (Noe, 2011) typically suggest four to six students; however, no refereed research was found on determining optimal
literature circle group size, and it is possible that the significance of group size may vary by grade level. Although further research is needed, teachers that observe unproductive peer-led discussions may try limiting the groups to three students.

In the classroom studied, reading ability was the most insignificant factor in the second regression ($p = .95$). This suggests that within the context of the selected classroom, reading ability had very little influence on the quality of verbal engagement although, it may have influenced comprehension. The quantitative analysis revealed that higher emotional stability predicted higher QVE, but a comprehension measure may help to understand the variability in QVE. The researchers did not employ an additional comprehension measure because the three-story intellect (Costa & Kallick, 2000) was created to teach educators how to help their students think (Fogarty & McTighe, 1993). The coding scheme used to categorize student contributions was considered evidence that the students were thinking at differing levels.

For example, a student discussing *Joey Pigza Swallowed the Key* (Gantos, 2000) predicted that Joey’s dad was going to take away Joey’s medication and let him run wild. The student made this prediction based on her analysis of Joey’s dad’s character, and her knowledge of how Joey behaves without his attention deficit hyperactivity disorder medication. It is safe to assume that her predication was based on her comprehension of the text. Another student mentioned that he thought the relationship between Joey and the father would be interesting because they are both crazy. Again, the inference was based directly on an accurate understanding of the text. Still, further research is needed to confirm the validity of using scores for verbal engagement as a proxy for levels of thinking and comprehension.

The coding scheme in the analysis also added to the extant research by Paradis, Chatton, Boswell, Smith, and Yovich (1991) who sought to develop a comprehension-coding matrix that could measure comprehension during discussions. In the current study, it was relatively easy to code the quality of contributions with the three-story intellect descriptors (Costa & Kallick, 2000), and it could be used by teachers to assess quality of students’ participation. Furthermore, teachers could use the descriptors during mini-lessons to teach students to contribute at higher levels. Teachers could plan lessons designed to practice, for example, applying principles. The lesson would help students identify and infer principles portrayed in the text, and how to facilitate discussion by verbally describing the principles found in the text, explaining how the reader identified or used inference to understand the principles, and how one might apply them to their own lives. The discussant, then, might be coached on how to invite other group members to engage in a similar process and share possible applications.

At times, the interraters (primary researcher and graduate student) in this study found Levels 2 and 3 somewhat troublesome to distinguish, so Levels 2 and 3 could be collapsed to eliminate confusion, if necessary. For example, surface-level contributions (Level 1) are text-based, while Levels 2 and 3 contributions require higher-order processing by the reader. While the three-story intellect may not have measured comprehension in a more typical way, it was helpful when discerning facilitative behaviors. In other words, are there behaviors that extend the discussion in worthwhile directions that teachers could model and promote among students?
Clearly, this research brings up several questions for further research. Do significant predictors vary among grade levels? Do smaller groups provide more opportunity for high-level discussions? In addition, the study questions whether choice should be a major factor when considering group configuration. If literature circles are lacking quality, teachers could try to place students in smaller groups with the facilitative tools necessary for a quality discussion, or perhaps, given further research of personality factors and student contributions, students could be intentionally placed in groups where personalities predict higher-level discussions. Even if the reading selections are the same, limiting the number of students in groups or configuring groups strategically would not entail any more time or materials. Future research could also consider time as a variable. Perhaps the investigation could reveal the optimum amount of time necessary for a quality discussion.

**Limitations and conclusion**

This study had several limitations. First, as noted in the results, because of the small number of participants, the study had relatively low power, and thus the results should not be generalized. However, given the complexities of quality literature circle discussions evidenced, further research with a larger sample is warranted.

Next, this research utilized the three-story intellect as a coding framework, and making the connection between verbal utterances and internalized reading comprehension is difficult. It is not clear whether the two are highly correlated. Perhaps future research could compare the quality of verbal engagement in a discussion with measured reading comprehension of a text used with literature circles.

Finally, the TIPI is a self-report instrument, and its results should be considered with caution. However, Gosling et al. (2003) reported acceptable convergence correlations between observed and self-reported results \((r = .65)\). Future research on the influence of personality factors could include both self-reported and observed personality traits.

This study augments the extant research on successful literature circles, at least in one intermediate grade. It appears that there are several variables that may contribute to students’ productive book discussions. Consideration of the influence of a range of factors, including purpose, roles, reading abilities, structure, gender, personality traits, and group size on literature circles may complicate our understanding of their effectiveness. However, some of these factors are more readily managed by busy classroom teachers. Group size and gender balance are more straightforward aspects that are easily determined and that could support reading comprehension and group dynamics. Consideration of gender and group size are practical issues that teachers can consider in forming and monitoring the function of their literature circles, while research further explores the impact of personality and interpersonal interactions.

**ORCID**

Chase Young [http://orcid.org/0000-0002-3331-9339](http://orcid.org/0000-0002-3331-9339)

**References**


Clarke, L. W., & Holwadel, J. (2007). Help! What is wrong with these literature circles and how can we fix them? Reading Teacher, 61(1), 20–29.


