Article

Effects of Collaborative Group Composition and Inquiry Instruction on Reasoning Gains and Achievement in Undergraduate Biology

Jamie Lee Jensen* and **Anton Lawson**[†]

*Brigham Young University, Provo, UT 84602; [†]Arizona State University, Tempe, AZ 85287

This study compared the effectiveness of collaborative group composition and instructional method on reasoning gains and achievement in college biology. Based on initial student reasoning ability (i.e., low, medium, or high), students were assigned to either homogeneous or heterogeneous collaborative groups within either inquiry or didactic instruction. Achievement and reasoning gains were assessed at the end of the semester. Inquiry instruction, as a whole, led to significantly greater gains in reasoning ability and achievement. Inquiry instruction also led to greater confidence and more positive attitudes toward collaboration. Low-reasoning students made significantly greater reasoning gains within inquiry instruction when grouped with other low reasoners than when grouped with either medium or high reasoners. Results are consistent with equilibration theory, supporting the idea that students benefit from the opportunity for self-regulation without the guidance or direction of a more capable peer.

INTRODUCTION

Classroom inquiry requires that learners engage in scientifically motivated questions, generate and test alternative explanations based on evidence, connect explanations with scientific knowledge, and communicate and justify their explanations (National Research Council [NRC], 2000). Research has shown that inquiry instruction, using the learning cycle, is an effective constructivist teaching method leading to greater conceptual understanding and scientific reasoning gains over a traditional lecture format (e.g., Heiss *et al.*, 1950; Renner *et al.*, 1973; Howard and Miskowski, 2005; Spiro and Knisely, 2008; Minner *et al.*, 2009; Rissing and Cogan, 2009). In addition, inquiry instruction improves student attitudes and motivation to learn (Gibson and Chase, 2002; Berg *et al.*, 2003).

Most inquiry instruction takes place in collaborative settings. As suggested by Bransford *et al.* (2000), social interactions can shape the learning process itself. Although collaborative learning has been heralded as an effective strategy to improve student learning, retain science, technology, engineering, and mathematics (STEM) students, and meet many of the National Science Education Standards (American Association for the Advancement of Science, 1989; Forman, 1989; Drew, 1996; NRC, 1996; Slavin, 1996; Lord, 1997; O'Donnell and King, 1999; Pratt, 2003), there exist differences of opinion regarding the most effective group composition (when composition is based on some form of cognitive ability). Evidence has been found in support of both homogeneous and heterogeneous group compositions. However, the causal mechanisms behind the success of one composition over another are less well defined. In this study, we attempt to resolve inconsistencies in the literature over the most effective composition of collaborative learning groups within inquiry instruction as well as within traditional didactic instruction.

Most research on group composition has been done at the K–12 education levels. Some researchers have found homogeneous groups to be more beneficial than heterogeneous groups in promoting achievement. For example, Hooper (1992) found that fifth- and sixth-grade students placed in

homogeneous groups helped each other more and had higher achievement than those in heterogeneous groups. Fuchs *et al.* (1998) found that, among third and fourth graders, high reasoners performed better in homogeneous groups. In addition, they found that, when high reasoners were placed in heterogeneous groups, lower reasoners tended to stop participating and allowed the higher reasoning student(s) to do all of the work. Additionally, some studies in secondary schools have found that medium reasoners do better in homogeneous groups (see Webb, 1991, for a review).

Alternatively, other researchers have found that heterogeneous groups outperform homogeneous groups. For example, Amaria *et al.* (1969) found that heterogeneous groups produced superior and more creative problem solutions, which led to greater group satisfaction than homogeneous groups. Simsek and Benhong (1992) also found a more positive attitude toward group work among fourth through sixth graders in heterogeneous groups. Further, several researchers have found that heterogeneous groups lead to greater achievement gains among low reasoners. Bracey (1994) and Carter and Jones (1994) found low reasoners (among fifth graders) to be more on task in heterogeneous groups, and Webb *et al.* (2002) found that the quality of eighth-grade student discussions was higher in heterogeneous groups. Bracey (1994) found that high reasoners (among fifth graders) practiced verbal elaboration more frequently when placed in heterogeneous groups. Nattiv (1994) and Hooper (1992) found the amount of verbal elaboration to be positively correlated to achievement among elementary students.

At the college level, several studies have documented the success of collaborative groups in terms of achievement, attitudes, and persistence in STEM subjects (McKinney and Graham-Buxton, 1993; Johnson *et al.*, 1998; Armstrong *et al.*, 2007; Doymus, 2008; Preszler, 2009; see Bowen, 2000, for a review). Only a few such studies have explored the effects of heterogeneous versus homogeneous group compositions. An early study by Laughlin (1978) found that high reasoners performed better in homogeneous groups. Lawrenz and Munch (1985) found homogeneous groups to have better physical science achievement and greater reasoning gains. Nevertheless, students' initial reasoning was a better predictor than group composition. Weld (1999) found that students working on abstract algebra problems in homogeneous groups gained greater understanding, presumably because they were forced to work through the problems together. In heterogeneous groups, one student usually took the lead and all others followed, regardless of the correctness of the lead student's strategies. In a college psychology course, Baer (2003) found that high and medium achievers had higher final exam scores in homogeneous groups than in heterogeneous groups. Although Watson and Marshall (1995a, 1995b) predicted that heterogeneous groups would promote better learning in the college classroom due to greater student interactions, they found no difference in achievement among the grouping strategies. Lawson (1992) found that medium reasoners were less motivated and more frustrated when paired with students of lower reasoning ability, and thus preferred to be paired with an equally matched or higher reasoning peer. In sum, the few college-level studies that have been done reveal no clear consensus regarding the better group composition.

Consideration of the relevant theories may offer additional insight into the reasons for the relative success of different group compositions. For example, equilibration theory may offer a causal mechanism for the success of homogeneous groups. According to the theory, as individuals interact with their environment and encounter contradictory (i.e., disequilibrating) experiences, prior mental structures are reorganized as gaps and contradictions are discovered and eventually resolved through the process of equilibration or self-regulation (Flavell, 1963; Piaget, 1985; Yackel *et al.*, 1991; Lawson, 2002; O'Donnell *et al.*, 2006). Social interaction can influence self-regulation by posing critical cognitive conflict that disturbs equilibrium and forces the individual to restructure his/her cognitive architecture (e.g., Pulaski, 1980; Damon, 1984; Doise and Mugny, 1984; Kubli, 1989; Lumpe, 1995). However, social interaction alone is presumably not sufficient to cause cognitive change. The process of equilibration is seen primarily as an individualized process that takes place without the interference or leading of more able peers. Only when the self-driven process is completed will it lead to the construction of new knowledge and the development of more complex cognitive structures. Thus, equilibration theory predicts that students should experience greater reasoning gains and higher science achievement in homogeneous groups where equilibration (self-regulation) is more likely to take place unimpeded.

Alternatively, more socially oriented theories of development propose that the social environment can initiate change and also shape the change itself (Vygotsky, 1981; Damon, 1984; Tudge and Rogoff, 1989). For example, Vygotsky believed that children could perform above their current level of development when collaborating with others of higher ability (to a limited extent, of course) (Vygotsky, 1978; Woolfolk, 2004). Called the zone of proximal development, students have a level of potential development beyond their actual developmental level. If helped by a more able peer, with time the child will internalize the skills he/she observes in the more capable peer and learn to use them on his/her own (Ormrod, 2000; O'Donnell *et al.*, 2006). Researchers have used Vygotsky's zone of proximal development to explain the success of collaboration, especially in heterogeneous groups, with respect to ability (e.g., Mugny and Doise, 1978; Tudge, 1990). Thus, Vygotsky's more socially oriented theory of development seems to predict that students should experience greater improvement in reasoning ability and science achievement in heterogeneous groups where these helping behaviors are more likely to take place.

Given these differing empirical results and theoretical perspectives and predictions, we evaluated the success of both homogeneous and heterogeneous groups within collaborative settings. Further, to determine whether the instructional method had an interactive effect on the varied success of each type of collaborative group, both types were integrated into two instructional methods: inquiry and didactic.

METHODS

Participants

The study was conducted at a Southwestern community college serving over 13,000 students annually. Approximately

67% of the students are White, non-Hispanic, and the remaining 33% are minority races including American Native, Asian/Pacific Islander, Black, and Hispanic. This study was approved by the Institutional Review Board under exempt status.

Eight sections of a college introductory biology course, Cellular & Molecular Biology 181, participated in the study. Each section consisted of approximately 20 students. Sections met for 2 h and 45 min, twice weekly, which included both lecture and lab time. The study was conducted over two semesters with a total of eight sections and 144 subjects (due to typical attrition rates between 10% and 20%, the number of subjects included in the final analysis varied slightly in each treatment condition; see Table 1).

Students ranged in age from 19 to 25 years old. A majority of students (67%) were declared Pre-Health majors (i.e., Pre-Medical, Pre-Dental, Pre-Pharmacy, Pre-Physical Therapy, Pre-Nursing, and a few others). Most students were in their first semester of their major; 47% of them were freshmen.

Experimental Design

A quasi-experimental nonequivalent groups design was used. Results were analyzed using two-way analysis of variance (ANOVA). Steps were taken to ensure as much group equivalence as possible among the four treatment groups (i.e., identical classrooms, laboratory materials, textbooks, resources, curriculum, and expected learning outcomes). Inquiry and didactic sections had equivalent gender distributions (34% male in inquiry; 35% male in didactic), initial average reasoning ability ($M_{Inquiry} = 14.0$; $M_{Didactic} = 13.2$), percentage of Pre-Health majors (67.4% in inquiry; 64.9% in didactic), and attrition rates (22.8% in inquiry; 20.0% in didactic). The four treatment conditions were as follows: 1) inquiry instruction with homogeneous collaborative groups, 2) inquiry instruction with heterogeneous collaborative groups, 3) didactic instruction with homogeneous collaborative groups, and 4) didactic instruction with heterogeneous collaborative groups (see Table 1). The collaborative interactions in each treatment condition most closely resembled peer collaboration techniques as described by Damon and Phelps (1989), group investigation described by Sharan and Sharan (1992), and base groups described by Johnson *et al.* (1998). Students were allowed to work together on in-class work, but they were assessed individually. Instructors watched for any issues of free-riding or dominating by any group members, and those behaviors were discouraged when necessary. Group activities, which generally consisted of laboratory activities, were similar in both instructional methods and were structured and intentional, rather than haphazardly suggested. Although presentation styles differed between instructional methods (see more specific descriptions below), both were structured identically with 1 h devoted to "instruction" and 1.5 h dedicated to collaborative laboratory activities. Classrooms were arranged so that students sat at tables in their groups for the entire class period, including lecture.

Instructional Methods. One instructor taught two sections each semester using the inquiry method, more specifically the learning cycle method as described by Lawson (2002). Another instructor taught two sections each semester using a traditional style of didactic teaching.

Within the inquiry sections, students began each topic with an exploration of some phenomenon followed by applicable term introduction and application activities. Generally, labs served as exploration activities in which students were allowed room to explore, raise causal questions, and generate and test multiple hypotheses. Term introduction used either PowerPoint or whiteboards. Occasionally, lab activities served as application activities, although more commonly, homework assignments or group activities did so.

The didactic treatment consisted primarily of lecture by PowerPoint during the first half of class with little student interaction. During the second half of class, students collaborated in groups on laboratory activities pertaining to the current day's lecture or the lecture of the previous day. Laboratories generally directed students on each step of the lab and hinted at the appropriate outcomes. Little room was allowed for student exploration or hypothesis generation and testing. Supplemental Material 1 contains example lesson plans for each instructional method illustrating the key differences. Supplemental Material 2 contains sample exam questions illustrating the differences in assessment strategies between the two instructional methods.

Group Composition. Within both inquiry and didactic sections, students were assigned to collaborative groups for the entire semester. Collaborative groups were organized as either homogeneous or heterogeneous based on students' initial reasoning abilities as determined by the Classroom Test of Scientific Reasoning (version 2000; Lawson, 1978). Homogeneous groups consisted of students with similar reasoning abilities. Heterogeneous groups consisted of at least one

Table 1. Experimental design for two-way ANOVA

| | | | Instructional method | | |
			Inquiry	Noninquiry	Total subjects in each group composition
Group composition	Homogeneous	High	12	11	67
		Medium	13	13	
		Low	9	9	
	Heterogeneous		39	38	77
	Total subjects in each instructional method		73	71	144

Numbers indicate the total number of subjects in each treatment condition.

student with low reasoning ability and at least one student with high reasoning ability. Each section contained six collaborative groups, three homogeneous and three heterogeneous, consisting of three to four students each (in some cases, attrition led to less than expected collaborative group sizes). The three homogeneous groups in each section consisted of a homogeneous high, homogeneous medium, and homogeneous low group. Some studies have found that students do not normally collaborate unless compelled to do so (Cohen, 1994; Webb and Palinscar, 1996). However, Bruffee (1995) found that, at college age, students were well adept at interdependence and therefore more apt to spontaneously collaborate. A guide to effective collaboration was designed, distributed, and discussed in each section to encourage collaboration. Instructors watched for and discouraged noncollaborative behaviors. In addition, most activities were designed so that collaboration was necessary to complete the task effectively and efficiently (i.e., lab activities with multiple tasks, or that asked for students to discuss their results within their groups).

Instruments

Achievement. Student achievement was assessed by a common course final exam. The exam, informed by several standardized biology exams, consisted of 30 multiple-choice items at the knowledge and comprehension levels of Bloom's Taxonomy (herein referred to as low-level Bloom items; Bloom, 1984) and 27 multiple-choice items at application level or above (herein referred to as high-level Bloom items). Thus, the exam assessed both declarative knowledge as well as reasoning abilities. Items were designed and then categorized into Bloom levels by two individuals trained in assessing levels of Bloom's Taxonomy. Items were discussed and modified until both raters came to an agreement on the Bloom's level. The Spearman-Brown split-half reliability coefficient was 0.83.

Initial Reasoning Ability and Reasoning Gains. A modified version of the Classroom Test of Scientific Reasoning (version 2000; Lawson, 1978) consisting of 24 items was used to assess initial reasoning ability and reasoning gains. Validity and reliability of the test have been established by several studies (Lawson *et al.*, 2000). Reasoning ability has been found to be a good predictor of future conceptual understanding and achievement (Lawson *et al.*, 2000), and thus should be an appropriate means of establishing collaborative groups. Students scoring between 0 and 8 were classified as "low reasoners"; scores between 9 and 14 were classified as "medium reasoners"; and scores 15 and above were classified as "high

reasoners." The reasoning test was administered as an in-class assignment at the beginning of the course. Students were not graded on their performance but were not made aware of this until after completing the test. The reasoning test was readministered at the end of the semester as part of the final exam. Reasoning gains were calculated by subtracting initial scores from final scores. Gains were reported as a positive number; digression was reported as a negative number.

Reasoning Transfer. Six reasoning transfer items not included in the reasoning pretest were included in the course final exam. These six items served as an independent assessment of reasoning abilities as they could not have been influenced by pretest exposure. Items were patterned after the Classroom Test of Scientific Reasoning. Two sample transfer items are presented in Supplemental Material 3.

Attitude Survey. A pre- and posttreatment attitude survey was administered. The survey consisted of 15 items probing attitudes about group work, learning style preferences, confidence in the subject matter, self-esteem, group functioning, and instructor and class organization. Multiple items were used to assess each attitude. A total score for each category (e.g., a group function score) was obtained by taking an average score of all related responses. For example, a group functioning score was calculated by averaging responses to items shown in Table 2. Other sample pretest items are presented in Supplemental Material 3. Posttest items were assessed using a four-point Likert scale asking for degree of agreement or disagreement with the items. Sample posttest items are also presented in Supplemental Material 3.

RESULTS

Data from the four sections in the first semester were combined with data from the four sections in the second semester to increase *N*, increase power, and decrease the chances of semester-specific spurious results. Both achievement scores and reasoning change scores were analyzed for approximate normality using q-q plots. Both sets of scores show approximate normality and do not appear to violate this assumption.

Achievement on High-Level Bloom Items

Instructional Method. The achievement measure was partitioned into high-level Bloom and low-level Bloom items. On high-level Bloom items, inquiry sections outperformed didactic sections ($n = 73$, $M_{inquiry} = 68.9\%$; $n = 71$, $M_{didactic} = 64.9\%$; $F = 4.15$, $p = 0.04$; see Table 3).

Table 2. Items from the post-attitudes survey assessing group function

Statements from post-attitudes survey	Group function assessed
1. My group worked very well together.	Degree of collaboration
2. There was one (or more) person in my group who did not participate much.	Free-riding
3. There was one (or more) person in my group who dominated most of the activities.	Dominating
4. I made significant contributions of knowledge and/or ideas to the group.	Free-riding
5. My group member(s) made significant contributions of knowledge and/or ideas to the group.	Free-riding
6. I studied with the members of my group outside of class.	Degree of collaboration

Items were graded on a four-point Likert scale (1 = strongly agree, 2 = somewhat agree, 3 = somewhat disagree, 4 = strongly disagree). To obtain an overall group function score, responses to questions 2 and 3 were reverse-coded, and then all responses were averaged.

Table 3. Mean scores for each condition on achievement and reasoning transfer

		Inquiry		Didactic	
		Homogeneous	Heterogeneous	Homogeneous	Heterogeneous
High-level achievement					
Instructional method		68.9 ± 0.13^a		64.9 ± 0.12	
Group composition		69.4 ± 0.12	68.6 ± 0.14	62.1 ± 0.12^b	67.3 ± 0.13
Initial reasoning abilities					
	High	72.5 ± 0.11	78.0 ± 0.11	70.1 ± 0.09	73.5 ± 0.10
	Medium	67.1 ± 0.13	64.7 ± 0.08	60.0 ± 0.09	67.4 ± 0.12
	Low	64.2 ± 0.08	56.3 ± 0.14	47.4 ± 0.09	59.0 ± 0.13^c
Low-level achievement					
Instructional method		61.6 ± 0.13		65.6 ± 0.16	
Group composition		62.4 ± 0.11	61.0 ± 0.14	58.5 ± 0.14	71.8 ± 0.14^d
Initial reasoning abilities					
	High	65.0 ± 0.10	68.4 ± 0.13	65.9 ± 0.11	78.8 ± 0.10
	Medium	58.9 ± 0.10	54.9 ± 0.13	56.9 ± 0.14	69.0 ± 0.16
	Low	65.6 ± 0.17	55.9 ± 0.13	44.4 ± 0.08	64.7 ± 0.15
Reasoning transfer					
Instructional method		2.89 ± 1.39^e		2.52 ± 1.32	
Group composition		3.00 ± 1.41	2.79 ± 1.38	2.18 ± 1.19	2.82 ± 1.37^f
Initial reasoning abilities					
	High	3.44 ± 1.26	3.65 ± 1.12	3.08 ± 0.86	3.38 ± 1.41
	Medium	2.33 ± 1.23	2.31 ± 1.38	1.67 ± 0.98	2.80 ± 1.14
	Low	4.00 ± 2.00^g	1.89 ± 0.93	1.40 ± 1.14	2.08 ± 1.24

Values are percentages (mean $\pm$ SD) for high-level Bloom's items and low-level Bloom's items. Reasoning transfer items are reported as total points (mean $\pm$ SD) out of a possible six items.

[a] Inquiry sections outperformed didactic sections ($p = 0.04$).

[b] Homogeneous groups within didactic instruction performed lower that homogeneous groups within inquiry instruction ($p = 0.02$).

[c] Low-reasoning students had higher scores in heterogeneous groups than homogeneous groups within didactic instruction ($p = 0.04$).

[d] Heterogeneous groups outperformed homogeneous groups within didactic instruction ($p < 0.01$).

[e] Inquiry sections significantly outperformed didactic sections ($p = 0.01$).

[f] Heterogeneous groups outperformed homogeneous groups in didactic sections ($p = 0.02$).

[g] Low reasoners in homogeneous inquiry groups outperformed low reasoners in heterogeneous inquiry groups ($p = 0.02$) and low reasoners in homogeneous didactic groups ($p < 0.01$).

Group Composition. Within the inquiry sections, no difference was seen between homogeneous groups and heterogeneous groups, overall ($n = 34$, $M_{homogeneous} = 69.4\%$; $n = 39$, $M_{heterogeneous} = 68.6\%$; $F = 0.08$, $p = NS$; see Table 3). Homogeneous groups within the didactic sections had the lowest scores, significantly lower than homogeneous groups in the inquiry sections ($n = 34$, $M_{inquiry} = 69.4\%$; $n = 33$, $M_{didactic} = 62.1\%$; $F = 5.69$, $p = 0.02$). (Homogeneous groups were lower than heterogeneous groups in the didactic condition, but the difference did not reach significance, $n = 33$, $M_{homogeneous} = 62.1\%$; $n = 38$, $M_{heterogeneous} = 67.3\%$; $F = 3.26$, $p = 0.08$.)

Low-reasoning students had higher scores in heterogeneous groups than homogeneous groups within the didactic sections ($n = 5$, $M_{homogeneous} = 47.4\%$; $n = 12$, $M_{heterogeneous} = 59.0\%$; $F = 4.44$, $p = 0.04$). However, within the inquiry sections, the average score for low reasoners in homogeneous and heterogeneous groups was statistically equal, although low reasoners had relatively higher scores in homogeneous groups, a trend that is repeated and more statistically significant for reasoning scores ($n = 3$, $M_{homogeneous} = 64.2\%$; $n = 9$, $M_{heterogeneous} = 56.3\%$; $F = 1.12$, $p = NS$; see Figure 1). Medium and high reasoners showed no statistically significant trends between group compositions or instructional methods.

Achievement on Low-Level Bloom Items

Instructional Method. Inquiry and didactic sections performed equally on low-level Bloom items ($n = 73$, $M_{inquiry} = 61.6\%$; $n = 71$, $M_{didactic} = 65.6\%$; $F = 2.35$, $p = NS$; see Table 3).

Group Composition. No significant differences were seen on low-level Bloom items between group compositions within the inquiry sections ($n = 34$, $M_{homogeneous} = 62.4\%$; $n = 39$, $M_{heterogeneous} = 61.0\%$; $F = 0.18$, $p = NS$). Within the didactic sections, heterogeneous groups significantly outperformed homogeneous groups ($n = 33$, $M_{homogeneous} = 58.5\%$; $n = 38$, $M_{heterogeneous} = 71.8\%$; $F = 22.02$, $p < 0.01$; see Table 3 and Figure 2). Didactic heterogeneous groups, however, showed no advantage over either group composition within inquiry (see Table 3).

Reasoning Gains and Reasoning Transfer

Instructional Method. No significant differences were found between the inquiry and didactic sections for overall reasoning gains ($n = 73$, $M_{inquiry} = 1.97$; $n = 71$, $M_{didactic} = 2.07$; $F = 0.01$, $p = NS$; reasoning gains were calculated by subtracting initial reasoning scores from final reasoning scores;
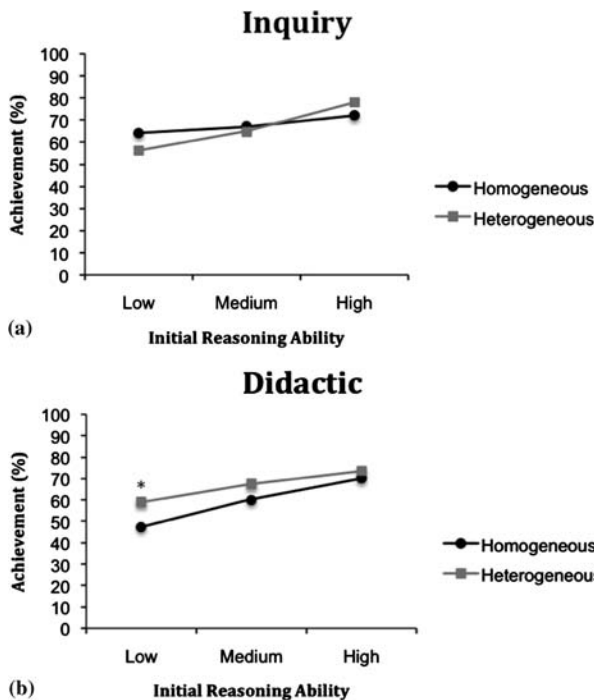
Figure 1. Achievement on high-level Bloom's Taxonomy items organized by the initial reasoning ability of the student and the group composition in which they were placed in inquiry instruction (a) and didactic instruction (b). Scores represent the average percentage correct out of the total number of high-level items administered. An asterisk (*) indicates that the difference between scores for low reasoners is significant ($p = 0.04$). Students in homogeneous groups within the didactic condition performed significantly lower than students in homogeneous groups in the inquiry condition ($p = 0.02$).
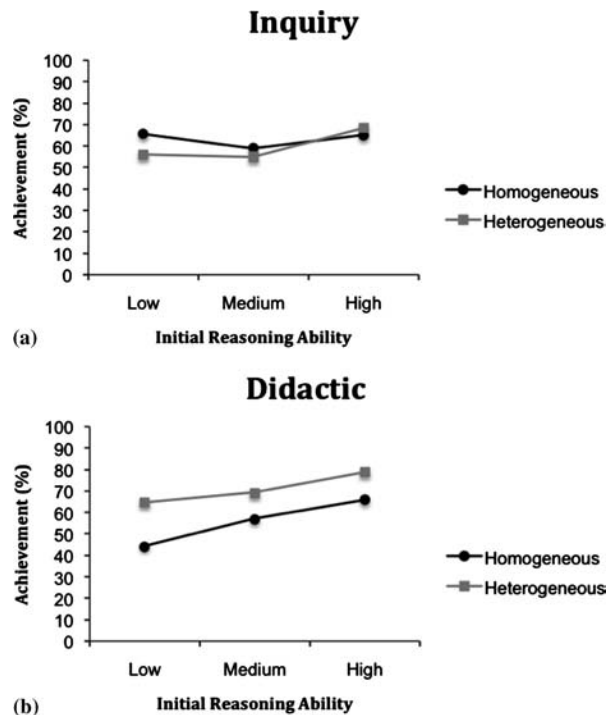

Figure 2. Achievement on low-level Bloom's Taxonomy items organized by the initial reasoning ability of the student and the group composition in which they were placed in inquiry instruction (a) and didactic instruction (b). Scores represent the average percentage correct out of the total number of low-level items administered. Overall, heterogeneous groups outperformed homogeneous groups ($p < 0.01$).

a maximum of 24 points was possible). However, the inquiry sections significantly outperformed the didactic sections on the reasoning transfer items ($n = 73$, $M_{inquiry} = 2.89$; $n = 71$, $M_{didactic} = 2.52$; $F = 6.42$, $p = 0.01$; see Table 3; reasoning transfer items were calculated based on a six-point scale). The most significant effects were found among low reasoners. Low reasoners in homogeneous groups within the inquiry sections significantly outperformed those within the didactic sections ($n = 3$, $M_{inquiry} = 4.00$; $n = 9$, $M_{didactic} = 1.40$; $F = 9.58$; $p < 0.01$; see Figure 3). In fact, within the inquiry sections, the students in homogeneous low groups equaled the performance of students in homogeneous high groups within inquiry ($n = 3$, $M_{homogeneous\ low} = 4.00$; $n = 16$, $M_{homogeneous\ high} = 3.44$; $p = $ NS). This was not the case in didactic homogeneous groups ($n = 5$, $M_{homogeneous\ low} = 1.40$; $n = 13$, $M_{homogeneous\ high} = 3.08$; $p < 0.01$).

Group Composition. Reasoning gains were most significant among low reasoners. These gains appear to be dependent on both the instructional method and the group composition ($F_{interaction} = 8.79$, $p < 0.01$). Within inquiry instruction, low reasoners had higher reasoning gains within homogeneous groups than heterogeneous groups ($n = 3$, $M_{homogeneous} = 8.67$; $n = 9$, $M_{heterogeneous} = 2.67$, $F = 5.22$, $p = 0.03$). Within didactic instruction, low reasoners had higher reasoning gains within heterogeneous groups ($n = 5$, $M_{homogeneous} = 2.20$; $n = 12$, $M_{heterogeneous} = 6.00$; $F = 3.39$, $p = 0.08$; note: differences were almost significant).

The same trend appears with reasoning transfer scores. Within inquiry instruction, low reasoners had much higher reasoning transfer scores when placed in homogeneous groups ($n = 3$, $M_{homogeneous} = 4.00$; $n = 9$, $M_{heterogeneous} = 1.89$; $F = 6.78$, $p = 0.02$; see Table 3). Medium and high reasoners had equal gains in both grouping conditions (see Table 3). Within the didactic sections, students in heterogeneous groups, overall, outperformed students in homogeneous groups ($n = 33$, $M_{homogeneous} = 2.18$; $n = 38$, $M_{heterogeneous} = 2.82$; $F = 5.75$, $p = 0.02$). However, low reasoners had no advantage in either group composition, having the lowest reasoning transfer scores in both conditions.

Attitudes Analyses

The attitudes survey revealed a significant positive correlation between the group functioning score of each group and the amount of helping behaviors that occurred within that group ($n = 144$; $r = 0.42$, $r^2 = 0.18$, $p < 0.01$; see Figure 4). Although not strong, there was also a significant positive correlation between the amount of helping behaviors in the group and average achievement for group members ($n = 142$; $r = 0.19$, $r^2 = 0.04$, $p = 0.02$; see Figure 4).

In addition, students in the inquiry sections expressed greater confidence in their own reasoning ability ($n = 73$, $M_{inquiry} = 3.35$; $n = 71$, $M_{didactic} = 3.08$; $F = 6.49$, $p = 0.01$). Inquiry sections also expressed better attitudes toward

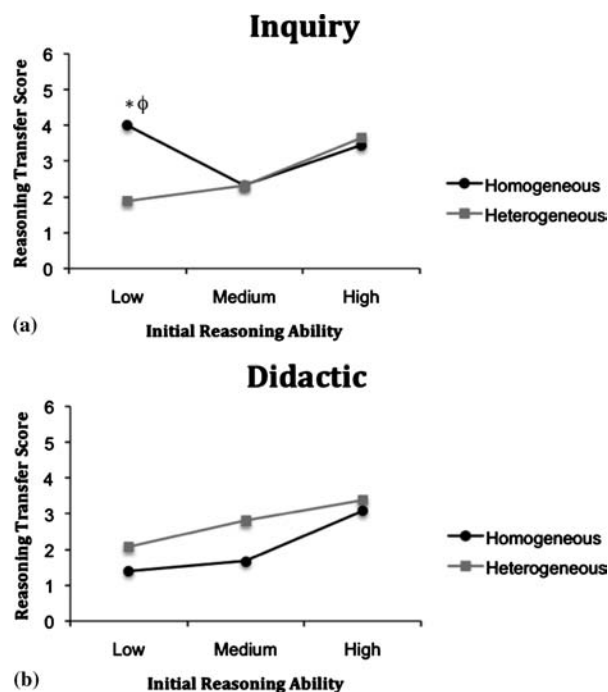**Inquiry**

Didactic

(a)

(b)

Figure 3. Scores on reasoning transfer items organized by the initial reasoning ability of the student and the group composition in which they were placed in inquiry instruction (a) and didactic instruction (b). Scores represent the average score on six reasoning transfer items administered as part of the final exam. An asterisk (*) indicates that the difference in reasoning transfer scores among low reasoners is significant ($p = 0.02$). ϕ, Low reasoners in homogeneous groups within the inquiry condition significantly outperformed low reasoners in homogeneous groups within the didactic section ($p < 0.01$) and performed equally as well as high reasoners in the inquiry condition ($p = NS$). Overall, heterogeneous groups outperformed homogeneous groups within the didactic condition ($p = 0.02$).

collaboration ($n = 73$, $M_{inquiry} = 3.50$; $n = 71$, $M_{didactic} = 3.30$; $F = 4.06$, $p = 0.05$).

DISCUSSION

When achievement was measured by high-level Bloom items, inquiry sections outperformed noninquiry sections. This result is consistent with previous studies, indicating that inquiry instruction leads to greater conceptual understanding (Heiss *et al.*, 1950; Renner *et al.*, 1973; Minner *et al.*, 2009). Instructional method had little effect on achievement when measured by items at the knowledge or comprehension levels of Bloom's Taxonomy (i.e., by items requiring recall or minimal understanding). This was expected as knowledge-level items should not be highly influenced by one's reasoning ability, although comprehension-level items have been found to be influenced somewhat by reasoning ability (Lawson *et al.*, 2000). Further, inquiry sections had higher scores on reasoning transfer items, with significantly higher performance within homogeneous groups, showing that inquiry instruction improves reasoning skills more effectively than noninquiry instruction. In addition to gaining better reasoning ability, students in the inquiry condition responded with better confidence in their ability to reason and more
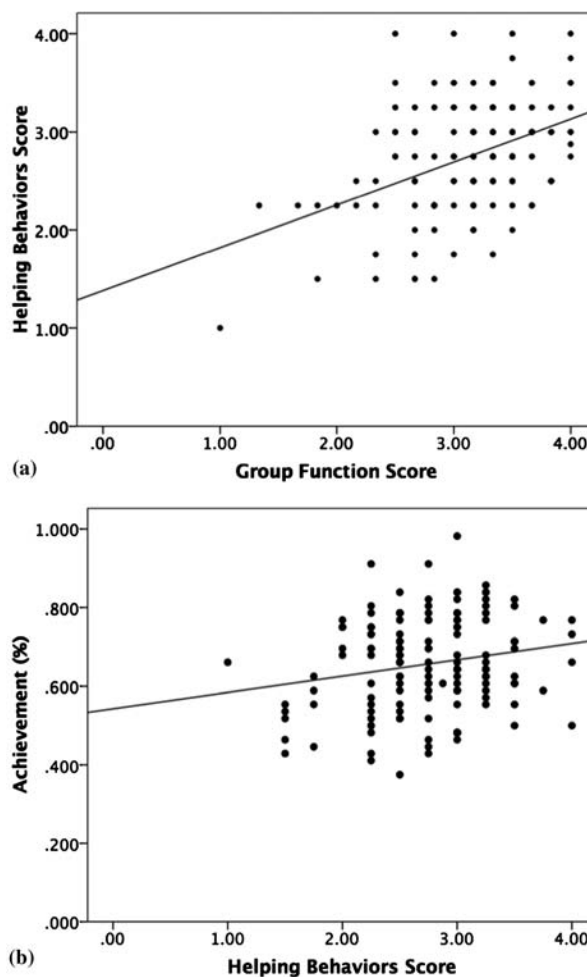

(a)

(b)

Figure 4. Attitudes analyses showed that (a) the amount of group functioning occurring is a significant predictor of the amount of helping behaviors that occurred within groups ($r^2 = 0.18$, $n = 144$, $p < 0.001$) and (b) the amount of helping behaviors is a significant predictor of the overall achievement in the course ($r^2 = 0.04$, $n = 142$, $p = 0.02$). Group functioning and helping behavior scores were obtained through a survey at the end of the course. Achievement scores were obtained on a common comprehensive final exam.

positive attitudes toward the collaborative experience. This result is also consistent with previous studies that have found that inquiry instruction improves overall attitudes and motivation toward learning (e.g., Gibson and Chase, 2002; Berg *et al.*, 2003). Because instructional conditions were taught by two different teachers, it is possible that there was an instructor effect. Although this variable is difficult to control, steps were taken to eliminate its effects as much as possible. These steps included using the same instructional materials (i.e., textbooks, lab manuals, lab supplies; see Supplemental Material 1 for sample lesson plans), teaching the curriculum in the same order and to the same depth, maintaining a similar teaching environment (i.e., classroom, group organization, seating arrangements), and maintaining active and open communication between instructors. However, an instructor effect cannot be entirely ruled out and should be taken into consideration when interpreting results.

Group composition (i.e., homogeneous vs. heterogeneous) appears to have a conditional effect on achievement on high-level items, depending on one's initial reasoning ability. Low reasoners tend to perform better when placed in homogeneous groups, whereas medium and high reasoners performed equally in both group compositions. On low-level items, an effect is only seen in the didactic condition, where heterogeneous groups outperformed homogeneous groups. Perhaps this is due to helping behaviors that occurred within the group since a correlation was seen between helping behaviors and average achievement. This finding is consistent with previous research (Webb *et al.*, 2002). Perhaps in an instructional situation where very little student-directed learning is taking place, having a high-reasoning student in the group to provide answers for memorization and recall may prove to be the most effective learning strategy. However, within inquiry, where more student-directed learning is taking place and less memorization is expected, the effect was not seen.

The effect of group composition on reasoning gains and reasoning transfer scores (and high-level achievement to a lesser extent) was dependent on both the student's initial reasoning level and instructional method. Consistent with equilibration theory, within inquiry instruction, low reasoners made significantly greater reasoning gains and had significantly higher reasoning transfer scores when placed in homogeneous groups. Thus, it appears that students with low reasoning ability are benefiting more when given the opportunity for self-regulation to occur without the guidance or direction of a more capable peer. In fact, within the inquiry sections, the two highest reasoning gains were seen in two low reasoners placed within a single homogeneous group. In contrast, low-reasoning students within the didactic sections did not experience such benefits, and in fact had the lowest scores in the study when placed in homogeneous groups. Possibly within didactic instruction, where processes that encourage cognitive development are not specifically encouraged, low reasoners do better in heterogeneous groups because they benefit from the helping behaviors and elaborative practices that are occurring or perhaps, more likely, because they receive the right answers from more capable peers. Medium- and high-reasoning students had equal reasoning gains and reasoning transfer scores in either group composition. Although we did not see a statistically significant difference in high-level achievement for low reasoners, the trend was the same. Low reasoners tended to do better in homogeneous groups. Because high-level achievement required domain-specific knowledge in addition to reasoning skills, it was a unique measure of reasoning ability not assessed by the reasoning test or reasoning transfer items.

The differing success of low reasoners seen in this study may offer an explanation for the differences in the previous literature regarding the relative success of homogeneous or heterogeneous groups. The benefit of each group composition seems dependent on the instructional approach used. When active, inquiry-based, constructivist teaching takes place, homogeneous groups are most beneficial, especially for students with low reasoning ability. Whereas if more traditional, didactic teaching occurs, where modes of assessment are focused more on memorization rather than understanding, heterogeneous groups appear to do better, although their knowledge appears to be based on memorization (as evidenced by their higher scores on low-level questions but lower scores on high-level questions), not on conceptual understanding (see Supplemental Material 2 for sample exam questions).

INSTRUCTIONAL IMPLICATIONS

With the decline in the number of students majoring in STEM disciplines and specifically in biology, teachers need to strive to make learning a positive and fruitful experience. Collaborative learning has been shown to be a beneficial strategy to increase student achievement. However, the best composition of collaborative groups has not been firmly established. This study finds that one of the possible keys to success in collaborative learning is to use an inquiry approach coupled with homogeneous group composition. This appears to be especially important for students with initially low reasoning abilities.

To apply these findings in the classroom, instructors should assess student reasoning ability prior to instruction. By doing so, instructors will be aware of the capabilities of their students and can tailor their instruction accordingly. In addition, initial assessment will allow instructors to form collaborative groups so that low-reasoning students are grouped together and are allowed to experience cognitive conflict and self-regulation without being interrupted and impeded by a higher-reasoning peer.

Applying the learning cycle to biology can be rather challenging, and many instructors are resistant to change. However, several resources are available to help instructors get started (e.g., Lawson, 1994, 2002; Bransford *et al.*, 2000; NRC, 2000). This and many other studies have found inquiry instruction to be superior to traditional lecture-style teaching in the development of scientific reasoning ability and, more specifically, in its application to higher-level biological conceptual understanding. Inquiry instruction also appears to foster greater confidence in student ability and more positive attitudes toward collaboration as a whole.

Collaboration is a foundational skill for success in the STEM disciplines as much of science is done in a collaborative atmosphere. Science should be taught in the manner in which science is practiced. Using collaborative groups not only helps students to better understand the nature of scientific endeavors, but also provides opportunities for cognitive conflict and self-regulation (Damon, 1984; Doise and Mugny, 1984). However, as this study found, collaborative learning groups are most effectively utilized in conjunction with inquiry-style instruction. To reap the benefits of collaborating with other students, students need to be given the opportunity to puzzle over explored phenomena, encouraged to devise possible explanations, and allowed to consider ways of testing them. During these activities, students experience cognitive conflict and engage in self-regulatory behaviors. Further, self-regulation should be allowed to take place, uninterrupted by a peer of higher ability. Through these activities, students are actively participating in the learning process and are gaining self-confidence in their own abilities. By keeping students actively involved, they are more likely to retain what they learn, progress in their ability to reason, and enjoy the process. Students may even be more likely to continue as STEM majors.

J. L. Jensen and A. Lawson

ACKNOWLEDGMENTS

REFERENCES

Amaria RP, Biran LA, Leith GOM (1969). Individual versus cooperative learning. Educ Res *11*, 95–103.

American Association for the Advancement of Science (1989). Science for All Americans: Project 2061, New York: Oxford University Press.

Armstrong N, Chang SM, Brickman M (2007). Cooperative learning in industrial-sized biology classes. CBE Life Sci Educ *6*, 163–171.

Baer J (2003). Grouping and achievement in cooperative learning. Coll Teach *51*, 169–174.

Berg AR, Bergendahl CB, Lundberg BKS (2003). Benefiting from an open-ended experiment? A comparison of attitudes to, and outcomes of, an expository versus an open-inquiry version of the same experiment. Int J Sci Educ *25*, 351–372.

Bloom BS (1984). Taxonomy of Educational Objectives, Boston: Allyn and Bacon.

Bowen CW (2000). A quantitative literature review of cooperative learning effects on high school and college chemistry achievement. J Chem Educ *77*, 116–119.

Bracey G (1994). Achievement in collaborative learning. Phi Delta Kappan *76*, 254–256.

Bransford JD, Brown AL, Cocking RR (eds.) (2000). How People Learn: Brain, Mind, Experience, and School, Washington, DC: National Academies Press.

Bruffee KA (1995). Sharing our toys: cooperative learning versus collaborative learning. Change *27*, 12–19.

Carter G, Jones MG (1994). Relationships between ability-paired interactions and the development of fifth graders' concepts of balance. J Res Sci Teach *31*, 847–856.

Cohen EG (1994). Restructuring the classroom: conditions for productive small groups. Rev Educ Res *64*, 1–35.

Damon W (1984). Peer education: the untapped potential. J Appl Dev Psychol *5*, 331–343.

Damon W, Phelps E (1989). Critical distinctions among three approaches to peer education. Int J Educ Res *13*, 9–19.

Doise W, Mugny G (1984). The Social Development of the Intellect, Oxford: Pergamon Press.

Doymus K (2008). Teaching chemical equilibrium with the jigsaw technique. Res Sci Educ *38*, 249–260.

Drew DE (1996). Aptitude Revisited: Rethinking Math and Science Education for America's Next Century, Baltimore: The Johns Hopkins University Press.

Flavell JH (1963). The Developmental Psychology of Jean Piaget, Princeton, NJ: D. Van Nostrand.

Forman EA (1989). The role of peer interaction in the social construction of mathematical knowledge. Int J Educ Res *13*, 55–70.

Fuchs LS, Fuchs D, Hamlett CL, Karns L (1998). High-achieving students' interactions and performance on complex mathematical tasks as a function of homogeneous and heterogeneous pairings. Am Educ Res J *35*, 227–267.

Gibson HL, Chase C (2002). Longitudinal impact of an inquiry-based science program on middle school students' attitudes toward science. Sci Educ *86*, 693–705.

Heiss ED, Obourn ES, Hoffman CW (1950). Modern Science Teaching, New York: Macmillan.

Hooper S (1992). Effects of peer interaction during computer-based mathematics instruction. J Educ Res *85*, 180–189.

Howard DR, Miskowski JA (2005). Using a module-based laboratory to incorporate inquiry into a large cell biology course. Cell Biol Educ *4*, 249–260.

Johnson DW, Johnson RT, Smith KA (1998). Active Learning: Cooperation in the College Classroom, Edina, MN: Interaction Book Company, 187–192.

Kubli F (1989). Piaget and interest in science subjects. In: Adolescent Development and School Science, ed. P Adey, J Bliss, and M Shayer, New York: Falmer Press.

Laughlin PR (1978). Ability and group problem solving. J Res Dev Educ *12*, 114–120.

Lawrenz F, Munch TW (1985). Aptitude treatment effects of laboratory grouping method for students of differing reasoning ability. J Res Sci Teach *22*, 279–287.

Lawson AE (1978). The development and validation of a classroom test of formal reasoning. J Res Sci Teach *15*, 11–24.

Lawson AE (1992). Using reasoning ability as the basis for assigning laboratory partners in nonmajors biology. J Res Sci Teach *29*, 729–741.

Lawson AE (1994). Biology: A Critical-Thinking Approach, Menlo Park, CA: Addison-Wesley Publishing.

Lawson AE (2002). Science Teaching and Development of Thinking, Belmont, CA: Wadsworth/Thompson Learning.

Lawson AE, Alkhoury S, Benford R, Clark BR, Falconer KA (2000). What kinds of scientific concepts exist? Concept construction and intellectual development in college biology. J Res Sci Teach *37*, 996–1018.

Lawson AE, Clark B, Cramer-Meldrum E, Falconer KA, Sequist JM, Kwon YJ (2000). Development of scientific reasoning in college biology: Do two levels of general hypothesis-testing skills exist? J Res Sci Teach *37*, 81–101.

Lord TR (1997). A comparison between traditional and constructivist teaching in college biology. Innov Higher Educ *21*, 197–216.

Lumpe AT (1995). Peer interaction in science concept development and problem solving. Sch Sci Math *95*, 302.

McKinney K, Graham-Buxton M (1993). The use of collaborative learning groups in the large class: is it possible? Teach Sociol *21*, 403–408.

Minner DD, Levy AJ, Century J (2009). Inquiry-based science instruction: what is it and does it matter? Results from a research synthesis years 1984 to 2002. J Res Sci Teach *47*, 474–496.

Mugny G, Doise W (1978). Socio-cognitive conflict and structure of individual and collective performances. Eur J Soc Psychol *8*, 181–192.

National Research Council (NRC) (1996). National Science Education Standards, Washington, DC: National Academies Press.

NRC (2000). Inquiry and the National Science Education Standards: A Guide for Teaching and Learning, Washington, DC: National Academies Press.

Nattiv A (1994). Helping behaviors and math achievement gain of students using cooperative learning. Elem School J *94*, 185–197.

O'Donnell AM, Hmelo-Silver CE, Erkens G (eds.) (2006). Collaborative Learning, Reasoning, and Technology, Mahway, NJ: Lawrence Erlbaum Associates.

O'Donnell AM, King A (eds.) (1999). Cognitive Perspectives on Peer Learning, Mahwah, NJ: Lawrence Erlbaum Associates.

Ormrod JE (2000). Educational Psychology: Developing Learners, Upper Saddle River, NJ: Prentice-Hall.

Piaget J (1985). The Equilibration of Cognitive Structures: The Central Problem of Intellectual Development, Chicago: University of Chicago Press.

Pratt S (2003). Cooperative learning strategies. Sci Teach *70*, 25–29.

Preszler RW (2009). Replacing lecture with peer-led workshops improves student learning. CBE Life Sci Educ *8*, 182–192.

Pulaski MAS (1980). Understanding Piaget: An Introduction to Children's Cognitive Development, New York: Harper and Row.

Renner JW, Stafford DG, Coffia WJ, Kellogg DH, Weber MC (1973). An evaluation of the Science Curriculum Improvement Study. School Sci Math *73*, 291–318.

Rissing SW, Cogan JG (2009). Can an inquiry approach improve college student learning in a teaching laboratory? CBE Life Sci Educ *8*, 55–61.

Sharan Y, Sharan S (1992). Expanding Cooperative Learning through Group Investigation, New York: Teachers College Press.

Simsek A, Benhong T (1992). The impact of cooperative group compositions on student performance and attitudes during interactive videodisc instruction. J Comput-Base Instr *19*, 86–91.

Slavin RE (1996). Research for the future: research on cooperative learning and achievement: what we know, what we need to know. Contemp Educ Psychol *21*, 43–69.

Spiro MD, Knisely KI (2008). Alternation of generations and experimental design: a guided-inquiry lab exploring the nature of the *her1* developmental mutant of *Ceratopteris richardii* (C-Fern). CBE Life Sci Educ *7*, 82–88.

Tudge J (1990). Vygotsky, the zone of proximal development, and peer collaboration: implications for classroom practice. In: Vygotsky and Education: Instructional Implications and Applications of Sociohistorical Psychology, ed. L Moll, New York: Cambridge University Press, 155–172.

Tudge J, Rogoff B (1989). Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In: Interaction in Human Development, ed. MH Bornstein and JS Bruner, Hillsdale, NJ: Erlbaum, 17–40.

Vygotsky LS (1978). Mind in Society: The Development of Higher Psychological Processes, Cambridge, MA: Harvard University Press.

Vygotsky LS (1981). The genesis of higher mental functions. In: The Concept of Activity in Soviet Psychology, ed. JV Wertsch, Armonk, NY: Sharpe, 134–240.

Watson SB, Marshall JE (1995a). Heterogeneous grouping as an element of cooperative learning in an elementary education science course. School Sci Math *95*, 401.

Watson SB, Marshall JE (1995b). Effects of cooperative incentives and heterogeneous arrangement on achievement and interaction of cooperative learning groups in a college life science course. J Res Sci Teach *32*, 291–299.

Webb NM (1991). Task-related verbal interaction and mathematics learning in small groups. J Res Math Educ *22*, 366–389.

Webb NM, Nemer KM, Zuniga S (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. Am Educ Res J *39*, 943–989.

Webb NM, Palinscar AA (1996). Group processes in the classroom. In: Handbook of Educational Psychology, ed. DC Berliner and RC Calfee, New York: MacMillan, 841–873.

Weld K (1999). Perfect problems and homogeneous groups enhance cooperative learning in abstract algebra. PRIMUS *9*, 355–364.

Woolfolk A (2004). Educational Psychology, 9th ed., Boston: Allyn and Bacon.

Yackel E, Cobb P, Wood T (1991). Small-group interactions as a source of learning opportunities in second-grade mathematics. J Res Math Educ *22*, 390.