

Learning Together While Designing: Does Group Size Make a Difference?

Xornam S. Apedoe · Michelle R. Ellefson ·
Christian D. Schunn

Published online: 19 February 2011
© Springer Science+Business Media, LLC 2011

Abstract As the use of project-based learning becomes more frequent in the K-12 science classroom, and in chemistry classrooms in particular, teachers have begun to identify practical questions about implementation that should be addressed empirically. One such question concerns whether there is an ideal group size that fosters individual student achievement. The current project was designed to assess how group size might impact student chemistry content learning in a project-based learning environment, and how well students are prepared to transfer this new knowledge to other relevant areas. The results indicated that particular conditions (e.g. advanced classrooms) interact with group size (a seemingly superficial feature) to differentially influence the depth and level of student learning related to the unit and student's ability to transfer his/her knowledge outside of the context of a project-based learning unit.

Keywords Group size · Project-based learning · High school chemistry · Design-based learning · Knowledge gain · High school science

X. S. Apedoe (✉)
Learning & Instruction, School of Education,
University of San Francisco, 2130 Fulton St,
San Francisco, CA 94117, USA
e-mail: xapedoe@usfca.edu

M. R. Ellefson
Psychology in Education, Faculty of Education, University
of Cambridge, 184 Hills Rd., Cambridge CB2 8PQ, UK

C. D. Schunn
Learning Research & Development Center, University of
Pittsburgh, 3939 O'Hara St, Pittsburgh, PA 15260, USA

Introduction

Chemistry is a challenging subject in the school curriculum, both to teach and to learn, because it is complex and abstract. A large part of students' difficulty understanding chemistry stems from the emphasis on teaching chemistry at the symbolic level (Gabel 1999). Not only do students have difficulty understanding the symbolic formalisms used in chemistry, they also have little or no motivation for understanding them. Gabel (1999) argues that incorporating hands-on practical work into the chemistry classroom is one way to make the symbolic formalisms more meaningful and strengthen the connections to the macroscopic and microscopic levels of representation.

One way to incorporate such hands-on practical work is through the use of project-based learning approaches. The enthusiasm and belief in the efficacy for project-based learning has waxed and waned over the years, resulting in only a handful of teachers adopting the approach consistently (Barron et al. 1998). However, there is strong evidence that project-based learning can be successful (e.g. Cognition and Technology Group at Vanderbilt (CTGV) 1994, 1997; Collins et al. 1991; Schauble et al. 1995), particularly within chemistry courses (e.g. Barak and Dori 2004; Schank and Kozma 2002).

There are a number of factors that are important for the design of project-based learning environments that can influence how successful they are for student learning. Krajcik et al. (1994) identified five essential features of project-based instruction wherein students: (a) investigate authentic questions or problems, (b) develop a series of artifacts or products that address the question/problem, (c) engage in investigations, (d) collaborate with peers, teachers and community members to inquire into the problem, and (e) use cognitive tools. Enacting each of these

features can be challenging for teachers, let alone orchestrating the enactment of each feature simultaneously (Krajcik et al. 1994). One feature that is particularly challenging for teachers to enact successfully is the collaboration feature.

Collaboration is an essential component of many project-based learning environments. It provides opportunities for students to extend their thinking, share ideas, and draw on the expertise of others (Krajcik et al. 1994). Having students simply work together to complete prescribed procedures for an investigation or task does not constitute collaboration (Marx et al. 1997). Rather, collaboration involves the construction of shared meanings (Webb and Palincsar 1996) through the exchange of ideas. Thus, collaboration requires that teachers establish classroom environments where students are held accountable for their work, respect each other's opinions and ideas, and are engaged in exchanging ideas and negotiating meaning.

However, the question remains whether all students receive the same learning benefits when engaged in collaborative work. The size of the collaborative groups, a seemingly superficial factor, can influence both group processes and individual learning outcomes. The size of the collaborative group also raises both pragmatic and theoretical concerns. Pragmatically, using smaller groups (such as pairs) rather than larger groups (three or more students) leads to more total groups in the classroom that need to be monitored by the teacher. A larger number of groups may limit the teachers' ability to adequately scaffold and facilitate each group, and individual students' learning. In addition, issues related to class management must be considered when deciding how many groups are feasible to monitor in a classroom by one teacher. Determining collaborative learning group size is based routinely on practical considerations such as instructor support, available learning resources, or anecdotal reports of positive instructional experiences with particular group sizes (Lohman and Finkelstein 2000).

While there are a number of studies that have examined the influence of group size on student learning or performance, there is no consensus regarding the optimal group size (Fuchs et al. 2000; Lohman and Finkelstein 2000; Wiley and Jensen 2006). Research studies from social and cognitive psychology suggests that pairs are rather different than groups of size three or larger (Levine and Moreland 2004; Wiley and Jensen 2006). Pairs are said to be optimal because each student has an increased opportunity to participate (Lohman and Finkelstein 2000). In addition, there are fewer group members in a pair to distract a student from their own thinking (Dugosh et al. 2000).

In contrast, larger groups (three or more students) may be more likely to think about multiple perspectives and critically evaluate both majority and minority statements in terms of evidence and justifications (Wiley and Jensen 2006). In addition, larger groups bring more intellectual

capital, diversity of thought, experiences, and viewpoints that allow for the active participation of all group members (Lohman and Finkelstein 2000).

The composition of the groups, i.e. high-ability versus low-ability students, may also have a significant impact on both group processes and the learning outcomes achieved by the group and individuals within the group. Previous evidence (e.g. Webb 1980; Lou et al. 1996) supports the use of heterogeneous ability groups, as they are beneficial for both low-ability and high-ability learners. Theories of children's concept development support the use of heterogeneous groups, because the high-ability students can facilitate lower-ability students' learning (Webb and Palincsar 1996). Contrary to common belief, high ability students do not suffer in these groups. Typically, they adopt a leadership and teaching role that consequently results in greater learning for the high-ability students (Webb 1982a, b; Webb and Palincsar 1996).

The ability to create collaborative groups of heterogeneous ability levels requires that teachers have classrooms of students with differing ability levels. In many schools, students are assigned to classes based on their ability levels (a process known as tracking), i.e. high-ability (advanced) or low-ability (mainstream). As a result, many classrooms do not have very broad diversity in the ability levels of the students, preventing teachers from using heterogeneous groups.

It is important to consider the depth of learning (e.g. automated skills versus more complex concepts and reasoning) that is expected from students as well as the depth of the measures of student learning (e.g. none, close, or far transfer measures). For example, larger group interactions may challenge superficial thinking that may show itself in improved close or far transfer measures. By contrast, larger groups might cover basic material more slowly, showing itself in reduced performance on measures of material directly taught.

Another critical concept in group learning is that individuals within a group often achieve different depths of learning. Beyond simple individual differences in ability to learn the given material given prior knowledge and cognitive ability differences, this inequality of learning may be due to one or more factors that can interact with group size: some students refuse to share the workload with team members, some students allow other team members to do most of the work, students divide work into subtasks that are not equal in learning opportunities, or subtasks are so finely divided that students lose track of the larger picture. A focus of the current study is the extent to which group size influences inequality of learning.

From the existent literature, it is not clear what size group should be expected to exhibit better overall group learning (pairs or groups of three or four), or how that might translate to individual learning outcomes, especially considering more basic and more advanced levels of knowledge gains. For cooperative or collaborative learning

in particular, groups of four are most often recommended, yet it is unclear what size group is most beneficial for learning (Webb and Palincsar 1996). In addition, the current evidence provides little insight regarding how the level of students (i.e. advanced versus mainstream) may interact with the size of the collaborative groups and consequently affect individual learning outcomes. Few studies have experimentally manipulated group size to examine its effects on student productivity while keeping other structural variables constant and using complex tasks (Fuchs et al. 2000; Webb and Palincsar 1996), so we seek to do this in the current study. Specifically, with this study we are trying to make progress on the question of *whom* will exhibit better group learning (pairs or groups of three or four), under what circumstances (advanced vs. mainstream classrooms), and for what kind of learning goals (basic vs. more advanced knowledge) is learning best achieved in chemistry classrooms that use project-based learning?

Goals of the Present Study

The purpose of the present study, therefore, was to explore the effect of group size on student learning in chemistry classrooms that used a project-based learning approach. This study was situated within high school chemistry classrooms in which students engaged in design-based learning, which is a particular form of project-based learning. In design-based learning, the activity that is meant to drive learning is a design-project; students are required to use and extend their knowledge of science and math to develop a technological solution to a problem using available resources. The main research question of interest was: how does group size affect individual students' chemistry content learning? We attempt to answer this question by first examining how well individual students *within* a group learn, by looking at their mean gain scores on the assessment measures. Next, we explore how the groups learned as a whole and examine within group variability on learning measures. Note that project-based learning in real classrooms typically involve a number of different modes of instruction and learning, ranging from individual worksheets to group-work to group presentations to the whole class to whole class discussions. Thus, a study of group size in project-based learning naturally involves studying an environment that is not purely group-work.

Method

Design

We used a pretest–posttest, quasi-experimental design with students assigned to specific group sizes. Based on the

literature suggesting that most cooperative or collaborative methods advocate using groups of four (e.g. Webb and Palincsar 1996), we chose to examine the difference in learning gains between pairs and groups of four. We counterbalanced the assignment of group size across the five teachers who participated in this study. That is, for each teacher that taught multiple class periods, each class period was assigned to a specific group size i.e. pairs or fours. Furthermore, group size was 'counterbalanced' across class type, i.e., mainstream or advanced. That is if a teacher taught two mainstream classes, one class was assigned a group size of pairs, while the other mainstream class was assigned a group size of fours. Occasionally, because of actual class sizes, some groups were comprised of three students. Groups with three students were combined with groups of four for the statistical analyses.¹

Context

This study was situated within high school chemistry classrooms in which students engaged in a design-based learning curriculum unit, the *Heating/Cooling System*. The *Heating/Cooling System* utilizes an engineering design approach to teach high school students the central and difficult chemistry concepts of atomic interactions, reactions and energy changes in reactions. In the unit, students work in groups, as they go through a design process to create a prototype of a heating or cooling system that relies on chemical energy and meets a personal need in their own life. This process is akin to the process used by engineering designers. For a detailed description of the unit see Apedoe et al. (2008).

The unit guides the students through three main parts: *Planning the Design*, *The Subsystems*, and *Presenting the Design*. For each of the subsystems, students follow a cycle from design goals to science goals and back to design goals (see Fig. 1). For this paper, we refer to the cycle as the *Learning Cycle*, and it is similar to the *Legacy Cycle* developed by Bransford and colleagues (see Brophy and Bransford 2001). The particular version of a learning cycle that is found in the unit was developed, in part, to integrate the advantages of group-work and whole class discussions.

Each pass through the *Learning Cycle* typically takes from 5 to 8 days to complete, allowing students to complete the cycle three times. Throughout this cycle, there are whole class discussions, group activities, and individual activities created to maximize learning of science content as well as learning design and science processes. Direct instruction is used when needed throughout the cycle at the

¹ Preliminary analyses indicated that there were no significant differences between group sizes of three and four on our assessment measures.

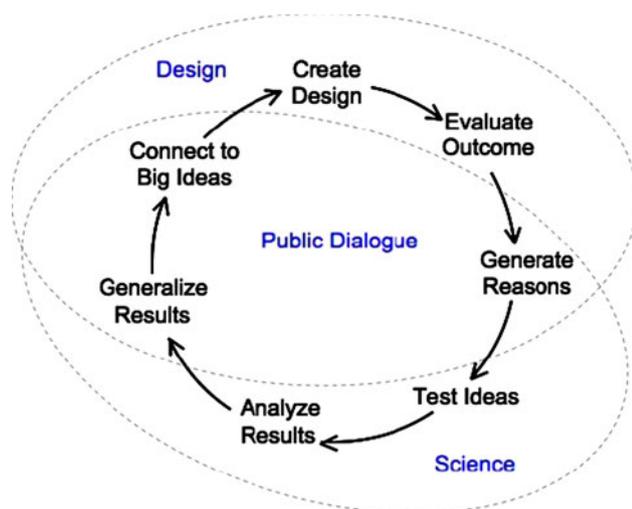


Fig. 1 The *Learning Cycle* structured learning for the students as they progressed through the *Heating/Cooling System Unit*

teacher's discretion to meet the needs of their particular students. However, it is important to note that the activities the students engage in throughout the *Learning Cycle* are designed to begin by eliciting the ideas and understanding currently held by students,

The types of activities included in the *Heating/Cooling System* unit range from requiring individual, group or whole class effort to complete. Examples of individual activities include: analyzing data, graphing experiment results, and responding to reflection prompts. Many of the individual activities also include a component that requires students to share their responses with their peers, either within their own group or with the whole class.

The majority of activities in the unit were designed to require group-work. Examples of group activities include: brainstorming ideas; modifying and testing prototypes; experimenting with chemicals and/or materials; and presenting data, results or prototype designs. The unit provided numerous opportunities for whole class work, mostly in the form of discussions where ideas are generated or shared, and broader scientific concepts discussed. Table 1 provides a characterization of the types of activities (e.g. individual, group or whole class) used most frequently during each phase of the *Learning Cycle*.

Participants

Teachers

As part of an effort to reform teaching and learning in high school science classrooms, teachers from a large urban area in the northeastern United States were invited to implement a design-based learning unit in their classrooms. This

implementation included professional development (20 h) around basic pedagogy (e.g. facilitating group work; leading class discussions) and unit particulars (e.g. facilitating specific lessons). The professional development was distributed across time (five sessions of 4 h each) to correspond with unit implementation. A total of seven teachers participated in the professional development sessions and implemented the unit, however only five teachers participated in the study.² These five teachers came from four different schools within two different school districts. All the participating teachers taught and implemented the unit in at least two classes or sections of chemistry (See Table 2). There were four female teachers and one male teacher, and their teaching experience ranged from 6 to 17 years. Of the five participating teachers, three teachers had prior experience with design-based learning, and had implemented a previous version of the unit.

Students

Students who were enrolled in the participating teachers classes were considered eligible for participation in this study. The student participants were drawn from urban high schools with a moderate proportion of traditionally underserved students (See Table 3). A total of 320 high school students participated in the design-based learning unit. Participants included 9th, 10th, 11th and 12th grade high school students enrolled in spectrum science, general chemistry, or advanced general chemistry courses. Inclusion in advanced courses was based on prior school achievement benchmarks determined by the school district. Across all classroom types, this course was the first year of chemistry for all students. As is common when working in urban schools, where attendance rates are typically well below 100%, some students were absent on the days the pre- or post-unit assessments were administered. Accounting for absences, a total of 271 students completed the unit and both the pre- and post-unit assessments. Attrition rates did not differ significantly across group size ($\chi^2 = 0.42$) or class type ($\chi^2 = 0.66$).

Measures

To address whether group size influenced chemistry content learning, students completed pre- and post-unit pencil-and-paper assessments. The assessments were relatively short (24 questions) because longer tests are difficult to embed in classrooms given all the other testing that now takes place in public schools in the USA. The assessment comprised 24 questions taken from the Chemical Concept Inventory (CCI; Mulford 1996) and the American

² Two teachers were unable to complete the unit due to other school schedule demands.

Table 1 Characterization of activities in the heating/cooling system unit

Learning cycle phase	Description of phase	Type of activity
Create design	Develop a design based on prior knowledge or experiences	Group
Evaluate outcome	Determine whether design works or does not work, based on specified requirements	Group
Generate reasons	Brainstorm all the possible reasons why design did or did not work	Individual
		Group
		Class
Test ideas	Systematically test one of the reasons (ideas) brainstormed in the <i>Generate Reasons</i> node	Group
Analyze results	Analyze the collected data to determine whether the results do or do not support the idea tested during <i>Test Ideas</i>	Individual
		Group
Generalize results	Determine whether there are patterns or trends in the data	Class (sharing group results)
Connect to big idea	Connect the patterns or trends to a scientific theory (idea). This scientific idea should be used to improve the performance of the original design	Class

Table 2 Teacher participants' demographic information

Teacher	Sex	Years teaching	Level of course (number of sections)	Prior experience with design-based learning?
1	F	7	Mainstream (2)	No
2	F	6	Mainstream (2)	Yes
			Advanced (1)	
3	F	10	Mainstream (2)	No
			Advanced (3)	
4	F	7	Mainstream (3)	Yes
5	M	17	Mainstream (2)	Yes
			Advanced (2)	

Table 3 Student participants' demographic information

Sex ^a (%)	Grade ^a (%)	Under-represented minority ^a (%)	Lunch subsidy ^b (%)	English language learner ^a (%)	Gifted ^a (%)
Female = 54	9th = 16	37	30	0.3	7
Male = 45	10th = 23				
	11th = 55				
	12th = 5				

^a Information unavailable for 0.3% of participants

^b Information unavailable for 19% of participants

Chemical Society's test item bank for high school chemistry (ACS; Eubanks and Eubanks 1993). The questions on the pre- and post-unit assessments were identical (See Appendix 1 for sample items).

The CCI is a multiple choice assessment instrument that contains non-mathematical conceptual questions that tap into student's understanding of chemical ideas. Many of the questions include representation at the macroscopic, microscopic (particulate) and symbolic level. Questions from the CCI were selected because they assess reasoning skills and common naïve conceptions. However, the CCI includes many difficult topics. Students who have completed one semester of college level chemistry had an average score of 50% on the CCI (Division of Chemical

Education 2008). Thus, it was expected that these questions would be difficult for the high school students and that analyses might be difficult if scores were low enough to encounter floor effects. However, project-based learning is supposed to target deeper conceptual learning and thus it was important for the study to include such questions.

Additionally, factual multiple-choice questions were selected from the ACS test bank based on their relevance to the *Heating/Cooling System* unit. The ACS Test-Item Bank was developed to provide high school instructors with assessment items for first-year chemistry courses, thus it was expected that the selected questions would be at a difficulty level more commonly assessed with high school students.

Item analyses for the assessment questions were conducted using Item Response Theory (Baker 2001). These analyses indicated that six of the items from the assessment had substantially skewed two-item characteristic curves during the post-unit assessment. These six items did not show the characteristic sigmoid response function (i.e., the response curve was flat), had high negative or positive difficulty scores (i.e., $> \pm 3$), and had either negative discrimination scores or low positive discrimination scores $< .30$. Therefore, data from these six items were excluded from the analyses reported here. However, the patterns reported here held for the full 24-item test as well. The remaining 18-item set had an average difficulty of 1.28 (SD = 1.67) and an average discrimination score of .87 (SD = .34).

The assessment questions were categorized by degree of transfer (i.e. none, close, or far). It was hoped that use of assessment items that differed in the degree of transfer required to answer the questions would allow us to better distinguish the depth and level of learning acquired by students in different classes (mainstream vs. advanced) as well as different group size (pairs vs. fours). Questions that addressed content directly taught in the unit were categorized as none. Close transfer questions were not necessarily directly taught in the unit, but were closely related to the unit content. For example, during the unit, students learn that chemical transformations involve changes in attractions (bonding) between particles, and that the energy released or absorbed during chemical transformations is dependent on the shape and structure of the particles involved in the transformation. Thus, an example of a close transfer assessment item is the following: “When gasoline is burned with oxygen in a car engine, energy is released. What is the source of this energy?” While combustion reactions are not directly addressed in the unit, the answer to this question (“The chemical bonds in the gasoline molecules”) is one that students could derive through reasoning and using their knowledge about bonding and chemical transformations. Far transfer questions were not taught in the unit and were not related closely to the unit content. However, through reasoning from taught material, the questions could be solved by the students. For example, students were asked “Which of the following must be the same before and after a chemical reaction?” The answer, (“The sum of the masses of all substances involved, and the number of atoms of each type involved”) deals with the law of conservation of mass/matter, which was not explicitly taught in the unit. The final 18-item test contained eight questions from the none-transfer, six questions from the close-transfer, and four questions from the far transfer category.

The internal consistency of the remaining 18-item test was moderate, with Chronbach’s alphas of 0.47 for the pretest and 0.58 for the post-test. Several factors significantly impacted the alpha levels. First, floor effects lowered the pretest alphas. It is reasonable to expect floor effects at pretest because the

students had not been exposed to this information at the time of testing. Second, the far transfer items, for which we had low accuracy rates during the pretest and posttest, reduced the overall alphas. For example, during the posttest, the alphas were 0.40, 0.40, and 0.23 for the none, close, and far transfer items. Finally, a test including a variety of transfer categories and chemistry topics is expected to produce moderate alphas because multiple chemistry concepts are being tested. Thus, the moderate alphas seen here are acceptable (Kline 2000).

Procedure

Prior to the start of the *Heating/Cooling System* unit, teachers administered the pretest to the students in their classes. Teachers completed the unit in their classrooms over the next 6–8 weeks. Following unit completion, teachers administered the posttest in their classroom and delivered the completed forms to the researchers for scoring.

Results and Discussion

Condition Effects on Gains

Students were nested within teacher groups. Not all teachers had students of both classroom types. Following standardized procedures, preliminary analyses using an unconditional multilevel linear model showed that there was significant heterogeneity *within* teacher groups ($Wald Z = 11.54, p < .001$), but not significant heterogeneity *between* teacher groups ($Wald Z = 1.32, p = .19$) indicating that there were not systematic differences between teacher groups and that ANOVA and ANCOVA are appropriate for subsequent analyses to assess conditional variance within teacher groups (Field 2009).

Following standard procedures for analyzing multiple effects and interactions, (Field 2009), data from the 271 students on the remaining 18-item pretest were subjected to a 2×2 between factors analysis of variance (ANOVA) with group size (two vs. three/four³) and class type

³ Including group sizes of threes and fours into one group was selected here because this allowed for the most equal distribution of numbers of participants across ability groups. However, we analyzed the data in other ways and found a consistent pattern of data compared to that reported in the manuscript. More specifically, when analyzing the three group sizes separately, we continue to find that the covariate (i.e., pretest) is a significant predictor of post-test scores, a significant main effect of class type and a class type \times group size interaction. Basically, these analyses revealed that that group sizes of threes and fours did not differ from each other in the advanced or mainstream groups. This pattern of results remained consistent when the data were analyzed without the threes group size.

(mainstream vs. advanced) as between subject factors. To account for individual differences at the pre-test, posttest data were subjected to a 2×2 analysis of covariance (ANCOVA) with group size and class type as between-subjects factors and pretest as a covariate. Cohen's d (1988) was used to calculate effect sizes for adjusted means as it is appropriate for comparing two means (Field 2009). Spectrum science and general chemistry courses were classified as mainstream and advanced general chemistry was classified as advanced. On average, students performed close to chance on the pretest, but still with significant variance across students ($M = .30$, $SD = .14$).

Although students from the advanced general chemistry courses had higher scores on the pretest, $F(1, 267) = 16.01$, $p < .0001$, $d = .57$, differences between the two group sizes were not significant, $F(1, 267) = 3.19$, $p > .05$, $d = .39$. Overall, pretest scores significantly co-varied with posttest scores, $F(1, 266) = 104.03$, $p < .0001$. Even after adjusting for the pretest differences, there was an effect for class type, $F(1, 266) = 42.93$, $p < .0001$, $d = .78$ (see Table 4). Students in advanced classrooms had higher posttest scores than students in mainstream classrooms (see Appendix 2 for scatter plots of pretest vs posttest scores for advanced and mainstream classrooms). There was a small marginal group size \times class type interaction, $F(1, 266) = 2.39$, $p = 0.12$, which will be unpacked further as it involves a further interaction with degree of transfer.⁴

Additional analyses were conducted for each transfer category to investigate whether group size and/or class type influenced performance on these subsets of questions. The pretest scores were significant covariates for each of the transfer categories. As one might expect, the far transfer item had the weakest performance and the no transfer items the best performance. In addition, advanced class students outperformed mainstream students in all three transfer categories (none, close, and far), $F_{\text{none}}(1, 266) = 34.57$, $p < .0001$, $d = .75$; $F_{\text{close}}(1, 266) = 42.01$, $p < .0001$, $d = .75$; $F_{\text{far}}(1, 266) = 7.64$, $p < .01$, $d = .37$ (see Table 5). In only the close and far transfer cases for the mainstream classes is there some concern of floor effects that may mask the effects of group size.

Although there was no overall effect of group size, $F(1, 266) < 1$, there was an effect of group size in some

Table 4 Adjusted mean posttest scores by class type and group size (SD in parentheses)

	Mean proportion correct overall (SD)
Class type	
Advanced ($n = 101$)	.42 (.12)
Mainstream ($n = 170$)	.31 (.16)
Group size	
Pairs ($n = 88$)	.37 (.12)
Threes/fours ($n = 183$)	.36 (.14)
Class type \times group size	
Mainstream—Pairs ($n = 34$)	.30 (.12)
Mainstream—threes/fours ($n = 136$)	.31 (.12)
Advanced—pairs ($n = 54$)	.44 (.12)
Advanced—threes/fours ($n = 47$)	.40 (.12)

Table 5 Adjusted mean posttest scores on transfer items by class type (SD in parentheses)

	Mean proportion correct		
	None (SD)	Close (SD)	Far (SD)
Advanced ($n = 101$)	.50 (.18)	.40 (.20)	.31 (.22)
Mainstream ($n = 170$)	.36 (.22)	.23 (.25)	.22 (.27)

contexts. The no transfer category and far transfer category showed no significant group size \times class type interactions, $F_{\text{none}}(1, 266) = 2.39$, $p = 0.12$; $F_{\text{far}}(1, 266) = 1.92$, $p = 0.17$. However, the close transfer category contained an additional group size \times class type interaction, $F_{\text{close}}(1, 266) = 9.27$, $p < 0.01$, with the differences between pairs and threes/fours significant only for the advanced classrooms ($M_{\text{pairs}} = .46$, $M_{\text{threes/fours}} = .34$, $d = .63$; see Fig. 2). In contrast, the opposite was true for students in mainstream classrooms; students from threes/fours performed slightly better than students in pairs ($M_{\text{pairs}} = .21$, $M_{\text{threes/fours}} = .24$, $d = .15$). However, this difference was not significant. These results indicate that students from advanced courses that are grouped in pairs are best able to transfer their knowledge to close transfer problems. Conversely, differences between group sizes were not important for mainstream classrooms. In sum, group size appears to only matter for the advanced classes on close transfer problems.

Thus, with regards to the question of *whom* (pairs or three/fours), exhibits better group learning, the results show that when ignoring classroom setting and collapsing across type of learning, there was no overall difference in chemistry content learning outcomes based on group size. This

⁴ In addition, we analyzed the results as a repeated-measures ANOVA with pretest and posttest scores as repeated measures and as an ANOVA with differences scores. The results were consistent across the different types of analyses. We've reported the ANCOVA here because it was the most appropriate analyses in this setting because it accounts for the differences between the advanced and mainstream groups at pretest (Field, 2009).

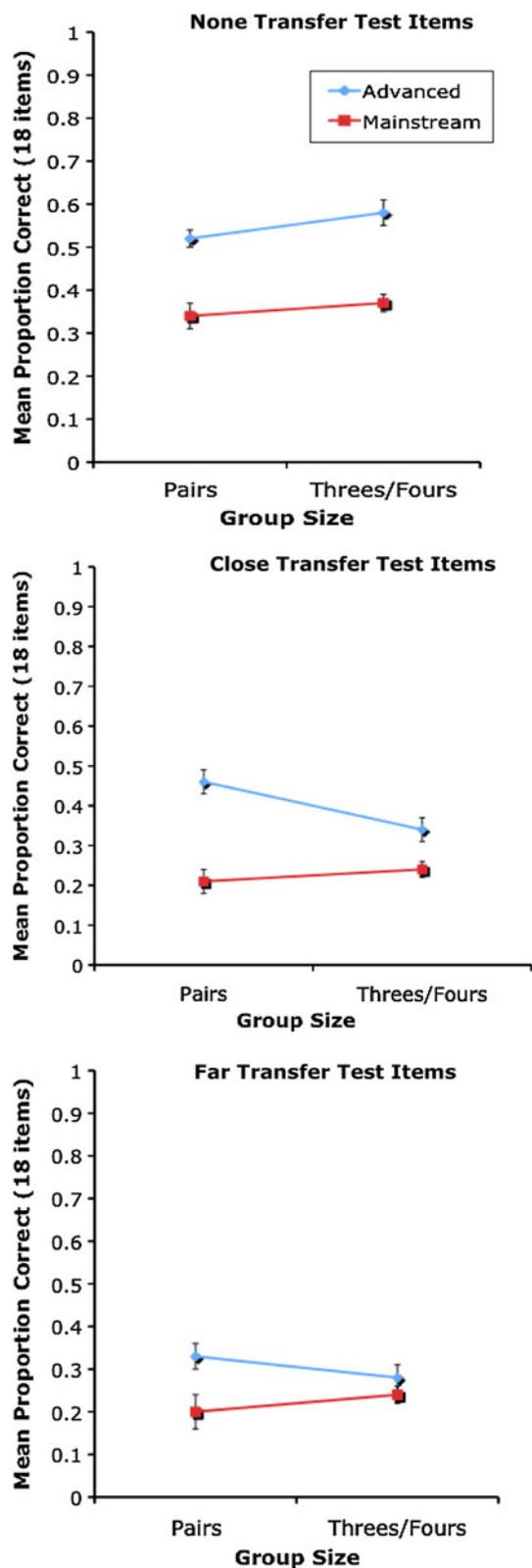


Fig. 2 Adjusted mean scores for the none, close, and far transfer test items by group size and class type

finding is contrary to prior research (e.g. Wiley and Jensen 2006) that suggests triads are better than pairs and individuals. Furthermore, there was no interaction between group size and class type.

But, when considering the *type of learning* (basic vs. advanced knowledge) we found that the *under what circumstances* question was important to ask. Specifically, there were differences in performance on a subset of test items for students in advanced classrooms. For the no transfer items, we did not necessarily expect a variation in performance, because students could likely obtain superficial learning results through simple observation in the group. A sufficient degree of learning is required for effective transfer to occur (Bransford and Schwartz 1999). As such, we expected that there would be variation in student performance on the close transfer test items, although perhaps not in the far transfer test items because students were not given additional learning resources along with the transfer items. In fact, close transfer items did show some effects of group size, with pairs within advanced classrooms outperforming groups of three/fours. In contrast, it appears that group size was not an important factor for individual student learning in mainstream classrooms. If anything, these results suggest that larger groups in mainstream classrooms learned more, instead of less, than smaller groups on the close transfer items.

Learning Inequalities

Many factors can produce heterogeneity of learning overall, including factors that do not vary with group size (such as variations in student abilities and prior knowledge) and factors that might vary with group size (such as unequal task sharing or sharing of tasks with differential learning opportunities). To explore the learning inequalities within groups, we categorized students as “minimal-learners” or “strong learners” within their group based on the gain scores from pretest to posttest on the chemistry assessment. If a student’s gain score was in the bottom quartile, they were classified as a minimal-learner. If a student had a gain score in the top quartile they were classified as a strong learner student. From the full sample of students, 29% were classified as minimal-learners, while 26% of students were classified as strong learners. Within advanced classrooms, 20% were classified as minimal-learners and 37% were classified as strong learners, whereas mainstream classrooms were 34% and 19% for minimal-learners and strong learners, respectively. Groups were identified as being *inequitable learning groups* if one student in the group was

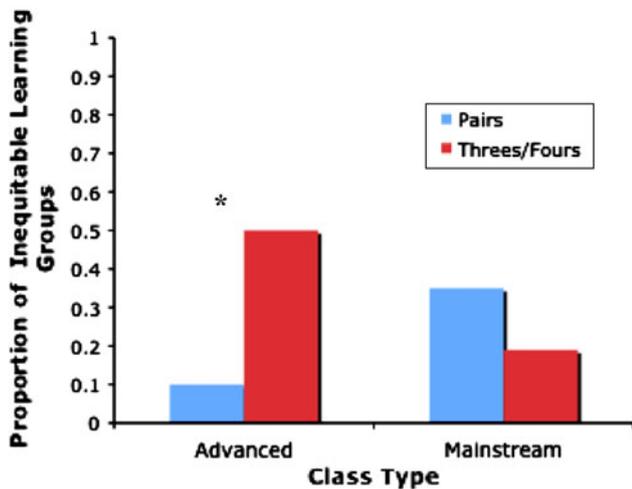


Fig. 3 The proportion of inequitable learning groups by group size and class type

characterized as a minimal-learner and another student in the group was characterized as a strong-learner. This method likely over-estimates the proportion of inequitable learning groups due to group effects, as there will be a number of minimal-learning/learning group combinations that simply reflect individual differences in student prior knowledge and ability—however, we can ignore this overestimation because we are only interested in relative differences across conditions. Figure 3 illustrates the proportion of groups that demonstrated learning inequalities.

In advanced classrooms, 46% of three/four groups and only 10% of pairs were characterized as inequitable learning groups; a difference that was statistically significant ($p = 0.01$) using a two-tailed Fisher's exact test. Fisher's exact test is appropriate here because it assesses whether the frequencies of two small data sets (i.e., the number of teams) are different (Field 2009). In contrast, in mainstream classrooms only 16% of three/four groups and 31% of pairs were inequitable learning groups ($p = 0.28$, two-tailed Fisher's exact test).

These results suggest that students in advanced classrooms may have difficulty sharing the intellectual workload in larger groups, but are able to do so within pairs, potentially explaining why pairs learned more overall than three/fours in the advanced classrooms. Students in these advanced classrooms may have preferred to learn individually, and thus had difficulty managing their learning in a larger group. Alternatively, the phenomenon of social loafing, wherein there is a decrease in motivation or effort when individuals work collectively as opposed to working individually (Karau and Williams 1993), may have played a role.

In contrast, the results of the learning inequalities analyses suggest that knowledge gains for students in mainstream classrooms is not dependent on group size, but that larger groups less frequently demonstrate learning inequalities within the group. Perhaps mainstream students, unlike the students in advanced classrooms, were able to distribute the intellectual workload more effectively and may have benefited from the dynamics of larger group sizes, including a greater diversity of perspectives.

Teacher Impressions

In a post-unit implementation survey, participating teachers were asked the following question: "If you used varying group sizes in your class (e.g. 2 s vs. 4 s) then what were your impressions?" Responses from the five participating teachers were overwhelmingly in favor of smaller groups (e.g. pairs or threes) as opposed to larger groups (e.g. fours). Groups of fours were viewed as being unproductive, as evidenced by one teacher's statement: "Twos groups worked better than 4 s groups. There is always somebody who has nothing to do or takes the free ride in 4 s groups. Too many free bodies in class." However, for one of the teachers, the practical issue of poor attendance overshadowed the potential benefit of using smaller groups:

I really think that if there wasn't such an attendance problem that my students would have gotten more out of the unit. Especially when they are only in pairs and sometimes neither partner was in the group. I had to rearrange some of the groups during the unit.

The remaining two teachers agreed that attendance could be an issue when using pairs, and suggested that groups of three students are ideal: "Groups of 3 worked best. Four was too many, and 2 of the students of the 4 didn't work. Groups of 2 worked OK except for absentee partners."

Conclusion and Recommendations

While there are a number of factors that instructors need to consider when assigning students to collaborative learning groups, the size of the group may be one of the most important. The size of collaborative learning groups used in classrooms should not be based solely on practical considerations (e.g. instructor support; available learning resources) but should also be informed by lessons learned

from research. While previous research has not provided a consensus as to the optimal group size for individual student learning, the results of our study provides insight into the question of *whom* (pairs or groups of three or four) will exhibit better group learning, *under what circumstances* (advanced vs. mainstream classrooms) and for *what kind of learning* (basic vs. more advanced knowledge).

It is important to note some of the limitations of our study, one being the use of knowledge gains as the only measure to determine optimal group size. While content knowledge gains are important to consider, understanding the gains in other cognitive skills such as problem-solving or group processing skills would also be valuable. Future research might examine the influence of group size on some of these more general cognitive skills.

Another limitation of our study is that there were fewer students from advanced classes than mainstream classes. This sample is representative of the proportion of students in mainstream and advanced classes at the schools involved in this first study exploring the influence of group size and ability grouping on learning. As the first study of its kind, the results are important, but additional evidence needs to be collected before strong conclusions can be made.

Similar to the previous point, future work should explore more directly whether differences exist between group sizes of three or four. Further analyses on this data indicate that there were not significant differences between group sizes of threes and fours (see footnote 3), but it would be interesting to explore this idea more directly.

Despite these limitations, the findings from our study is the first to suggest that group size cannot be considered in isolation, when thinking about how to structure collaborative learning groups. Rather, the class type (e.g. mainstream vs. advanced) and the type of knowledge desired (basic vs. advanced) must also be considered when determining the optimal size group for student learning. If basic knowledge (i.e. requiring no transfer) is the goal, the size of the group is likely not important. However, if more advanced knowledge (i.e. requiring close transfer) is

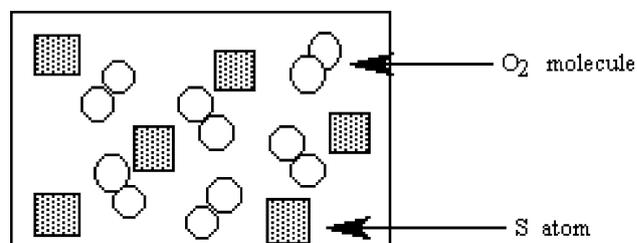
desired, then students in mainstream classrooms may benefit most from groups of three/fours, while students in advanced classrooms are best served by working in pairs. Future research with advanced level classrooms might include an individual learning condition to be contrasted with groups of various sizes. In addition, future research might examine whether these findings hold true for different chemistry content, and in different science disciplines.

In conclusion, it is important for researchers to conduct studies that include theoretical implications *and* address practical concerns. As the use of project-based learning approaches continues to increase in K-12 science classrooms, there are a number of practical concerns that will be raised by teachers. Understanding what factors are important to consider when assigning students to collaborative groups is just one of those issues. Our study provides evidence that, given particular conditions (e.g. advanced classrooms), group size (a seemingly superficial feature) can differentially influence the depth and level of student learning.

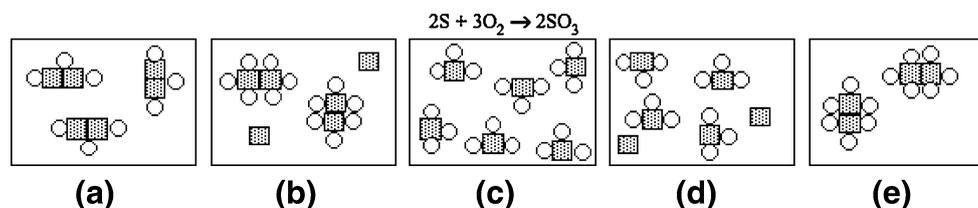
Appendix 1

Chemical Concept Inventory Sample Item 1

The diagram represents a mixture of S atoms and O₂ molecules in a closed container.



Which diagram shows the results after the mixture reacts as completely as possible according to the equation:



Chemical Concept Inventory Sample Item 2

100 ml of water at 25°C and 100 ml of alcohol at 25°C are both heated at the same rate under identical conditions. After 3 min the temperature of the alcohol is 50°C. Two minutes later the temperature of the water is 50°C. Which liquid received more heat as it warmed to 50°C?

- a. The water.
- b. The alcohol.
- c. Both received the same amount of heat.
- d. It is impossible to tell from the information given.

What is the reason for your answer to the question above?

- a. Water has a higher boiling point than the alcohol.
- b. Water takes longer to change its temperature than the alcohol.
- c. Both increased their temperatures 25°C.
- d. Alcohol has a lower density and vapor pressure.
- e. Alcohol has a higher specific heat so it heats faster.

ACS Sample Item 1

In any chemical reaction energy is

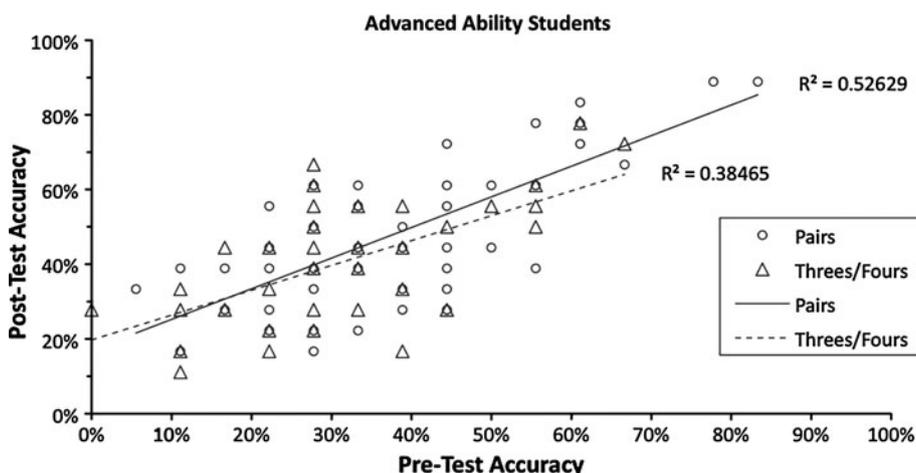
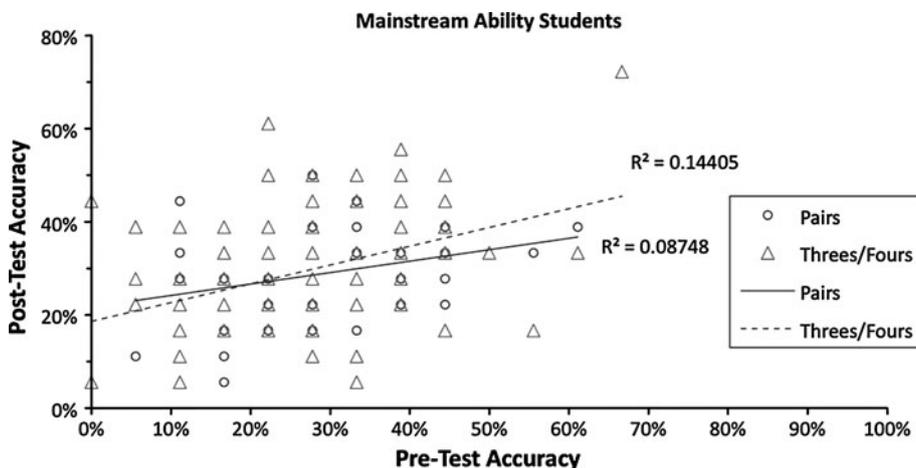
- a. Always absorbed.
- b. Always released.
- c. Released as heat.
- d. Usually not involved.
- e. Either absorbed or released.

ACS Sample Item 2

When KCl dissolves in distilled water, which of the following, other than water molecules, are in the solution?

- a. KCl molecules.
- b. K and Cl atoms.
- c. K⁺ and Cl⁻ ions.
- d. KCL⁻ ions.

Appendix 2



References

- Apedoe XS, Reynolds B, Ellefson MR, Schunn CD (2008) Bringing engineering design into high school science classrooms: the heating/cooling unit. *J Sci Educ Tech* 17(5):454–465
- Baker FB (2001) The basics of item response theory, 2nd edn. Eric Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD
- Barak M, Dori YJ (2004) Enhancing undergraduate students' chemistry understanding through project-based learning in an IT environment. *Sci Educ* 89:117–139
- Barron BJS, Schwartz DL, Vye NJ, Moore A, Petrosino A, Zech L, Bransford JD (1998) Doing with understanding: lessons from research on problem- and project-based learning. *J Learn Sci* 7:271–311
- Bransford JD, Schwartz DL (1999) Rethinking transfer: a simple proposal with multiple implications. *Rev Res Educ* 24:61–100
- Brophy SP, Bransford JD (2001). Design methods for instructional modules in bioengineering. In: Proceedings of the 2001 American Society of Engineering Education. Retrieved from <http://vanth.org/Publications.html>
- Cognition and Technology Group at Vanderbilt (1994) From visual word problems to learning communities: changing conceptions of cognitive research. In: McGilly K (ed) *Classroom lessons: integrating cognitive theory and classroom practice*. MIT Press/Bradford, Cambridge, pp 157–200
- Cognition and Technology Group at Vanderbilt (1997) The Jasper project: lessons in curriculum, instruction, assessment, and professional development. Lawrence Erlbaum, Mahwah
- Cohen J (1988) *Statistical power analysis for the behavioral sciences*, 2nd edn. Lawrence Erlbaum Associates, Hillsdale
- Collins A, Hawkins J, Carver SM (1991) A cognitive apprenticeship for disadvantaged students. In: Means B, Chelemer C, Knapp MS (eds) *Teaching advanced skills to at-risk students*. Jossey-Bass, San Francisco, pp 216–243
- Division of Chemical Education, American Chemical Society (2008) Conceptual questions: chemical concept inventory. Retrieved from Journal of Chemical Education website: <http://jchemed.chem.wisc.edu/JCEDLib/QBank/collection/CQandChP/CQs/ConceptsInventory/CCIIntro.html>
- Dugosh KL, Paulus PB, Roland EJ, Yang H (2000) Cognitive stimulation in brainstorming. *J Pers Soc Psychol* 79:722–735
- Eubanks ID, Eubanks LP (1993) ACS test—item bank for high school chemistry. American Chemical Society Division of Chemical Education Examinations Institute, Milwaukee
- Field A (2009) *Discovering statistics using SPSS*, 3rd edn. Sage, London
- Fuchs LS, Fuchs D, Kazdan S, Karns K, Calhoon MB, Hamlett CL, Hewlett S (2000) Effects of workgroup structure and size on student productivity during collaborative work on complex tasks. *Elementary School J* 100:183–212
- Gabel D (1999) Improving teaching and learning through chemistry education research: a look to the future. *J Chem Educ* 76:548–554
- Karau SJ, Williams KD (1993) Social loafing: a meta-analytic review and theoretical integration. *J Pers Soc Psychol* 65:681–706
- Kline P (2000) *The handbook of psychological testing*, 2nd edn. Routledge, London
- Krajcik JS, Blumenfeld PC, Marx RW, Soloway E (1994) A collaborative model for helping middle grade science teachers learn project-based instruction. *Elementary School J* 94:483–497
- Levine JM, Moreland RL (2004) Collaboration: the social context of theory development. *Pers Soc Psychol Rev* 8:164–172
- Lohman MC, Finkelstein M (2000) Designing groups in problem-based learning to promote problem-solving skill and self-directedness. *Instr Sci* 28:291–307
- Lou Y, Abrami P, Spence JC, Poulsen C, Chambers B, d'Apollonia S (1996) Within-class grouping: a meta-analysis. *Rev Educ Res* 66:423–458
- Marx RW, Blumenfeld PC, Krajcik JS, Soloway E (1997) Enacting project-based science. *Elementary School J* 97:341–358
- Mulford DS (1996) An inventory for measuring college students' level of misconceptions in first semester chemistry. Unpublished master's thesis, Purdue University
- Schank P, Kozma R (2002) Learning chemistry through the use of a representation-based knowledge building environment. *J Comput Math Sci Teach* 21:253–279
- Schauble L, Glaser R, Duschl RA, Schulze S, John J (1995) Students' understandings of the objectives and procedures of experimentation in the science classroom. *J Learn Sciences* 4:131–166
- Webb NM (1980) A process-outcome analysis of learning in group and individual settings. *Educ Psychol* 15:69–83
- Webb NM (1982a) Group composition, group interaction, and achievement in cooperative small groups. *J Educ Psychol* 74:475–484
- Webb NM (1982b) Peer interaction and learning in cooperative small groups. *J Educ Psychol* 74:642–655
- Webb NM, Palincsar AS (1996) Group processes in the classroom. In: Berliner DC, Calfee RC (eds) *Handbook of educational psychology*. Prentice Hall, New York, pp 841–873
- Wiley J, Jensen M (2006). When three heads are better than two. In Sun R (ed) *The Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum, Mahwah, pp 2375–2380