Understanding the Effect of Group Variance on Learning

Sinan Yavuz, Michael Forkner, Dana Gnesdilow, Catherine L. Dornfeld Tissenbaum, Jee-Seon Kim, Sadhana Puntambekar
syavuz@wisc.edu, forkner@wisc.edu, gnesdilow@wisc.edu, catdoti@illinois.edu, jeeseonkim@wisc.edu, puntambekar@education.wisc.edu
Department of Educational Psychology, University of Wisconsin-Madison

Abstract: Given that group composition is a key factor that affects learning in CSCL environments, it is important to study how students in groups with homogeneous or heterogeneous levels of prior knowledge collaborate. This study investigated the potential differences in students' learning outcomes from participating in a 13-week design-based unit. We used the pre- and post-test data from 361 eighth grade students (102 groups) and performed a hierarchical linear model analysis to examine how the convergence or divergence in the students' level of prior knowledge affected students' learning outcomes. We found that students in homogenous groups with similar levels of initial prior knowledge scored significantly higher on their post-test when their pre-test was used as a covariate, than students in heterogeneous groups. Implications of these findings are discussed along with directions for future research.

Introduction

Group composition is a key factor affecting learning in CSCL (Puntambekar & Young, 2003). Often, group members come with different levels of prior knowledge, bringing convergence or divergence of ideas (Weinberger, Stegmann, Fischer, 2007). It is therefore important to study how student groups with homogeneous or heterogeneous levels of prior knowledge collaborate. Vygotsky's (1978) Zone of Proximal Development (ZPD) framework assumes that there is a "more capable other" who can scaffold a learner to accomplish more with assistance than alone. The ZPD therefore implies that for peers to support and scaffold learning in group interactions, there has to be an inherent asymmetry in the group's knowledge. Stahl (2004) also suggested that divergent ideas are an essential mechanism for the exploration of ideas and negotiation of knowledge during group collaborations.

However, in classrooms, groups are often composed of students with more homogeneous than heterogeneous levels of prior knowledge. It may also be the case that heterogeneous groups mimic homogeneous groups, as the more capable other does not actually provide the proper support to the other members. Because of the homogeneous nature of these groups, multiple students may collectively scaffold each other, rather than only the more capable peer providing the scaffolding. For example, Fernández, Wegerif, Mercer and Rojas-Drummond (2001) found that students' dialogue during collaboration in symmetrical (i.e., homogeneous) groups provided enough support to help students solve problems, and thus argued for a reconceptualization of the relationship described by Vygotsky's ZPD. Further, other researchers have claimed that peers may not intentionally try to scaffold each other; but, by working together, peers can solve a problem or complete a task that they could not accomplish when working alone (Wells, 1999; Zuckerman, 2003).

Prior research that examined homogeneous and heterogeneous group composition identified learning benefits when students were placed in heterogeneous groups (Csanadi, Kollar, & Fischer, 2016; Gijilers & De Jong, 2005; Patchan, Hawk, Stevens, & Schunn, 2013; Wiedmann, Leach, Rummel, & Wiley, 2012; Zhao et al., 2018). These benefits may be due to the processes that occur as students work together during collaborative knowledge-building activities, as peers with greater knowledge bring important issues and resources, while peers with less knowledge play an important role by raising questions and asking for clarifications, which the peers with greater knowledge may address (Scardamalia & Bereiter, 1994).

Yet other researchers have found that, in some cases, homogeneous group composition is preferable. Webb, Nemer, and Zuniga (2002) found that high-ability students performed better in homogeneous groups, whereas low-ability students performed better when they had a more capable other in their heterogeneous group. Along these lines, Lou et al. (1996) found that high-ability students benefited equally from both homogeneous and heterogeneous groups, while medium-ability students benefited most from homogeneous groups, and low-ability students benefited most from heterogeneous groups. Other research has identified that same-ability dyads were better at metacognitively regulating their collaborative process to reach their goals than their heterogeneous partners (Zillmer & Kuhn, 2018).

To further understand the relationship between groups with similar or different levels of prior knowledge, our study examined 102 groups of middle-school students learning science over 13 weeks. Students worked with the same group as they engaged in CSCL tasks each school day, which gave us the unique opportunity to examine how group composition, based on levels of prior knowledge, affected students' conceptual learning outcomes, using a larger sample over a longer time frame than many prior studies.

Methods

Participants and instructional context

This study took place during the 2016-2017 academic year, with seven science teachers and their 515 eighth-grade students (229 female and 286 male). All students and teachers were from one of three middle schools in the same urban school district in the U.S. Midwest. This district served about 2,066 middle school students, with about 53% of them identified as being economically disadvantaged. Students in all classes participated in a design-based unit called "Make Your Own Compost!". The curriculum challenged students to create a compost that would break down quickly and contain nutrients while minimizing landfill waste and other negative effects of conventional fertilizers on the environment. Students collaborated in the same group of three to six students (mean group size is 3.54) to learn about ecosystems, energy transformations, matter cycling, and human impacts over the entire 13 weeks of the unit. Students participated in a variety of science activities, such as experiments, and worked collaboratively using computers to conduct research using an online digital text. They also ran multiple compost simulations to help them to build the necessary knowledge to solve the challenge over the course of the unit. All activities in the unit were designed to help students to solve the challenge and write a final report to their principal to propose their design of a composting program for the school. For this study, we examined the pre- and post-tests from 150 groups of students. However, due to missing data and varying group sizes, we only included groups for which we had both pre- and post-test information from at least three students in a group. The results of this study are based on data from 102 groups (361 students), who were assigned by the teacher. Each student took the pre-test prior to being introduced to unit content and activities. Students took the post-test after finishing the unit. Groups were categorized as homogeneous, medium, and heterogeneous based on their pre-test score variances.

Data sources and analysis

Pre- and post-test measures

The "Make Your Own Compost!" unit focused on helping students to build science understanding about ecosystems and humans' impact on them. The test was designed by the research team and consisted of 24 questions that assessed students' understanding of concepts and relationships related to biotic and abiotic factors in ecosystems; organisms' roles and relationships in ecosystems; the flow of energy and cycling of matter in ecosystems; and human impacts on ecosystems. Four of the 24 questions were open-ended items, and 20 questions were in a multiple choice (MC) format. Three of the 24 questions had multiple parts. Overall test reliability was calculated using Cronbach's alpha. We found the test to be reliable with alpha values of .863 for the pre-test and .874 for the post-test. After analysis of the open-ended items, a conflict was determined between the pre- and post-test scores due to incomplete responses on the post-test. Therefore, we only included MC items in our analyses. The maximum score students could earn on the MC items was 23.

Hierarchical Linear Models

In this study, students were nested in groups, and we could not assume that students' learning gains were independent, as the intervention was applied in a group setting. Thus, students' learning gains within the cluster were expected to be correlated, and the dependency between students needed to be considered (Kim, Anderson, & Keller, 2013). Therefore, using classical statistical methods, such as linear regression or ANOVA on the student-level data, while ignoring the group clustering effect, would lead to inaccurate results and interpretation. By using hierarchical linear models (HLM), we could represent each level by its own submodel (Raudenbush & Bryk, 2002) to consider the multilevel nature of the data. We performed all the analyses using R software (R Core Team, 2017) and used the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) for HLM and *ggplot2* for producing graphs (Wickham, 2009). We considered student level as level-1 and group level as level-2.

Results

Since we were interested in investigating group variability, we only analyzed groups with three to six students who completed both the pre- and post-test. After eliminating groups that did not fit our criteria, we were left with 361 students who were nested in 102 groups. As mentioned before, we only analyzed the students' total score for the MC questions. Descriptive statistics of students' pre- and post-test item scores are shown in Table 1.

Table 1: Level-1 Variables (N = 361 Students)

Variable	N	Mean	Std Dev	Median	Max. Poss.
Pre-Test Scores	361	15.05	4.43	15.5	23
Post-Test Scores	361	19.46	3.82	21	23

We divided the student groups into three group types based on their pre-test score variance, which ranged from 0.25 to 72.58. Thirty-three percent of the groups with the lowest variance were classified as homogenous; the middle 34% were classified as medium; and 33% of the groups with the highest variance were classified as heterogenous. The variance of within-group prior knowledge distribution is shown in Figure 1. The descriptive statistics of the groups are shown in Table 2.

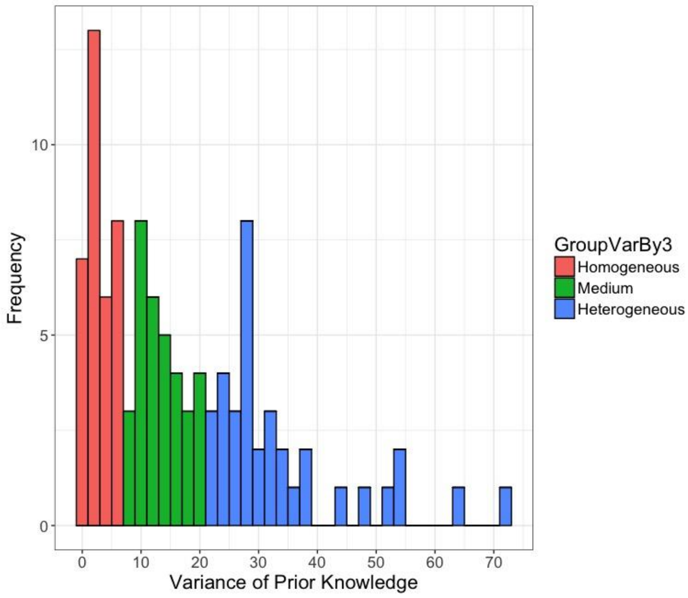

Figure 1. The histogram of within group prior knowledge variance.

Table 2: Level-2 Variables (N = 102 Groups)

Variable	Levels	N	Mean	Std Dev	Median
Pre-Test Scores	Homogeneous	34	15.21	3.31	15.19
	Medium	33	15.82	2.70	16.33
	Heterogeneous	35	14.07	2.08	14.42
Post-Test Scores	Homogeneous	34	19.97	2.53	20.35
	Medium	33	19.87	2.05	19.94
	Heterogeneous	35	18.60	2.02	18.12

Figure 2a shows student-level pre- and post-test scores based on their group variability. From this graph, we can see that students whose groups we designated as homogeneous showed a steeper line from pre- to post-test, meaning they had the highest learning gains of all three group types. Figure 2b shows group level means for homogeneous, medium, and heterogeneous groups. Again, we see that the homogenous groups have comparatively steeper lines, showing they had higher average learning gains than the other two groups. These differences were found to be statistically significant, which we describe below.

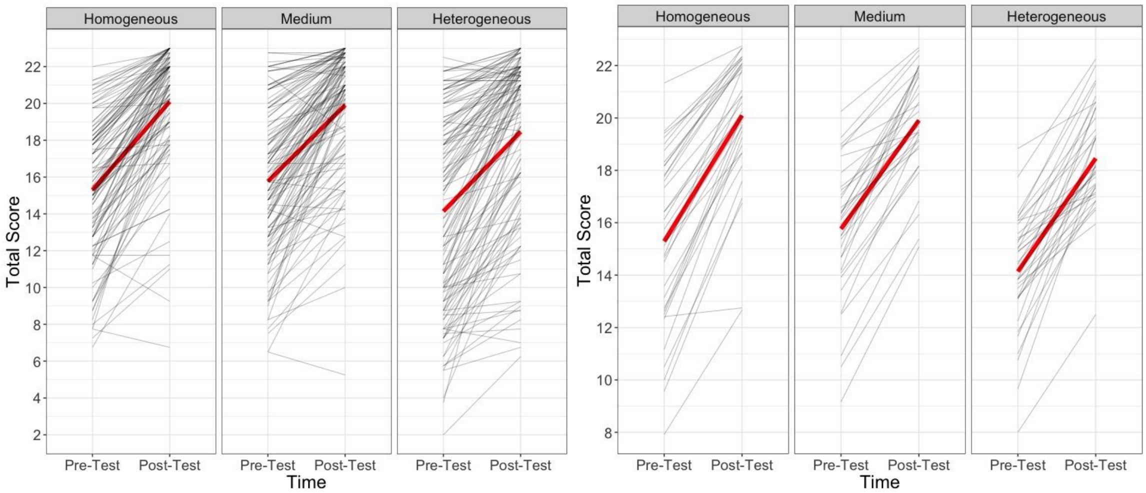

Figure 2 a & b. Student- (left) and group- (right) level pre- and post-test scores.

HLM analyses

We ran HLM analyses to investigate whether being in a homogenous, medium, or heterogenous prior knowledge group at the start of the unit affected students' learning gains. We used students' pre-test scores as a covariate and post-test scores as an outcome variable. Then we added group variability as an independent variable to the model so that we could investigate whether being in a homogenous, medium, or heterogenous prior knowledge group affected students' learning gains.

Table 3: HLM Model coefficients, standard errors

Fixed Effect	Coefficient (SE)	t (df)	p
(Intercept) - γ_{00}	10.28 (0.54)	19.17 (309)	0.0000
PreScore - γ_{10}	0.64 (0.03)	20.76 (355.45)	0.0000
Medium - γ_{20}	-0.58 (0.36)	-1.58 (100.48)	0.1163
Heterogenous - γ_{30}	-0.72 (0.36)	-1.99 (99.96)	0.0495
Variance Components	Estimate		
Residual – σ^2	6.01		
(Intercept) – τ_{02}	0.49		

In this model, we added groups' homogeneity level as an independent variable, using dummy coding. The model can be written as follows:

Level-1:	$PostScore_{ij} = \beta_0 + \beta_1 PreScore_{ij} + \beta_2 Medium + \beta_3 Heterogeneous + R_{ij}$
Level-2:	$\beta_0 = \gamma_{00} + U_{0j}$ $\beta_1 = \gamma_{10}$ $\beta_2 = \gamma_{20}$ $\beta_3 = \gamma_{30}$
Mixed Model:	$PostScore_{ij} = \gamma_{00} + \gamma_{10} PreScore_{ij} + \beta_2 Medium + \beta_3 Heterogeneous + R_{ij} + U_{0j}$

Based on the results in Table 3, we rewrote the model equation as:

$$PostScore_{ij} = 10.28 + 0.64 PreScore_{ij} - 0.57 Medium - 0.72 Heterogeneous$$

After controlling for pre-test scores, the average adjusted post-test score for students in homogeneous groups was 10.28. Additionally, students' average adjusted post-test score in the medium and heterogeneous groups were 9.70 and 9.56 respectively. We found that students in homogeneous group performed significantly better on the post-test than the students in the heterogeneous group ($t(99.96) = -1.99$, $p = 0.0495$), but not the medium group ($t(100.48) = -1.58$, $p = 0.1163$). Analysis shows intraclass correlation as .08, which means that group membership explains 8% of the variance in the post-test scores.

Because we were interested in knowing whether students with lower levels versus higher levels of prior knowledge benefitted similarly from being in a homogenous group, we ran further analyses to examine the difference within the homogeneous group. To do this, we examined the range of all students' pretest scores and then divided the homogenous groups into thirds to make three pre-test score categories: low (n = 11 groups), medium (n = 11 groups), and high (n = 12 groups). The adjusted average post-test scores of the students in these three categories were 10.11, 10.30, and 10.07 respectively. After running linear regression analyses, we found no significant differences between the three categories of students within the homogeneous group, when controlling for pre-test scores.

Discussion

We conducted a quantitative study to explore students' learning based on whether they differed in their levels of prior knowledge, in order to understand whether group composition based on students' prior knowledge was a factor affecting collaboration (Puntambekar & Young, 2003; Weinberger, Stegmann, &Fischer, 2007). Based on Vygotsky's conceptualization of the ZPD, many educators have historically believed that students should be placed in heterogeneous groups so that more capable students are able to support their less capable peers. More capable students in heterogeneous groups can provide knowledge and resources and address questions that less capable peers raise (Scardamalia & Bereiter, 1994). However, we found that students in similar-prior knowledge groups (i.e., homogeneous) had significantly higher learning gains than students in mixed-prior knowledge groups (i.e., heterogeneous) after working together during a design-based unit over 13 weeks. These findings contrast with those of other researchers who have found that students in heterogeneous groups showed greater learning outcomes than students in homogeneous groups (e.g., Csanadi et al., 2016; Gijilers & De Jong, 2005; Patchan et al., 2013; Zhao et al., 2018).

Yet, our findings also indicated benefits for students working in homogenous groups. We found that students in homogenous groups with low, medium, or high prior knowledge benefited equally from collaborating with similar level peers. These results differ from both Webb, Nemer, and Zuniga (2002) and Lou et al. (1996), who found that low prior knowledge students performed better in heterogeneous groups, while high prior knowledge students performed better, or equally as well, in homogenous groups. However, we only analyzed students' pre- and post-test scores on a content test in our study. In the absence of analysis of the discourse among group members, we cannot explain why these results occurred. However, one possible explanation, based on Zillmer and Kuhn's (2018) findings, is that students with similar prior knowledge levels may be more capable of providing metacognitive support to their peers, switching roles dynamically as needed. Zillmer and Kuhn (2018) also pointed out that students with similar levels of prior knowledge may have been better at metacognitively scaffolding one another because they worked together for a greater amount of time, which helped them establish greater intersubjectivity. In our study, students worked in groups for 13 weeks, which is a relatively long period of time. It could be the case that students in homogenous groups were better able to establish intersubjectivity earlier and maintain it as they collaborated throughout the unit.

Another possible explanation is that the more capable peers often do not intentionally scaffold other group members (Wells, 1999; Zuckerman, 2003). Analysis of students' discourse will help us understand the extent to which students in homogeneous and heterogeneous groups provided scaffolding, or not. Our findings lend support to Fernández and colleagues' (2001) argument that Vygotsky's ZPD framework should be reconceptualized to capture the kinds of scaffolding that occur in symmetrical groups. For example, different types of dialogue can provide support for symmetrical groups, which may help them to pool the groups' intellectual resources and establish a common context to collaboratively solve a problem that they would not be able to solve individually. In symmetrical groups, different group members might provide scaffolding at different times, in a continuously unfolding, reciprocal, and ever-changing process, as participants discursively contribute to the collaboration.

Future research

We plan on analyzing students' discourse from the video and audio data collected during the 13-week unit. This will help us to understand the types of contributions and support that students provided each other within their groups and try to determine why homogenous groups had higher learning gains than the heterogeneous groups in our study. Further, Zillmer and Kuhn (2018) suggested that the length of time students spend collaborating impacts the quality of their interactions. While we collected data over a longer time period than many previous studies, future research could also examine how the length of collaboration, a few sessions to weeks or months of collaboration, affects students' learning outcomes based on whether they are in a homogenous or heterogenous prior knowledge group. Finally, future research could also examine how student's learning gains differ within the groups at the end of the unit. This would help us to identify the extent to which students starting out with different levels of prior knowledge are benefitting from participating in each type of group and whether students' understanding becomes more convergent over time. Information from each of these lines of research could provide practical guidance to teachers as they are strategically forming groups in the classroom.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Csanadi, A., Kollar, I., & Fischer, F. (2016). Scientific Reasoning and Problem Solving in a Practical Domain: Are Two Heads Better Than One?. *Singapore: International Society of the Learning Sciences.*

Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2001). Re-conceptualizing "scaffolding" and the zone of proximal development in the context of symmetrical collaborative learning. *The Journal of Classroom Interaction*, 40-54.

Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching*, 42(3), 264-282.

Kim, J. S., Anderson, C. J., & Keller, B. (2013). Multilevel analysis of assessment data. Rutkowski, L. von Davier, M., & Rutkowski, D. (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 389-425.

Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. London, England: Sage Publication.

Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, *66*(4), 423-458.

Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, 41(2), 381-405.

Puntambekar, S. & Young, M. F. (2003). Moving toward a theory of CSCL. In U. Hoppe, B. Wasson, & S. Ludvigson, (Eds.) *Computer Supported Collaborative Learning 2003: Designing for change in networked learning.* pp. 503-512. IOS Press: Amsterdam. [Presented in the closing plenary session at CSCL 2003].

R Core Team (2017). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (Vol. 1). (2nd ed.). Thousand Oaks: Sage Publications.Sage Publication.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* Oxford, England: Oxford University Press.

Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The journal of the learning sciences*, *3*(3), 265-283.

Stahl, G. (2004). Building collaborative knowing. In J.W. Strijbos, P.A. Kirschner, & R.L. Martens (Eds.), *What we know about CSCL* (pp. 53-85). Springer, Dordrecht.

Vygotsky, L. S. (1978*). Mind in society: The development of higher mental process*.

Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, *39*(4), 943-989.

Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17(4), 416-426.

Wells, G. (1999). *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge University Press.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis Springer-Verlag New York.

Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention?. *Instructional Science*, 40(4), 711-730

Zillmer, N., & Kuhn, D. (2018). Do similar-ability peers regulate one another in a collaborative discourse activity?. *Cognitive Development*, 45, 68-76.

Zhao, J., Lin, L., Sun, J., Zheng, X., & Yin, J. (2018). Students' engagement in a science classroom: Does knowledge diversity matter?. *The Journal of Educational Research*, 1-8.

Zuckerman, G. (2003). The developmental path toward reflection. In A. Kouzulin, B. Gindis, V,S. Ageyev, & S.M. Miller (Eds.), *Vygotsky's educational theory in cultural context*, *177*. Cambridge.

Acknowledgments

The research reported in this study is supported by a NSF-DRL grant (#1418044) to the last author. We thank all of the teachers and students who participated in this study.