Part 20 / Strand 20

Methods and Methodological Aspects in Science Education Research

Co-editors: *Kason Ka Ching Cheung & Adrienne Lorelei Traxler*

Part 20 / Strand 20 Methods and Methodological Aspects in Science Education Research

Foundational aspects and debates of epistemology, ontology and axiology of methods and methodologies of science education research.

Sub-themes:

1) Qualitative Methods in Science Education Research
2) Quantitative Methods in Science Education Research
3) Mixed Methods Approaches in Science Education Research
4) Ethical Considerations in Science Education Research
5) Critical perspectives on methods and methodologies
6) Emerging technologies and Methodological Aspects in Science Education Research

Contents

Strand 20: Methods and Methodological Aspects in Science Education Research

Adrienne L. Traxler[1] and Kason Ka Ching Cheung[2]
[1]University of Copenhagen, Denmark
[2]The Education University of Hong Kong, China

Foreword

The strand Methods and Methodological Aspects in Science Education Research (or informally, "the methodology strand") is a slim but important part of the ESERA conference. Other strands employ a rich array of techniques and arguments in their studies, but for this strand, the primary focus is on the methods. Authors may use their work to showcase a methodological advance, to promote an existing technique to the community, or to invite critical discussion of their analysis structure and choices. Method is the key focus of this theme, while it can be integrated into a range of science education contexts set up in other themes.

In this year's proceedings, the methodology strand has two chapters. The first, by Sundstrom and collaborators, uses temporal exponential random graph models to quantify the existence and growth of structure in student collaboration networks in university science courses. The second, by El Karkri and collaborators, uses a Solomon Four-Group Design to probe for the effect of pre-test sensitisation in secondary school science courses. Both studies are quantitative in focus, and both recommend further use of their tools in the science education research community. Temporal network models allow for the adaptation of popular pretest/post-test designs to incorporate network effects, and the Solomon Four-Group Design gives principled attention to the often-neglected question of pretest effects that may damage the precision or reliability of findings.

Integrating advanced methodology and data analyses make an important contribution to science education issues. We envisage that these advanced methodological considerations will gain increasing attention among researchers who work in different issues and disciplines in science education. For instance, the temporal exponential random graph models, which show student networks in physics education research in this conference proceeding, will interest researchers in chemistry and biology education. The Solomon Four-Group Design, which attends to students' scientific reasoning in different experimental conditions in this proceeding, will interest scholars working on experimental design that advances students' nature of science and scientific practices. We believe that papers in this strand will extend our discussion on how to advance science education using methods. In the future, science education researchers will continually derive methodological frameworks based on the designs discussed in the two chapters. We thank the authors for their work, and look forward to the next stage in the conversation of methodological advances and practices in the field of science education research.

Measuring Classroom Network Dynamics with Temporal Exponential Random Graph Models

Meagan Sundstrom[1], Adrienne Traxler[2] and Eric Brewe[1]
[1]Drexel University, United States of America
[2]University of Copenhagen, Denmark

Interacting with peers has been shown to positively correlate with undergraduate science students' performance and sense of belonging. Correspondingly, a growing number of physics education research studies have examined patterns of peer interactions using social network analysis. Time-dependent student networks are of particular interest, as they illuminate the ways in which student communities form. Existing studies of time-dependent student networks, however, rely on descriptive measures and/or simple statistical tests that do not account for other variables that may explain classroom network dynamics. In this study, therefore, we demonstrate the use of temporal exponential random graph models (TERGMs), a form of multiple regression for time-dependent network data. We apply TERGMs to pre- and post-semester peer interaction networks from two introductory physics courses in the United States taught using different active learning pedagogies: Peer Instruction (i.e., small groups of students answer poll questions during lecture) and SCALE-UP (i.e., small groups of students work on problem solving and laboratory activities in a classroom specifically designed for group work). We find that students form connections with small groups of three peers in both courses. In SCALE-UP, however, students are also unlikely to form connections that create long chains. These results are not extractable from the descriptive network measures alone, suggesting that TERGMs (or other similar models) are necessary for measuring classroom network dynamics.

Keywords: Network Analysis, Collaborative Learning, Pedagogy

Introduction

According to the participationist framework, learning occurs as people engage with their surrounding community (Packer & Goicoechea, 2000). Collaborating with others allows for learners to revise upon their own thinking and co-construct understanding. In university-level physics classrooms in particular, researchers have demonstrated that social interactions with peers positively correlate with student outcomes such as academic performance, attitudes, and persistence (Bruun & Brewe, 2013; Dou et al., 2016; Zwolak et al., 2017).

Social network analysis (SNA) is an increasingly popular tool for analyzing peer interactions among physics education researchers (Grunspan et al., 2014). Many SNA studies administer a network survey to students, asking which peers they interact with about their physics course, at one point in time and use a combination of descriptive statistics, regression methods, and exponential random graph models (ERGMs) to make claims about patterns in the resulting interaction network (Bruun & Brewe, 2013; Commeford et al., 2019; Sundstrom et al., 2022b; Crossette et al., 2023; Pulgar et al., 2023). Other SNA studies examine how student interaction networks evolve over time by administering several network surveys throughout a physics course and analyzing changes in the resulting networks (Brewe et al., 2010; Bruun & Bearden, 2014; Commeford et al., 2021; Sundstrom et al., 2022a, Wolf et al., 2022). Such "time-dependent" student networks illuminate the ways in which student communities develop in a classroom and how classroom instruction may impact student network development.

The existing studies of time-dependent student networks, however, have exclusively relied on descriptive statistics (e.g., network density) and/or simple statistical tests (e.g., t-tests) to make claims. These methods are limited because they do not allow for hypothesis testing: they measure the effect of one variable at a time without accounting for other variables that may explain why networks change over time. More sophisticated tools, therefore, are necessary to make more nuanced claims about time-dependent student networks.

In this study, we expand upon the SNA methods that have been applied in physics education research to meet the methodological needs of time-dependent networks described above. We use temporal exponential random graph models (TERGMs)–a form of multiple regression for time-dependent network data–to measure network dynamics in two university-level, introductory physics courses taught using different active learning pedagogies. We aim to address the following research question: What types of peer interaction network structures form in different active learning environments in physics?

Methods

Data Collection

We use data from two courses at different institutions in the United States that were collected as part of a prior research project (Commeford et al., 2021). The first was an introductory physics course at Drexel University containing 116 students. The course contained one combined lecture section and several smaller laboratory and recitation sections. The lecture component was taught in a large lecture hall using "Peer Instruction," where the instructor asked students to discuss and respond to clicker questions. The second was an introductory physics course at North Dakota State University containing 71 students. The course was taught using the "SCALE-UP" pedagogy, where all students were in one course section that blended lecture, laboratory, and recitation activities. The course took place in a large room designed for group work: there were several round tables where students sat, and whiteboards were placed along the perimeter of the room.

In each course, the instructor administered an online network survey at the beginning and end of the semester. Students responded to the following prompt: "Please choose from the list of people that are enrolled in your physics class the names of any other student with whom you had a meaningful interaction in class during the past week, even if you were not the main person speaking." Below the prompt was an alphabetized course roster with associated check boxes, from which students could select as many names as they wanted. For the Peer Instruction course, the pre- and post-survey had response rates of 42% and 58% of enrolled students, respectively. For the SCALE-UP course, the pre- and post-survey had response rates of 70% and 66% of enrolled students, respectively. Responses made by students who did not consent to participating in the research were excluded from our analysis.

Data Analysis

We created undirected networks using the survey responses, with nodes representing students and edges representing a reported interaction between two students (regardless of which student reported the interaction on the survey), to reduce possible impacts of missing data due to non-respondents. We calculated several descriptive measures for each network, as in prior work (Commeford et al., 2021), to characterize its structure:

- Density: the proportion of all possible edges in the network that we observed
- Transitivity: the number of two-paths (e.g., student A is connected to student B and student B is connected to student C) that are closed (e.g., students A and C are connected)
- Giant component: the number of nodes that are contained in the largest connected cluster of nodes.

We then applied TERGMs to the data to measure whether different network structures (e.g., transitivity) are more (or less) likely to form in our observed post-semester networks than we would expect if the pre-semester networks evolved randomly. We used separable TERGMs in particular, as they model two distinct processes of network evolution: (1) formation–the tendency for new edges to form over time and (2) persistence–the tendency for edges formed early on to remain in the network over time (Krivitsky & Handcock, 2014). In line with our research goal of identifying the types of network structures that form in active learning classrooms, we only examine the formation process in this study.

Table 1. Predictor variables included in TERGMs.

Variable	Description	R syntax
Edges	The number of edges in the observed network.	edges
Chains	The number of two-paths (e.g., student A is connected to student B and student B is connected to student C) that are open (e.g., students A and C are not connected).	gwdsp(decay=0.5, fixed=TRUE)
Transitivity	The number of two-paths (e.g., student A is connected to student B and student B is connected to student C) that are closed (e.g., students A and C are connected).	gwesp(decay=0.5, fixed=TRUE)
Degree Distribution	The distribution of the number of edges connected to each node.	gwdegree(decay=0.5, fixed=TRUE)

We ran one TERGM for each of the two courses using the predictor variables listed in Table 1 and the "tergm" package in R. We included these particular variables because we expected them to vary between the Peer Instruction and SCALE-UP classrooms based on prior work (Commeford et al., 2019; Commeford et al., 2021). For example, we expected that chains would be more likely to form in Peer Instruction classes, as students typically sit in long rows of chairs in stadium style lecture halls, than in SCALE-UP classes, where students work in small groups at a round table.

Both models exhibited sufficient goodness-of-fit: the observed networks and a set of networks simulated from the resulting models had similar structural properties (Hunter et. al., 2008).

Results

Figures 1 and 2 show the pre- and post- semester networks for the Peer Instruction and SCALE-UP courses, respectively. Table 2 contains descriptive measures for each network. The Peer Instruction network does not appear to change much over time. In both networks for this course, the majority of students are connected in the giant component, which contains a combination of chains and closed triangles. Quantitatively, the density and transitivity values are similar for the pre- and post-networks. The SCALE-UP pre-network contains some isolated small groups of students and a few edges that bridge together these groups. The SCALE-UP post-network, on the other hand, contains more edges (the density increases) and more strongly connected small groups (the transitivity increases) than the pre-network.

Figure 3 and Table 3 show the coefficient estimates for the TERGMs applied to the observed networks. Odds ratios that are greater (less) than one and significant indicate that the network structure represented by the predictor variable is more (less) likely to form in our observed network than we would expect if the pre-network evolved randomly.

Figure 1. Pre- (left) and post- (right) peer interaction networks for an introductory physics course taught using Peer Instruction. Nodes represent students and edges represent self-reported, in-class peer interactions.

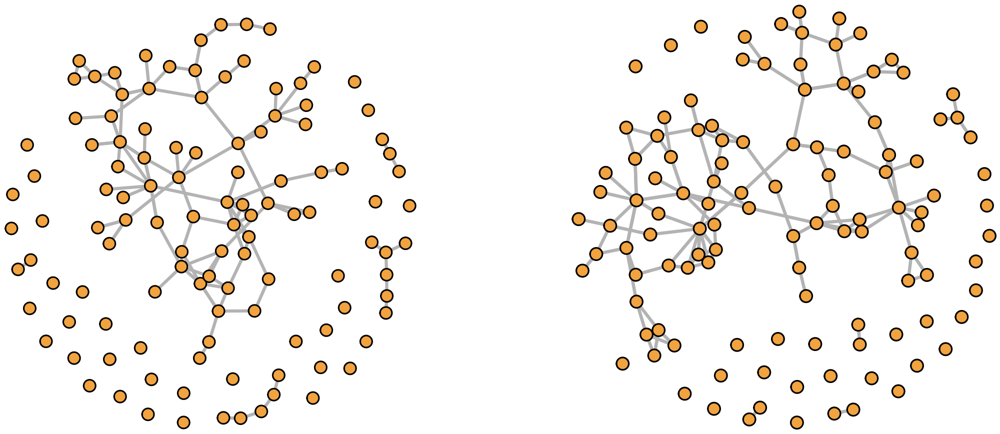


Figure 2. Pre- (left) and post- (right) peer interaction networks for an introductory physics course taught using SCALE-UP. Nodes represent students and edges represent self-reported, in-class peer interactions.

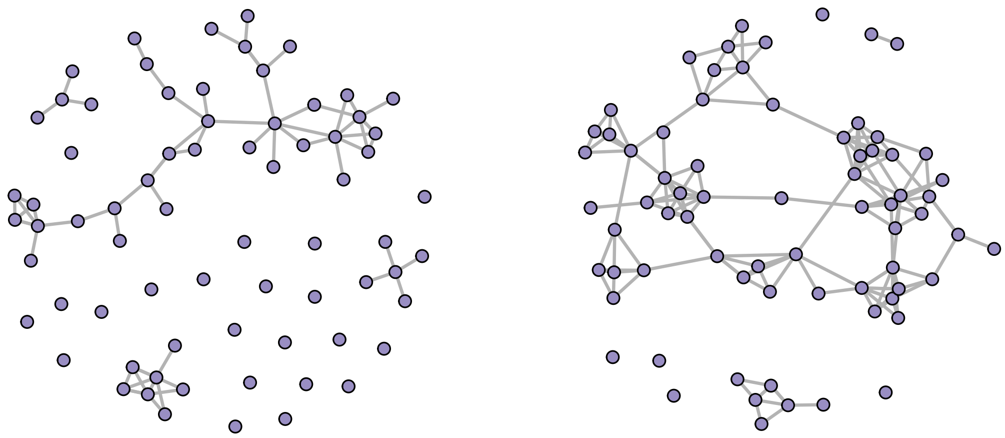


Table 2. Descriptive measures for the four networks.

	Peer Instruction		SCALE-UP	
	Pre	Post	Pre	Post
Nodes	116	116	71	71
Edges	135	159	71	176
Density	0.020	0.024	0.029	0.071
Transitivity	0.209	0.231	0.327	0.517
Giant component	67	80	34	58

In the Peer Instruction course, the only significant effect is transitivity: edges forming closed triangles are significantly more likely to form in our observed network than we would expect at random. Notably, this result is after controlling for the other effects included in the model and despite the raw transitivity values of the pre- and post-networks being similar (Table 2). The non-significant effect of degree distribution means that edges are not more likely to form if they create either a highly skewed degree distribution (i.e., with many students having few connections and a few students having many connections) or a very uniform degree distribution (i.e., with all students having a similar number of connections).

In the SCALE-UP course, we find that chains are significantly less like to form than we would expect at random and that edges that form closed triangles are significantly more likely to form than we would expect at random. Similar to the Peer Instruction course, we find no significant effect of degree distribution.

Figure 3. Coefficient estimates for TERGMs. Error bars represent 95% confidence intervals and asterisks indicate statistical significance (**p* < 0.001).**

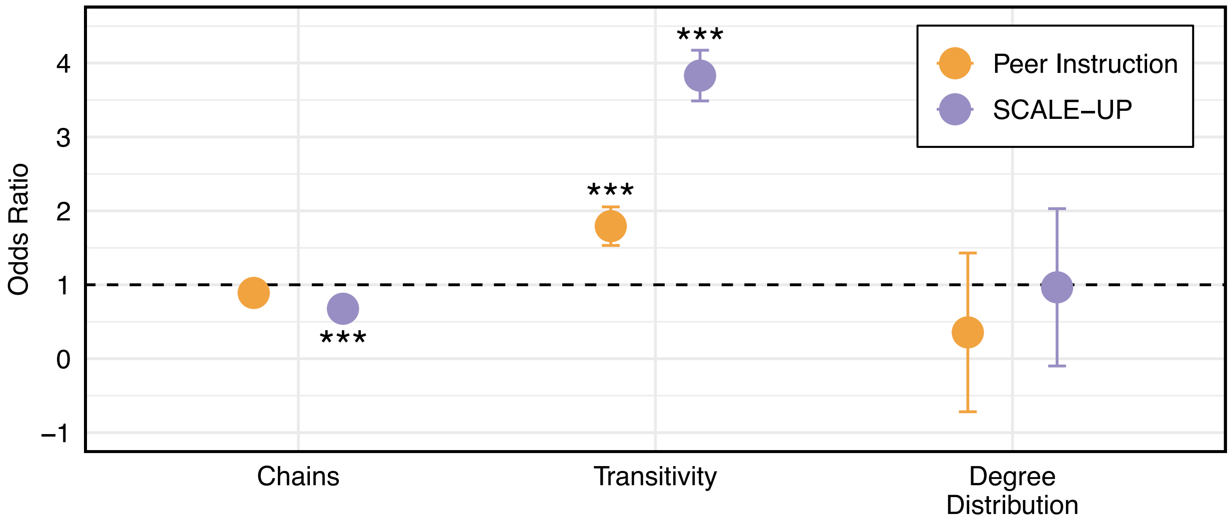


Table 3. Coefficient estimates for TERGMs. SE indicates standard error and asterisks indicate statistical significance (**p* < 0.001).**

	Peer Instruction			SCALE-UP		
	Log-odds	SE	Odds ratio	Log-odds	SE	Odds ratio
Edges	-3.872***	0.653	0.021	-2.454***	0.559	0.086
Chains	-0.117	0.083	0.890	-0.392***	0.056	0.676
Transitivity	0.584***	0.133	1.793	1.343***	0.175	3.831
Degree Distribution	-1.033	0.548	0.356	-0.035	0.543	0.966

Discussion and Conclusion

We have characterized the dynamics of student networks using a statistical tool not yet applied in physics education research: temporal exponential random graph models. Results indicate structural differences in the types of connections students form with their peers in two introductory physics courses taught using different active learning pedagogies. In a Peer Instruction course, students form connections with small groups of three peers. In a SCALE-UP course, students also form small group of three peers, but they are unlikely to form connections that create open chains.

We recommend for other education researchers to use TERGMs or similar models (e.g., stochastic actor-oriented models) when conducting time-dependent network studies, moving beyond descriptive and/or simple statistical approaches. In this study, for example, we found that after controlling for other effects in the model, some changes in network structure identified in the TERGMs differed from the descriptive statistics alone.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 2111128. We thank Kelley Commeford for collecting the data.

References

Brewe, E., Kramer, L. H., & O'Brien, G. E. (2010). Changing participation through formation of student learning communities. In *AIP Conference Proceedings*, *1289*(1), 85-88. https://doi.org/10.1063/1.3515255

Bruun, J., & Brewe, E. (2013). Talking and learning physics: Predicting future grades from network measures and Force Concept Inventory pretest scores. *Physical Review Special Topics—Physics Education Research*, *9*(2), 020109. https://link.aps.org/doi/10.1103/PhysRevSTPER.9.020109

Bruun, J., & Bearden, I. G. (2014). Time development in the early history of social networks: Link stabilization, group dynamics, and segregation. *PLoS One*, *9*(11), e112775. https://doi.org/10.1371/journal.pone.0112775

Commeford, K., Brewe, E., & Traxler, A. (2019). Characterizing active learning environments in physics: network analysis of Peer Instruction classroom using ERGMs. Paper presented at Physics Education Research Conference 2019, Provo, UT. https://doi.org/10.1119/perc.2019.pr.Commeford

Commeford, K., Brewe, E., & Traxler, A. (2021). Characterizing active learning environments in physics using network analysis and classroom observations. *Physical Review Physics Education Research*, *17*(2), 020136. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.17.020136

Crossette, N., Carr, L. D., & Wilcox, B. R. (2023). Correlations between student connectivity and academic performance: A pandemic follow-up. *Physical Review Physics Education Research*, *19*(1), 010106. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.19.010106

Dou, R., Brewe, E., Zwolak, J. P., Potvin, G., Williams, E. A., & Kramer, L. H. (2016). Beyond performance metrics: Examining a decrease in students' physics self-efficacy through a social networks lens. *Physical Review Physics Education Research*, *12*(2), 020124. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.12.020124

Grunspan, D. Z., Wiggins, B. L., & Goodreau, S. M. (2014). Understanding classrooms through social network analysis: A primer for social network analysis in education research. *CBE—Life Sciences Education*, *13*(2), 167-178. https://doi.org/10.1187/cbe.13-08-0162

Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, *103*(481), 248-258. https://doi.org/10.1198/016214507000000446

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *76*(1), 29-46. https://doi.org/10.1111/rssb.12014

Packer, M. J., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist*, *35*(4), 227-241. https://doi.org/10.1207/S15326985EP3504_02

Pulgar, J., Ramírez, D., & Candia, C. (2023). Enhanced social connectivity in hybrid classrooms versus academic centrality in online settings. *Physical Review Physics Education Research*, *19*(2), 020155. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.19.020155

Sundstrom, M., Wu, D. G., Walsh, C., Heim, A. B., & Holmes, N. G. (2022a). Examining the effects of lab instruction and gender composition on intergroup interaction networks in introductory physics labs. *Physical Review Physics Education Research*, *18*(1), 010102. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.18.010102

Sundstrom, M., Schang, A., Heim, A. B., & Holmes, N. G. (2022b). Understanding interaction network formation across instructional contexts in remote physics courses. *Physical Review Physics Education Research*, *18*(2), 020141. https://doi.org/10.1103/PhysRevPhysEducRes.18.020141

Wolf, S. F., Sault, T. M., Suda, T., & Traxler, A. L. (2022). Social network development in classrooms. *Applied Network Science*, *7*(1), 24. https://doi.org/10.1007/s41109-022-00465-z

Zwolak, J. P., Dou, R., Williams, E. A., & Brewe, E. (2017). Students' network integration as a predictor of persistence in introductory physics courses. *Physical Review Physics Education Research*, *13*(1), 010113. https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.13.010113

The Critical Role of Pre-test Sensitisation and the Solomon Four-Group Design in Science Education Research: A Transition From Uncertainty to Necessity

Mourad El Karkri, Antonio Quesada, and Marta Romero-Ariza
University of Jaén, Spain

Pre-test sensitisation (PS) is a critical concern in science education research, as it can influence participants' responses to interventions and potentially distort study outcomes. When students are pretested, they might gain insights or knowledge that allow them to answer correctly during subsequent assessments, thereby affecting the validity of the results (Pan & Carpenter, 2023). Two studies, undertaken a Spanish-Moroccan team of researchers, were conducted in secondary schools over two consecutive academic years (2021–2022 and 2022–2023), both utilising the Solomon Four-Group Design (SFGD) as a tool to detect PS (Campbell & Stanley, 1966; Solomon, 1949). Despite this shared framework, a notable divergence emerged in the results: the first study revealed no evidence of PS, while the second clearly shows sensitization. This difference led to new considerations and perspectives, raising essential questions about the role of PS in science education research. Should it be regarded as a prestigious methodological element, or is its inclusion always necessary? These findings challenge established perceptions and highlight the need for a critical reflect on the methodological significance of PS.

Keywords: Pre-test sensitisation, Solomon Four-Group Design, Quasi-experimental studies.

Theoretical Background

Numerous studies across different fields provide evidence supporting the presence of PS, demonstrating its impact on learning processes, attitudes, mental health interventions, and behavioural intentions (Hoogstraten, 1980; Lana & King, 1960; Spence et al., 2009; Spirito et al., 1988; Swanborn & De Glopper, 1999).

Conversely, many studies have reported the absence of PS, showing no significant effects on attitude change, behavioural measures, family interaction patterns, self-efficacy, or stress management scores (Jordaan et al., 2018; Kazdin, 1973; Lana, 1959; Parsons & Alexander, 1973; Rosa-Castillo et al., 2023; Rubel et al., 2010; Spence & Blanchard, 2001).

A nuanced perspective on pre-testing is offered by Hollins et al. (2023), who challenged the assumption that pre-testing universally enhances memory. Their findings suggest that pre-testing benefits are limited to specific guessed information, providing a more detailed understanding of its role in memory. Similarly, Pan & Carpenter (2023) demonstrated that the pre-questioning or pre-testing effect can enhance learning, even when initial tests include incorrect answers, through the use of various educational materials.

Taken together, these studies illustrate that while PS may not always manifest, its potential to influence participant responses remains substantial. This is particularly relevant in contexts where pre-tests and post-tests share identical formats, differing only in content and numerical measurements. Although the evidence does not provide definitive conclusions, it strongly indicates that PS is a plausible factor warranting careful consideration in research design and interpretation.

Research Questions

This study seeks to address two interrelated questions concerning the phenomenon of PS in science education research through two implementations conducted in different secondary schools. First, it investigates whether PS might occur in science education contexts, examining its potential influence on research outcomes. Second, it explores whether PS should be considered as another aspect while research design, something that must be considered and accounted for, whenever possible to ensure methodological rigour.

Research Design And Methods

The Solomon Four-Group Design (SFGD) received its name from Richard Lester Solomon (born in 1918, passed away in 1995) and consists of four different groups of participants, each subjected to a different experimental condition, as outlined in Table 1:

Table 1: Structure of the Solomon four-group design (Campbell & Stanley, 1966).

Groups	Assignment	Pre-test	Treatment	Post-test
Group 1 (Experimental & Pretested)	Randomly	O1	Yes	O2
Group 2 (Control & Pretested)	Randomly	O3	No	O4
Group 3 (Experimental and Non-Pretested)	Randomly		Yes	O5
Group 4 (Control and Non-Pretested)	Randomly		No	O6

The treatment in both studies was planned as five sessions focusing on getting students familiar with reason patterns related to making a fair test with variable control. These lessons were inspired by those offered within the Cognitive Acceleration through Science Education, a broadly implemented program in various countries across Europe, America, Australia, and Asia (Adey et al., 2004). The pre-test and post-test used in the studies were composed of 10 questions extracted from activities, examples, and applications from CASE program, designed to be completed within 1 hour.

Findings And Interpretation

Tables 2 and 3 present the SFGD for each study. Table 2 features a larger number of participants with a more balanced distribution across the four groups compared to Table 3. This difference arises from the availability of the study samples.

Despite the less balanced sample in the second study, normality was verified before proceeding with parametric analysis, using IBM SPSS Statistics (version 26). The Shapiro-Wilk test confirmed normality in Study 1 for both the treated group ($p = 0.469$) and the untreated group ($p = 0.240$). Similarly, in Study 2, it indicated normality for the treated group ($p = 0.496$) and the untreated group ($p = 0.254$), supporting the assumption of normality and the use of parametric methods.

Table 2. Distribution and composition of the Solomon four-group design in the 2023 Study.

Groups	Frequency	Percentage
Experimental 1	31	25 %
Control 1	23	18 %
Experimental 2	38	30 %
Control 2	34	27 %
Total	126	100 %

Table 3. Distribution and composition of the Solomon four-group design in the 2022 study.

Groups	Frequency	Percentage
Experimental 1	9	10 %
Control 1	25	29 %
Experimental 2	17	19 %
Control 2	37	42%
Total	88	100 %

To visualise the interaction between the variables, we generated Two-Way ANOVA interaction plots, which are not commonly used or shown in comparable studies. In this approach, the quantitative variable Post-test Scores were set as the dependent variable, while the two qualitative variables, Pre-test Scores and Treatment, were assigned as fixed factors. This method produced the two interaction graphs displayed below (Figures 1 & 2), illustrating the effects of pre-testing and treatment on the Post-test Scores.

Figure 1. Interaction between pre-testing and treatment on post-test scores in study 1 (2023).

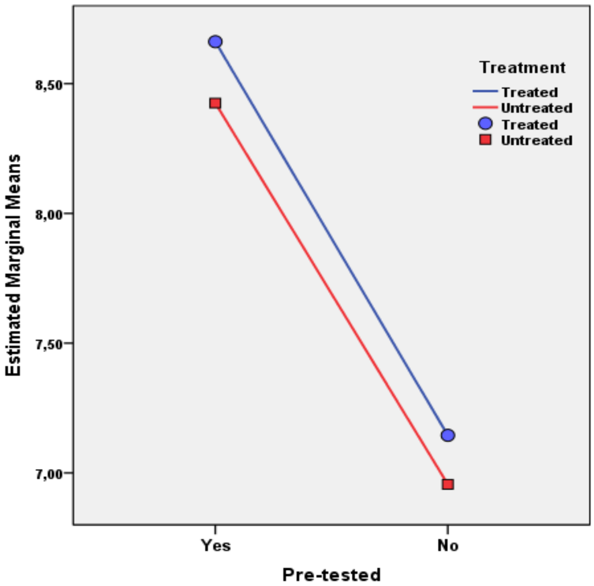
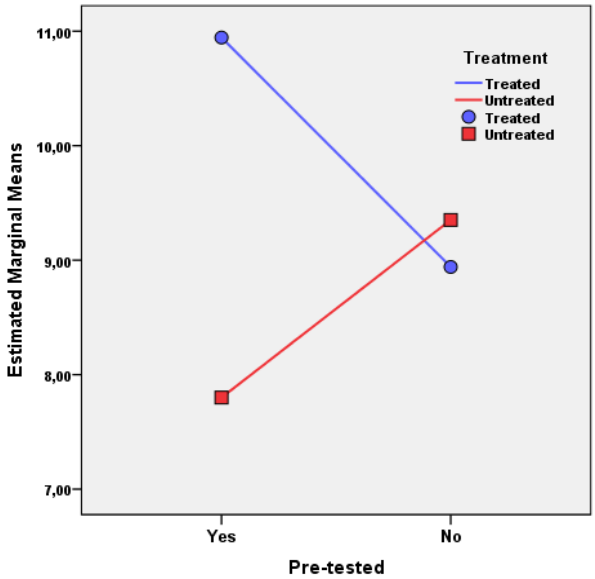


Figure 2. Interaction between pre-testing and treatment on post-test scores in study 2 (2022).



Before giving interpretation of plots, some details need to be clarified. A line in an interaction plot does not represent a mathematical function. Instead, it is a visualization of group averages calculated from a statistical model. Each line connects these averages to make trends easier to interpret. The end points of the line correspond to the estimated marginal means for each condition, providing a clear summary of the data for different groups.

Each line shows how the dependent variable (Post-test Scores) changes with the independent variable (Pre-test Scores) for a specific group (Treated or Untreated). The lines connect estimated marginal means from the statistical model and are straight, reflecting assumed linear relationships. This simplifies group comparisons and emphasizes trends, even if real-world relationships are not strictly linear.

When lines in an interaction plot intersect, it means there is an interaction effect between variables. The effect of the treatment on the outcome depends on the condition of another variable (pre-test). Crossing lines show that the effect changes or reverses between conditions, while parallel lines indicate no interaction, meaning the effect is consistent across all conditions.

Based on these reflections, In Study 1 (Figure 1), the lines are nearly parallel (in general, no intersection), indicating that there is no interaction between pre-testing and treatment. This shows that treatment effects were consistent and unaffected by pre-testing.

In contrast, in Study 2 (Figure 2), the lines are non-parallel (in general, they intersect), indicating a significant interaction between pre-testing and treatment. This demonstrates how pre-testing influenced treatment outcomes, highlighting the variability of PS effects.

Future Directions In Adoption

By repeating the same experiment with the SFGD, we gain a broader perspective on the phenomenon of PS itself. This repetition allows us to determine whether PS consistently occurs, consistently does not occur, or varies depending on specific conditions. If the findings suggest variability, it implies that PS cannot be generalised across studies. Consequently, this variability underscores the necessity of accounting for PS in every research design to ensure methodological rigour and accurate interpretation of results.

Historically, the Pre-test-Post-test Control Group Design, now viewed as a fundamental element of the "traditional" or "usual" approach to two-group experimental studies, was once considered a prestigious and innovative methodology (Huck & McLean, 1975). Similarly, the Solomon Four-Group Design, often described as prestigious due to its ability to detect and account for PS, deserves a reassessment of its role in educational research.

It is time to recognise this design not just as a specialised tool for select studies but as an essential approach for designing and conducting experiments in education. Adopting the Solomon Design as a standard methodology could significantly enhance the precision and reliability of findings, especially in contexts where the effects of pre-testing cannot be ignored.

The absence of PS, as observed in Study 1 does not imply that researchers should revert to the usual two-group design of. Instead, it may encourage the adoption of a more rigorous approach, such as parametric analysis and meta-analysis, as proposed by Braver & Braver (1988). However, this lack of PS should not lead to a disregard for the Solomon Four-Group Design. Despite its perceived complexity, this design remains invaluable, particularly in fields like educational research, where the influence of PS should not be overlooked.

Reflections Following Conference Feedback

In light of the notes and discussion generated by attendees during the presentation of this work at the ESERA Copenhagen 2025 conference, it became clear that the topic of pre-test sensitisation (PS) continues to raise important methodological questions in science education research. The central purpose of the present article is therefore to foreground the significance, and in some contexts the practical necessity, of considering PS when designing and interpreting experimental or quasi-experimental interventions. Even when PS is not consistently observed across studies, its potential to shape participants' responses can influence the magnitude and interpretation of treatment effects. For this reason, the article argues that PS should not be treated as a marginal issue, but rather as a plausible source of variation that warrants explicit attention whenever research conditions make it possible.

The discussion with attendees also brought out a practical question that many researchers face, which is how the analysis should proceed when a study includes a pretest, when it does not, or when it includes both conditions, as in the Solomon Four-Group Design. Put simply, attendees wanted to know what changes in the analytic strategy from one situation to another, and how to choose the most appropriate route for estimating treatment effects without drawing conclusions that the design cannot support. This ambiguity was subsequently addressed more extensively in a methodological contribution published in Review of Education (El Karkri, Quesada, & Romero-Ariza, 2025a), where the procedures and decision steps are clarified in a systematic manner, including guidance on how to proceed under the presence or absence of PS and how to select appropriate analyses for different types of outcome variables.

Attendees also commented on the uneven distribution of participants across groups in Study 2 (Table 3). The concern was not only about balance in principle, but about what such imbalance might mean for statistical power, the reliability of estimates, and the interpretation of the interaction term. Rather than treating this as a minor limitation, it was used to sharpen the methodological discussion and to explain more clearly how these issues can be handled and reported transparently. This line of critique was taken forward in a later empirical paper published in Brain Sciences (El Karkri, Quesada, & Romero-Ariza, 2025b), which examines the second study in greater depth and discusses its implications in a more detailed way. Taken together, the conference feedback and the follow-up publications help to clarify why PS deserves attention, how it can be analysed carefully, and how design-related concerns can be strengthened through systematic follow-up work. It was also suggested that Table 1 should clarify the random allocation of participants to groups. Attendees also noted that Table 1 should clarify the random allocation of participants to groups, and this has now been reflected in the revised version.

Acknowledgement

We are grateful to Dr Martin Riopel (Department of Didactics, Faculty of Education Sciences, University of Quebec in Montreal, Canada) for his comments during the presentation and for subsequent discussions afterwards, which contributed to enriching this work. We also acknowledge Dr Ahmad Qablan (Faculty of Education, United Arab Emirates University, Al Ain, United Arab Emirates).

References

Adey, P., Hewitt, G., John, Hewitt, & Landau, N. (2004). *The Professional Development of Teachers: Practice and Theory*. Kluwer Academic Publishers. https://doi.org/10.1007/0-306-48518-4

Braver, M. W., & Braver, S. L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin*, *104*(1), 150–154. https://doi.org/10.1037/0033-2909.104.1.150

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Wadsworth.

El Karkri, M., Quesada, A., & Romero-Ariza, M. (2025a). Methodological aspects of the Solomon four-group design: Detecting pre-test sensitisation and analysing qualitative and quantitative variables in education research. *Review of Education*, *13*(1), e70050. https://doi.org/10.1002/rev3.70050

El Karkri, M., Quesada, A., & Romero-Ariza, M. (2025b). The Dual Impact of Pretest Sensitisation and the Cognitive Acceleration Through Science Education Programme in the Solomon Four-Group Design. *Brain Sciences*, *16*(1), 64. https://doi.org/10.3390/brainsci16010064

Hollins, T. J., Seabrooke, T., Inkster, A., Wills, A., & Mitchell, C. J. (2023). Pre-testing effects are target-specific and are not driven by a generalised state of curiosity. *Memory*, *31*(2), 282–296. https://doi.org/10.1080/09658211.2022.2153141

Hoogstraten, Joh. (1980). The Reactive Effect of Pretesting in Attitude Change Research: General or Specific? *Applied Psychological Measurement*, *4*(1), 39–42. https://doi.org/10.1177/014662168000400105

Huck, S. W., & McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: A potentially confusing task. *Psychological Bulletin*, *82*(4), 511–518. https://doi.org/10.1037/h0076767

Jordaan, J., Beukes, R., & Esterhuyse, K. (2018). The Development and Evaluation of a Life Skills Programme for Young Adult Offenders. *International Journal of Offender Therapy and Comparative Criminology*, *62*(10), 3077–3096. https://doi.org/10.1177/0306624X17737682

Kazdin, A. E. (1973). The effect of suggestion and pretesting on avotdance reduction in fearful subjects. *Journal of Behavior Therapy and Experimental Psychiatry*.

Lana, R. E. (1959). Pretest-treatment interaction effects in attitudinal studies. *Psychological Bulletin*, *56*(4), 293–300. https://doi.org/10.1037/h0044646

Lana, R. E., & King, D. J. (1960). Learning factors as determiners of pretest sensitization. *Journal of Applied Psychology*, *44*(3), 189–191. https://doi.org/10.1037/h0040275

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and Pretesting Effects: A Review of Empirical Research, Theoretical Perspectives, and Implications for Educational Practice. *Educational Psychology Review*, *35*(4), 97. https://doi.org/10.1007/s10648-023-09814-5

Parsons, B. V., & Alexander, J. F. (1973). Short-term family intervention: A therapy outcome study. *Journal of Consulting and Clinical Psychology*, *41*(2), 195–201.

Rosa-Castillo, A., García-Pañella, O., Roselló-Novella, A., Maestre-Gonzalez, E., Pulpón-Segura, A., Icart-Isern, T., & Solà-Pola, M. (2023). The effectiveness of an Instagram-based educational game in a Bachelor of Nursing course: An experimental study. *Nurse Education in Practice*, *70*, 103656. https://doi.org/10.1016/j.nepr.2023.103656

Rubel, S. K., Miller, J. W., Stephens, R. L., Xu, Y., Scholl, L. E., Holden, E. W., Stroud, L. A., & Volk, R. J. (2010). Testing the Effects of a Decision Aid for Prostate Cancer Screening. *Journal of Health Communication*, *15*(3), 307–321. https://doi.org/10.1080/10810731003686614

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, *46*(2), 137–150. https://doi.org/10.1037/h0062958

Spence, J. C., & Blanchard, C. (2001). Effect of Pretesting on Feeling States and Self-Efficacy in Acute Exercise. *Research Quarterly for Exercise and Sport*, *72*(3), 310–320. https://doi.org/10.1080/02701367.2001.10608966

Spence, J. C., Burgess, J., Rodgers, W., & Murray, T. (2009). Effect of pretesting on intentions and behaviour: A pedometer and walking intervention. *Psychology & Health*, *24*(7), 777–789. https://doi.org/10.1080/08870440801989938

Spirito, A., Overholser, J., Ashworth, S., Morgan, J., & Benedict-Drew, C. (1988). Evaluation of a Suicide Awareness Curriculum for High School Students. *Journal of the American Academy of Child & Adolescent Psychiatry*, *27*(6), 705–711. https://doi.org/10.1097/00004583-198811000-00007

Swanborn, M. S. L., & De Glopper, K. (1999). Incidental Word Learning While Reading: A Meta-Analysis. *Review of Educational Research*, *69*(3), 261–285. https://doi.org/10.3102/00346543069003261