

# Undertaking Experiments in Social Sciences: Sequential, Multiple Time Series Designs for Consideration

Huy P. Phan<sup>1</sup> · Bing H. Ngu<sup>1</sup>

Published online: 28 April 2016

© Springer Science+Business Media New York 2016

**Abstract** In social sciences, the use of stringent methodological approaches is gaining increasing emphasis. Researchers have recognized the limitations of cross-sectional, non-manipulative data in the study of causality. True experimental designs, in contrast, are preferred as they represent rigorous standards for achieving causal flows between variables. The Solomon four-group design, for example, is ideal for its positioning to account for, and factor out, confounded influences of predictors on outcomes. However, in daily life settings, it is often difficult to emulate true experimental conditions. Identified limitations include financial resources, logistic difficulties, time constraint, and small sample sizes in social science research settings. There are, of course, other experimental designs that are noteworthy for consideration. Time series and single-case designs, quasi in nature, are effective alternatives for educators and researchers to consider in their research foci. This article examines the different experimental designs that may be implemented in naturalistic classroom settings. In particular, one important inquiry of our theoretical discussion pertains closely to the conceptualization of two innovative designs that we have made, consequently as a result of our research development and examination of the literature: a *sequential, multiple time series multi-group design* and a *multi time series, multi-group single-case design*. These experimental designs are innovative and enable comparisons for *within* and *between* differences under different experimental conditions.

**Keywords** True experimental designs · Quasi experimental designs · Multiple time series · Motivational research

---

✉ Huy P. Phan  
hphan2@une.edu.au

<sup>1</sup> School of Education, The University of New England, Armidale, NSW 2351, Australia

## Introduction

One area of inquiry in motivational research focuses on the establishment of causal flows between variables. For example, some researchers are interested in the causal impact of appropriate instructional designs on the improvement of problem solving skills in mathematics learning (Ngu et al. 2015a; Ngu and Yeung 2013). Other researchers are interested in understanding the causal relationship between self-concept and academic performance (Marsh and Yeung 1997, 1998; Skaalvik and Rankin 1998). Often the case, from a quantitative perspective, the use of experiments and correlational studies to validate causal directions is quite pronounced. Experimental studies are known to be more stringent and enable sound justification for confirming methodological issues of causality (Rogosa 1979). Cross-sectional and longitudinal designs, involving the use of non-manipulative data, may sometimes serve the purpose, but they are somewhat limited and do not enable the confirmation of causal determination of one variable onto another (e.g.,  $X \rightarrow Y$ ). However, it is often difficult for researchers to undertake true experimental works in many real-life settings (e.g., in classrooms within a school), given a variety of logistic difficulties, limited financial resources, time constraint, etc. Quasi experiments, as an alternative, may offer compensatory measures and provide grounding for researchers to investigate causal patterns in the social sciences.

The purpose of this article is to revisit existing experimental designs and to introduce an innovative quasi experimental design that we have recently conceptualized and implemented in our own research studies. Our *sequential, multiple time series design*, quasi experimental in nature, focuses on sequential experimental treatments (e.g., self-efficacy at time 1, self-concept at time 2, etc.) with multiple cohorts of individuals from a longitudinal point of view (e.g., 3 cohorts  $\times$  2 treatments  $\times$  6 observations). This structured design (e.g.,  $O_1 \rightarrow O_2 \rightarrow O_3 \rightarrow X_1 \rightarrow O_4 \rightarrow O_5 \rightarrow O_6 \rightarrow X_2 \rightarrow O_7 \rightarrow O_8 \rightarrow O_9$ ) is advantageous, especially in terms of cross-cohort comparisons of treatments (e.g., treatment  $X_1$  for group A versus group B) and growth modeling of individuals' behavioral and motivational patterns across time (Bollen and Curran 2006; McArdle and Nesselroade 2003). We will demonstrate why there is credence to consider this time series design for usage in social science research.

As part of our discussion, for research that has the purpose of juxtaposing the effects of two or more treatments, we will also illustrate a counterbalanced method to our proposed sequential, multiple time series design. This extended design is more interesting because it enables researchers to consider the methodological issue of the sequencing of multiple treatments (e.g., treatment  $X_1 \rightarrow$  treatment  $X_2$  versus treatment  $X_2 \rightarrow$  treatment  $X_1$ ). Sequencing, in this sense, is a pervasive matter that may often influence the final interpretation of effects. Hence, is it essential to ascertain that the final outcome would not be different whether treatment  $X_n$  preceded treatment  $X_y$ , or vice versa? Our discussion that follows is anticipated to shed further insights into the operational nature of quasi experimental designs.

## Establishment of Causality: Theoretical Overview

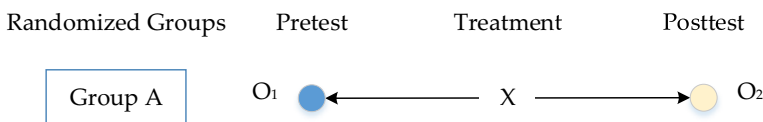
Research in social sciences has applied quantitative methodological approaches. Common approaches include *cross-sectional*, *longitudinal*, *quasi-experimental*, and *experimental* approaches. The main focus of inquiry, in this case, entails the establishment of associative patterns in relationships and the possibility of causal flows (Marsh and Yeung 1997; Rogosa 1979). In educational contexts, the notion of causal establishment between motivational

variables, in particular, is of relevance and importance to educators and researchers. For example, in the context of academia, researchers have often queried the extent to which heightened self-efficacy beliefs (Bandura 1986, 1997) would cause an improvement in performance outcomes in a subject matter. This inquiry (e.g., self-efficacy  $\rightarrow$  academic achievement) is significant for potential in-class interventions to enhance learning quality and academic achievements from a positive perspective. From a deficit point of view, validation of causality in motivational research may assist in the identification of extraneous influences that cause work-avoidance behaviors and detrimental outcomes (e.g., anxiety: Pajares and Kranzler 1995).

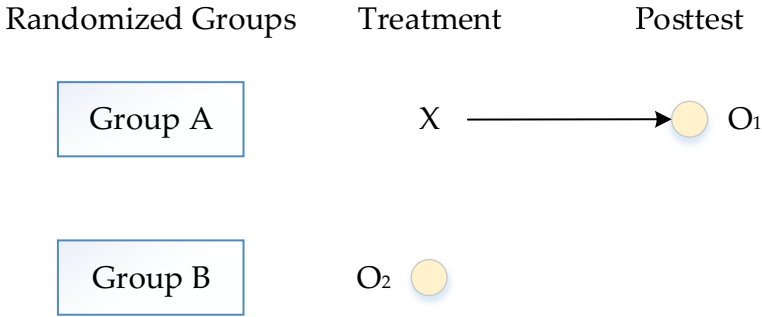
One methodological issue, which is relatively pervasive in the social sciences areas, is the affirmation of causality in associations. Early career researchers and researchers with methodological inexperience, often the case, make unwarranted assumptions and/or inaccurate interpretations of causal relationships between variables. This issue may be observed in descriptions and inference of empirical findings. For example, in a previous study of student learning, Phan (2008) conceptualized and inferred causality between various cognitive and motivational variables, using correlational data. Phan (2008) claimed: “Central to this investigation was the amalgamation of two major strands of research inquiry—the causal relationship between achievement goals, study strategies, and effort, and the causal relationships between study strategy and reflective thinking—within one theoretical and conceptual framework” (p. 336). This claim of causality is somewhat erroneous, given that the methodological design was non-experimental, in nature.

Experimental designs, in contrast, enable empirical validation of causality (Creswell 2003; Fraenkel and Wallen 2006). Does “X” cause “Y”? Even simplistic experimental designs, such as a *one-group pretest-posttest design* (Fig. 1), a *two-group design* (Fig. 2), a *pretest-posttest control group design* (Fig. 3), and/or a *posttest-only control group design* (Fig. 4), are sufficed in permitting researchers to study causal effects of cognitive, motivational, and social variables. For example, an experimental design would enable a test of the use of attributional feedback (Schunk 1982, 1983) in enhancing personal self-efficacy for academic learning (Bandura 1997). Of course, such simplistic experimental designs would not overcome unaccounted influences. We could not be definitive that, in this instance, a teacher’s *effort attributional feedback* actually accounted for an increase in academic self-efficacy scores across time (e.g., time 1  $\rightarrow$  time 2).

In exploring the causal impact of effort attributional feedback, consider a pretest-posttest control group design by Schunk (1982, 1983). It is for us to surmise the following: (i) a comparison in self-efficacy measures for both pretest and posttest (e.g., control versus experimental), and (ii) a comparison in self-efficacy measures from pretest to posttest, which could then determine the impact of effort attributional feedback as a treatment. Assume that a repeated *t* test showed a statistically significant difference at, say,  $p < .05$ . How can we be sure, though, that the increase in self-efficacy scores from pretest to posttest is the result of the



**Fig. 1** A one-group pretest-posttest design. Note: A one-group pretest-posttest design indicates no control grouping, and a particular unit (e.g., self-efficacy, academic performance) is measured twice, denoted as  $O_1$  and  $O_2$ . The treatment effect (TE) =  $O_2 - O_1$

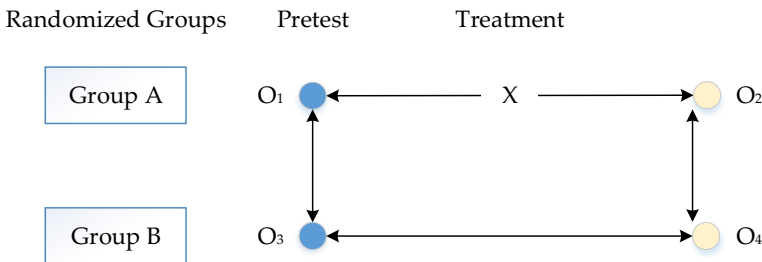


**Fig. 2** A two-group design. Note: A two-group design that shows measurements being made only after the experimental treatment, X, as denoted by O<sub>1</sub> and O<sub>2</sub>. The treatment effect (TE)=O<sub>1</sub> – O<sub>2</sub>

treatment, itself? It is plausible to suggest that, in this instance, other extraneous influences could have also accounted for this difference in mean scores. We will discuss this limitation in depth in the following sections.

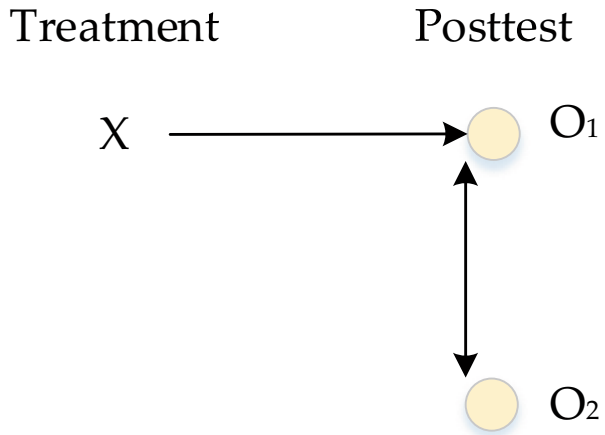
### True Experimental Designs: a Case for the Solomon Four-Group Design

The study of student learning in motivational contexts (e.g., an improvement in academic performance), as previous research studies have attested, may involve the inclusion of various cognitive and/or motivation processes. For example, situated within the framework of social cognition (Bandura 1986, 1997), one major inquiry undertaken by researchers has focused on the explanatory and predictive power of *personal self-efficacy beliefs* on academic achievement (Fast et al. 2010; Fenollar et al. 2007; Pajares and Miller 1994). Does self-efficacy for academic learning predict and/or cause a change in improvement outcomes? In a similar vein, in the area of cognitive load theory (Paas et al. 2003; Sweller 2005), one major emphasis has been on the impact of appropriate instructional designs to reduce cognitive load (Ngu et al. 2015b; Ngu and Yeung 2013). Extensive research, in this sense, has used correlational designs (e.g., cross-sectional data with Likert-scale inventories) to identify and/or to test out hypothesized associations between motivational variables (e.g., self-efficacy → performance-approach goals: Liem et al. 2008; academic buoyancy → self-efficacy: Martin et al. 2010; performance-approach goals → interest-based studying: Senko and Miles 2008).



**Fig. 3** Pretest-posttest control group design. Note: A pretest-posttest control group design indicates a control group, and a pretreatment measure is taken on each group. Posttest results between the two groups are possible, enabling researchers to determine the effectiveness of the experimental treatment, X. The treatment effect (TE)=(O<sub>2</sub> – O<sub>1</sub>) – (O<sub>4</sub> – O<sub>3</sub>)

**Fig. 4** Posttest-only control group design note: A posttest-only control design is similar to the pretest-posttest control group design, with the exception of premeasurement. The treatment effect (TE) =  $O_1 - O_2$

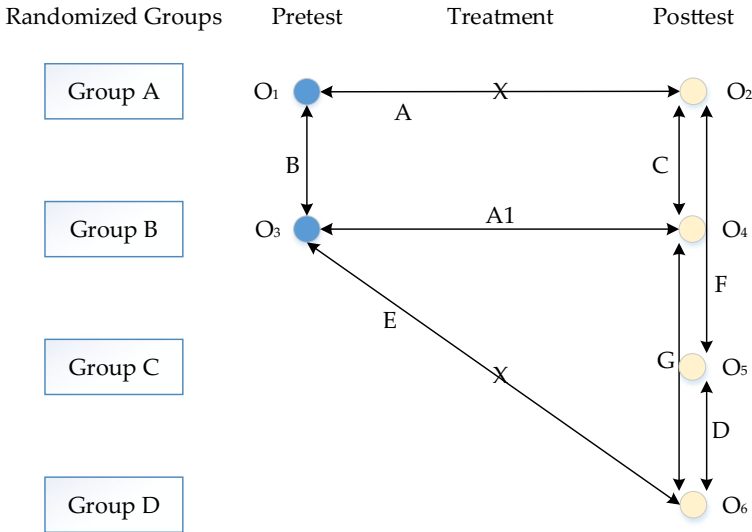


Research studies have, likewise, used simplistic experimental designs to explore relationships between variables. For example, in the early to mid-1980s, Schunk and colleagues (e.g., Schunk 1982, 1983; Schunk et al. 1987) conducted a number of experiments to determine the potent effects of *modeling* and *attributional feedbacks* on achievements and self-efficacy beliefs for academic learning (e.g., which type of attributional feedback is most effective?). Likewise, Ngu and colleagues (Ngu et al. 2015a, 2015b) have explored the effects of instructional practices on effective learning. Of an interesting nature, and consistent across many of the experimental research investigations, is the use of a pretest-posttest control group design, whereby an *intervention* group is compared to that of a *control* group (Fig. 3). This true experimental design with participants randomly assigned to experimental and control groups, although somewhat limited, is efficient and enables researchers to execute the following: (i) the randomization precludes effects of background differences; (ii) a pre-treatment measure is taken for each group; (iii) the treatment effect (TE) is measured as  $(O_2 - O_1) - (O_4 - O_3)$ ; (iv) it ensures that bias is eliminated via means of randomization; and (v) the experimental result is obtained by  $(O_2 - O_1) - (O_4 - O_3) = TE + IT$ , whereby IT is the interactive testing effect.

Analysis of the pretest-posttest control group design indicates that there are notable limitations, notwithstanding the fact that interactive testing effects are not controlled. For example, referring to our aforementioned example of effort attributional feedback, it is possible to affirm that (i)  $TE_{(Self-efficacy)} = (O_2 - O_1) - (O_4 - O_3)$ , where O is self-efficacy measure at time  $n$  and (ii)  $O_1$  versus  $O_3$  and  $O_2$  versus  $O_4$  comparisons are possible, enabling researchers to determine the effectiveness of the self-efficacy treatment, X. However, it is also possible to contend that students' inadvertent social comparison with others have strengthened their self-efficacy beliefs (Bandura 1997; Schunk and Hanson 1985, 1989). We cannot be certain then, in this analysis, that  $O_2 - O_4$  is in fact a derivative of the effort attributional treatment.

An alternative to this experimental design, of course, is the *Solomon four-group design*. The Solomon four-group design is more advantageous for the following indicators, as shown in Fig. 5. It enables researchers to determine the following:

1. Whether pretesting could, in fact, influence the results, as attested by the comparison of posttest results of group C and group D (i.e., denoted by line "D"). If the difference between the posttest results of groups C and D differs from the groups A and B difference



**Fig. 5** A Solomon four-group design note: A Solomon four-group design differs from other experimental designs, as it enables the factoring of confounding influences. A snapshot of this complex design shows the following: (i) Group A and group B represent a pretest-posttest control group design, and (ii) group C and group D represent a posttest-only control design. Source: <https://explorable.com/solomon-four-group-design?gid=1580>

- (i.e., denoted by line “C”), then a researcher can assume that the pretest has had some effect upon the results.
- 2. Whether any external factors could have caused a temporal distortion. This possibility is ensured by a comparison between the group B pretest and the group D posttest (i.e., denoted by line “E”).
- 3. Whether pretesting could have an effect on the actual treatment. If the posttest results between group A and group C (i.e., denoted by line “F”) differ, then the pretest has had some effect upon the treatment, and consequently, the experiment is flawed.
- 4. Whether pretesting itself could actually influence the outcome, independent in this case of the treatment. If the posttest results between group B and group D (i.e., denoted by line “G”) are significantly different, then the act of pretesting has influenced the overall results and refinement is therefore needed.

In our discussion of the Solomon four-group design, we refer to Schunk’s (1982) seminal research question: *does effort feedback enhance elementary school children’s academic self-efficacy beliefs for academic learning?* This research question aligns closely to Bandura’s (1997) social cognitive theory, whereby verbal discourse is a potent source of information in the formation of personal self-efficacy. Verbal praises such as “This is a scholarly piece of work, Thomas” (Phan 2012a) have been theorized to heighten one’s self-efficacy beliefs for academic learning. A number of researchers have used a non-experimental, correlational approach to study the impact of verbal discourse on the formation of academic self-efficacy. Analyses of responses to Likert-scale inventories (e.g., sources of the Mathematics Self-efficacy Scale: Lent et al. 1991) indicate, in some cases, the positive association between verbal persuasion and self-efficacy beliefs (Pajares et al. 2007; Phan 2012b). Methodologically, of course, correlational data without any experimental manipulation are somewhat limited

and do not permit the identification and confirmation of causal effects of the different types of verbal discourse (e.g., does effort attributional feedback cause elementary school children to feel good about themselves?).

From an experimental point of view, the Solomon four-group design provides much stronger grounding for establishing empirical yields. How would researchers go about achieving this feat of validating the mentioned hypothesis? Consider, say, 40 sixth grade students, randomized into one of four experimental groups ( $N=10$ ): group A=pretest-intervention-posttest, group B=pretest-posttest, group C=intervention-posttest, and group D=posttest only. The intervention, in this case, involves teachers providing effort attributional feedback to students (e.g., “I can see you’re working very hard, Mary”: Phan 2012a). Self-efficacy measures for academic learning (e.g., “I feel confident that I have the perceived competence to solve...”) are made at the pretest and posttest occasions. We note that, overall, there is a change in self-efficacy scores from pretest to posttest (e.g.,  $O_2-O_1$ , +ve in scores). The question then is whether intervention has, in fact, contributed to the change in self-efficacy scores. In terms of analyses and methodological implications, we can consider and postulate the following:

1. We note that the difference between posttest results of self-efficacy of groups C and D (i.e., denoted as line “D” in Fig. 5) does not differ from the groups A and B (i.e., denoted as line “C”) difference, and hence, we can assume that pretesting has had no influence.
2. Pretesting of the self-efficacy measure for group B and self-efficacy measure for group D posttest (i.e., denoted as line “E”) show no difference, and hence, there is no indication of temporal distortion.
3. The posttest self-efficacy measures for group A and group C (i.e., denoted as line “F”) indicate no difference, suggesting that pretesting does not have an effect.
4. The posttest self-efficacy measures for group B and group D (i.e., denoted as line “G”) do not differ, indicating that pretesting independent of intervention does not have any effect.

Furthermore, aside from the above mentioned, we note that the pretest self-efficacy measures for group A and group B (i.e., denoted as line “B”) do not differ, indicating no major problem with the process of randomization. Comparison of posttest self-efficacy measures from groups A and B (i.e., denoted as line “C”), however, indicated a statistically significant difference, suggesting the overall effectiveness of the intervention itself. Finally, attesting to the impact of effort attributional feedback provided by the teacher, there is increase in self-efficacy measures from pretest to posttest for group A (i.e., denoted as line “A”).

Overall then, from the aforementioned Solomon four-group design example, it can be determined that effort attributional feedback has made a causal contribution to the enhancement of self-efficacy beliefs. This testament is based on the use of experimental intervention and subsequent statistical analyses of the collected data (e.g., repeated  $t$  tests). Having said this, though, there is one major question for consideration: *how realistic is the use of the Solomon four-group design in school settings?* This experimental design, as we have noted, is significant for its methodological and statistical power, enabling researchers to address both internal and external validity issues. Its complexity, however, is a pervasive drawback and limits full implementation in social sciences research. Schools, for example, do not have flexibility or time to allow researchers to assign four groups randomly. In the senior years of schooling, in particular, such a design would interfere and disrupt normal classroom practices. In a similar vein, schools in regional and rural areas are relatively small in terms of student population and,

as such, may not necessarily have adequate sample sizes for the use of the Solomon four-group design. Like many other researchers, we have experienced this difficulty ourselves in previous experimental studies, whereby small sample sizes, time constraint, and limited financial resources available deterred us from undertaking true experimental research (e.g., using the Solomon four-group design). In the majority of cases, where appropriate, we would depend and use more simplistic experimental designs (e.g., pretest-posttest control group design).

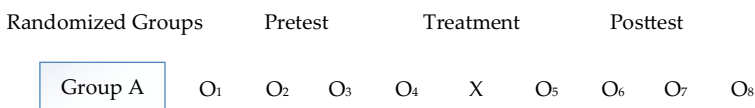
In sum, on the one hand, a Solomon four-group design is more stringent for its ability to reduce the influence of confounding variables (e.g., the potential influence of social comparison with other students). On the other hand, logistic difficulties in school contexts (e.g., time constraint) may also deter many educators from using the Solomon four-group experimental approach in their research investigations. *What are the alternatives* to the Solomon four-group design, especially when one wishes to explore the issue of causality in educational and/or social contexts? Quasi experimental designs, in this sense, are a possibility that could in fact be used in place of true experiments. In the following section, we discuss an example of a quasi-experimental approach that we have recently introduced and tried out in our own studies.

### Time Series Design: a Different Experimental Design

Given there are limitations to the use of the Solomon four-group experimental design, we could consider other quasi-experimental alternatives, notably, for example, the *time series* (Fig. 6) and *multiple time series* (Fig. 7) designs. A time series design, in this case, involves periodic measures (e.g., observations:  $O_n, n = 1, 2, \dots$ ) for a cohort of individuals, both before and after an intervention, X (Gottman et al. 1969). Time series experimental studies are important as they enable an examination of the intervention or treatment effect (Gribbons and Herman 1997; Morgan et al. 2000). For example, referring back to our inquiry, we could use time series experimental approach to study the effect of verbal discourse on elementary school children’s self-efficacy beliefs for academic learning. We could, in this instance, administer a self-efficacy scale once before the intervention (i.e.,  $O_1$ ), and then 1 month after the intervention (i.e.,  $O_2$ ), and again 3 months following the intervention (i.e.,  $O_3$ ). Interventions, X, in this case, may involve the injections of effort attributional feedback.

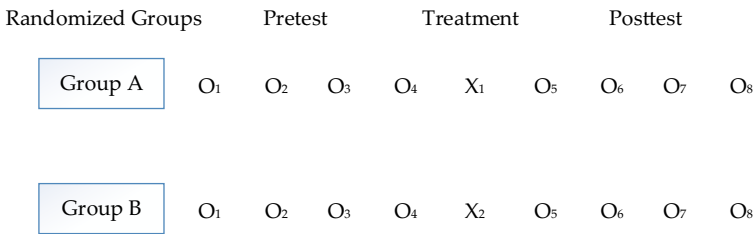
### Multiple Time Series Design: an Alternative Approach

A multiple time series design differs somewhat and adds a control group to the study for the purpose of comparison. Comparing the two groups, researchers could explore whether treatment has actually made an impact on subsequent measures ( $O_5$  to  $O_8$  in Fig. 7). It is also possible for the control group to receive some treatment ( $X_2$ ) to compare with  $X_1$  in group A. For example, referring to our previous discussion, a multiple time series study could include



**Fig. 6** Time series design note: X = treatment,  $O_1$  = measurement on dependent variable. A time series design that shows no control grouping.  $O_1, O_2, O_3,$  and  $O_4$  are considered as the baseline. A unit or a set of units can also be measured, posttest, on multiple occasions. A researcher may not necessarily have control over the timing of treatment or which test units are exposed to the treatment (e.g., self-efficacy versus academic performance)

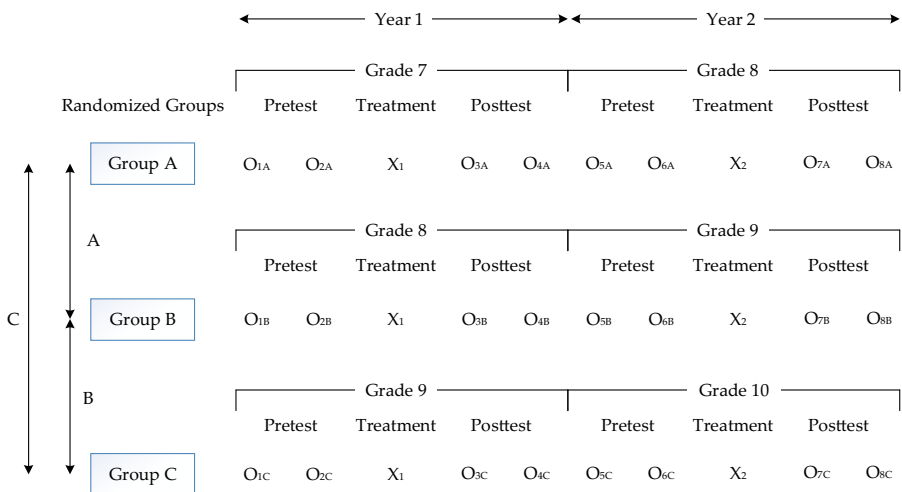




**Fig. 7** Multiple time series design note: X = treatment, O<sub>i</sub> = measurement on dependent variable. A multiple series design has the advantage of having a second group, which could be either be control or having a second experimental treatment

the following: (i) a cohort, group A, which receives both ability + effort attributional feedbacks as treatment (i.e., denoted as “X<sub>1</sub>”) and (ii) another cohort, group B, which receives only ability attributional feedback as treatment (i.e., denoted as “X<sub>2</sub>”). Four observation measures are made for baseline assessment, and likewise, four posttest measurements are made.

We recently modified the multiple time series design, shown in Fig. 8, to accommodate our own research studies. This modified design involves a three-group comparison that spans a 2-year period: group A versus group B versus group C. As a brief introduction, our focus of inquiry involved an examination of personal self-efficacy (Bandura 1997; Pajares 1996) and self-concept (Marsh 1990; Shavelson and Marsh 1986) as potential enhancers of academic performance in mathematics learning. This postulation coincides with previous theoretical tenets and findings regarding the potency of self-efficacy and self-concept in human agency. Empirical evidence (e.g., see Bong and Skaalvik 2003; Pietsch et al. 2003), as attested previously, indicates that both self-constructs positively associate with various educational outcomes (e.g., self-efficacy → academic buoyancy,  $\beta = .22, p < .001$ ; Martin et al. 2010; self-efficacy → performance outcome,  $\beta = .35, p < .05$ ; Pajares and Kranzler 1995).



**Fig. 8** A sequential, multiple time series multi-group design note: X = treatment, O<sub>i</sub> = measurement on dependent variable. This sequential, multiple time series multi-group design differs from the multiple time series design with its stipulation of three groups. There are two experimental treatments that are sequenced in a structured order, namely, X<sub>1</sub>, year 1 → X<sub>2</sub>, year 2

Given the consistency and clear evidence shown, we decided to explore both academic self-efficacy and self-concept as causal determinants of students' learning experiences (e.g., problem solving skills) and cognitive load imposition (e.g., a self-efficacious student is likely to rate low mental effort as an indicator of cognitive load on a problem). This inquiry is of significance for its major emphasis on the amalgamation of both non-cognitive (e.g., self-efficacy: Bandura 1997; self-concept: Shavelson and Marsh 1986) and cognitive (e.g., cognitive load imposition: Sweller et al. 2011) components in the teaching and learning processes.

To explore the research questions posed (e.g., self-efficacy  $\rightarrow$  academic achievement), we developed a conceptual framework that consequently resulted in the proposition of a *sequential, multiple time series multi-group design*: three groups  $\times$  2 years two treatments. As shown in Fig. 8, the project spanned the course of 2 years, denoted as year 1 and year 2. All three groups, grade 7, grade 8, and grade 9, started in year 1 so that by the end of year 2, grade 7 students would have completed grade 8, grade 8 students would have completed grade 9, and grade 9 students would have completed grade 10. This structured sampling is innovative and enables developmental inference via means of cross-sectional (e.g., grade 7 versus grade 8 at time 1) and sequential (e.g., grade 7  $\rightarrow$  grade 8 for group A) comparisons. There were two treatments that we implemented, sequentially in year 1 and year 2:  $X_1$  = self-efficacy in year 1 and  $X_2$  = self-concept in year 2 (e.g., grade 7 students received self-efficacy treatment in year 1 and received self-concept treatment in year 2, when they were in grade 8). Methodologically, our proposed experimental design enables examination of the following:

1. Between-group cross-sectional comparisons of different educational levels for self-efficacy and self-concept treatments, namely, grade 7 versus grade 8 versus grade 9 for self-efficacy treatment at time 1, and grade 8 versus grade 9 versus grade 10 for self-concept treatment at time 2. This examination enables inference and validation of the self-efficacy ( $X_1$ ) and self-concept ( $X_2$ ) treatment effects (TE) at time 1 and time 2, respectively. In this case, we have the following for consideration: (i)  $TE_{(X_1\text{-group A})} = O_{3A} - O_{2A}$ ,  $TE_{(X_1\text{-group B})} = O_{3B} - O_{2B}$ , and  $TE_{(X_1\text{-group C})} = O_{3C} - O_{2C}$  for self-efficacy and  $TE_{(X_2\text{-Group A})} = O_{7A} - O_{6A}$ ,  $TE_{(X_2\text{-group B})} = O_{7B} - O_{6B}$ , and  $TE_{(X_2\text{-group C})} = O_{7C} - O_{6C}$  for self-concept, (ii) cross-sectional comparisons of TEs( $X_1, X_2$ ), such as  $TE_{(X_1\text{-group A})}$  versus  $TE_{(X_1\text{-group B})}$  versus  $TE_{(X_1\text{-group C})}$  at time 1 and  $TE_{(X_2\text{-group A})}$  versus  $TE_{(X_2\text{-group B})}$  versus  $TE_{(X_2\text{-group C})}$  at time 2. Note, of course, that we also have the following measurements, for example,  $O_{4A} - O_{3A}$  and  $O_{8A} - O_{7A}$  for group A,  $O_{4B} - O_{3B}$  and  $O_{8B} - O_{7B}$  for group B, and  $O_{4C} - O_{3C}$  and  $O_{8C} - O_{7C}$  for group C for examining sustainability.
2. Within-group longitudinal comparisons across a 2-year period, namely, grade 7, self-efficacy treatment (year 1)  $\rightarrow$  grade 8, self-concept treatment (year 2), grade 8, self-efficacy treatment (year 1)  $\rightarrow$  grade 9, self-concept treatment (year 2), and grade 9, self-efficacy treatment (year 1)  $\rightarrow$  grade 10, self-concept treatment (year 2). In this case, across the 2 years, for each group, we would be able to find  $TE_{(X_2)} - TE_{(X_1)}$  for groups A, B, and C, whereby  $TE_{(X_1)}$  = self-efficacy treatment at time 1 and  $TE_{(X_2)}$  = self-concept treatment at time 2. For this purpose, it is important to consider the following:  $(O_{8A} - O_{7A}) - (O_{4A} - O_{3A})$  for group A,  $(O_{8B} - O_{7B}) - (O_{4B} - O_{3B})$  for group B, and  $(O_{8C} - O_{7C}) - (O_{4C} - O_{3C})$  for group C. Evidence as such would indicate two possible patterns. Using group C as an example, there may be (i) an increase in mean outcome scores (e.g., achievement) for  $(O_{8C} - O_{7C}) - (O_{4C} - O_{3C})$ , indicating the potency of self-

concept over that of self-efficacy, or (ii) a decrease in mean achievement scores for  $(O_{8c} - O_{7c}) - (O_{4c} - O_{3c})$ , indicating the potency of self-efficacy over that of self-concept.

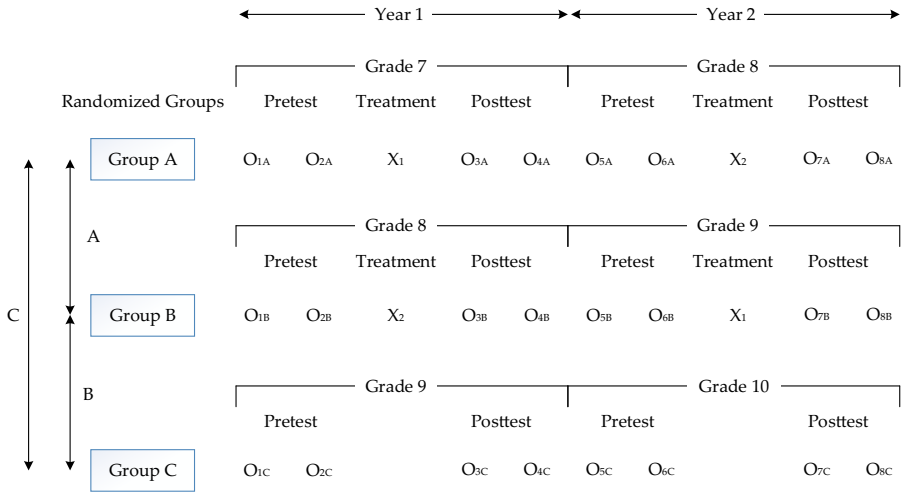
Our sequential, multiple time series multi-group design reflects some innovative insights for consideration. Notably, there is justification to warrant usage of this quasi-experimental approach to the following advantages: (i) inference of possible developmental changes regarding an experimental treatment (e.g., the effectiveness of self-efficacy treatment in its causal prediction of students' optimal best achievement in different grades), using a cross-sectional comparative approach (e.g., comparing posttest measures for two educational groups,  $O_{1A}$  versus  $O_{1B}$ ), and (ii) explore comparative effectiveness of different treatments for a cohort (e.g., does self-efficacy treatment surpass that of self-concept in causing an improvement in academic performance for grade 7 students?) over the course of time (e.g., comparing  $X_1 \rightarrow O_3$  versus  $X_2 \rightarrow O_7$ ). The advantage of this experimental inquiry is that it aligns closely with existing quasi-experimental tenets and does not require the use of control grouping. Furthermore, this sequential, multiple time series multi-group design may also accommodate small sample sizes (e.g., school in rural/remote areas that has a limited sample size) and enable in-depth longitudinal examination of individuals' learning across time.

The examination of students across 2 years of academic study provided information regarding effectiveness and potency of different experimental treatments on various educational outcomes (Phan et al. 2016). An inspection of the results indicated, interestingly, that similar patterns (e.g.,  $O_{3A} - O_{2A}$  with treatment  $X_1$  (i.e., an increase in score) for group A is similar to that  $O_{3B} - O_{2B}$  with treatment  $X_1$  for group B) were observed between the three groups (denoted in the subscripts). Within analyses, similarly, evidence of effectiveness of the two experimental treatments is shown. For example, in relation to grade 7 students, we noted that academic performance outcome measured after treatment  $X_1$  ( $O_{2A} \rightarrow O_{3A}$ ) and treatment  $X_2$  ( $O_{6A} \rightarrow O_{7A}$ ) increased from the pretest measures. Interestingly, however, the difference between  $(O_{7A} - O_{6A})$  and  $(O_{3A} - O_{2A})$  was not statistically significant, indicating that the two experimental treatments did not differ in their effectiveness (i.e., neither  $X_1 > X_2$  nor  $X_2 > X_1$ ).

## Extending the Time Series Design

Having discussed our recent implementation of a sequential, multiple time series design, we now propose an alternative to quasi-experimental approach for consideration. This alternative sequential, multiple time series multi-group design, depicted in Fig. 9, shows three cohorts of secondary school students: group A consists of grade 7 students, group B consists of grade 8 students, and group C consists of grade 9 students in year 1 of the study. The experimental study, for the sake of illustration, spans the course of 2 years and involves two pretest and two posttest measures in year 1, and likewise, the same pattern in year 2. Each cohort of students, in this case, experienced a transitioning from one educational level to another (e.g., group A: grade 7 in year 1  $\rightarrow$  grade 8 in year 2). There are, however, two major differences between this design and our previous design: (i) group C does not receive any experimental treatment, and (ii) there is a counterbalanced emphasis, whereby group A received self-efficacy treatment ( $X_1$ ) in year 1 followed by self-concept ( $X_2$ ) in year 2 (i.e.,  $X_1 \rightarrow X_2$ ), whereas group B received self-concept ( $X_2$ ) in year 1 followed by self-efficacy ( $X_1$ ) in year 2 (i.e.,  $X_2 \rightarrow X_1$ ).

Why is this alternative worthy of consideration? In our previous use of the sequential, multiple time series design, we noted one major caveat: the order of treatments across time (e.g.,  $X_1 \rightarrow X_2$ ) does play a part in shaping participants' responses, consequently as a result of



**Fig. 9** An alternative sequential, multiple time series multi-group design note: X = treatment, O<sub>i</sub> = measurement on dependent variable. This alternative sequential, multiple time series multi-group design differs from the sequential, multiple time series design with its stipulation of a control group. The sequencing of experimental treatments is also counterbalanced, namely, X<sub>1</sub>, year 1 → X<sub>2</sub>, year 2 for group A and X<sub>2</sub>, year 1 → X<sub>1</sub>, year 2 for group B

fatigue and/or influences from outside factors. For example, in relation to the empirical literature (e.g., Bong and Skaalvik 2003; Pajares and Miller 1994; Pietsch et al. 2003), it has been noted that both self-efficacy and self-concept are distinctive self-constructs, despite the fact that they both share similar attributes. What is unclear, however, is whether self-efficacy should precede self-concept (i.e., X<sub>1</sub> → X<sub>2</sub>) or whether self-concept should precede self-efficacy (i.e., X<sub>2</sub> → X<sub>1</sub>) for optimal intervention effects. There is no research that we know of, at present, that has explored this avenue of inquiry regarding the effectiveness of different self-belief treatments within one experiment (i.e., self-efficacy → self-concept versus self-concept → self-efficacy).

By using a counterbalanced design, we contend that it is possible to explore the extent to which *sequencing of experimental treatments could adversely affect the final results*. In our previous research investigation, depicted in Fig. 8, we stipulated a specific pattern, whereby self-efficacy treatment in year 1 was followed by self-concept in year 2. Self-efficacy is presumably more potent in predictive power, given its specific and contextual nature (Bandura 1997; Pajares 1996), as opposed to self-concept, which is domain-specific and involves social comparisons (Bong and Skaalvik 2003). Despite this theoretical tenet, the question of the sequencing of self-efficacy and self-concept treatments is unanswered and requires further development. To address this experimental issue of sequencing of treatments, our proposed counterbalanced methodological design addresses the following: (i) identify if crossover effects are, indeed, asymmetrical (e.g., treatment X<sub>1</sub> → treatment X<sub>2</sub> being stronger than that of X<sub>2</sub> → treatment X<sub>1</sub>, or vice versa); (ii) determine the effectiveness of sequencing of experimental treatments by comparing with a control group, for example, group A (e.g., X<sub>1</sub> → X<sub>2</sub>) versus group C (control) and group B (e.g., X<sub>2</sub> → X<sub>1</sub>) versus group C (control)—this comparison would enable us to consider the importance of the ordering of treatments—and (iii) determine the effectiveness of a treatment for different educational levels, for example, X<sub>1</sub> for grade 7 in year 1 versus X<sub>1</sub> for grade 9 in year 2.

Notwithstanding the limitations of quasi-experimental designs, it is possible for us then to determine, comparatively, the effectiveness of sequencing the two types of self-beliefs. This inquiry is of significance for its educational potential. Optimal academic performance and learners' personal well-being experiences at school (e.g., daily functioning: Phan and Ngu 2015; social integration: Van Damme et al. 2002) that stem from quality learning are important educational outcomes for consideration. The specific sequencing of self-efficacy and self-concept enhancing interventions may lead to the enhancement of different educational outcomes. Without an optimal experimental design to delineate the specific effects of each treatment at various stages of intervention, the relative effects of treatments will be hard to identify. For other researchers to replicate our approach, we have provided a mapping of the sequencing of self-efficacy and self-concept treatments across 2 years in Table 1. As shown in the table, the following educational outcomes are measured across eight occasions ( $O_1 \rightarrow O_8$ ): *academic performance*, *cognitive load imposition*, *liking for school experience*, and *daily functioning*. In essence, Table 1 shows two experimental conditions that could cause changes in the four educational outcomes, for example,  $X_1 \rightarrow X_2 \rightarrow O$  and  $X_2 \rightarrow X_1 \rightarrow O$ . By comparing these with group C, we would be able to tease out which condition would be more potent for improving educational outcomes? Hence, Table 1 serves to summarize a practical application of the design to investigate a substantively important research question.

In summary, the alternative sequential, multiple time series design that we proposed is significant for its emphasis on the following: (i) ensuring a sound methodological approach to the study of causality without utilizing true experimental designs (e.g., Solomon four-group); (ii) provide opportunities for identification and validation of longitudinal trajectories of variables (e.g., academic performance), taking into consideration the impact of experimental treatments; and (iii) provide opportunities for identification and counterbalancing of the impact of sequencing of experimental treatments. Having said this, however, we also note that our proposed alternative sequential, multiple time series design has a number of limitations. Firstly, in naturalistic classroom settings, small sample sizes may limit us from having a control group, especially when we acknowledge the multiple time points of data collection (e.g.,  $O_1 \rightarrow O_6$ ). If this is the case, then it is unfeasible for us to conclude and generalize whether an experimental intervention (e.g., the use of self-efficacy) has an impact. Secondly, time series designs, in general, are relatively stringent in terms of multiple occasions of data collection. This "longitudinal" approach, regardless of timeframe, faces an important problem of attrition.

### Data Analyses for Consideration

Ultimately, drawn from the aforementioned proposition, the issue for consideration here entails possible complex, stringent analyses of repeated data. A pretest-posttest control group design, for example, may involve a simple repeated  $t$  test (e.g.,  $Mn_{(O_1)} - Mn_{(O_2)}$ , where  $Mn$  = mean score) to determine whether treatment X, say self-efficacy (Bandura 1997), has made an impact on improvement of academic performance ( $O_1 - O_2$ ). This statistical approach, methodologically, is somewhat simplistic, in nature. Our alternative sequential, multiple time series design, as proposed, is more advantageous, enabling the use of latent growth modeling (LGM) (Bollen and Curran 2006; McArdle and Nesselroade 2003). LGM, an expansion of structural equation modeling (SEM), is more rigorous and provides grounding for analyses of longitudinal trajectories of variables.

Consider, in this instance, an experimental group, group A, that has been subject to the following: (i) receiving treatment  $X_1$ , self-efficacy, and  $X_2$ , self-concept, and (ii) observational

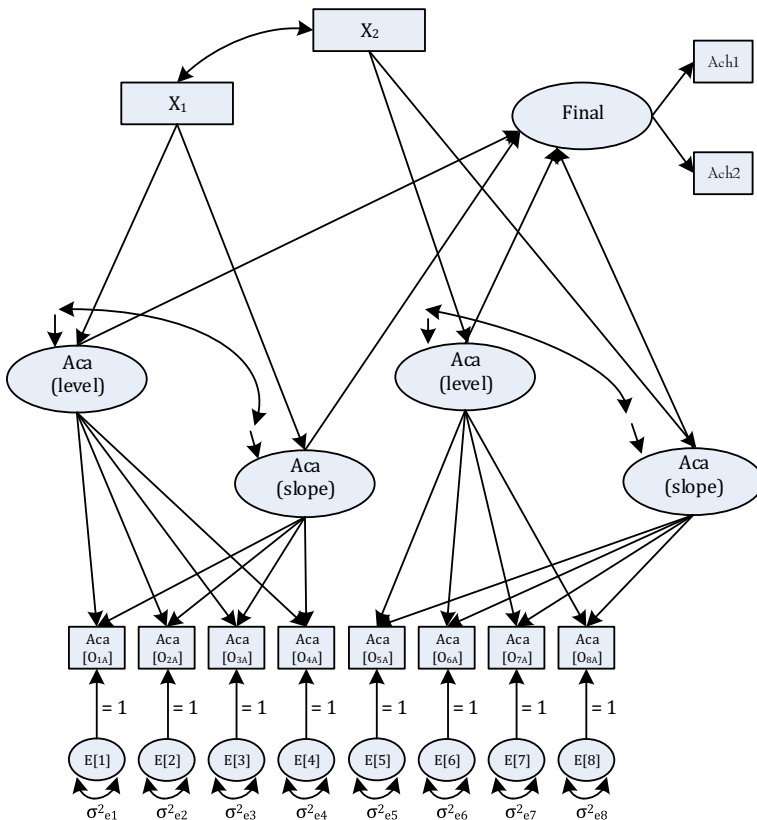
**Table 1** Example of sequencing of self-efficacy and self-concept treatments on educational outcomes

	Year 1				Year 2				
	O <sub>1</sub>	O <sub>2</sub>	O <sub>3</sub>	O <sub>4</sub>	O <sub>5</sub>	O <sub>6</sub>	O <sub>7</sub>	O <sub>8</sub>	
Group A	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Treat X <sub>2</sub> Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily
Group B	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Treat X <sub>1</sub> Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily
Group C	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily	Control Aca Per Cog Load Liking Daily	Control Aca Per Cog Load Liking Daily	Aca Per Cog Load Liking Daily

*Aca Per* academic performance, *Cog Load* cognitive load imposition, *Liking* liking for school, *Daily* daily functioning, *Treat X<sub>1</sub>* self-efficacy treatment, *Treat X<sub>2</sub>* self-concept treatment

measurements,  $O_{1A}$  and  $O_{2A}$  (pre- $X_1$ ) and  $O_{3A}$  and  $O_{4A}$  (post- $X_1$ ), and  $O_{5A}$  and  $O_{6A}$  (pre- $X_2$ ) and  $O_{7A}$  and  $O_{8A}$  (post- $X_1$ ). Multiple occasions of data collection, especially more than three times, are of significance for allowing us to identify longitudinal trajectories and growth of variables, in this case, academic performance, understanding of problem solving, engagement, etc. Crucial, though, in this statistical approach is the possible inclusion of experimental treatments (e.g.,  $X_1$  and  $X_2$ ) as extraneous variables that could explain and account for growth patterns in variables (e.g., academic performance). Figure 10 shows an example of a possible growth analysis that we could undertake to determine the extent to which an experimental treatment,  $X_1$  and/or  $X_2$ , could explain for growth patterns in terms of slope and level.

In relation to this example, we could focus on the following growth analyses: (i) the tracing of longitudinal trajectory of academic performance, accounted for by treatment  $X_1$  (Note: dummy coded—1 (i.e., receiving treatment) and 2 (i.e., no treatment)); (ii) the tracing of longitudinal trajectory of academic performance, accounted for by treatment  $X_2$ ; (iii) the tracing of longitudinal trajectory of academic performance, accounted for by both treatments  $X_1$  and  $X_2$ ; and (iv) the impact of longitudinal trajectory of academic performance, accounted for by both treatments  $X_1$  and  $X_2$ , on a final outcome (e.g., final grade). The possible latent growth analyses, attempting to discern and/or identify longitudinal trajectories of variables (e.g., academic performance), in this case, may involve two contrasting permutations: the



**Fig. 10** Latent growth modeling of alternative sequential, multiple time series multi-group design note:  $X$  = treatment,  $O_i$  = measurement on dependent variable. *Aca* academic performance, *E* error

inclusion of treatment(s) as an extraneous variable (e.g.,  $O_1 \rightarrow O_2 \rightarrow X_1 \rightarrow O_3 \rightarrow O_4$ ) versus the exclusion of treatment(s) as an extraneous variable (e.g.,  $O_1 \rightarrow O_2 \rightarrow O_3 \rightarrow O_4$ ).

LGM, compared to other multivariate techniques, is more advantageous as it enables us to study the notion of change in variables (Bollen and Curran 2006; McArdle and Nesselroade 2003). Change may, of course, depict a number of permutations and possibilities: *linear, non-linear, positive, negative*, etc. Importantly, of course, with this study of change, we are able to identify factors that could account, mitigate, optimize, and enhance longitudinal changes—for example, does the impact of an intervention involving self-efficacy (e.g., using verbal discourse) account for non-linear changes in academic performance of year 9 mathematics? Having said this, though, LGM is a sophisticated cause modeling technique that requires fulfillment of a number of stringent and rigorous criteria (e.g., appropriate sampling in terms of subject/parameter ratio). Limited resources (e.g., small sample size), in this analysis, may deter educators and researchers from using this complex statistical approach.

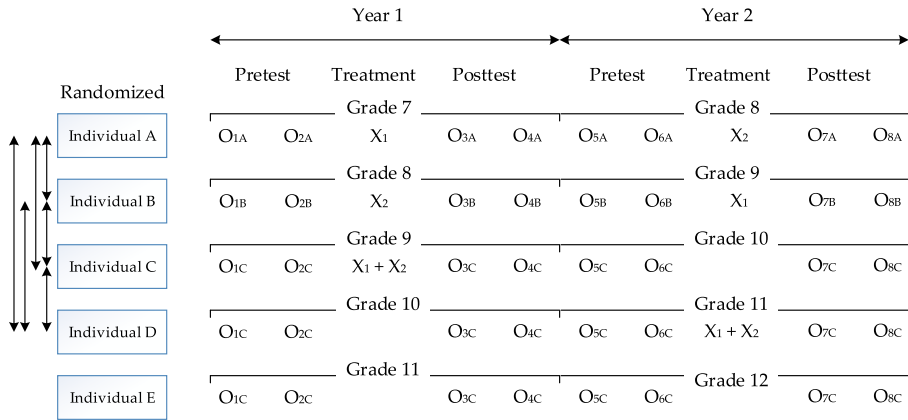
### **An Alternative to a Sequential, Multiple Time Series Design: a Single Case Design?**

Our alternative, multiple time series design is effective for the reasons that we outlined. Having said this, however, it is also a testament that this proposed experimental design has one major drawback, namely, the issue of attrition and possible extensive resources that may be needed. We acknowledge this limitation, especially in light of the fact that some schools in rural and remote areas may not necessarily have adequate sample sizes to implement a three-group sequential, multiple time series design.<sup>1</sup> Differently from our methodological approach, a number of scholars (Hammond and Gast 2010; Horner et al. 2005; Manolov et al. 2014) in the area of applied research in social sciences have recommended the use of a *single-case experimental design*. Single-case experimental research is prominent in the area of special education and “has proven particularly relevant for defining educational practices at the level of the individual learner” (Horner et al. 2005, p. 165).

Does a single-case experimental design have methodological merits for educators and researchers to consider? We contend that the single-case experimental design, combined with our proposed sequential, multiple time series design, may in fact yield an innovative alternative for in-class research development. Figure 11 depicts a proposed *sequential, multiple time series single-case design*. Adopting from our previous alternative sequential, multiple time series design (Fig. 9), the proposed sequential, multiple time series single-case design indicates the following: (i) five individuals, sequentially chosen (e.g., individual A in grade 7 in year 1, individual B in grade 8 in year 1, etc.) subject to one of five experimental conditions, namely, individual A with treatment  $X_1$  (year 1)  $\rightarrow$  treatment  $X_2$  (year 2), individual B with treatment  $X_2$  (year 1)  $\rightarrow$  treatment  $X_1$  (year 2), individual C with both treatments  $X_1 + X_2$  in year 1, individual D with treatments  $X_1 + X_2$  in year 2, and individual E with no treatment; (ii) identify if crossover effects are, indeed, asymmetrical for an individual (e.g., treatment  $X_1 \rightarrow$  treatment  $X_2$  being stronger than that of  $X_2 \rightarrow$  treatment  $X_1$ , or vice versa); (iii) determine the effectiveness of sequencing of experimental treatments by comparing with an individual who receives both treatments, simultaneously, and with an individual who does not receive any treatment, for example, individual A (e.g.,  $X_1 \rightarrow X_2$ ) versus individual C (e.g.,  $X_1 + X_2$ ) versus individual

<sup>1</sup> Note: this is also an issue that was raised by one of the reviewers in the first draft of this manuscript.

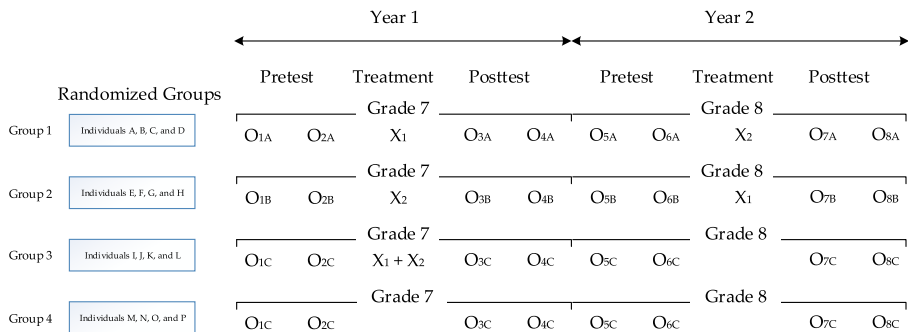




**Fig. 11** A proposed sequential, multiple time series single-case design note: X = treatment,  $O_t$  = measurement on dependent variable. This proposed sequential, multiple time series single-case design differs from the sequential, multiple time series multi-group designs for its comparison of individuals, and not comparison of groups. The sequencing of experimental treatments is also counterbalanced, namely,  $X_1$ , year 1  $\rightarrow$   $X_2$ , year 2 for individual A and  $X_2$ , year 1  $\rightarrow$   $X_1$ , year 2 for individual B

E (Control), and individual B (e.g.,  $X_2 \rightarrow X_1$ ) versus individual C (e.g.,  $X_1 + X_2$ ) versus individual E (control)—this comparison would enable us to consider the importance of combination as well as the ordering of treatments—and (iv) determine the effectiveness of a treatment for different educational levels, for example,  $X_1$  for grade 7 in year 1 versus  $X_1$  for grade 9 in year 2.

It is also possible, of course, for us to consider both within- and between-individual comparisons for individuals who are of the same educational level (i.e., not sequentially structured). This experimental approach is similar to our previous sequential, multiple time series single-case design, with the exception of individuals of one educational level that may span the course of a year, 2 years, etc. Between-individual comparison, for example, may likewise involve four experimental conditions, namely,  $X_1$ ,  $X_2$ ,  $X_1 + X_2$ , or no treatment for four year 7 students, for example. How does single-case experimental research hold up,



**Fig. 12** A proposed combination of a multiple time series, multi-group single-case design note: X = treatment,  $O_t$  = measurement on dependent variable. This proposed combination of a multiple time series, multi-group single-case design emphasizes on the importance of between-group, between-individual and within-group, within-individual comparisons. The sequencing of experimental treatments is also counterbalanced, namely,  $X_1$ , year 1  $\rightarrow$   $X_2$ , year 2 for individual A and  $X_2$ , year 1  $\rightarrow$   $X_1$ , year 2 for individual B

overall, especially in terms of theoretical, methodological, and empirical contributions? Single-case research addresses the shortcomings of limited resources (e.g., financial constraint) and enables educators to focus on the study of experimental interventions in naturalistic settings over the course of time. This focus is paramount, especially for individuals who may have extreme special needs and require an immediate intervention (e.g.,  $X_1 \rightarrow O \rightarrow X_1$ ). Despite the fact that single-case studies provide methodological grounding for the establishment of cause and effect, it is still noted though that this experimental design has a number of flaws, namely, (i) the issue of generalization to a wider population since the focus of inquiry is based on experimentation and observation of an individual; (ii) data analyses that rely on the use of visualizations (Kennedy 2005; Lane and Gast 2014) to gather information regarding trend, level, and stability; and (iii) multiple observations of a participant can influence his/her responses.

### A Proposed Combination of a Multi-time Series, Multi-group Single-Case Design

In this final section of the article, we propose an alternative experimental design: a *multi-time series, multi-group single-case* experiment (Fig. 12). This proposed experimental design takes into consideration our previous discussion and considers examination of both groups (i.e., group 1, group 2, group 3, and group 4) and individuals (e.g., individual A, individual B, individual C, and individual D in group 1). This proposal is more innovative, methodologically, enabling us to focus on the following:

1. Between-group cross-sectional comparisons, for example, group 1 ( $X_1$  in year 1) versus group 2 ( $X_2$  in year 1) versus group 3 ( $X_1 + X_2$  in year 1) versus group 4 (control)
2. Between-individual, but within-group cross-sectional comparisons, for example, individual A, group 1 ( $X_1$  in year 1) versus individual B, group 1 ( $X_1$  in year 1) versus individual C, group 1 ( $X_1$  in year 1) versus individual D, group 1 ( $X_1$  in year 1)
3. Between-individual and between-group cross-sectional comparisons, for example, individual A, group 1 ( $X_1$  in year 1) versus individual E, group 2 ( $X_2$  in year 1) versus individual I, group 3 ( $X_1 + X_2$  in year 1) versus individual M, group 4 (control)
4. Within-individual and within-group longitudinal comparisons across a 2-year period, for example, individual A, group 1, year 7,  $X_1$  treatment (year 1)  $\rightarrow$  individual A, group 1, year 8,  $X_2$  treatment (year 2); individual E, group 2, year 7,  $X_2$  treatment (year 1)  $\rightarrow$  individual E, group 2, year 8,  $X_1$  treatment (year 2); individual I, group 3, year 7,  $X_1 + X_2$  treatment (year 1)  $\rightarrow$  individual I, group 3, year 8, no treatment (year 2); and individual M, group 4, year 7, no treatment (year 1)  $\rightarrow$  individual M, group 4, year 8, no treatment (year 2)
5. Within-group longitudinal comparisons across a 2-year period, for example, group 1, year 7,  $X_1$  treatment (year 1)  $\rightarrow$  group 1, year 8,  $X_2$  treatment (year 2); group 2, year 7,  $X_2$  treatment (year 1)  $\rightarrow$  group 2, year 8,  $X_1$  treatment (year 2); group 3, year 7,  $X_1 + X_2$  treatment (year 1)  $\rightarrow$  group 3, year 8, no treatment (year 2); and group 4, year 7, no treatment (year 1)  $\rightarrow$  group 4, year 8, no treatment (year 2)

The multi-time series, multi-group single-case experimental design provides a balance in terms of addressing a number of methodological and statistical issues. More importantly, however, the proposed integration of having both individual- and group-based levels of sampling is integral to the testing of between and within comparisons. We contend that this experimental design is sound, logically, and may provide complementary results in the study of in-class interventions.

## Conclusion

Theorizing and research design go hand in hand. To build cutting-edge theory, we need cutting-edge research design. The study of associative and causal patterns between variables has, to date, involved the use of cross-sectional, longitudinal, and experimental designs. Correlational analyses of cross-sectional and longitudinal data, as attested in numerous quantitative studies, are advantageous methodologically, for establishing associative patterns for further validation. True experimental research is a potent avenue for investigative measures of causality—for example, does heightened self-efficacy cause an improvement in problem solving performance in mathematics? There are different experimental approaches for usage, especially in the area of social sciences. Often the case, because of school-based and in-class limitations (e.g., small sample population), there is tendency for educators to use a two-group, pretest-posttest experimental design to explore causal relationships (e.g., Ngu and Yeung 2013; Ngu et al. 2015a). This experimental inquiry, as we have explained, poses a number of limitations, hence resulting in other alternatives for consideration.

There is acknowledgment that, in general, there are caveats in educational contexts, which may prevent the use of true experiment designs. It is possible, in this analysis, for educators and researchers to consider on alternative quasi-methodological approaches. In this article, we have proposed and explored a number of different possibilities, for example, the alternative sequential, multiple time series multi-group design and the multiple time series, multi-group single-case design. We recommend that researchers explore this line of inquiry (e.g., the impact of self-efficacy on cognitive load imposition) or other similar inquires in methodological investigations. Social sciences research cannot, in this sense, rely predominantly on cross-sectional, non-manipulative data. With the advent of sophisticated statistical techniques, it is important for researchers to consider the use of stringent methodological approaches in their research undertakings. One clear possibility, in this case, may include growth analyses of quasi-experimental data measured across time.

### Compliance with Ethical Standards

**Ethical Approval** This article does not contain any studies with human participants performed by any of the authors.

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Bandura, A. (1986). *Social foundations of thought and action: a social cognitive theory*. Englewood Cliffs: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: W. H. Freeman & Co.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: a structural equation perspective*. Hoboken: Wiley.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: how different are they really? *Educational Psychology Review*, 15(1), 1–40.
- Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed methods approaches*. Thousand Oaks: Sage Publications, Inc.
- Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010). Does math self-efficacy mediate the effect of perceived classroom environment on standardized math performance? *Journal of Educational Psychology*, 102(3), 729–740.

- Fenollar, P., Román, S., & Cuestas, P. J. (2007). University students' academic performance: an integrative conceptual framework and empirical analysis. *British Journal of Educational Psychology*, 77(Pt 4), 873–891. doi:10.1348/000709907X189118.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). New York: McGraw Hill.
- Gottman, J. M., McFall, R. M., & Barnett, J. T. (1969). Design and analysis of research using time series. *Psychological Bulletin*, 72(4), 299–306.
- Gribbons, B., & Herman, J. (1997). *True and quasi-experimental designs*. Washington: ERIC Clearinghouse on Assessment and Evaluation.
- Hammond, D., & Gast, D. L. (2010). Descriptive analysis of single subject research designs: 1983–2007. *Education and Training in Autism and Developmental Disabilities*, 45(2), 187–202.
- Homer, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, 71, 165–179.
- Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn & Bacon.
- Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: brief review and guidelines. *Neuropsychological Rehabilitation*, 24(3–4), 445–463.
- Lent, R. W., Lopez, F. G., & Bieschke, K. J. (1991). Mathematics self-efficacy: sources and relation to science-based career choice. *Journal of Counseling Psychology*, 38(4), 424–430. doi:10.1037//0022-0167.38.4.424.
- Liem, A. D., Lau, S., & Nie, Y. (2008). The role of self-efficacy, task value, and achievement goals in predicting learning strategies, task disengagement, peer relationship, and achievement outcome. *Contemporary Educational Psychology*, 33(4), 486–512. doi:10.1016/J.Cedpsych.2007.08.001.
- Manolov, R., Gast, D. L., Perdices, M., & Evans, J. J. (2014). Single-case experimental designs: reflections on conduct and analysis. *Neuropsychological Rehabilitation*, 24(3–4), 634–660.
- Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: theoretical and empirical justification. *Educational Psychology Review*, 2, 77–172.
- Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: structural equation models of longitudinal data. *Journal of Educational Psychology*, 89(1), 41–54.
- Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: gender differences in the development of math and English constructs. *American Educational Research Journal*, 35(4), 705–738.
- Martin, A. J., Colmar, S. H., Davey, L. A., & Marsh, H. W. (2010). Longitudinal modelling of academic buoyancy and motivation: do the 5Cs hold up over time? *British Journal of Educational Psychology*, 80(3), 473–496.
- McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: research methods in psychology* (pp. 447–480). New York: Wiley.
- Morgan, G. A., Gliner, J. A., & Harmon, R. J. (2000). Quasi-experimental designs. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39(6), 794–796.
- Ngu, B. H., & Yeung, A. S. (2013). Algebra word problem solving approaches in a chemistry context: equation worked examples versus text editing. *The Journal of Mathematical Behavior*, 32(2), 197–208.
- Ngu, B. H., Chung, S. F., & Yeung, A. S. (2015a). Cognitive load in algebra: element interactivity in solving equations. *Educational Psychology*, 35(3), 271–293.
- Ngu, B. H., Yeung, A. S., & Phan, H. P. (2015b). Constructing a coherent problem model to facilitate algebra problem solving in a chemistry context. *International Journal of Mathematical Education in Science and Technology*, 46(3), 388–403.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent development. *Educational Psychologist*, 38, 1–3.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543–578.
- Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem solving. *Contemporary Educational Psychology*, 20(4), 426–443.
- Pajares, F., & Miller, D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: a path analysis. *Journal of Educational Psychology*, 86(2), 193–203.
- Pajares, F., Johnson, J., & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English*, 42(1), 104–120.
- Phan, H. P. (2008). Unifying different theories of learning: theoretical framework and empirical evidence. *Educational Psychology*, 28(3), 325–340. doi:10.1080/01443410701591392.
- Phan, H. P. (2012a). Informational sources, self-efficacy, and achievement: a temporally displaced approach. *Educational Psychology*, 32(6), 699–726. doi:10.1080/01443410.2012.708320.
- Phan, H. P. (2012b). Relations between informational sources and self-efficacy: a developmental approach. *Educational Psychology*, 32(1), 81–105. doi:10.1080/01443410.2011.625612.
- Phan, H. P., & Ngu, B. H. (2015). Validating personal well-being experiences: a quantitative approach. *Education*, 136(1), 34–52.

- Phan, H. P., Ngu, B. H., & Williams, A. (2016). Introducing the concept of Optimal Best: theoretical and methodological contributions. *Education, 136*(3), 312–322.
- Pietsch, J., Walker, R., & Chapman, E. (2003). The relationship among self-concept, self-efficacy, and performance in mathematics during secondary school. *Journal of Educational Psychology, 95*(3), 589–603. doi:10.1037/0022-0663.95.3.589.
- Rogosa, D. (1979). Causal models in longitudinal research: rationale, formulation, and interpretation. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behaviour and development* (pp. 263–302). New York: Academic Press.
- Schunk, D. H. (1982). Effects of effort attributional feedback on children's perceived self-efficacy and achievement. *Journal of Educational Psychology, 74*, 548–556.
- Schunk, D. H. (1983). Ability versus effort attributional feedback: differential effects on self-efficacy and achievement. *Journal of Educational Psychology, 75*, 848–856.
- Schunk, D. H., & Hanson, A. R. (1985). Peer models: influence on children's self-efficacy and achievement. *Journal of Educational Psychology, 77*(3), 313–322.
- Schunk, D. H., & Hanson, A. R. (1989). Influence of peer-model attributes on children's beliefs and learning. *Journal of Educational Psychology, 81*(3), 431–434.
- Schunk, D. H., Hanson, A. R., & Cox, P. D. (1987). Peer-model attributes and children's achievement behaviors. *Journal of Educational Psychology, 79*(1), 54–61.
- Senko, C., & Miles, K. M. (2008). Pursuing their own learning agenda: how mastery-oriented students jeopardize their class performance. *Contemporary Educational Psychology, 33*(4), 561–583. doi:10.1016/J.Cedpsych.2007.12.001.
- Shavelson, R. J., & Marsh, H. W. (1986). On the structure of self-concept. In R. Schwarzer (Ed.), *Anxiety and cognitions*. Hillsdale: Lawrence Erlbaum.
- Skaalvik, E. M., & Rankin, R. J. (1998). *Self-concept, self-efficacy and achievement in mathematics: a test of causal relations*. San Diego: Paper presented at the Annual Meeting of the American Educational Research Association.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.
- Van Damme, J., De Fraine, B., Van Landeghem, G., Opdenakker, M.-C., & Onghena, P. (2002). A new study on educational effectiveness in secondary schools in Flanders: an introduction. *School Effectiveness and School Improvement, 13*(4), 383–397.