TERRY BYERS

A QUASI-EXPERIMENTAL AND SINGLE-SUBJECT RESEARCH APPROACH AS AN ALTERNATIVE TO TRADITIONAL POST-OCCUPANCY EVALUATION OF LEARNING ENVIRONMENTS

INTRODUCTION

The past decade has seen a resurgence in the literature concerning the effectiveness of physical learning environments. A worrying characteristic of this research has been a lack of rigorous experimental methodology (Brooks, 2011; Painter et al., 2013). This may be due to the difficulties associated with randomly assigning students and staff to specific settings and problems associated with accounting for the complex intervening variables that come to play within the educative experience (Byers, Imms, & Hartnell-Young, 2014). Nevertheless Brooks (2011) lamented the disparity between the ‘potential’ of contemporary learning spaces and the seeming lack of empirical evidence concerning the impact of these spaces on teaching and learning.

This is especially true for Secondary school settings. A prominent literature review by Blackmore, Bateman, O'Mara, and Loughlin (2011) indicated that there is little empirical evidence addressing the connections between physical learning spaces, teacher pedagogical practice and student learning experiences and outcomes. Blackmore et al. suggested that much of the research has focused on the design and physical attributes of buildings. Indeed, there is strong empirical evidence connecting the effects of the physical attributes of the built environment for example air quality, temperature, and noise on student learning (Higgins, Hall, Wall, Woolner, & McCaughey, 2005). Yet, like Uptis (2009), Higgins et al. (2005) argued that there is little known about how and why the physical attributes of a given space influence the teaching and learning process.

This chapter explores the development of a quasi-experimental and single-subject research approach to investigate the effectiveness of physical learning environments in primary and secondary school settings. This approach is put forward as an alternative to traditional methodologies used in the post-occupancy evaluation of learning environments.

CONTEXT

In general, the few empirical studies that have been conducted around the effectiveness of learning spaces have bolstered claims about the positive effects of technologically enabled or active learning spaces on student learning outcomes. Much of this work has centred on two influential projects: North Carolina State University's Student-Centred Activities for Large Enrolment Undergraduate Programs (SCALE-UP) project, and Massachusetts Institute of Technology's Technology Enabled Active Learning (TEAL) project. Both incorporated a redesign of the course, curriculum and pedagogies, in addition to redesigning the

learning spaces in which introductory physics courses were held (Dori & Belcher, 2005). These projects found that students in these ‘studio physics’ learning environments had lower failure rates and higher levels of conceptual understanding compared to students taking the same course in traditional lecture-based environments (Dori et al., 2003). However, Brooks (2011) is of the opinion that both studies suffered from methodological issues that have detracted from their effectiveness in linking attributes of the learning space with student learning.

The multiple changes to assessment, curriculum and pedagogical approaches, in addition to the changes made to the formal learning space in the SCALE-UP and TEAL projects is a source of methodological concern. Brooks (2011) was of the opinion that these multiple changes translated to a lack of sufficient control of confounding variables. For instance, Brooks (2011) identified that neither study accounted for a host of exogenous factors related to student body composition (i.e. student cognitive ability) and endogenous factors related to different instructors and therefore pedagogical approaches and changes in assessment methodologies between the ‘experimental’ and control groups. It has been argued that this lack of control obscured the relationships between changes to the learning environment and consequential effects on learning outcomes (Brooks, 2011).

It is suggested here that a more rigorous and systematic research design is required to empirically bolster the nascent link between contemporary learning spaces and effects on teaching and learning.

RESEARCH DESIGN

The objective of this paper is to outline a viable and methodologically robust research design that can quantitatively examine and evaluate the causal links between changes in the physical learning environment and the subsequent effects on teaching and learning. This focus on causality would suggest that randomised experimental studies which represent the ‘gold standard’ of systematic evidence would be the ideal approach to adopt (Clegg, 2005). However, the nuances and ethical considerations of the schooling and tertiary education environments rarely support the requisite random assignment and absolute variable control of a randomised experimental study (Shadish & Cook, 1999). An alternative, the subject of this paper, is a synthesis of quasi-experimental and Single Subject Research Design (SSRD) research methodologies.

Even though quasi-experimental and SSRD are defined as different methodologies, they are similar in their aim, approach and means of analysis. Both are well-established approaches to non-randomised intervention studies in the applied and clinical health sciences and have been used extensively in particular rehabilitation studies (e.g. Harris et al., 2006; Johnston, Ottenbacher, & Riechardt, 1995). Unlike randomised experimental studies, both approaches place greater emphasis on the design of the study rather than statistics alone to facilitate causal inference (Shadish & Cook, 1999). A key facet of the proposed combination is its ability to control the spuriousness effect/s of confounding variables and to then isolate and measure the effect of a single intervention (Coryn, Schröter, & Hanssen, 2009; Robson, 2011). Controlling all other confounding variables improves both within-subject variability and internal validity (Rassafiani & Sahaf, 2010). These improvements are essential to enhance the rigour and reliability

around the claimed causality between the intervention and desired outcomes (Harris et al., 2006; Mitchell & Jolley, 2012; West & Thoemmes, 2010).

METHODOLOGY AND METHODS

The collective work of Shadish, Cook, and Campbell (2002), Shadish and Cook (1999), Campbell and Stanley (1963) and Campbell (1957) is considered seminal in the fields of experimental and quasi-experimental design. An important theme of their work around the concept of causality is that ‘design rules, not statistics’ (Shadish and Cook, 1999). A thoughtful design is a proactive means of identifying and accounting for and the ‘plausible’ threats to internal validity and the spurious effect of confounding variables (West & Thoemmes, 2010). This ‘priori’ focus seeks to moderate alternative interpretations or competing explanations for the observed effects (Coryn et al., 2009; Shadish & Cook, 1999). At the heart of this synthesised quasi-experimental and SSRD approach is a strong design focus that moderates plausible threats to the validity of the study.

Quasi-experimental methods

Campbell and Stanley (1963) identified a number of research designs that can be classified as quasi-experiments. Like true experiments, quasi-experiments test a hypothesis about the effects of interventions (treatments) that have been actively employed or manipulated to achieve an effect (Shadish & Luellen, 2012). However, unlike true experiments, the quasi-experimental design skirts the requirements of random assignment in treatment and non-treatment comparisons (Mitchell & Jolley, 2012; Robson, 2011). This lack of randomised assignment is the major weakness of quasi-experimental design (Harris et al., 2006). The result, according to Johnston, Ottenbacher, and Reichardt (1995) is reduced internal confidence of outcomes.

To moderate the impact of this weakness, the ‘logic’ of the quasi-experimental research design is paramount to ensure that the intervention preceded (temporal precedence) and caused (covariance) the achieved effect, with no plausible alternative explanation (Shadish & Luellen, 2012). Campbell and Stanley (1963) suggested that these ‘alternative plausible explanations’, or *threats to internal validity*, can cast doubt on the determination and justification of causal effect. The key threats of internal validity identified by Cook and Campbell (1979) are history, maturation, testing, instrumentation, statistical regression, selection, experiential morality and diffusion of treatment. West and Thoemmes (2010) suggested that the identified plausible threats should directly shape the research design, collection methods and analysis techniques. These design implications support the claim of successful integration of quasi-experimental and SSRD to develop a viable and methodologically sound quantitative research approach suited to learning environments research.

Single-subject research design methods

A key facet of a SSRD is the selection of an individual, group or class of students, acting as their own control, baseline and unit of analysis (Cakiroglu, 2012; Horner, Swaminathan, & George, 2012). Comparing and contrasting each individual, group

or class against themselves negates between-subject variability (Horner et al., 2005). It should be noted that in the applied health sciences the unit of analysis is traditionally an individual. However, this selection can be difficult to facilitate in school settings due to ethical issues surrounding the requisite anonymity of individual students. However, Kinugasa, Cerin, and Hooper (2004) are of the opinion that a group of individuals can become the unit of analysis if the sample size is large enough to meet the statistical power requirements to detect a statistically significant effect. For example, to achieve an adequate statistical power (greater than 0.8) to conduct a two-tailed hypothesis test with probability level ($p = 0.05$) and estimated medium Cohen's d effect size ($d = 0.5$) requires a sample size of 128 (Faul, Erdfelder, Lang, & Buchner, 2007).

Underlying the quality and validity of a SSRD is the need to establish a stable baseline data set. Byiers, Reichle, and Symons (2012) described how the stable baseline phase was critical in establishing the benchmark against which the individual's/group's behaviour in subsequent conditions can be compared and contrasted against. An example of the importance and effect of baseline stability is evident in the Byers et al. (2014) study. The stable baseline data set was critical in attributing any statistically significant change in student perception of their learning experiences attributable to the intervention. The intervention in this study was the change in classroom layout from traditional to contemporary, with all other variables (i.e. teacher, class, subjects, etc.) kept constant. Figure 1 provides a visual comparison of the difference between an unstable (class 7.2) and stable (all other classes) baseline data set. The degree of change throughout the baseline data set for class 7.2 indicates a high within-subject variability. This signified the presence of unforeseen and extraneous confounding variables not accounted for in the initial research design (Byiers et al., 2012; Casey et al., 2012). This raised the problem of an inability to confidentially attribute the changes in student perception solely to the effect of the intervention (Creswell, 2005; Mitchell & Jolley, 2012; Shadish et al., 2002).

Domain B - Effect of the Formal Learning Space on Teacher Pedagogy

Figure 1. Byers et al. (2014) study and the effect of a stable baseline data set in the determination of statistically significant change

In this example, comparison of the class 7.2 data against the stable baseline data set for the other five classes made the range in which future data points fell more predictably. (Byers et al., 2012). This improved level of predictability indicated the lack of extraneous variables and therefore reduced within-subject variability. Therefore, the changes in student perceptions of their learning experiences can be attributed to the intervention (Creswell, 2005; Mitchell & Jolley, 2012; Shadish et al., 2002).

Data analysis techniques

The data from quasi-experimental studies is normally analysed through pre- and post-test comparison through inferential statistical analysis, while SSRD studies are evaluated through visual analysis (Beeson & Robey, 2006). The suggestion of a synthesised methodology extends to the synthesis of the methods of data analysis. By combining the strengths of visual analysis with quantitative effect size

calculations, the strengths of one analysis approach is able to respond to the limitation of the other. The aim of this combined approach is to improve the statistical rigour behind the conclusions made.

Visual analysis is a proven mechanism for observing changes in level, trend and variability within and between the baseline and intervention periods (Byiers et al., 2012). Bobrovitz and Ottenbacher (1998) claimed that visual analysis has a proven ability to identify and derive a functional relationship between the independent and dependent variables. Generally, visual analysis has centred on the use of single point analysis. However, the addition of 95% confidence intervals to group means, can identify both inter- and intra-intervention trends (Nourbakhsh & Ottenbacher, 1994). Baguley (2009) was of the opinion that the addition of the confidence intervals is a superior approach to single point analysis as it indicates the plausible range of values that the 'true' effect might take.

Given the identified strengths associated with visual analysis, its limitations are also well documented. Beeson and Robey (2006) and Kromrey and Foster-Johnson (1996) outlined a common criticism directed towards visual analysis is its subjective nature. This subjectivity can lead to interpreter disagreement and resultant concerns around external validity (Bobrovitz & Ottenbacher, 1998). It has been suggested that an additional layer of quantitative statistical analysis is required to mitigate the subjective nature of visual analysis (Johnston et al., 1995; Kromrey & Foster-Johnson, 1996).

This paper is recommending the use of effect size calculations, as the additional layer of quantitative analysis. This recommendation is supported by the research of Beeson and Robey (2006) and Kromrey and Foster-Johnson (1996). Both utilised the use of effect size calculations to justify the outcomes of the visual analysis. The application of Cohen's d (mean shift) has the ability to mitigate the potential of Type 1 errors and reduce the sole reliance on interpreter judgement (Beeson & Robey, 2006; Kromrey & Foster-Johnson, 1996; Shadish, Hedges, & Pustejovsky, 2014).

Findings and Results

Byers et al. (2014) provided an example of synthesised SSRD and quasi-experimental design. Their study employed such an approach to examine the difference in variables between two educational settings – 'traditional' classrooms, and 'New Generation Learning Spaces' (NGLS). The impetus behind this design was to address the paucity of systematic, empirical evidence, especially in Primary and Secondary school settings, connecting the impact of the learning space on teaching and learning (Blackmore et al., 2011; Brooks, 2011). It took its lead from the studies of Brooks (2011), Walker, Brooks, and Baepler (2011) and Whiteside, Brooks, and Walker (2010), all concerned with the University of Minnesota's Active Learning Classrooms (ALC) project. These studies applied a quasi-experimental design to measure the effect of the change of the learning space (single variable intervention) on teaching and learning in a tertiary setting. These designs were able to moderate the effects of confounding variables such as the instructor, curriculum and assessment.

Study design

The aim of the Byers et al. (2014) study was to determine if changing the learning space had any effect on students' learning experiences, and their levels of engagement. It utilised a SSRD in an attempt to control all factors (curriculum, student ability, class construction, assessment and the teacher) except for the 'intervention' (type of learning space). The intervention involved a retrofit of six classrooms, changed from a traditional classroom to the NGLS illustrated in Figure 2 below.

Figure 2. New Generation Learning Space (NGLS) layout (Byers et al., 2014)

A baseline/intervention (A/B) design determined the effect of the intervention of the change in learning space (independent variable) on student learning experiences and their levels of engagement (dependent variables). Student attitudinal data was collected through a repeated measures survey. Each of the six participating classes acted as their own baseline and unit of analysis. Student data was summed and treated as one subject across three baseline measures and four post-intervention measures. This number of measures was well within the parameters set by Vickers (2003) to ensure the desired statistical power (0.8).

Data collection and preparation

The Linking Technology, Pedagogy and Space (LPTS) instrument focused on measuring the effect of the spatial intervention on the dependent variables of teacher pedagogy, learning activities, and student engagement. The survey Likert-scale items assigned to underlying scales represented each of the dependent variables. The LPTS questions relating to student learning experiences were derived from elements of the Tamim, Lowerison, Schmid, Bernard, and Abrami (2011) longitudinal study. The Tamim et al. (2011) study was chosen as it focused on active and more student-centred learning. The survey instrument was based on the APA Learner-Centred principles developed by Lambert and McCombs (1988).

Questions relating to engagement were adapted from the *Motivated Strategies for Learning Questionnaire* (MSLQ) – a uni-dimensional student self-report questionnaire that focuses on motivational beliefs and self-regulated learning strategies domains (Fredricks et al., 2011; Pintrich & De Groot, 1990). The MSLQ has a proven track record in its use to evaluate instructional-based interventions (Fredricks et al., 2011; Pintrich & De Groot, 1990).

The study involved the participation of six Middle Years classes. The sample size ($n = 164$) represented a significant proportion of (97.1%) of students from the participating classes. Across the baseline and intervention data collection phases a high retention rate (96.7%) was achieved and met the parameters set by Vickers (2003) to ensure the desired statistical power (0.8). However, Jenson, Clark, Kircher, and Kristjansson (2007) recommend that a complete data set will ensure the statistical power is maintained.

To maintain the sample size, a mechanism to deal with the ‘missingness’ was implemented. It was initially assumed that any missing data was classified as Missing Completely at Random (MCAR), due to random factors such as student illness or appointments at data collection times. This assumption was verified by Little’s MCAR test score greater than 0.05 (0.94) (Peugh & Enders, 2004). This result enabled a missingness approach to produce a complete data set.

A complete data set was produced in each instance through the implementation of the Maximum Likelihood Estimation (MLE) process and the use of expectation-maximization algorithm (EM). MLE is a robust technique for estimating missing data as it does not discard or try to ‘fix’ or ‘fill in’ the data (i.e. mean substitution) like more traditional methods (Little & Rubin, 2002; Peugh & Enders, 2004). The EM searches the available data for the parameters that will yield the best fit to the observed data, with the inclusion of the incomplete cases assisting towards a more accurate estimation. The work of Peugh and Enders (2004) suggested that unlike mean substitution and linear regression, the MLE approach does not artificially truncate the variance and covariance around the mean. This truncation would unduly bias the visual analysis process by decreasing the spread of the 95% confidence intervals. This decrease in confidence intervals would increase the likelihood of increased instances of ‘false’ statistically significant differences or Type 1 errors.

The adequate sample size and high retention rate enabled post-hoc reliability analysis through Cronbachs Alpha. The aim was to determine the construct validity for each domain of the survey instrument. Cronbach’s Alpha for the summative score for each class in each of the domains was calculated based on the suggestions of Gliem and Gliem (2003). The Cronbach’s Alpha for the ‘Student Learning Experiences’ ($\alpha = 0.88$) and ‘Student Engagement’s ($\alpha = 0.86$) domains calculated coefficients indicated a very high level (when $\alpha > 0.80$) of internal consistency across the multiple items. This signified an adequate level of reliability for the purposes of this study (Gliem & Gliem, 2003).

Data analysis

Statistical analysis of student attitudinal data was undertaken through visual analysis, justified by Cohen’s d effect size calculations. A split-middle method of visual analysis derived a functional relationship between the independent and

dependent variables (Nourbakhsh & Ottenbacher, 1994). Inter- and intra-intervention trends were identified through the projection of baseline data trends through the intervention phase, along with the application of 95% confidence intervals of class means (Nourbakhsh & Ottenbacher, 1994). Bobrovitz and Ottenbacher (1998) suggested that this is equitable to statistical analysis such as *t*-tests, but more suited to sample sizes such as found in this study.

In the Byers et al. (2014) study, the process outlined by Horner et al. (2012) was utilised to determine 'significant' and 'non-significant' statistical differences through non-overlapping confidence intervals between the baseline and intervention phases. This approach incorporated the criterion of level, trend and variability within and between the baseline and intervention periods. The approach concluded that in five out of the six classes (Figure 1) there was a clear statistical difference in student attitudes associated with the change from a traditional to NGLS classroom (see Table 1). All six classes indicated a clear statistical difference in student engagement in the NGLS compared to the traditional classroom.

To circumvent the criticism that surrounded the potential subjective nature of visual analysis, the Byers et al. (2014) study justified its conclusion through the application of effect size calculations. The thresholds suggested by Cohen (1998) were employed to elicit if the degree of the effect size correlated to a statistically significant effect. The Cohen's *d* (mean shift) that was computed in the Byers et al. (2014) study is summarised in Table 1 below. Statistically significant differences identified through the visual analysis process correlated to large (0.8 to 1.3) to very large (greater than 1.3) effect sizes (Cohen, 1998). This is significant as effect sizes are considered the exact equivalent of *z*-scores (standard deviations above the mean) (Jenson et al., 2007). This would suggest that statistically significant differences identified through visual analysis were corroborated by an improvement of between 1 to 2 standard deviations when compared against the baseline data.

Table 1. Summary table provided in Byers et al. (2014) study that compared visual analysis and effect size calculations for student learning experiences and engagement

Class	<u>Student learning experiences</u>		<u>Student engagement</u>	
	Visual analysis	Cohen's <i>d</i> effect size	Visual analysis	Cohen's <i>d</i> effect size
7.1	Significant	1.50	Significant	1.63
7.2	Non-significant	1.04	Significant	1.38
8.1	Significant	1.01	Significant	1.34
8.2	Significant	1.31	Significant	1.73
8.3	Significant	1.35	Significant	1.16
8.4	Significant	2.01	Significant	2.48

The comparison between effect size calculations and the outcomes from the visual analysis resulted in some interesting findings. For example, class 7.2 achieved a slightly larger effect size in student learning experiences domain than class 8.1. Interestingly the visual analysis process identified only a statistically significant effect in class 8.1. This suggested that the ability for visual analysis to

distinguish between results based on changes in level, trend and variability throughout the baseline and intervention phases signifies its statistical robustness.

CONCLUSIONS

The aim of this synthesis of quasi-experimental and SSRD was to elicit initial empirical evidence around the effect of the formal learning space on school-age students. The Byers et al. (2014) study was able to moderate a number of confounding variables, such as the teachers, class composition and subject type. The analysis indicated that the method was able to correlate improvements in student attitudes around their learning experiences and engagement to the change from a traditional classroom to a NGLS. This study, while small in scale, provided findings that demonstrate causal linkages between contemporary schooling spaces and changes in teaching and learning. Furthermore, it validated the claim that a synthesised quasi-experiment and SSRD has the potential to be a statistically robust method for exploring this topic.

Quasi-experimental and SSRD research design do possess inherent limitations centred around threats to external validity through their predominant focus on internal validity (Cakiroglu, 2012). The focus on a single unit of analysis conducted under a particular set of conditions in a particular context can lead to difficulties in generalization to other contexts or settings (Creswell, 2005; Shadish et al., 2002). In addition, context specific gender, ethnic and socio-economic characteristics also present generalizability problems. Another threat to external validity is the ability to generalise findings to past and future situations, depending on the timing of the treatments (Creswell, 2005; Mitchell & Jolley, 2012).

Even though the synthesised research design has proved to be successful, direct and systematic replication across a wider population is needed (Cakiroglu, 2012). The generalizability of the results of this model across different and multiple sites is required to validate its methodology in an educational context. The use of multiple and different class-types as the unit of analysis is needed to encompass the full range of student abilities and prevent selection bias. The replication of this approach can then facilitate both within- and between-subject comparison (Cakiroglu, 2012; Shadish et al., 2002). This degree and depth of data analysis would be expected to further reinforce the power and reliability of this synthesised approach.

IMPLICATIONS FOR TEACHERS AND DESIGNERS

In the context of research in an educational context, the synthesis of a quasi-experimental and SSRD methodologies can provide a simple yet effective method of identifying the effect of an intervention. The key to this approach is a well-thought-out research design. Where and when possible, within the nuances of the schooling context, the design must focus on a singular intervention or variable change. The focus on both the means of data collection and methods of analysis can attempt to moderate the spuriousness effect/s of confounding variables.

This approach has the potential to add dramatically to current learning environment research. It can make use of conceptual frameworks that are central to the assertion that the physical learning environment can act as a mechanism to hinder or support teacher pedagogical practices and therefore student learning

experiences and outcomes. At this same time, this methodology can further generalise and validate the importance of the physical learning space. The empirical focus of this approach holds potential to illuminate the causal effects of contemporary learning environments on teaching and learning practices and outcomes.

REFERENCE LIST

- Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, 100, 603-617. doi: 10.1348/000712608X3771 17
- Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review*, 16(4), 161-169. doi: 10.1007/s11065-006-9013-7
- Blackmore, J., Bateman, D., O'Mara, J., & Loughlin, J. (2011). *Research into the connection between built learning spaces and student outcomes: Literature review*. Melbourne: Victorian Department of Education and Early Childhood Development Retrieved from http://www.eduweb.vic.gov.au/edulibrary/public/publ/research/publ/blackmore_learning_spaces.pdf.
- Bobrovitz, C. D., & Ottenbacher, K. J. (1998). Comparison of visual inspection and statistical analysis of single-subject data in rehabilitation research. *American Journal of Physical Medicine and Rehabilitation*, 77(2), 94-102.
- Brooks, D. C. (2011). Space matters: The impact of formal learning environments on student learning. *British Journal of Educational Technology*, 42(5), 719-726. doi: 10.1111/j.1467-8535.2010.01098.x
- Byers, T., Imms, W., & Hartnell-Young, E. (2014). Making the case for space: The effect of learning spaces on teaching and learning. *Curriculum and Teaching*, 29(1), 5-19. doi: 10.7459/ct/29.1.02
- Byiers, B. J., Reichle, J., & Symons, F. J. (2012). Single-subject experimental design for evidence-based practice. *American Journal of Speech-Language Pathology*, 21(4), 397-414. doi: 10.1044/1058-0360(2012/11-0036)
- Cakiroglu, O. (2012). Single subject research: Applications to special education. *British Journal of Special Education*, 39(1), 21-29.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54(4), 297-312. doi: 10.1037/h0040950
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research on teaching*. Chicago: Rand McNally.
- Casey, L. B., Meindl, J. N., Frame, K., Elswick, S., Hayes, J., & Wyatt, J. (2012). Current trends in education: How single-subject research can help middle and high school educators keep up with the zeitgeist. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 85(3), 109-116.
- Clegg, S. (2005). Evidence - based practice in educational research: A critical realist critique of systematic review. *British Journal of Sociology of Education*, 26(3), 415-428. doi: 10.1080/01425690500128932
- Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum: New Jersey.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Coryn, C. L. S., Schröter, D. C., & Hanssen, C. E. (2009). Adding a time-series design element to the success case method to improve methodological rigor an application for nonprofit program evaluation. *American journal of evaluation*, 30(1), 80-92. doi: 10.1177/1098214008326557
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (2nd ed.). Boston: Pearson.
- Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts? *The Journal of the Learning Sciences*, 14(2), 243-279.
- Dori, Y. J., Belcher, J., Bessette, M., Danziger, M., McKinney, A., & Hult, E. (2003). Technology for active learning. *Materials Today*, 6(12), 44-49. doi: 10.1016/S1369-7021(03)01225-2
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, 39(2), 175-191.
- Fredricks, J. A., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). Measuring student engagement in upper elementary through high school: A description of 21 instruments (pp. 88). Regional Educational Laboratory Southeast.
- Gliem, J. A., & Gliem, R. R. (2003). *Calculating, interpreting, and reporting cronbach's alpha reliability coefficient for likert-type scales*. Paper presented at the Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education, The Ohio State University, Columbus.
- Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association*, 13(1), 16-23. doi: 10.1197/jamia.M1749

- Higgins, S., Hall, E., Wall, K., Woolner, P., & McCaughey, C. (2005). The impact of school environments: A literature review. Newcastle, United Kingdom: University of Newcastle.
- Horner, 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(2), 165-179.
- Horner, R. H., Swaminathan, H. S., & George, S. K. (2012). Considerations for the systematic analysis and use of single-case research. *Education & Treatment of Children*, 35(2), 269.
- Jenson, W. R., Clark, E., Kircher, J. C., & Kristjansson, S. D. (2007). Statistical reform: Evidence-based practice, meta-analyses, and single subject designs. *Psychology in the Schools*, 44(5), 483-493.
- Johnston, M. V., Ottenbacher, K. J., & Reichardt, C. S. (1995). Strong quasi-experimental designs for research on the effectiveness of rehabilitation. *American Journal of Physical Medicine and Rehabilitation*, 74, 383-392.
- Kinugasa, T., Cerin, E., & Hooper, S. (2004). Single-subject research designs and data analyses for assessing elite athletes' conditioning. *Sports Medicine*, 34(15), 1035-1050.
- Kromrey, J. D., & Foster-Johnson, L. (1996). Determining the efficacy of intervention: The use of effect sizes for data analysis in single-subject research. *The Journal of Experimental Education*(1), 73-93.
- Lambert, N. N., & McCombs, B. L. (Eds.). (1988). *How students learn: Reforming schools through learner-centered education*. Washington, DC: APA Books.
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, N.J: Wiley.
- Mitchell, M. L., & Jolley, J. M. (2012). *Research design explained* (8th ed.). Belmon: Wadsworth Cengage Learning.
- Nourbakhsh, M. R., & Ottenbacher, K. J. (1994). The statistical analysis of single-subject data: A comparative examination. *Physical Therapy*, 74(8), 768-776.
- Painter, S., Fournier, J., Grape, C., Grummon, P., Morelli, J., Whitmer, S., & Cevetello, J. (2013). Research on learning space design: Present state, future directions *The Perry Chapman Prize*: Society for College and University Planning.
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525-556. doi: 10.3102/00346543074004525
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 82(1), 33-40. doi: 10.1037/0022-0663.82.1.33
- Rassafiani, M., & Sahaf, R. (2010). Single case experimental design: an overview. *International Journal of Therapy & Rehabilitation*, 17(6), 285-289.
- Robson, C. (2011). *Real world research: A resource for social scientists and practitioner-researchers* (3rd ed.). Chichester;Hoboken: Wiley-Blackwell.
- Shadish, W. R., & Cook, T. D. (1999). Comment-design rules: More steps toward a complete theory of quasi-experimentation. *Statistical Science*, 14(3), 294-300. doi: 10.2307/2676764
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of school psychology*, 52(2), 123-147. doi: 10.1016/j.jsp.2013.11.005
- Shadish, W. R., & Luellen, J. K. (2012). Quasi-experimental design. In J. L. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 539-550). United States of America: Routledge.
- Tamim, R. M., Lowerison, G., Schmid, R. F., Bernard, R. M., & Abrami, P. C. (2011). A multi-year investigation of the relationship between pedagogy, computer use and course effectiveness in postsecondary education. *Journal of Computing in Higher Education*, 23(1), 1-14. doi: 10.1007/s12528-010-9041-4
- Uptis, R. (2009). Complexity and design: How school architecture influences learning. *Design Principles and Practices: An International Journal*, 3(2), 1-14.
- Vickers, A. (2003). How many repeated measures in repeated measures designs? Statistical issues for comparative trials. *BMC medical research methodology*, 3(1), 22.
- Walker, J. D., Brooks, D. C., & Baepler, P. (2011). Pedagogy and space: Empirical research on new learning environments. *EDUCAUSE Quarterly*, 34(4). <http://www.educause.edu/ero/article/pedagogy-and-space-empirical-research-new-learning-environments>
- West, S. G., & Thoemmes, F. (2010). Campbell's and rubin's perspectives on causal inference. *Psychological methods*, 15(1), 18-37. doi: 10.1037/a0015917

Whiteside, A. L., Brooks, D. C., & Walker, J. D. (2010). Making the case for space: Three years of empirical research on learning environments. *EDUCAUSE Quarterly*, 33(3).

*Terry Byers,
Melbourne Graduate School of Education,
The University of Melbourne*