International Journal of Learning, Teaching and Educational Research
Vol. 24, No. 9, pp. 1-19, September 2025
https://doi.org/10.26803/ijlter.24.9.1
Received Apr 6, 2025; Revised May 31, 2025; Accepted Jun 4, 2025

Examining the Factors Behind Experimental Group Superiority in Quasi-Experimental Research: A Mixed-Methods Analysis of MA Theses

Sheikha Ali Al-Buraiki*
Sohar University
Sohar, Sultanate of Oman

Salama Saif Al-Siyabi
Ministry of Education
Muscat, Sultanate of Oman

Ali Salim Alghafri
Sohar University
Sohar, Sultanate of Oman

Abstract. When random assignment is not feasible in education research, quasi-experimental designs are used to examine the impact of instructional interventions on other variables. Notably, a recurring pattern in such studies is the outperformance of the experimental groups over the control groups, which raises questions about the validity and interpretation of these outcomes. This study investigated the factors that lead to the consistent outperformance of experimental groups over control groups in quasi-experimental research. By adopting a mixed-methods approach, the researchers analysed 100 MA theses of the Department of Curriculum and Instruction at Sultan Qaboos University in the Sultanate of Oman. This analysis was supported by a survey that was distributed to lecturers at four universities in the country to elicit their insights on issues that include selection bias, intervention effectiveness, implementation differences, and the Hawthorne effect, as potential contributors to this trend. Based on the survey responses of the lecturers, the primary contributing factor to the superior performance of experimental groups compared to control groups was the Hawthorne effect, followed by the effectiveness of the intervention and selection bias as secondary factors, accounting for the observed trend. The findings offer a comprehensive explanation of the underlying causes and present

*Corresponding author: Sheikha Ali Al-Buraiki; *shikaalbriki@gmail.com*

practical recommendations for improving the design and interpretation of quasi-experimental studies in education research.

Keywords: Quasi-experimental design; Sultan Qaboos University; Experimental group; Control group

1. Introduction

Quasi-experimental research designs refer to the type of research used to investigate the causal relationships between a treatment and an expected result or impact. Randomized controlled trials (RCTs) are considered to be the "gold standard" for measuring the causal effects of interventions. In RCTs, the differences between control and experimental groups can be attributed, with a high level of certainty, to the treatment. Hence, RCTs are "deemed the gold standard by which decisions related to the design of products are evaluated" (Jobson et al., 2024). However, quasi-experimental research differs from RCTs or true experimental research because it deals with existing groups. Here, the participants are assigned either to one or more experimental groups or to a control or comparison group. This means that full random sampling is not applied (Andrade, 2021; Cook & Campbell, 1979).

Quasi-experimental research designs are widely utilized in education research because of their practicality in settings where random assignment is either impossible or would be unethical (Andrade, 2021; Harris et al., 2006; Reichardt, 2019). These designs enable researchers to examine the effects of interventions in real-world educational environments, thereby providing valuable insights that can inform practice and policy. However, a recurring observation in quasi-experimental studies is that experimental groups often outperform control groups (Kim & Steiner, 2016; Shadish et al., 2002), which raises important questions about the validity and generalizability of such findings. Several reasons have been cited for the outperformance of experimental groups in quasi-experimental research (Trochim, 2006, 2020), namely selection bias, intervention effectiveness, the Hawthorne effect, differences in implementation, regression to the mean (RTM), and measurement timing.

2. Literature Review

Quasi-experimental research studies are informed by a pragmatic philosophical view, and acknowledge the complexity of real-world educational contexts, in which the completely random assignment of participants is often unfeasible or even unethical. Pragmatism encourages research that addresses practical problems, and which allows for methodological flexibility. This means that research from a pragmatic viewpoint focuses on what works in a specific context (Biesta, 2010). Thus, quasi-experimental research operates from a pragmatic approach that aims to determine the impact of an intervention on another variable in cases where full experimental control is not possible.

Random assignment of subjects ensures that the two groups are similar and that, if any differences exist, then they are due only to chance (Cook & Campbell, 1979). However, in the case of quasi-experimental studies, randomization is neither feasible nor ethical (Andrade, 2021; Siedlecki, 2020), and participants are not randomly assigned to experimental and control groups. This lack of randomization could result in selection bias (Andrade, 2021) and could mean that the groups differ at the outset of the study. In the absence of randomization, internal threats are introduced (Maciejewski, 2018). For example, individuals who choose to participate in the intervention (experimental group) might be more motivated or have other characteristics that predispose them to perform better than members in the control group.

Furthermore, the intervention being tested might genuinely be effective, leading to better outcomes in the experimental group. If the intervention is well-designed and targets specific skills, knowledge areas, or the behaviors it aims to improve, then it is likely to produce measurable improvements. Randomized controlled trails provide strong evidence of effectiveness of an intervention by producing measurable and meaningful differences among participants (Biddle & Curchin, 2024). Hence, researchers should, first, consider the feasibility of implementing RCTs to examine program effectiveness that "ensures that treatment and control group do not differ except for receipt of the intervention" (Scher et al, 2015, p. 1).

RCTs are considered to be the gold standard for evaluating the effectiveness of a program and quasi-experimental research studies may be subject to various confounding factors, among which selection bias (Biddle & Curchin, 2024). The effectiveness of an intervention in an experimental group can be influenced not only by the quality of the intervention itself but also by external psychological factors, such as the Hawthorne effect. While the selection involved in RCTs helps ensure that any differences in outcomes are the result of the intervention, the awareness of being observed can still affect the participant's performance, potentially confounding the results unless controlled through techniques such as blinding.

The Hawthorne effect refers to the tendency of people to perform differently when they know that they are being observed by a researcher, observer, or supervisors (Vannan, 2021). The first studies that gave rise to the name Hawthorne effect were undertaken at a telephone manufacturing factory at Hawthorne between 1924 and 1933, where an increase in productivity was observed among workers when they were supervised by their managers as part of a research program. Thus, based on the meaning of the Hawthorne effect, the participants in an experimental group might perform better simply because they know they are part of an experiment.

This heightened awareness can lead to increased motivation and effort and contribute to better outcomes. If participants act differently because they are aware of being part of an intervention or under observation, the impact of the intervention may be the result of their belief and awareness and not because of the treatment (Gillespie, 1993), which represents a threat to the internal validity of the experiment.

However, this phenomenon is not guaranteed. Levitt and List (2011) reanalyzed the original illumination experiment data of the Hawthorne plant and found that claims about the Hawthorne effect may have been overstated or incorrect. The data, previously thought to be lost, did not, as widely reported, show dramatic productivity increases. Instead, only subtle indications of behavior changes owing to observation were found, suggesting that the effect was less significant than originally believed. One way to prevent this from happening is through a process called blinding (also called masking), which means that participants are unaware of the treatment they are receiving (Seltman, 2013), and thus they do not alter their behaviors.

Sometimes, the experimental group receives more attention, resources, or support during the intervention, which may possibly affect the effectiveness of the treatment that is being implemented. In education, the effectiveness of an intervention can be affected by teacher differences during teaching (El Soufi & See, 2019), which can be attributed to personal teaching characteristics (Kelchtermans, 2009) and the teachers' beliefs about the theories that drive the way they teach (Mansour, 2009; Richardson, 2005). Teachers' teaching differs, not only in comparison to other teachers, but also in different teaching situations.

This fact is reflected in a personal account of his teaching experience by Palmer (2007, p. 10), who explains, "I have taught thousands of students … But when I walk into a new class, it is as if I am starting over the techniques I have do not disappear, but neither do they suffice." This difference in the implementation of an intervention because of teaching differences can lead to different performance by the experimental group than the control group. In some cases, if the participants of the experimental group were chosen because of extreme scores (e.g., those who performed poorly on a pre-test), their scores may naturally improve over time because of statistical factors, rather than program intervention.

Kahan et al. (2015) found that, of 152 eligible published trials, only 3% employed randomized participant selection. Properly conducted RCTs ensure an equitable distribution of both identified and unidentified variables that the treatment groups may encounter. In most cases, an improvement could be mistakenly attributed to the intervention itself. Thus, the observed improvement is often attributed to RTM, which is often misinterpreted as an impact of the implemented intervention (Trochim, 2020). RTM is a statistical phenomenon that can affect quasi-pre-experimental designs that rely on analyzing data from participants who have been selected based on extreme low, or sometimes high, pre-test scores (Cook & Campbell, 1979). RTM can be reduced in two ways: random assignment to comparison groups, and selecting subjects on multiple measurements (Barnett et al., 2004).

The timing of measurements can significantly influence study outcomes and affect the interpretation of findings. When post-tests are conducted immediately after an intervention, the experimental group may demonstrate short-term improvements that may not persist in the long term, which makes it challenging to determine the intervention's lasting impact. Furthermore, test effects can confound the results, because participants could recall test questions or experience increased awareness and learning elicited by the pre-test (Marsden &

Torgerson, 2012). In other words, the phenomenon of test effects may artificially inflate the post-test results, to make it quite difficult to determine the true influence of the intervention (Kazdin, 2019). To ascertain whether improvements are genuinely attributable to the intervention rather than the test itself, Shadish et al. (2002) propose a Solomon four-group design. This approach encompasses four participant groups: "pre- and post-test no intervention, pre- and post-test with intervention, post-test only no intervention, and post-test only with intervention" (Marsden & Torgerson, 2012, p. 585). This design enables researchers to differentiate between the intervention effects and potentially test-induced improvements.

Quasi-experimental designs are crucial when true randomization is not possible, though these factors can influence the outcomes and should be carefully considered when the results are being interpreted (Cook & Campbell, 1979). While the effectiveness of the intervention is a primary factor, other elements, such as selection bias, differences in implementation, and the Hawthorne effect, could also play significant roles. Understanding these factors is crucial for accurately interpreting research outcomes and improving the design of future studies.

Because of ethical concerns and feasibility, quasi-experimental studies are often used in education settings to examine the impact of interventions on certain variables, for instance, performance and attitudes. However, a recurring pattern in quasi-experimental studies is the outperformance of the experimental groups over the control groups. This trend raises a critical question: Does the outperformance of the experimental group reflect the true effectiveness of the intervention or is it a result of confounding factors such as selection bias, the Hawthorne effect, differences in implementation, RTM, or the timing of measurement (Trochim, 2006, 2020)? Problematizing this trend is crucial for accurately interpreting the findings of quasi-experimental studies and emphasizing the application of careful study designs.

The present study investigated the factors that contribute to the experimental group's superiority in quasi-experimental studies, particularly in education-related studies. The study pursued answers to the following questions:

1. To what extent do the experimental groups outperform control groups in quasi-experimental studies?
2. According to university lecturers' perspectives, what are the main causes of the superiority of the experimental groups over the control groups in quasi-experimental research theses?
3. Are there any statistically significant differences in the responses of the lecturers regarding the factors behind experimental group superiority in quasi-experimental research attributable to each of the following demographic variables: gender, affiliation, and academic rank?

This study's mixed-methods approach, which combined the analysis of quasi-experimental research with the insights of experienced educators, was expected to yield a comprehensive understanding of the factors that contribute to the superior performance of experimental groups in quasi-experimental studies.

3. Methodology

3.1 Research Design

This study employed a mixed-methods approach that combined, as a first phase, quantitative content analysis of quasi-experimental MA theses, and in the second phase, the collection of quantitative and qualitative data through a questionnaire distributed to university lecturers. By combining a quantitative analysis of completed research with qualitative insights provided by experienced academics, this study sought to provide a comprehensive understanding of the mechanisms behind experimental group superiority in quasi-experimental studies. The findings were expected to offer valuable recommendations for enhancing the rigor and credibility of quasi-experimental research and ultimately contributing to the advancement of education research methodologies.

3.2 Sample Selection of the Theses

The sample for this phase consisted of 100 quasi-experimental MA theses from the Department of Curriculum and Instruction at Sultan Qaboos University in the Sultanate of Oman; to ensure the relevance of the findings, the theses involved quasi-experimental research undertaken in the last 10 years (2014–2023). The theses were on five subjects: Arabic, English, science, mathematics and social studies. The studies reported in the theses involved a clear comparison between an experimental group and a control group. These theses represent a rich source of data that reflects the application of quasi-experimental methods in the Omani education context. The sample of the theses was selected according to certain criteria, such as the type and duration of the treatment and the nature of the groups in terms of selection and size, in addition to their study findings.

3.3 Participants as Questionnaire Respondents

The participants of this phase of the study were 25 university lecturers who had considerable experience of supervising or conducting quasi-experimental research. A questionnaire was administered to the university lecturers to gather their insights on the potential factors influencing the performance of experimental groups versus comparison groups. The questionnaire comprised four dimensions: selection bias, intervention effectiveness, implementation differences and the Hawthorne effect. Each dimension contained three statements. The respondents were drawn from the Department of Curriculum and Instruction at four Omani universities: Sohar University, Sharqiya University, Nizwa University and Sultan Qaboos University.

3.4 Data Collection Tools and Procedure

The questionnaire was distributed electronically to lecturers, with follow-up reminders to maximize the response rate. Participation was voluntary, and the responses were anonymized to ensure confidentiality. The questionnaire used multiple-choice questions and a 5-point Likert scale to obtain quantitative data on the factors that contributed to the superiority of the experimental group over the control group.

Additionally, qualitative data was collected through two open-ended questions. The first question explored the possibility of factors in addition to the four factors addressed in the questionnaire, while the second invited respondents to propose solutions to avoid this superiority. Overall, the questionnaire was designed to capture the lecturers' perspectives on the four key dimensions of quasi-experimental design: selection bias, intervention effectiveness, implementation differences, and the Hawthorne effect. The reliability scores of the current study instrument, according to Cronbach's alpha value, is .80, indicating a high level of internal consistency when applied to the actual study sample.

3.5 Data Analysis

The data presented in the theses were examined using descriptive statistics to identify trends and patterns that highlighted the key factors linked to the superiority of the experimental group. Quantitative data gathered from the questionnaire were analyzed through descriptive statistics to uncover trends in the lecturers' responses. Additionally, qualitative data from open-ended questions were thematically analyzed to extract recurring themes and insights regarding the factors that affect the performance of the experimental group. The qualitative findings were then combined with the quantitative analysis from Phase 1 to offer a comprehensive understanding of the research questions.

3.6 Ethics Considerations

Ethics approval for the study was obtained from the relevant institutional review board at Sultan Qaboos University. Participants were informed of the study purpose. The anonymity and confidentiality of all participants and their responses were strictly maintained.

4. Findings and Discussion

This study investigated factors that contributed to the outperformance of experimental groups over control groups in quasi-experimental research. This section discusses the main findings that address the three research questions.

4.1 Question 1: Extent of Outperformance of Experimental Groups

To answer Question 1, To what extent do the experimental groups outperform control groups in quasi-experimental studies? The researchers used descriptive statistics as the primary method of analysis to summarize and organize the findings. The study samples of the theses were analyzed. Table 1 illustrates the trends and patterns among the theses.

Table 1: Degrees of Outperformance of Experimental groups over the Control Groups per Subject

Subject	Number of theses	Study Findings		
		Outperformance of experimental groups	Partial outperformance of experimental groups	No significant **differences** between the two groups
English	20	90%	5%	5%
Arabic	20	80%	10%	10%
Science	20	70%	30%	0%
Social studies	20	85%	15%	0%
Maths	20	90%	10%	0%

According to the data displayed in Table 1, it is evident that the theses reporting on studies in mathematics and English evidence the highest levels of outperformance by experimental groups over control groups: in 90% of the theses in these two specializations, the experimental groups outperformed the control groups in all the investigated skills. However, in mathematics, 10% of the theses report partial outperformance, and, in English, 5% report partial outperformance.

The remaining 5% indicates equal results for the experimental and control groups. Social studies had the next level of outperformance: 85% of the experimental groups outperformed the control groups, 15% report partial outperformance, and no studies indicated an absence of significant differences. The results for theses reporting on the subject of Arabic language is that, in 80%, the experimental groups outperformed control groups, 10% report partial outperformance, and another 10% report no significant differences between the two groups. The lowest outperformance of the experimental groups was observed for theses that investigated science subjects, in which 70% of studies reported outperformance by experimental groups, 30% reported partial outperformance, and no studies reported non-significant differences between the two groups for the skills being investigated.

It is noticeable that the studies conducted in Arabic and science scored the lowest in terms of the outperformance of the experimental groups over the control groups in tests and the measurement of perspectives and attitudes. This finding can, to some extent, be attributed to the number of subskills investigated by each study. In Arabic, for example, researchers typically target various subskills to assess the impact of a certain strategy or model. Consequently, investigating the effect of a single strategy on multiple subskills increases the likelihood of finding statistically insignificant differences between the two groups in one or two subskills, which could result in the partial outperformance of an experimental group. A similar justification can be given for the partial outperformance of the experimental groups reported in theses with quasi-experimental design in science.

To conclude the findings relating to Question 1, the overall analysis of the theses shows that the experimental groups generally outperformed control groups for all subjects, with mathematics and English achieving the highest rates of full outperformance, of 90%, followed by social studies (85%), Arabic (80%) and science (70%). Partial outperformance was more common for science (30%) and Arabic (10%), with some studies in Arabic (10%) and English (5%) reporting no significant differences. The relatively lower full outperformance for Arabic and science may be the result of the broader range of subskills assessed by studies in these subjects, which increases the likelihood of mixed results. Overall, the findings highlight the effectiveness of experimental teaching strategies, particularly for mathematics and English.

4.2 Question 2: Lecturers' perspectives of the reasons behind the frequent outperformance of experimental groups over control groups

To answer Question 2, What are the main causes of the superiority of the experimental groups over the control groups in quasi-experimental research based on university lecturers' perspectives? a questionnaire was administered to university lecturers that explored the factors that could account for the observed differences between the performance of experimental and control groups in quasi-experimental studies. The questionnaire comprised four dimensions, each targeting a specific aspect of the research. These dimensions are 1) Selection bias; 2) Intervention effectiveness; 3) Implementation differences; and 4) Hawthorne effect.

Table 2: Mean, Standard Deviation and Frequency

Item	N	Mean	SD	Rank
Dimension 1	20	3.77	.718	1
Dimension 2	20	3.67	.496	3
Dimension 3	20	3.70	.611	2
Dimension 4	20	3.48	.587	5
Overall	20	3.65	.458	4

Table 2 lists the means and standard deviations for the four dimensions of the study questionnaire and reveals that the means of all dimensions are high, thereby illustrating a high level of agreement with the questionnaire items. The mean of Dimension 1 is quite a bit higher than the other dimensions, while that of Dimension 4 is the lowest. While all four factors are recognized as important, selection bias is perceived as the most critical threat to internal validity in quasi-experimental research.

4.2.1 Selection Bias

Selection bias refers to the possibility of relating the superiority of the experimental group over the control group to the differences in the characteristics of the participants. To make their work more convenient, some researchers tend to select a sample of students who are more active and skillful in a particular school subject. Furthermore, when these students are informed that they have been chosen for a study, it may motivate them to work harder and prove to the teacher that they meet the teacher's expectations. Table 3 presents the percentages of the lecturer responses related to the factor of selection bias.

Table 3: Lecturer's Responses to Selection Bias Procedures

Items	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD
Lack of randomization	40%	40%	15%	5%	0%	4.15	0.875
Motivation and prior knowledge of the experi-mental groups	15%	60%	25%	0%	0%	3.80	0.695

As shown in Table 3, about 80% of respondents agreed that the absence of randomization might mean there are pre-existing differences between the experimental and control groups. The means and standard deviations (M = 4.15, SD = 0.88) indicate that the respondents reported a high level of agreement on how a lack of randomization allows for pre-existing differences to affect the intervention outcomes. This also indicates that most researchers experience selection bias during the sample selection process. This finding is supported further by responses to the second item, to which 75% (M = 3.80, SD = 0.695) of respondents agreed that the motivation of the experimental groups and their prior knowledge of being selected for a specific task might contribute to their superiority. Only 25% of respondents were neutral, and none of them disagreed with this issue.

4.2.2 Intervention Effectiveness

Arguments have arisen among researchers who either utilize RCTs or prefer quasi-experimental studies regarding the effectiveness of interventions. The intervention for the experimental group involves being subjected to a specific treatment or teaching strategy to investigate its effectiveness. Therefore, experimental groups are provided with additional materials that control groups do not receive. Consequently, in most cases, these extra materials lead to better learning and achievement of the target outcomes, especially when an intervention is well-designed. Table 4 displays the percentages of agreement and disagreement that relate to intervention effectiveness as expressed by the lecturers.

Table 4: Lecturers' Responses to Intervention Effectiveness Procedures

Items	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD
(theoretical) Effectiveness of intervention itself	15%	40%	30%	15%	0%	3.50	.946
Well-designed intervention leads to improvement in experimental groups	5%	70%	25%	0%	0%	3.65	.587
(demonstrated) Genuine effectiveness of the intervention	10%	70%	15%			3.85	.671

According to the results shown in Table 4, it is obvious that the effectiveness of interventions plays a significant role in the outperformance of experimental groups. About 70% of the respondents agreed that a well-designed intervention typically leads to a noticeable improvement in the final achievement of experimental groups, especially when the two—experimental and control groups—are taught the same unit of a textbook but the experimental group receives the treatment as extra material. Doing so leads to better teaching and a wider range and variety of learning opportunities.

4.2.3 Implementation Differences

During the implementation period, regular teaching may be influenced, because the experimental group could receive different treatment from the teacher, including the use of additional learning resources and variations in the way the intervention was being implemented in the treatment groups compared to the control group. Moreover, the teaching methods used for the experimental group might be more closely monitored and structured than those used for the control group. Table 5 lists the percentages of degree of agreement of the lecturers regarding differences in implementation.

Table 5. Lecturers' Responses Regarding Different Implementation Procedures

Items	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD
Extra attention and resources	15%	65%	15%	10%	0%	3.75	1.020
Differences in implementation	25%	70%	5%	0%	0%	3.80	0.523

The findings in Table 5 show the lecturers' perceptions of the implementation differences that affect the effectiveness of the interventions in quasi-experimental studies. The respondents reported that differences in the way interventions are implemented between experimental and control groups frequently contribute to outcome differences (M = 3.55, SD = 0.76). There was also a strong belief that students in experimental groups often receive more attention or resources, which could influence their performance (M = 3.75, SD = 1.02).

The importance of such implementation differences was rated as important (M = 3.80, SD = 0.52). Thus, most of the participants acknowledged the significant impact of implementation differences on the performance of experimental groups, with 15% strongly agreeing and 65% agreeing, resulting in a total of 80% being in agreement. Only 10% of lecturers disagreed about the influence of this issue.

4.2.4 Hawthorne Effect

In many cases where experimental groups outperform control groups, researchers make significant efforts to ensure that the two groups are equivalent. However, participants may become aware of the intervention, which can unconsciously motivate them to perform better in the given tasks. This dimension captured the lecturers' views on whether an awareness of being observed would influence the participants' performance and behavior. Table 6 provides the percentages of degree of agreement of the respondents regarding the impact of the Hawthorne effect in quasi-experimental research.

Table 6: The Influence of the Hawthorne Effect on the Results

Items	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Mean	SD
Awareness of being observed	20%	45%	15%	20%	0%	3.65	1.040
General influence of the Hawthorne effect	20%	75%	5%	0%	0%	3.70	0.571
Researchers always take the impact of the Hawthorne effect into account	0%	30%	45%	25%	0%	3.10	0.788

Table 6 reveals that approximately 65% of participating lecturers agreed that the awareness of being observed contributed to improved performance by an experimental group. Furthermore, about 95% believed that the Hawthorne effect significantly influenced experimental groups' superior performance. However, only 30% indicated that researchers consistently considered the Hawthorne effect when implementing their interventions.

To conclude the findings in relation to Question 2, it was obvious that the lecturers identified four key factors that contributed to the frequent outperformance of experimental groups in quasi-experimental research: selection bias, intervention effectiveness, implementation differences, and the Hawthorne effect. The majority (80%) agreed that selection bias, particularly the absence of randomization and the motivation of selected students, played a significant role in creating pre-existing advantages for the experimental group. Additionally, 70% of respondents referred to the importance of well-designed interventions and indicated that providing extra materials enhanced the learning outcomes of the experimental groups.

Differences in implementation, such as experimental groups receiving more structured teaching and additional resources, were acknowledged by 80% of lecturers as being a major contributor to their superior performance. Furthermore, 85% acknowledged the Hawthorne effect, where the awareness of being observed motivates participants to perform better, as a significant influence, although only 30% believed that researchers consistently considered this factor. These findings underscore the interplay of these factors in driving experimental groups' success.

The Hawthorne effect was found to have the most influence of all the factors. The Hawthorne effect refers to changes in participants' behavior because they realize that they are being observed and monitored (Cook, 1967). Most participants (95%) believed in the general influence of the Hawthorne effect. This result differs from that reported by Levitt and List (2009). They had analyzed the original data of the illumination experiments conducted at the Hawthorne plant. Their findings suggest that many of the claims about the Hawthorne effect might have been exaggerated or even incorrect. They discovered that the data did not show the dramatic productivity increases that had been widely reported. Instead, they found only subtle hints of a Hawthorne effect. This means that, while some behavioral changes might have occurred as a result of observation, they were not as significant as originally believed.

The second factor was selection bias: About 80% of lecturers agreed that selection bias played a big role in the validity of the results and led to the outperformance of the comparison groups. Kahan et al. (2015) found that, out of 152 eligible published trials, only 3% of the trails used randomized selection to select participants. Well-conducted RCTs can guarantee the balance of both known and unknown factors that treatment groups might run into.

Avoiding the risk of selection bias is possible when the probability of the participants being selected does not rely on their probability of being assigned to a particular treatment group. Because of the risk of selection bias, Infante-Rivard (2018) raises a flag about published quasi-experimental research studies that "may be reporting entirely spurious associations" (p. 557). In their research, Infante-Rivard (2018) examined examples of selection bias in a variety of fields where it had resulted in unreliable and misleading findings.

4.3 Qualitative Findings

The respondents were asked two open-ended questions. The first question explored the possibility of other factors contributing to the superiority of experimental groups over control groups in quasi-experimental studies, in addition to those mentioned in the questionnaire, and the second invited respondents to propose solutions to avoid this superiority. One of the respondents said that "The proposed programs are not in consistent structures," which implies that an intervention might not be well-structured and could lack coherence. This view aligns with other research findings, which indicate that poorly designed interventions can skew results (Durlak & DuPre, 2008).

Another respondent pointed to the problems associated with the implementation of interventions, by stating that "researchers in psychological field do not have the skills to apply the program they claim to apply," and a third respondent explicated that "most experimental research provides statistics without a precise analysis of the numbers and their meanings. Psychological interpretation is largely missing." Some researchers expressed that a lack of training or expertise in conducting interventions can negatively affect internal validity and lead to misleading conclusions (Century et al., 2010).

Regarding the second question, one of the lecturers proposed "documenting and standardizing the implementation to help mitigate this issue." This suggestion is supported by the literature, which advocates rigorous protocol development to reduce researcher bias (Durlak & DuPre, 2008). One of the lecturers indicated that "universities and researchers should discuss the acceptance of results as it appears without having pre-assumptions that the experimental groups should be better", thereby emphasizing the role of universities and researchers. Thus, in research, it is important to ensure transparency and unbiased interpretation (Nosek et al., 2012).

4.4 Question 3: Factors Behind Experimental Group Superiority Attributable to Demographic Variables

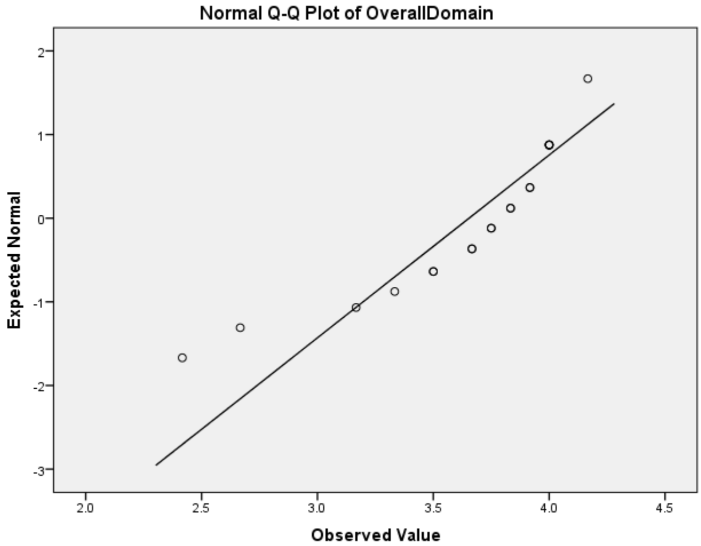

Figure 1: Normality of study variable

The Q–Q plot values indicate that the study variable is not normally distributed. Table 7 shows that the results of the Kolmogorov-Smirnov value is .02 and Shapiro-Wilk is .002, both of which are below the significance level (.05). This indicates that the distribution of the study variable does not follow normality. The analysis used non-parametric statistics.

Table 7: Tests of Normality of the Study Variable

	Kolmogorov-Smirnov			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Overall, of all domains	.211	20	.020	.822	20	.002

To answer the third study question, are there any statistically significant differences in the responses of the lecturers regarding the factors behind experimental group superiority in quasi-experimental research attributable to each of the following demographic variables: gender, affiliation and academic rank?', the Mann-Whitney test and Kruskal Wallis were used. The results are as shown in Tables 8, 9 and 10.

Table 8: Mann-Whitney Test Results for the Differences in Response by Gender

Item	Result
Mann-Whitney U	16.000
Wilcoxon W	19.000
Z- score	-.254
Asymptotic significance. (2-tailed)	.799
Exact significance. [2*(1-tailed Sig.)]	.853

Table 8 shows the results for the probability value (Sig = .853) following the Mann-Whitney test; this is a value greater than the significance level (.05), which shows that there are no statistically significant differences; thus, no differences between the responses of the sample in terms of gender. Male and female lecturers held similar views about the factors that contribute to the outperformance of experimental groups over control groups in quasi-experimental studies.

Table 9: Kruskal Wallis Test Results for the Differences in Response by University Affiliation

Item	Result
Chi-square	.325
df	3
Asymp. Sig.	.955

Table 9 shows the results of the probability value (Sig = 0.955) of the Kruskal Wallis test, which is a value greater than the significance level (.05). There are no statistically significant differences between the responses of the lecturers with respect to affiliation. In other words, lecturers from different universities shared a common perspective on the underlying factors contributing to the superior performance of experimental groups over control groups in quasi-experimental studies.

Table 10: Kruskal Wallis Test Results for the Differences in Response by Academic Rank

Item	Result
Chi-Square	.644
df	**3**
Asymp. Sig.	.092

Table 10 shows the result of the probability value (Sig = 0.092) of the Kruskal Wallis test, which is a value greater than the significance level (.05). There are no statistically significant differences between the responses of the sample in terms of academic rank. The agreement across the different academic ranks strengthens the credibility and cohesion of the findings further. This implies that the perceived issues are not limited to specific levels of experience but are acknowledged throughout the academic spectrum.

5. Implications of the Study

In the field of education, there is room for a variety of research methodologies, including quantitative, qualitative, correlational and mixed-methods research studies (Cook & Cook, 2016). However, true experiments using randomized assignments are underutilized in education. There are several reasons why true experimental studies are not feasible to implement, including the inability to limit the study participants to a selected number because of their pre-existing formation, such as intact classes and late starts which make randomization infeasible (Scher et al., 2015). To achieve a comparison group that is a mirror image of the treatment group, researchers need to consider designing quasi-experimental studies that are similar to true experiments (Rosenbaum, 2010).

To address the design-related concerns elucidated in this study, researchers could implement experimental studies established through random selection and assignment. This methodology could result in both groups being similarly affected by factors that compromise the internal validity of quantitative research (Marsden & Torgerson, 2012). The random assignment of participants to either experimental or control groups is beneficial to ensure equal distribution across groups. Consequently, any observed changes in the pre- and post-test scores can be more accurately attributed to the intervention itself.

In the context of the quasi-experimental research design, it is imperative to explicitly address internal validity threats in the study design section. Strategies should be delineated to mitigate these threats through design, measurement, and statistical approaches. Critical design considerations include the criteria for sample selection and procedures for identifying and characterizing both the experimental and control groups. Measurement aspects should focus on evaluating unobserved confounding factors, pre-intervention outcomes, and outcomes that are not anticipated to be affected by treatment (non-equivalent outcomes).

To mitigate internal validity threats, researchers can employ various statistical techniques, including multivariable regression, propensity score methods, and instrumental variable analyses. To identify and address potential internal validity issues, researchers should conceptualize an ideal randomized trial with unlimited resources, access, and time, and subsequently contrast their proposed study against this benchmark (Maciejewski, 2018).

6. Limitations of the Study

Some limitations reflect methodological, data-related and interpretive concerns. The findings may not be generalized to other contexts, given that the study focused only on approximately 100 MA theses from a single department (Curriculum and Instruction) at Sultan Qaboos University. Another limitation that should be considered relates to the self-reported lecturer survey. The perceived causes of superiority of experimental groups over control groups were based on self-reported views of university lecturers whose responses may have been affected by their personal biases, theoretical orientations or differing levels of research experience, which could have affected the objectivity of the interpretations.

7. Conclusion

This study explored the factors that contribute to the superiority of experimental groups over control groups in quasi-experimental research. It started by examining the extent of this superiority by analyzing 100 master's theses from the Department of Curriculum and Instruction at the College of Education, Sultan Qaboos University, Oman; the theses were in five disciplines, and 20 theses were selected from each: mathematics, science, social studies, Arabic language, and English language. The analysis reveals that, in 70% to 90% of theses across all subjects, experimental groups outperformed controls. According to the participants, the primary factors contributing to this outperformance were the Hawthorne effect, intervention effectiveness, selection bias, and variations in implementation.

The participants' awareness of an experiment being executed with them as the focus could have artificially enhanced their performance, and if an intervention is not properly designed, it may fail to produce genuine results or could lead to questionable outcomes. Selection bias contributed to the skewness of the results, as could variations in the implementation itself, which may also complicate the interpretation of the findings of such studies. To ensure valid comparisons and reliable conclusions, these confounding factors should be addressed via training, monitoring, and setting up standardized intervention protocols.

8. References

Andrade, C. (2021). The limitations of quasi-experimental studies, and methods for data analysis when a quasi-experimental research design is unavoidable. *Indian Journal of Psychological Medicine*, *43*(5), 451–452. https://doi.org/10.1177/02537176211034707

Barnett, A. G., Van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Biddle, N., & Curchin, K. (2024). Randomised controlled trials and quasi-experimental designs. In B. M. Smyth, M. A. Martin, & M. Downing (Eds.), *The Routledge handbook of human research ethics and integrity in Australia* (pp. 187–198). Routledge. https://doi.org/10.4324/9781003319733-20

Biesta, G. (2010). "Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education, 29*, 491–503. https://doi.org/10.1007/s11217-010-9191-x

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. https://doi.org/10.1177/1098214010366173

Cook, B. G., & Cook, L. (2016). Research designs and special education research: Different designs address different questions. *Learning Disabilities Research & Practice*, *31*(4), 190–198. https://doi.org/10.1111/ldrp.12110

Cook, D. (1967). *The impact of the Hawthorne effect on experimental designs in educational research*. Final report No. BR-5-0725. Washington DC Bureau of Research.

Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Houghton Mifflin.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*, 327–350. https://doi.org/10.1007/s10464-008-9165-0

El Soufi, N., & See, B. H. (2019). Does explicit teaching of critical thinking improve critical thinking skills of English language learners in higher education? A critical review of causal evidence. *Studies In Educational Evaluation*, *60*, 140–162. https://doi.org/10.1016/j.stueduc.2018.12.006

Gillespie, R. (1993). *Manufacturing knowledge: A history of the Hawthorne experiments*. Cambridge University Press.

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. https://doi.org/10.1197/jamia.M1749

Infante-Rivard, C., & Cusson, A. (2018). Reflection on modern methods: selection bias—a review of recent developments. *International Journal of Epidemiology*, *47*(5), 1714−1722. https://doi.org/10.1093/ije/dyy138

Jobson, D., Li, Y., Nishimura, N., Ohashi, K., Yang, J., & Matsumoto, T. (2024). Covariate ordered systematic sampling as an improvement to randomized controlled trials. Submitted to the *33rd ACM International Conference on Information and Knowledge Management*, October. https://doi.org/10.1145/3627673.3679892

Kahan, B. C., Rehal, S., & Cro, S. (2015). Risk of selection bias in randomised trials. *Trials*, *16*, 1–7. http://dx.doi.org/10.1186/s13063-015-0920-x

Kazdin, A. E. (2019). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

Kelchtermans, G. (2009). Who I am in how I teach is the message: Self-understanding, vulnerability and reflection. *Teachers and Teaching: Theory and Practice*, 15, 257–272. http://dx.doi.org/10.1080/13540600902875332

Kim, Y., & Steiner, P. (2016). Quasi-experimental designs for causal inference. *Educational Psychologist*, *51*(3–4), 395–405. https://doi.org/10.1080/00461520.2016.1207177

Levitt, S. D., & List, J. A. (2011). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. *American Economic Journal: Applied Economics*, *3*(1), 224–238. https://doi.org/10.1257/app.3.1.224

Maciejewski, M. L. (2018). Quasi-experimental design. *Biostatistics & Epidemiology*, *4*(1), 38–47. https://doi.org/10.1080/24709360.2018.1477468

Mansour, N. (2009). Science teachers' beliefs and practices: Issues, implications and research agenda. *International Journal of Environmental & Science Education*, *4*, 25–48.

Marsden, E., & Torgerson, C. J. (2012). Single group, pre-and post-test research designs: Some methodological concerns. *Oxford Review of Education*, *38*(5), 583–616. https://doi.org/10.1080/03054985.2012.731208

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615–631. https://doi.org/10.1177/1745691612459058

Palmer, P. J. (2007). *The courage to teach: Exploring the inner landscape of a teacher's life*. Jossey-Bass. http://dx.doi.org/10.1111/j.1744-1722.2000.tb00290.x

Reichardt, C. S. (2019). *Quasi-experimentation: A guide to design and analysis*. Guilford Publications.

Richardson, J. T. E. (2005). Students' approaches to learning and teachers' approaches to teaching in higher education. *Educational Psychology*, 25, 673–680. http://dx.doi.org/10.1080/01443410500344720

Rosenbaum, P. R. (2010). *Design of observational studies*. Springer.

Scher, L., Kisker, E., & Dynarski, M. (2015). *Designing and conducting strong quasi-experiments in education* (Version 2). Decision Information Resources.

Seltman, H. J. (2013, December 3). *Experimental design and analysis*. http://www.stat.cmu.edu/~hseltman/309/Book/Book.pdf

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Siedlecki, S. L. (2020). Quasi-experimental research designs. *Clinical Nurse Specialist*, *34*(5), 198-202. https://doi.org/10.1097/NUR.0000000000000540

Trochim, W. M. K. (2006). *The qualitative debate: Research methods knowledge base*. Conjointly. http://www.socialresearchmethods.net/kb/qualmeth.php

Trochim, W. M. K. (2020). *Research methods: The essential knowledge base*. Cengage Learning.

Vannan, K. (2021). *History of the Hawthorne effect*. https://doi.org/10.1002/9781119111931.ch47