Examining Mindfulness Training for Teachers: Theoretical and Methodological Extensions of Intervention Effectiveness

Choles, Jaiya Rae
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Let us know how access to this document benefits you.

Recommended Citation
https://doi.org/10.15760/etd.7380

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.
Examining Mindfulness Training for Teachers:
Theoretical and Methodological Extensions of Intervention Effectiveness

by

Jaiya Rae Choles

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Applied Psychology

Dissertation Committee:
Andrew Mashburn, Chair
Joel Steele
Ellen Skinner
Pat Burk

Portland State University
2020
Abstract

This dissertation seeks to extend the field of mindfulness-based interventions (MBIs) for teachers, both theoretically and methodologically. The first study is a systematic review conducted of randomized controlled trials examining the effectiveness of MBIs for school teachers. The purpose of Study 1 was to determine theoretical and methodological next steps for the field. Results of the theoretical review indicate more empirical evidence is needed examining mindfulness practices used in the interventions, distal impacts on classrooms and students, and mediation tests connecting proximal and distal outcomes. The methodological review indicates that more studies should focus on measuring the fidelity of implementation of the intervention, include longer follow-up times of measurement and larger sample sizes, and utilize modern missing data methods to manage missing data due to attrition and non-response.

The second study addresses a theoretical gap elucidated in Study 1 by testing whether random assignment to an MBI after baseline data collection (T1) significantly increases teachers’ use of mindfulness practices both during an 8-week intervention (T1.5) and in the summer months after the intervention is over (T2.5); as well as whether frequency of mindfulness practices explains teachers’ development of mindfulness-related skills immediately following the intervention (T2) and at a 4-month follow-up (T3). Using a sample of 173 teachers, results indicate teachers assigned to the MBI group engaged in mindfulness practices significantly more frequently than control teachers at both time points ($\beta_{T1.5} = .58, p<.001; \beta_{T2.5} = .41, p<.001$). Furthermore,
mediation structural equation models demonstrated that teachers’ frequency of
mindfulness practices explained their short- and long-term improvements in
occupational self-compassion (IDE\textsubscript{T2} = .17, \( p < .01 \); IDE\textsubscript{T3} = .12, \( p < .01 \)) and long-term
improvements in mindfulness (IDE\textsubscript{T3} = .13, \( p < .05 \)).

Study 3 addresses a more general gap in the methodological literature by
comparing modern missing data methods, namely, two types of multiple imputation—
combined multiple imputation (CMI) and separate group imputation (SGI)—and full
information maximum likelihood (FIML), in producing unbiased estimates of
intervention effectiveness. These methods were compared using simulated intervention
data that varies the effect size of the intervention impact, the sample size of the study,
and the amount of missing data present in the study, to determine which methods
perform best under common applied intervention research conditions. Results indicate
all three methods produce similar and unbiased results under conditions of little missing
data (10%) and large intervention impacts (an effect size of .80). However, under
smaller intervention effects (.20 and .50 effect sizes) and higher rates of missing data
(25% and 50%) both SGI and FIML are less biased than CMI, particularly in small sample
sizes. Results of this study aim to be beneficial in guiding the use of such methods in
future intervention studies.
Dedication

I dedicate this dissertation to several important people in my life. To my dad, Peter Choles, for introducing me to mindfulness and for continuing to remind me of what's most important in life—to be happy. To my dear friend, Emilee Naylor, for sharing my love of research and all things science, I will carry your spirit with me always. And to my partner, Jim Goodman, for standing by me through this entire process, for knowing when to provide words of encouragement, and for bringing my life immense joy.
Acknowledgements

I would like to thank my advisors and committee, Drs. Andrew Mashburn, Joel Steele, Rob Roeser, Ellen Skinner, and Pat Burk for their support and guidance on these varied projects. I also thank my undergraduate advisors, Drs. Lesa Ellis and Russ Costa, for their mentorship and for igniting my passion for scientific inquiry. I also thank my parents Tami and Peter, my sister Liani, and my family, friends, cohort, and lab mates, both near and far, for their support and encouragement throughout this challenging endeavor. Most importantly, I would like to acknowledge my partner, Jim Goodman, for his unwavering support and continued belief in me always.
# Table of Contents

Abstract ............................................................................................................................... i
Dedication ........................................................................................................................... iii
Acknowledgements .......................................................................................................... iv
List of Tables .................................................................................................................... vii
List of Figures .................................................................................................................. viii
Chapter 1: Introduction ................................................................................................... 1

Chapter 2: Study 1 ........................................................................................................... 9

Theoretical and Methodological Extensions of Mindfulness-Based Interventions for Teachers: A Systematic Review of Randomized Controlled Trials .................................... 9

Abstract ........................................................................................................................... 10
Methods ............................................................................................................................. 15
  Search Procedure ........................................................................................................... 15
  Selection Criteria .......................................................................................................... 15

Theoretical Review ......................................................................................................... 20
  Developing Mindfulness-Related Skills ....................................................................... 21
  Developing Occupational Health and Wellbeing .......................................................... 31
  Extending Impacts into the Classroom Context ............................................................ 41
  Improving Student Outcomes ....................................................................................... 47

Methodological Review .................................................................................................. 49
  Research Design Features ............................................................................................. 51
  Intervention Implementation ......................................................................................... 58

Factors that Influence Conclusions of Intervention Effectiveness ................................ 68
Discussion ........................................................................................................................ 80

Chapter 3: Study 2 ......................................................................................................... 83

The Explanatory Role of Mindfulness Practice in the Development of Mindfulness Skills: Results from Studies of a Mindfulness-Based Intervention with Teachers ..................... 83

Abstract ........................................................................................................................... 84
  Mindfulness: What it is and How it Works ................................................................. 86
  Research on Mindfulness Practices ............................................................................ 90

Study Purpose .................................................................................................................. 91
Methods ............................................................................................................................. 94
  Samples ......................................................................................................................... 95
  Intervention .................................................................................................................. 98
  Measures ....................................................................................................................... 99

Baseline Equivalence of Study Samples and Intervention Groups ................................ 101
Analyses .......................................................................................................................... 104
Mediation Analyses ....................................................................................................... 106
Results .............................................................................................................................. 107
  Preliminary Results ...................................................................................................... 107
  Intervention Impact on Teachers’ Frequency of Mindfulness Practices .................... 110
Mediation Results ......................................................................................................... 110
Table of Contents

Discussion ................................................................................................................. 117
Limitations ............................................................................................................... 120
Conclusion .............................................................................................................. 123

Chapter 4: Study 3 ................................................................................................. 124
Contrasting Missing Data Methods in Applied Experimental Research: Simulated
Studies to Inform Best Practices ........................................................................... 124
Abstract .................................................................................................................. 125
Missing Data Mechanisms ...................................................................................... 129
Missing Data Methods ............................................................................................ 132
  Full Information Maximum Likelihood .................................................................. 132
  Multiple Imputation ............................................................................................... 133
Method ..................................................................................................................... 136
  Simulation Specification ......................................................................................... 137
  Imputation Specification ......................................................................................... 139
  Analysis Specification ............................................................................................ 139
Results ...................................................................................................................... 141
  Raw Bias ................................................................................................................ 142
  Efficiency ................................................................................................................ 142
  Standardized Bias .................................................................................................. 143
Discussion ................................................................................................................ 148
  Limitations and Future Directions ....................................................................... 152

Chapter 5: Discussion ........................................................................................... 155
Summary of Current Studies ................................................................................... 155
  Study 1 – Review of Mindfulness Interventions for Teachers ................................. 155
  Study 2 – The Role of Mindfulness Practice in the Development of Mindfulness-
  Related Skills ........................................................................................................... 158
  Study 3 – Informing the Use of Missing Data Methods in Intervention Research .. 162
Future Directions ..................................................................................................... 168
  Personal Future Directions .................................................................................... 168
  Additional Future Directions for the Field ............................................................ 174
Conclusion ............................................................................................................... 182
References ............................................................................................................... 183

Appendix A – Study 3 Simulation Code ................................................................. 203
Appendix B – Study 3 Supplementary Tables ......................................................... 212
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Overview of Studies – Demographic Information</td>
<td>18</td>
</tr>
<tr>
<td>2.2</td>
<td>Review of Studies – Mindfulness Practices and Mindfulness-Related Skills</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>Review of Studies – Occupational Health and Wellbeing</td>
<td>37</td>
</tr>
<tr>
<td>2.4</td>
<td>Review of Studies – Classroom Impacts</td>
<td>45</td>
</tr>
<tr>
<td>2.5</td>
<td>Review of Studies – Research Design Features</td>
<td>57</td>
</tr>
<tr>
<td>2.6</td>
<td>Review of Studies – Intervention Implementation</td>
<td>65</td>
</tr>
<tr>
<td>2.7</td>
<td>Review of Studies – Factors that Influence Conclusions of Intervention Effectiveness</td>
<td>77</td>
</tr>
<tr>
<td>3.1</td>
<td>Descriptive Statistics by Project Sample and Intervention Condition</td>
<td>97</td>
</tr>
<tr>
<td>3.2</td>
<td>Curricular Components of the Mindfulness Training for Teachers</td>
<td>99</td>
</tr>
<tr>
<td>3.3</td>
<td>Baseline Comparisons of Demographics and Outcomes by Study Condition and Research Site</td>
<td>103</td>
</tr>
<tr>
<td>3.4</td>
<td>Effect Sizes by Project Sample and Intervention Condition</td>
<td>109</td>
</tr>
<tr>
<td>4.1</td>
<td>Average Estimates from Small Effect Size (.20) Intervention Impact Simulations</td>
<td>145</td>
</tr>
<tr>
<td>4.2</td>
<td>Average Estimates from Medium Effect Size (.50) Intervention Impact Simulations</td>
<td>146</td>
</tr>
<tr>
<td>4.3</td>
<td>Average Estimates from Large Effect Size (.80) Intervention Impact Simulations</td>
<td>147</td>
</tr>
<tr>
<td>4.4</td>
<td>Method Recommendations for Best Practice by Study Condition</td>
<td>150</td>
</tr>
<tr>
<td>5.1</td>
<td>Applied Example Comparing Missing Data Methods</td>
<td>168</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Mindfulness-Based Intervention for Teachers Theory of Change</td>
<td>21</td>
</tr>
<tr>
<td>2.2</td>
<td>Methodological Factors Examined That Influence Conclusions of Effectiveness</td>
<td>51</td>
</tr>
<tr>
<td>3.1</td>
<td>Mindfulness-Based Intervention (MBI) Theory of Change</td>
<td>94</td>
</tr>
<tr>
<td>3.2</td>
<td>Intervention Impact on Teachers’ Post-Program Mindfulness Mediated by Frequency of Mindfulness Practice During the Intervention</td>
<td>112</td>
</tr>
<tr>
<td>3.3</td>
<td>Intervention Impact on Teachers’ Post-Program Occupational Self-Compassion Mediated by Frequency of Mindfulness Practice During the Intervention</td>
<td>113</td>
</tr>
<tr>
<td>3.4</td>
<td>Intervention Impact on Teachers’ Follow-Up Mindfulness Mediated by Frequency of Mindfulness Practice After the Intervention</td>
<td>115</td>
</tr>
<tr>
<td>3.5</td>
<td>Intervention Impact on Teachers’ Follow-Up Occupational Self-Compassion Mediated by Frequency of Mindfulness Practice After the Intervention</td>
<td>116</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

The use of mindfulness-based interventions (MBIs) as a method for supporting teachers in education settings has markedly increased since the turn of the twenty-first century. The development of these programs for teachers has stemmed, most predominantly, from reports indicating the need to support teachers in reducing occupational stress and the high rates of teachers leaving the profession that often follow (Kyriacou, 2001; Montgomery & Rupp, 2005). Indeed, teaching has been indicated as one of the leading high-stress professions, with reported levels commensurate to those of nurses and physicians (Gallup, 2014). If left unmanaged, teacher stress can have wide-spread deleterious effects—not only negatively impacting teachers’ personal health and wellbeing, but also their ability to provide a high-quality teaching environment in which their students can learn and thrive (Roeser, 2014; Skinner & Beers, 2016).

Fortunately, research examining the effectiveness of such programs for teachers has demonstrated positive effects on many key outcomes. Specifically, research suggests MBIs may be particularly useful tools for bolstering teacher wellbeing, with evidence supporting improvements in teacher occupational stress and burnout (Anderson, et al., 1999; Roeser et al., 2013; Taylor et al., 2016); mindfulness and self-compassion (Benn, et al., 2012; Garner et al., 2018; Harris et al., 2016; Kemeny et al., 2012; Roeser et al., 2013); cognitive and self-regulatory abilities (Garner et al., 2018; Jennings et al., 2017; Jennings et al., 2013; Roeser et al., 2013); negative emotional
experiences of anxiety, depression, and rumination (Anderson et al., 1999; Harris et al., 2016; Jennings et al., 2017; Kemeny et al., 2012; Roeser et al., 2013); physical symptoms of health (Harris et al., 2016; Jennings et al., 2013); quality and quantity of sleep (Crain et al., 2016); and job satisfaction (Crain et al., 2016). Of even greater note, more recent research has begun to demonstrate that benefits of these programs for teachers may extend to the classroom context through positive changes in the quality of teacher-student interactions (Flook et al., 2013; Jennings et al., 2017).

Much of this research reported to date has utilized randomized controlled trial (RCT) methodology to test the efficacy of these MBI programs for teachers. The RCT design, in which individuals are randomly assigned to study conditions as a means to ensure groups do not differ on any measured or unmeasured characteristics, is particularly well suited to determine the important question of whether an MBI “works” (Shadish et al., 2002). Understanding if an MBI is effective in supporting teachers, from studies that provide trustworthy data, can aid administrators, policy makers, and other academic decision makers in determining which professional development programs would be best in supporting teachers on a district-wide or larger scale (Gersten & Hitchcock, 2009). In addition, the benefits of the RCT methodology also extend to the greater field of MBIs, as a way to provide methodologically rigorous evidence that can help to counteract the extensive amount of misinformation (and unsubstantiated claims) about mindfulness that also exists (van Dam et al., 2018).
Despite the apparent strengths in the RCT design to answer important questions of causality, many have criticized it for its inability to extend beyond the “does it work” question and address questions that could be considered much more theoretically rich and interesting. Contrary to these claims, the RCT design can be used for more than answering the “does it work” question, and extend to areas such as those testing causal mechanisms of action within the intervention or examining differential intervention impacts within subgroups of participants (Kraemer et al., 2002; Skinner et al., 2019). In fact, due to its ability to infer causality, it is particularly well suited to testing causal mechanisms of action—if only researchers have the foresight to design high quality RCT trials that seek to examine the black box of the intervention, as much as they wish to determine its effectiveness.

At the same time, however, there are numerous ways, that are less often discussed, in which the RCT design’s ability to make strong causal inferences can be threatened. Factors such as insufficient recruitment of a large sample, incomplete randomization of participants to groups (where some participants are randomized but others are not), participant attrition, and low-quality implementation of the intervention can all threaten the foundation in which the RCT design is built upon (Bickman & Reich, 2015; Wilson & Lipsey, 2001). However, the RCT methodology’s strong reputation as the “gold standard” research design often makes it easy to overlook these many ways in which the design can be vulnerable. These vulnerabilities do not discount the use of RCTs as valuable methods for determining program
effectiveness, but it is extremely important for researchers to know these vulnerabilities exist, as well as what to do about them if they do occur, in order to produce high quality studies that maintain the strong credibility in which the RCT methodology innately commands. As such, producing optimal studies examining MBIs, or any program’s effectiveness, requires a strong theoretical foundation as well as careful consideration of the methods employed.

This dissertation uses the interwoven theoretical and methodological perspective described above as a means to improve our understanding of the theory and methods used to examine the efficacy of mindfulness interventions for teachers using the RCT methodology. Specifically, the purposes of this dissertation were twofold: (1) to conduct a comprehensive theoretical and methodological review of studies of MBIs for teachers to determine the current state of the field and future directions and (2) to utilize information derived from the literature review to guide two additional studies that extend the field: one that addresses a theoretical gap in the literature of MBIs for teachers and another that addresses a more general methodological gap in the greater field of intervention research. Below each of these three studies are summarized.
Study 1 – Review of Mindfulness Interventions for Teachers

Study 1 serves the first purpose of this dissertation and represents a formal review of studies examining MBIs for teachers. The aim of this paper was to conduct a comprehensive review of the literature on RCT studies from both theoretical and methodological perspectives. Reviewing the field from both of these perspectives allows for a more comprehensive understanding of where research in the field of MBIs for teachers has been strongest, and which areas are in need of greater expansion. Additionally, narrowing the review to studies that incorporate the RCT framework allows for a close examination of the current evidence base within studies that allow for strong causal claims to be made about program effectiveness.

In a general sense, this first study represents a summary of much of what I have learned throughout the course of my graduate studies—not only highlighting my focus on understanding if and how mindfulness interventions impart benefits to teachers, classrooms, and students—but also taking what I have learned regarding methodological factors that can enhance the quality of research studies and subsequently increase our certainty regarding conclusions made about intervention effectiveness. This latter point, although often overshadowed by the RCT’s pristine reputation, is essential in applied research to ensure intervention studies are optimally able to benefit the communities and individuals they seek to support.
Study 2 – The Role of Mindfulness Practice in the Development of Mindfulness-Related Skills

Study 2 addresses a theoretical gap in the literature regarding the paucity of research examining the role of mindfulness practices in teachers’ development of mindfulness-related skills. Regular engagement in mindfulness practices has been discussed as a key causal link in the teacher MBI theory of change (see Roeser et al., 2012), however, no research is known to have explicitly tested this hypothesis in samples of teachers participating in MBI programs uniquely tailored to their profession. Specifically, this study tests whether an 8-week MBI significantly increases the frequency with which teachers engage in mindfulness practices, both during the intervention and in the summer months after the intervention is over. Furthermore, this study examines whether teachers’ frequency of mindfulness practices helps explain their increases in mindfulness-related skills, namely, occupational self-compassion and mindfulness, both at post-program and follow-up.

This study represents one way in which researchers can go beyond the “does it work” question of an RCT study and examine causal mechanisms of action that are theorized to be the driving forces of change within the intervention. It is often the case that researchers skip over this step in testing the active ingredients of the program being tested, and instead simply infer these processes took place if they find significant impacts at the end of their study. In my own work, I believe these process-based questions to be the bread and butter of intervention research. Not only does answering
“why” questions satisfy intellectual curiosity regarding how interventions work, but they are also crucial for program improvement and theory refinement as well.

**Study 3 – Informing the Use of Missing Data Methods in Intervention Research**

Study 3 addresses a more general gap in the methodological literature regarding the use of recommended missing data methods to reduce bias created through attrition and measurement non-response in intervention studies. Specifically, using simulated data of intervention impact estimates, Study 3 compares two types of multiple imputation (combined multiple imputation and separate group imputation) and full information maximum likelihood methods under a series of model specifications that mimic common applied intervention conditions. This study examines the conditions under which each of these methods perform best by simulating data of various intervention effect size estimates, sample sizes, and missing data rates within a study. Determining differences in method performance can help inform the general field of intervention studies, and the more specific field of MBIs for teachers, in understanding when each of these methods are recommended for use and under what conditions each method may be more likely to bias conclusions made about intervention effectiveness.

**Summary**

It is often the case that theoretically focused and methodologically focused studies are published separately. While it is well known that each informs the other, it is less typical to contextualize the quality of one of these factors based on the other and vice versa. However, producing the highest quality studies requires both a strong
theoretical foundation and thoughtfully constructed and implemented methods. Given this perspective, this dissertation seeks to combine these seemingly separate worlds, to push forward the field of MBIs for teachers in meaningful ways. Doing so will not only deepen our understanding of how mindfulness interventions work, but also provide us with the greatest chance in positively impacting teachers and their communities through our applied intervention efforts.
Chapter 2: Study 1

Theoretical and Methodological Extensions of Mindfulness-Based Interventions for Teachers: A Systematic Review of Randomized Controlled Trials
Abstract

Research examining the impacts of mindfulness-based interventions (MBIs) for teachers has demonstrated positive benefits for teachers in a variety of domains including their occupational health and wellbeing, job satisfaction, mindfulness and self-compassion, and self-regulation (Hwang et al., 2017; Klingbeil & Renshaw, 2018). With this foundation of evidence arises the need to assess the current state of the field to elucidate areas of research that need greater focus in the future. As such, a review was conducted of all randomized controlled trial MBI studies of pre-service or in-service school teachers to date. Twelve studies were included in the review and were examined for empirical evidence supporting theoretical hypotheses established in the field and methodological decisions used to determine intervention effectiveness. Results of the review indicate that more empirical evidence is needed examining whether the impact of MBIs for teachers extends into classrooms and improves student outcomes. More evidence is also needed examining the hypothesized causal linkages connecting proximal and distal MBI teacher, classroom, and student outcomes. Assessing these same twelve MBI studies for teachers also indicated that several methodological improvements are needed, including designing studies with larger sample sizes and longer follow-up times of measurement. It was also found that researchers in this field should be utilizing modern missing data methods more frequently to account for attrition and missing data more generally, and that more measures of fidelity of
implementation are also warranted. The results of this review seek to guide future research in this field.

*Keywords:* Mindfulness, Intervention, Teachers
Chapter 2: Study 1

Theoretical and Methodological Extensions of Mindfulness-Based Interventions for Teachers: A Systematic Review of Randomized Controlled Trials

Mindfulness-based interventions, popularized by Jon Kabat-Zinn in the 1990’s in the form of mindfulness-based stress reduction (MBSR; Kabat-Zinn, 1990), are frequently structured as 8-week programs that focus on teaching participants a variety of mindfulness practices meant to be engaged in frequently and applied to everyday life situations to improve wellbeing. Through this lens, mindfulness is often defined as “paying attention in a particular way, on purpose, to the present moment, and non-judgmentally” (Kabat-Zinn, 1990, p. 4). Defined as such, mindfulness can be thought of as a skill that can be developed and refined by practicing intentional attention regulation and an open and accepting attitude toward all present moment experiences (Bishop et al., 2004).

Common mindfulness practices learned in MBIs consist of: directed attention exercises, in which participants focus their attention on their breath or a mental object and purposefully redirect their attention back to their attentional anchor when the mind wanders; body scan exercises, that focus on directing attention through the body while practicing non-reactivity to all sensations, thoughts, and feelings that arise; open monitoring practices, which emphasize a broad focus of attention on the body holistically; and loving kindness practices, which focus on evoking feelings of commonality and kindness towards the self and others (Carmody & Baer, 2008; Davidson & Kaszniak, 2015; Lutz et al., 2008). With regular repetition, each of these
mindfulness practices is thought to help an individual to improve their non-reactivity, attention, regulatory, and compassion capacities respectively, in a manner that allows these skills to be applied more readily to difficult situations that arise in everyday life (Shapiro et al., 2006).

MBI programs developed specifically for teachers share many of the same characteristics with general MBI programs (see Cullen & Pons, 2015; Cullen & Wallace, 2010; Jennings & Greenberg, 2009). Teachers participating in these programs are asked to attend a pre-determined number of in-person sessions over the course of weeks or months, during which they learn many, if not all, of the mindfulness practices described above. Similarly, in-person sessions are supplemented by encouragement to engage in the practices learned during the intervention at home and in the classroom, on a daily or otherwise regular basis. The main difference between MBIs generally, and those created for teachers specifically, is a program focus aimed at helping teachers apply the skills they learn during the intervention to the teaching profession and its unique demands. A teaching-specific focus is integrated into these programs in a variety of ways, including conversations during in-person sessions that help teachers to re-perceive an interaction with a challenging student; homework to apply skills learned in the classroom; practices aimed at forgiving a colleague; and in some cases, ongoing coaching to aid teachers in implementing aspects of the program in the classroom (Jennings et al., 2013; Roeser et al., 2013). In this way, MBIs developed for teachers serve to provide teachers with a series of skills that can be used to buffer against the
stressors associated with their occupation, bolster their general wellbeing, and that can be applied in the classroom context directly to benefit students (Skinner & Beers, 2016).

Fortunately, previous research examining the effectiveness of MBIs for teachers is promising, and has demonstrated significant program impacts on a variety of outcomes. Specifically, research suggests MBIs are successful strategies in reducing teachers’ occupational stress and burnout (Anderson et al., 1999; Roeser et al., 2013; Taylor et al., 2016) and experiences of psychological distress (Anderson et al., 1999; Harris et al., 2016; Jennings et al., 2017; Kemeny et al., 2012; Roeser et al., 2013); and that these programs can be promotive of positive skills and experiences including enhanced mindfulness and self-compassion (Benn et al., 2012; Garner et al., 2018; Harris et al., 2016; Kemeny et al., 2012; Roeser et al., 2013); improved cognitive and self-regulatory abilities (Garner et al., 2018; Jennings et al., 2017, 2013; Roeser et al., 2013) and job satisfaction (Crain et al., 2016). Additionally, research has demonstrated positive program impacts on more distal outcomes as well, including teachers’ physical symptoms of health (Harris et al., 2016; Jennings et al., 2013); quality and quantity of sleep (Crain et al., 2016); and even the quality of interactions between teachers and their students in the classroom (Flook et al., 2013; Jennings et al., 2017).

Given this promising foundation of previous work, the goals of this review were to understand the strengths of previous research examining the impacts of MBIs for teachers and to elucidate areas within the field that need greater focus in the future. The choice to review the current evidence base from both theoretical and
methodological perspectives stemmed from the notion that the quality of the theoretical contribution of a study is largely determined by the methodological decisions that researchers make in planning and implementing their research. As such, it is not only important to understand if and how MBIs impart benefits to teachers, classrooms, and students; but also, to assess the potential ways in which methodological improvements can be made in future intervention studies to increase our collective certainty regarding conclusions made about MBI effectiveness for teachers.

**Methods**

**Search Procedure**

Articles used in this review were identified using two methods: first, I examined recent systematic reviews and meta-analyses of mindfulness training for educators (e.g., Hwang et al., 2017; Klingbeil & Renshaw, 2018; Lomas et al., 2017). Second, the list of studies identified through meta-analysis was supplemented with others found through personal literature searches of relevant key words (e.g., mindfulness intervention, mindfulness training, teachers, education). Using both of these methods, a collection of articles was identified and screened further to assess whether they met the inclusion criteria established for this project.

**Selection Criteria**

Articles were examined based on the following criteria: (1) articles had to be written in English and published in a peer-reviewed journal; (2) studies had to have
employed an experimental, randomized controlled trial design. This required researchers to have designed and implemented the project using, at minimum, a pre-post design (baseline and post-intervention times of measurement), included a comparison (control) group, and used an acceptable method of randomization to assign participants to groups; (3) studies were required to have a large enough sample size to utilize between-group differences statistical tests of significance (e.g., t-tests, regression). A liberal inclusion benchmark of at least 15 participants per group (n = 30 total) was used in order to incorporate as many studies as possible – including promising and well-designed pilot studies or those sampling more than one population (e.g., teachers and parents); (4) studies were included only if they tested the effectiveness of a mindfulness- and/or yoga-based program. However, yoga programs were only included if they had an explicit emphasis on mindfulness; and (5) study populations were limited to pre-service (college students soon entering the teaching profession) or in-service teachers who taught pre-school (i.e., pre-kindergarten) through secondary school (i.e., high school). As such, studies involving college students or school administration populations were excluded from this review.

The total number of published studies that met the above criteria was 12 (see Table 2.1 for study demographics). Within these 12 studies, eight different mindfulness-based programs were evaluated. More specifically, five programs have been evaluated once, one program has been evaluated three times (i.e., CARE), and one program has been evaluated four times (i.e., SMART). It is also important to note that, of the SMART
studies, three publications utilized variants of the same project data examining different outcomes.
<table>
<thead>
<tr>
<th>Research Study</th>
<th>Program Name</th>
<th>Study Information</th>
<th>Sample Size</th>
<th>Grades Taught</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td>Standardized Meditation (SM)</td>
<td>MT for educators and parents of children with special needs</td>
<td>91</td>
<td>Elementary, Middle, &amp; High School</td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>Stress Management and Relaxation Techniques (SMART)</td>
<td>38 educators, 32 parents</td>
<td></td>
<td>Special Education</td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td>SMART</td>
<td>2 Sites (Canada + US)</td>
<td>113</td>
<td>Elementary &amp; Middle School</td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td>Flow Meditation</td>
<td></td>
<td>68</td>
<td>Middle School</td>
</tr>
<tr>
<td>Garner et al., 2018</td>
<td>Mindful Awareness and SEL</td>
<td>Preservice Program</td>
<td>87</td>
<td>Preservice</td>
</tr>
<tr>
<td>Harris et al., 2015</td>
<td>Community Approach to Learning Mindfully (CALM)</td>
<td>Block randomization of two middle schools</td>
<td>42 teachers, 22 other staff</td>
<td>Middle School</td>
</tr>
<tr>
<td>Jennings et al., 2011</td>
<td>Cultivating Awareness and Resiliency in Education (CARE)</td>
<td>Block (Suburban/Semi-Rural) Randomized Pilot</td>
<td>32 pre-service, 11 in-service</td>
<td>Pre-service &amp; In-service mentors</td>
</tr>
<tr>
<td>Jennings et al., 2013</td>
<td>CARE</td>
<td>Matched Pairs</td>
<td>53</td>
<td>Preschool - High School</td>
</tr>
<tr>
<td>Research Study</td>
<td>Program Name</td>
<td>Study Information</td>
<td>Sample Size</td>
<td>Grades Taught</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>-------------------------------------------------------------</td>
<td>-------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>CARE</td>
<td>Cluster Randomized Trial (Schools as natural blocks)</td>
<td>224</td>
<td>Elementary</td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td>Cognitive and Emotional Balance (CEB)</td>
<td>School teachers living with an intimate partner</td>
<td>82</td>
<td>Not Specified</td>
</tr>
<tr>
<td>Roeser et al., 2013</td>
<td>SMART</td>
<td>2 Sites (Canada + US)</td>
<td>113</td>
<td>Elementary &amp; Middle School</td>
</tr>
<tr>
<td>Taylor et al., 2016</td>
<td>SMART</td>
<td>Canada sample from Roeser et al., 2013</td>
<td>59</td>
<td>Elementary &amp; Middle School</td>
</tr>
</tbody>
</table>
Theoretical Review

To review the current state of the field from a theoretical perspective, I drew on theoretical frameworks put forth by experts in the field. Both Roeser and colleagues (Roeser et al., 2012) and Jennings and colleagues (Jennings & Greenberg, 2009) have proposed detailed theoretical frameworks that illustrate how mindfulness training translates into positive outcomes for teachers and later extends to benefit students in the classroom context. While these frameworks differ slightly in detail, both include the same four phases: (1) through the learning and regular practice of program components teachers will develop a set of mindfulness-related skills; (2) through the proximal development of mindfulness-related skills, and directly through participation in the program, teachers will experience enhanced occupational health and wellbeing; (3) teachers’ increased mindfulness skills and greater wellbeing will extend into the classroom by improving teachers’ behaviors and interactions with students while teaching; and (4) teachers’ improved teaching and quality of relationships with students will promote positive social-emotional and academic changes in students. These four phases are depicted in Figure 2.1.

In the sections below, I elaborate on each of the four components highlighted in Figure 2.1, providing first a description of each component in greater detail. Then, I summarize the research about MBI impacts that has been conducted relating to each component, using studies compiled for this review, to assess which areas have a strong foundation of evidence, and where future work should be focused. Additionally,
because well-designed pilot studies were included in this review, I delineate where needed between statistically significant results and those that were not statistically significant (but still noteworthy) given the small sample size of the study.

Figure 2.1

Mindfulness-Based Intervention for Teachers Theory of Change

Note. Theory of change depicts how teachers in the mindfulness intervention, who engage in the program, develop a set of mindfulness-related skills that promote benefits for teachers, their classrooms, and their students. Figure adapted from Roeser et al., 2012.

Developing Mindfulness-Related Skills

As the first step in the theory of change, teachers are hypothesized to develop a set of mindfulness-related skills by learning and engaging in the program, and in particular, by engaging regularly in the mindfulness practices that are emphasized as necessary for intervention effectiveness (Roeser et al., 2012). Mindfulness-related skills is a term used here to define proximal program outcomes associated with an individual’s attention, awareness, regulation, compassion, and/or discernment (van Dam et al., 2018). In this context, mindfulness can be thought of as a skill that can be
developed in a manner that is similar to strengthening the body during physical exercise, such that it requires frequent practice to enact positive change (Tang et al., 2015). Similar to MBIs more generally, MBIs for teachers have incorporated a variety of practices shown to enhance mindfulness-related skills, including those utilized in MBSR — such as attention tasks focused on the breath, body scan practices, and mindful eating and walking exercises. These practices are then supplemented with additional information and exercises focused on emotion, compassion, and forgiveness practices (Cullen & Pons, 2015). Teachers are introduced to these various exercises during in-person program sessions lead by a program facilitator and then encouraged to continue practicing what they learn through a daily home practice that they are asked to establish and maintain throughout the program (Roeser et al., 2013).

Teachers’ active and regular engagement in these practices is seen as an integral component of the program and necessary for teachers to gain the insight and skills that underlie improvements in their mindfulness-related skills and wellbeing. This is because practicing skills such as focused attention, non-judgment, and compassion during a formal mindfulness practice is thought to promote the use of such skills in everyday life situations (Carmody & Baer, 2008). For instance, teachers may experience improvements in their ability to pause before reacting when a student is disruptive with enhanced self-regulatory and attentional capacities (Davidson et al., 2012; Teper et al., 2013). They may also become better able to notice and detect ongoing physiological experiences of stress, which may help them to tune in and initiate changes in their
thoughts or behaviors in a manner that reduces the overall impact of the stressful situation (Farb et al., 2015; Williams et al., 2009). Teachers may also begin to shift their repertoire of coping strategies to rely more heavily on positive and adaptive strategies through enhanced meta-cognitive awareness and decision-making abilities (Gross, 1998; Skinner & Beers, 2016). Last, by practicing compassion and kindness during mindfulness practice, teachers may be more likely to approach difficult interactions with students from a place of shared common experience rather than from one that imposes judgment and criticism (Neff, 2003; Neff & Knox, 2017). In sum, the skills teachers cultivate during formal mindfulness practice both during the intervention, as a group and at home as a personal practice, are thought to bolster the resources available to teachers to help them more successfully manage the high emotional demands of their profession (Chang, 2009; Hargreaves, 2000; Skinner & Beers, 2016).

**Review of Research on Mindfulness-Related Skills.** A review of the studies regarding the current evidence base on mindfulness practices and mindfulness-related skills can be found in Table 2.2. Specifically, each of the 12 MBI studies with teachers was assessed regarding: (1) whether they outline home mindfulness practice requirements for participating teachers; (2) whether they measured and reported teachers’ engagement in those mindfulness practices; (3) whether the studies included measures of mindfulness-related skills as proximal outcomes of their intervention; and (4) whether significant impacts on those mindfulness-related skills were found.
Mindfulness Practice Requirements and Reports. Of the 12 studies reviewed, five studies outlined home practice expectations for their program; five studies mention home practice as a component of the program, but do not describe expectations; and two studies did not mention a practice requirement at all. Of the five studies that discuss home practice requirements, expectations for teachers ranged from 15 minutes to 40 minutes of daily practice. In terms of teachers’ reports of actual practice, five studies reported teachers’ average engagement in practices, with two reporting on teachers’ frequency of practice per week (1 time per week and 2-5 times per week), two providing average minutes spent per day (average of 10 and 16 minutes), and one describing total practice time for the program (12.8 hours). Additionally, only one study reported engagement levels that met program expectations (average of 16 minutes per day with expectations at 15 minutes per day; Roeser et al., 2013).

Measures of Mindfulness-Related Skills. Nine of the 12 studies measured at least one mindfulness-related skill. Specifically, mindfulness itself was measured in nine studies, self-compassion or compassion for others was measured in four studies, and emotion competence or emotion regulation was measured in seven studies. Several other measures, utilized less frequently, are also defined here as mindfulness-related skills, including forgiveness, empathic concern, executive functioning, and coping.

Impacts on Mindfulness-Related Skills. In order to assess the impacts of each MBI on the mindfulness-related skills measured, I calculated percentages that represent the number of studies that found statistically significant impacts on each outcome, over
the total number of studies that measured that outcome. Results indicate significant intervention impacts on teachers’ mindfulness in 78% (7/9) of studies, emotion competence or emotion regulation in 71% (5/7) of studies, and compassion in 75% (3/4) of studies. Significant intervention impacts were also found on forgiveness in two studies and executive functioning in one study.

**Summary.** This brief review of mindfulness practices and mindfulness-related skills in teacher MBIs to date illuminates several important features of the current state of the field. First, there is a need for MBIs with teachers to clarify the role that mindfulness practices have in the mindfulness program being tested. Over half of the studies reviewed did not provide details of the program’s mindfulness practice expectations for participants outside of the intervention sessions and several did not even mention regular practices as a key component of the program. It will be important in the future for researchers to explicate whether or not engaging in mindfulness practices regularly at home is considered a key program component for participant success and whether or not continued practice is required to see lasting intervention benefits.

Second, when mindfulness practice is reported, there is a lot of variation in the manner in which it is described. Some studies report how frequently teachers practiced, but not how long; whereas others reported how long teachers practiced, but not how frequently. This variation in reporting makes it challenging to compare practice across studies. It is also unclear in some cases what is meant by practice – with some studies
including a well outlined definition of home practice requirements and some that may be referring only to practices taught and engaged in during formal intervention sessions. Additionally, as the mindfulness practice component is considered key for maximizing effects in most programs, it is important for future studies to directly study whether the MBI being tested actually promotes teachers’ use of these practices, as well as to better understand specific practice to outcome relations (Parsons et al., 2017).

Third, most of these MBI studies conducted with teachers include measures of mindfulness-related skills. As this is the first key outcome hypothesized to change through participation in a mindfulness intervention, it is important to study explicitly. It is also promising to see that the majority of teacher studies that examined mindfulness-related skills found significant intervention impacts in this area, as this empirical evidence supports the proposed theory of change (Jennings & Greenberg, 2009; Roeser et al., 2013). In contrast, this trend in measuring mindfulness-related skills does not translate as easily to studies of MBIs in general populations. In a meta-analysis of MBIs for nonclinical adult populations, Khoury and colleagues (2015) found that less than half of the studies reviewed included mindfulness as an intervention outcome, despite the importance of this link theoretically.

Lastly, this review of teacher studies has demonstrated that there is a paucity of work in teacher MBI studies that directly tests the link between engaging in mindfulness practices and the development of mindfulness-related skills. Only one study reviewed examined these relations in any manner. Specifically, Kemeny and colleagues (2012)
examined the effect of frequency of practice on outcomes and found significant effects of practice on reductions in teachers’ trait anxiety and increases in teachers’ mindfulness. The authors also found that frequency of practice predicted reductions in teachers’ blood pressure and respiration sinus arrhythmia during a Trier Social Stress Test. It is important to note, however, that Kemeny and colleagues (2012) examined these relations only within the MBI group. As such, it will be important to continue examining these links directly in future work, especially in ways that allows the comparison of both treatment and control groups, to better rule out potential confounding causes of these effects.
### Table 2.2

*Review of Studies – Mindfulness Practices on Mindfulness-Related Skills*

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Daily Practice Requirements</th>
<th>Daily Practice Reported</th>
<th>Measures of Mindfulness-Related Skills</th>
<th>Impacts on Mindfulness-Related Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td>Twice daily for 20 minutes</td>
<td>2-5 times per week</td>
<td>Mindfulness, Self-Compassion, Forgiveness, Empathic Concern, Emotion Regulation Self-Efficacy</td>
<td>Forgiveness*</td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>Homework discussed, but not detailed</td>
<td>Average of 10 minutes formal practice</td>
<td>Mindfulness, Self-Compassion, Forgiveness, Empathic Concern</td>
<td>Mindfulness, Self-Compassion, Empathic Concern¹, Emotion Regulation Self-Efficacy¹</td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td>Homework discussed, but not detailed</td>
<td>Not reported</td>
<td>Mindfulness</td>
<td>Mindfulness*</td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td>40 minutes per day</td>
<td>Not reported</td>
<td>Mindfulness, Emotional Competence, Emotional Display Rules</td>
<td>Mindfulness*, Emotional Competence*, Emotional Display Rules*</td>
</tr>
<tr>
<td>Garner et al., 2018</td>
<td>Not reported</td>
<td>N/A</td>
<td>Mindfulness, Emotional Competence, Emotional Display Rules</td>
<td>Mindfulness*, Emotional Competence*, Emotional Display Rules*</td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
### Table 2.2

*Continued*

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Daily Practice Requirements</th>
<th>Daily Practice Reported</th>
<th>Measures of Mindfulness-Related Skills</th>
<th>Impacts on Mindfulness-Related Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jennings et al., 2011</td>
<td>Not reported</td>
<td>N/A</td>
<td>Mindfulness, Interpersonal Mindfulness, Mindfulness in Teaching</td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2013</td>
<td>Homework discussed, but not detailed</td>
<td>Not reported</td>
<td>Mindfulness, Emotion Regulation</td>
<td>Mindfulness*, Emotion Regulation*</td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>Home practice discussed, but not detailed</td>
<td>Not reported</td>
<td>Mindfulness, Emotion Regulation</td>
<td>Mindfulness*, Emotion Regulation*</td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td>25 minutes per day</td>
<td>Not reported</td>
<td>Mindfulness, Computerized Emotion Battery, Computerized Compassion Task, Marital Interaction Task (hostility + compassion)</td>
<td>Mindfulness*, Computerized Emotion Battery*, Computerized Compassion Task*, Marital Interaction Task*</td>
</tr>
<tr>
<td>Roeser et al., 2013</td>
<td>15 minutes per day</td>
<td>Average of 16 minutes per day</td>
<td>Mindfulness, Occupational Self-Compassion, Executive Functioning</td>
<td>Mindfulness*, Occupational Self-Compassion*, Executive Functioning*</td>
</tr>
</tbody>
</table>

*Note. * indicates a statistically significant intervention impact and ° indicates a notable effect that was not statistically significant.*
### Table 2.2
*Continued*

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Daily Practice Requirements</th>
<th>Daily Practice Reported</th>
<th>Measures of Mindfulness-Related Skills</th>
<th>Impacts on Mindfulness-Related Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taylor et al., 2016</td>
<td>16 hours total</td>
<td>Average of 12.8 hours total</td>
<td>Emotion Regulation Efficacy, Self-Compassion, Forgiveness, Work Forgiveness Efficacy, Qualitative measures of: Coping at Work, Coping Strategies, Compassion Towards Challenging Student</td>
<td>Emotion Regulation Efficacy*, Work Forgiveness Efficacy*, Forgiveness*</td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and † indicates a notable effect that was not statistically significant.
Developing Occupational Health and Wellbeing

In the next step of the theory of change, teachers are hypothesized to see improvements in their occupational health and wellbeing directly through participation in the program, as well as indirectly through their development of mindfulness-related skills (Roeser et al., 2013). Improving teachers’ occupational health and wellbeing has been the primary target of many of these MBI studies due to the need to support teachers in managing the high emotional demands of the teaching profession (Chang, 2013; Jennings et al., 2013; Skinner & Beers, 2016). Sources of stress frequently faced by teachers include student disruptions, workload, and challenging social interactions with students, parents, administrators, and other colleagues (Kyriacou, 2001; Montgomery & Rupp, 2005). Meeting these demands successfully requires teachers to allocate substantial attentional and emotional resources in order to both effectively teach their pedagogical requirements and simultaneously attend to the needs of individual students as well as the whole class collectively (Roeser et al., 2012). As might be expected, managing such demands from day-to-day can substantially decrease a teacher’s energy and enthusiasm for teaching and simultaneously increase their likelihood of experiencing high levels of negative emotions and stress that, when left unmanaged, can lead to chronic experiences of burnout and job dissatisfaction (Chang, 2009).

In the same way that a lack of attentional, emotional, and pedagogical resources can lead teachers down a negative spiral of greater classroom chaos and depleted energy to manage issues as they arise, improvements in teachers’ available resources
through enhanced mindfulness-related skills can set a positive cascade of events in motion (Skinner & Beers, 2016). Directly, MBI program participation may expand teachers’ available positive coping resources to include new ways of managing stress, such as relaxation exercises (Cullen & Pons, 2015). Additionally, the program may support teachers in developing additional coping resources, as is described in the mindfulness-related skills section above. For example, by improving a teacher’s ability to notice physiological indicators of stress earlier in a stress reaction phase, by enhancing attention to and regulation of experiences of stress, by providing teachers with new positive ways of coping and problem-solving, and by reducing the total number of experiences teachers appraise as stressful or negative (Davidson et al., 2012; Skinner & Beers, 2016).

**Review of Research on Occupational Health and Wellbeing.** A summary of the measures and impacts of MBIs on teacher occupational health and wellbeing can be found in Table 2.3. Each of the 12 studies reviewed were assessed by describing (1) measures of occupational health and wellbeing that were included in each study and (2) the impacts of the MBI being tested on those outcomes. Occupational health and wellbeing outcomes were broken into two categories: those that focus on the reduction of negative psychological and physical distress and those that focus on the enhancement of psychological wellbeing. An additional column was also included that assessed whether any studies tested relations between mindfulness-related skills and occupational health and wellbeing outcomes using mediation methods. The purpose of
this last column was to examine whether studies in the field have explicitly tested the connections between these two adjacent steps of the theoretical framework proposed.

**Measures of Psychological and Physical Distress.** All studies, with the exception of one, included at least one measure of teacher psychological or physical distress. The most common construct measured, in 10 of the 12 studies, was psychological distress measured in various ways (e.g., negative affect, anxiety, depression). Stress was the second most common, measured in half of the studies. Other common measures included burnout (5/12 studies), time urgency (4/12 studies), physical symptoms of stress (4/12 studies), and sleep (3/12 studies). Several studies also included more distal physiological aspects of stress including work absences, blood pressure, and cortisol.

**Impacts on Psychological and Physical Distress.** Taken together, results examining the impact of teacher MBIs on occupational health and wellbeing outcomes seem very promising. Specifically, using the same method as before to calculate the number of studies with significant program impacts on outcomes over the total number of studies that included the measure, results indicate strong impact trends across stress (5/6 or 83%), burnout (4/5 or 80%), and psychological distress (8/10 or 80%), with additional noteworthy effects indicated as well. However, results examining more distal or physiological indicators of wellbeing were more mixed, indicated by results for outcomes such as physical symptoms (2/4 or 50%) and sleep (2/3 or 67%).

**Measures of Psychological Wellbeing.** Unlike psychological distress, positive psychological wellbeing constructs have been measured much less frequently. Five of
the 12 studies reviewed did not include a single measure of psychological wellbeing. Of
the seven studies that did include such measures, the most common one was positive
affect (6/12 studies). Additionally, personal growth, life/job satisfaction, and relational
trust were all measured in one study each.

**Impacts on Psychological Wellbeing.** Results examining the effect of MBIs on
psychological wellbeing indicate significant impacts on positive affect in two of the six
studies (33%), a significant impact on life/job satisfaction, and additional
marginal/noteworthy but not statistically significant impacts on some studies examining
positive affect, personal growth, and relational trust.

**Tests of Mediation.** Four studies tested for mediation, examining whether
mindfulness-related skills explain improvements in teacher occupational health and
wellbeing. Specifically, Benn and colleagues (2012) found that levels of post-program
mindfulness mediated program impacts on teachers’ follow-up stress, anxiety, negative
affect, and personal growth. They also examined self-compassion as a mediator of these
effects, but results were not statistically significant. In contrast, Roeser and colleagues
(2013) found both mindfulness and self-compassion at post-program were significant
mediators of teachers’ follow-up stress, burnout, anxiety, and depression. Crain and
colleagues (2016) found significant mediation effects as well, finding that post-program
mindfulness explained teachers’ reduction in negative moods at home and marginally
explained teachers’ improved sleep quality. And lastly, Taylor and colleagues (2016)
found evidence for dispositional forgiveness and marginally for efficacy for forgiving
students as mediators of teacher stress, but found no support for the proposed mediators of efficacy for regulating emotions or situational forgiveness.

Summary. The review of MBIs for teachers regarding the measure and impacts of occupational health and wellbeing in MBIs for teachers indicates a strong evidence base for reducing negative psychological symptoms. As MBIs for teachers are often advertised as stress-reducing programs, it is encouraging to see researchers measuring and finding significant improvements for teachers in these areas. Furthermore, some researchers in the field are examining more distal physiological indicators of stress, such as cortisol, blood pressure (e.g., Roeser et al., 2013), and sleep disruptions (e.g., Crain et al., 2016), and evidence of improvements in these areas is also promising. However, these more distal aspects of psychological distress are still measured infrequently. As such, future work examining MBI impacts on teacher psychological distress should continue to examine the more distal physiological indicators of distress.

Unlike psychological distress, the impact on positive aspects of psychological wellbeing remains understudied in MBIs for teachers. While measuring positive affect, defined as the extent to which individuals regularly experience positive moods such as energy, enthusiasm, and alertness (Watson et al., 1988), is a good starting place, it will be important for future studies to expand measures to include other aspects of positive wellbeing. Other measures of positive psychological wellbeing that have been measured, such as life and job satisfaction, personal growth, and relational trust, should continue to be examined. It would also be useful to expand these measures to include
other aspects of wellbeing, such as teachers’ happiness (Hervás & Vázquez, 2013) and work-specific energy and enthusiasm (Seppälä et al., 2009).

Additionally, several studies have examined and found evidence to support improved mindfulness-related skills as mechanisms through which MBIs bolster teacher occupational health and wellbeing. These results provide empirical support for the connection between mindfulness-related skills to occupational health and wellbeing. However, future research should continue to examine these relations, as the four studies that tested mediating effects examined differing mechanisms and outcomes, and conclusions regarding mediating mechanisms were mixed.
Table 2.3  
Review of Studies – Occupational Health and Wellbeing

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Measures of Psychological &amp; Physical Distress</th>
<th>Impacts on Psychological &amp; Physical Distress</th>
<th>Measures of Psychological Wellbeing</th>
<th>Impacts on Psychological Wellbeing</th>
<th>Tests of Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td>Stress, Burnout, Anxiety</td>
<td>Stress*, Burnout*, Anxiety*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>Stress, Anxiety, Depression, Negative Affect</td>
<td>Stress*, Negative Affect*, Anxiety¹, Depression¹</td>
<td>Personal Growth, Positive Affect</td>
<td>Personal Growth¹, Positive Affect¹</td>
<td>Post-program Mindfulness mediated follow-up Stress*, Anxiety*, Negative Affect*, and Personal Growth*.</td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td>Job Rumination, Bad Mood at Work &amp; Home, Sleep Quality, Quantity, &amp; Sleepiness</td>
<td>Job Rumination*, Bad Mood at Work*, Sleep Quality*, Sleep Quantity*, Sleepiness*</td>
<td>Life &amp; Job Satisfaction, Good Mood at Work &amp; Home</td>
<td>Life &amp; Job Satisfaction*</td>
<td>Post-program Mindfulness mediated Negative Mood at Home* and Sleep Quality¹</td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td>Psychological Distress</td>
<td>Psychological Distress*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
<table>
<thead>
<tr>
<th>Research Study</th>
<th>Measures of Psychological &amp; Physical Distress</th>
<th>Impacts on Psychological &amp; Physical Distress</th>
<th>Measures of Psychological Wellbeing</th>
<th>Impacts on Psychological Wellbeing</th>
<th>Tests of Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Garner et al., 2018</td>
<td>Perceived Stress, Burnout, Negative Affect, Distress Tolerance, Time Urgency, Physical Symptoms, Sleep-related Impairment, Blood Pressure, Cortisol</td>
<td>Distress Tolerance*, Physical Symptoms*, Diastolic BP*, Cortisol*, Time Urgency¹, Burnout¹</td>
<td>Positive Affect, Relational Trust</td>
<td>Positive Affect* Relational Trust¹</td>
<td></td>
</tr>
<tr>
<td>Harris et al., 2015</td>
<td>Depression, Negative Affect, Time Urgency, Physical Symptoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2011</td>
<td></td>
<td></td>
<td>Positive Affect</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
### Table 2.3

**Continued**

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Measures of Psychological &amp; Physical Distress</th>
<th>Impacts on Psychological &amp; Physical Distress</th>
<th>Measures of Psychological Wellbeing</th>
<th>Impacts on Psychological Wellbeing</th>
<th>Tests of Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jennings et al., 2013</td>
<td>Burnout, Depression, Negative Affect, Time Urgency, Physical Symptoms</td>
<td>Burnout*, Time Urgency*, Physical Symptoms*</td>
<td>Positive Affect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>Stress, Burnout, Anxiety, Depression, Negative Affect, Time Urgency, Physical Symptoms, Sleep Disturbance</td>
<td>Psychological Distress (Stress, Negative Affect, Depression, Anxiety, Sleep Disturbance, and Burnout composite)<em>, Time Urgency</em></td>
<td>Positive Affect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td>Anxiety, Depression, Negative Affect, Ruminition, Trier Social Stress Test (Blood Pressure, Respiration, Respiration Arrhythmia)</td>
<td>Anxiety*, Depression*, Negative Affect*, Rumination*, Blood Pressure*</td>
<td>Positive Affect</td>
<td>Positive Affect*</td>
<td></td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
<table>
<thead>
<tr>
<th>Research Study</th>
<th>Measures of Psychological &amp; Physical Distress</th>
<th>Impacts on Psychological &amp; Physical Distress</th>
<th>Measures of Psychological Wellbeing</th>
<th>Impacts on Psychological Wellbeing</th>
<th>Tests of Mediation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roeser et al., 2013</td>
<td>Stress, Burnout, Anxiety, Depression, Blood Pressure, Heart Rate, Cortisol, Absences from Work</td>
<td>Stress*, Burnout*, Anxiety*, Depression*</td>
<td>Absences from Work¹</td>
<td></td>
<td>Post-program Mindfulness and Self-Compassion mediated follow-up Stress*, Burnout*, Anxiety*, and Depression*</td>
</tr>
<tr>
<td>Taylor et al., 2016</td>
<td>Stress</td>
<td>Stress*</td>
<td></td>
<td></td>
<td>Dispositional Forgiveness* and Efficacy for Forgiving Students¹ mediated Stress</td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
Extending Impacts into the Classroom Context

The third step in the proposed theoretical framework suggests that teachers’ own self-improvement will extend into the classroom climate through positive changes in their teaching and their interactions with students (Jennings & Greenberg, 2009; Roeser et al., 2012; Skinner & Beers, 2016). Specifically, it is hypothesized that with added coping resources (in the form of mindfulness-related skills) and improved occupational wellbeing (less job-related stress) teachers will be able to more readily allocate attentional and regulatory resources to the teaching task at hand (Roeser et al., 2013; Skinner & Beers, 2016). First, with additional resources “free” to allocate elsewhere, teachers may see improvements in their ability to effectively organize and manage their classroom. Specifically, teachers may increase their use of proactive, rather than reactive, classroom management strategies with greater frequency by anticipating issues before they arise (Roeser et al., 2012).

Second, MBIs for teachers may help to enhance positive relationships between teachers and students. With greater confidence, efficacy, and enjoyment of teaching, teachers may be better able to support the needs of their students, both emotionally and academically (Jennings & Greenberg, 2009; Skinner & Beers, 2016). For instance, teachers’ greater feelings of non-judgment and non-reactivity may help them to view student misbehavior or other challenges that occur in the classroom with greater patience and understanding. Enhanced understanding and clarity during moments of stress may also increase the likelihood that a teacher responds to chaos or disruption in
the classroom with kindness and compassion, rather than reacting in a harsh or critical manner. Additionally, greater attentional resources may help teachers to more successfully attend to the needs of individual students and the whole class collectively at any given time (Jennings & Greenberg, 2009; Roeser et al., 2012).

**Review of Research on Classroom Contexts.** The review of studies examining the impacts of MBIs on classroom outcomes can be found in Table 2.4. Each study was summarized in terms of measures used and MBI impacts on teaching-specific skills (e.g., teaching efficacy) and classroom climate (e.g., the quality of teacher-student interactions). All studies were also examined for mediation analyses testing teacher mindfulness and/or occupational health and wellbeing outcomes as mechanisms that explain improved classroom outcomes. However, only one study was found to have tested such relations. This single mediation report is discussed in-text.

**Teaching Skills.** Only three measures of teaching skills were used as outcomes across all studies reviewed. Specifically, five studies examined teaching efficacy, or a teacher’s belief in their own abilities to effectively teach students. Additionally, one study examined impacts on teachers’ levels of autonomy support, specifically indicating how supportive and/or controlling they are of students. And one study examined teachers’ perceptions of the cognitive and social costs associated with student misbehavior on student development. Impact reports on these teaching outcomes indicate that only two of the five studies (40%) examining teaching efficacy found significant program effects, with two additionally reporting noteworthy effects that did
not reach statistical significance. Significant intervention effects were reported in both studies examining autonomy support and costs of student misbehavior. Garner and colleagues (2018) also examined whether changes in teachers’ mindfulness and emotional competence explained teachers’ increased perceptions of the costs of misbehavior on students’ development, however, the results of these mediation tests were not significant.

**Classroom Climate.** Two studies have examined the greater classroom climate using the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). The CLASS is an observational tool designed to assess the quality of a teacher’s interactions with their students using the three domains of instructional support, emotional support, and classroom management. Of the two studies examining the impact of an MBI on the CLASS, only one study found statistically significant effects. Specifically, Jennings and colleagues (2017) found significant improvements in the emotional support domain and marginal improvements in the instructional support domain of the CLASS.

**Summary.** Evidence examining intervention impacts on teacher behaviors in the classroom context is promising, however, generally more work is needed in this area. Only six of the 12 studies reviewed examine the potential impact of the MBI on the classroom context and results that are reported are mixed. It will be important for future work to continue focusing on examining these distal program impacts, to better understand how MBIs affect teachers’ behaviors in the classroom, and whether and how MBIs impact the classroom climate. Additionally, broadening conceptualizations of
teaching skills and the types of observational measures that are used in the classroom
to measure teachers’ behaviors will provide researchers with an opportunity to better
understand whether and how mindfulness manifests in the classroom at a more
nuanced level.
Table 2.4

Review of Studies – Classroom Impacts

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Measures of Teaching Skills</th>
<th>Impacts on Teaching Skills</th>
<th>Measures of Classroom Interactions</th>
<th>Impacts on Classroom Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>Teaching Efficacy</td>
<td>Teaching Efficacy¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garner et al., 2018</td>
<td>Costs of Student Misbehavior to Development</td>
<td>Costs of Student Misbehavior to Development*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harris et al., 2015</td>
<td>Teaching Efficacy</td>
<td>Teaching Efficacy*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
### Table 2.4
*Continued*

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Measures of Teaching Skills</th>
<th>Impacts on Teaching Skills</th>
<th>Measures of Classroom Interactions</th>
<th>Impacts on Classroom Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jennings et al., 2011</td>
<td>Teaching Efficacy,</td>
<td>Autonomy Support*</td>
<td>Classroom Assessment Scoring System</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Autonomy Support</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Teaching Efficacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2013</td>
<td>Teaching Efficacy</td>
<td>Teaching Efficacy*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>Teaching Efficacy</td>
<td></td>
<td>Classroom Assessment Scoring System</td>
<td>Classroom Assessment Scoring System*</td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roeser et al., 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor et al., 2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. * indicates a statistically significant intervention impact and ¹ indicates a notable effect that was not statistically significant.
Improving Student Outcomes

As the last step in the theory of change, MBIs are hypothesized to influence students’ emotional and academic development in positive ways via several pathways. First, students may be impacted directly by the teacher’s ability to model positive coping strategies, such as emotion regulation, proactive problem-solving, and compassion for others (Jennings & Greenberg, 2009). Each time a teacher applies these skills and strategies to challenging situations in the classroom represents an opportunity for students to learn successful coping strategies and begin developing these same skills within themselves. The more teachers are able to confront ongoing challenges with calmness, clarity, and compassion, the more they scaffold student’s own development of these social-emotional skills—skills that can then support students in feeling more efficacious when faced with their own emotional and academic challenges (Jennings & Greenberg, 2009; Rickert, Skinner, & Roeser, 2019; Roeser et al., 2012).

Second, MBIs are also hypothesized to positively affect students’ social-emotional and academic development indirectly, through a teacher’s ability to establish and maintain well organized and managed classrooms that are supportive of students. Specifically, instructional support behaviors, such as scaffolding and noticing when students need assistance, can help students to cultivate necessary skills and strategies (e.g., problem-solving, persistence) that support their learning and academic success (Wentzel, 2010). Additionally, high levels of organization and structure in the classroom, such as clearly stated expectations that are consistently enforced and explicit
communications of relevance, can help students to feel efficacious and invested in meeting and exceeding learning expectations (Wentzel, 2010). Lastly, provisions of emotional support, such as expressions of kindness, respect, and compassion promote closeness and trust among teachers and students. These close relationships in turn help students to value learning and be intrinsically motivated to work hard in achieving academic success (Wentzel, 2010; Wigfield et al., 2015).

**Review of Research on Student Outcomes.** No studies reviewed have examined impacts of MBIs for teachers on student outcomes. Future work is needed to test whether MBIs targeting teachers can have far reaching effects on students as is proposed and discussed in the theory of change. Specifically, it will be important to examine the impact of MBIs on both social-emotional and academic outcomes for students, as well as potential lasting effects of these programs in supporting students’ long-term success in school.
Methodological Review

Next, I conducted a methodological assessment of the studies included in this review to complement the theoretical analysis reported in the previous section. The purpose of this second review was to evaluate each study based on several methodological features of RCTs that can influence conclusions made about program effectiveness. In many ways RCTs are considered the “gold standard” research design because they include a process of randomizing participants to conditions and high internal control—both of which enhance our ability to make strong causal claims about intervention effectiveness (Shadish et al., 2002; Wilson & Lipsey, 2001). These strengths of design increase confidence in determining whether a program “works” by helping to rule out any potential confounding factors that might systematically bias results and conclusions.

Unfortunately, the RCT design’s strong reputation makes it easy to overlook the many factors that can threaten a study once it is underway. Factors such as a study’s design (e.g., the comparison group used, the sample size collected), the way in which an intervention is implemented, and the choices researchers make in analyzing data, can all influence conclusions about a program’s effectiveness (Bickman & Reich, 2015; Durlak, 2010; Mark & Lenz-Watson, 2011; Wilson & Lipsey, 2001). The factors discussed here are those that can threaten an RCT study’s integrity and limit its strengths, however, there are also limitations to the design itself (Skinner et al., 2019). General RCT design limitations – such as an RCT’s lack of focus on answering process-based “how” and
“why” questions, participant noncompliance, and potential restricted generalizability (Skinner et al., 2019) – are outside the scope of factors evaluated here.

In order to maximize confidence in conclusions made and minimize mistakes in this high-stakes context, it is important to understand the underlying assumptions of the RCT design, the numerous ways in which its strengths can be threatened, and recommendations of best practices for managing threats when they do occur. Increased certainty regarding the effectiveness of intervention programs is important, as decisions about whether a program “works” can meaningfully impact individuals, groups, and society (Cizek & Rosenberg, 2011). For instance, if a program is incorrectly determined to be effective when in fact it is not, society may waste resources on a program that, at best, does no harm to participants. However, it is also problematic if a promising program is discarded. This latter mistake can potentially impact society on a larger scale, by impeding innovation and progress in solving complex societal problems in new and innovative ways (Leviton, 2011).

As such, each of the studies included in this review were assessed based on methodological factors that can influence an intervention before, during, and after an intervention is over. Specifically, Figure 2.2 depicts all factors that were examined and these factors were broken into three categories: (1) research design features, (2) research implementation, and (3) other factors that influence conclusions made about intervention effectiveness. Studies were reviewed in each of these three areas in order to understand the field’s strengths and what areas still need to be addressed in future
work to better align with current methodological recommendations for conducting high quality RCT studies.

**Figure 2.2**

*Methodological Factors Examined That Influence Conclusions of Effectiveness*

![Diagram showing methodological factors examined]

*Note.* Methodological factors can influence conclusions made about intervention effectiveness before, during, and after an intervention is over. Factors included in this review were considered an important subset of factors, rather than comprehensive.

**Research Design Features**

The way an intervention is designed can influence conclusions made about its effectiveness in numerous ways. The study design itself holds a series of assumptions regarding the nature of the phenomena that is supposed to change and what questions are asked about that phenomena (Mark & Lenz-Watson, 2011). For instance, whether an intervention is designed as a waitlist-controlled trial or an active waitlist-controlled trial alters the very nature of the question the intervention is meant to address (Miller, 2007). Changing from a waitlist-controlled trial to an active waitlist-controlled trial shifts the question from “does the intervention work better than business as usual?” to “does the intervention work better than another similar intervention?” The latter design helps
to elucidate unique active ingredients that might be driving changes due to intervention participation, for instance, asking, “is it relaxation or uniquely mindfulness practices that cause teachers to feel less stress through mindfulness training?”.

Deciding on a study’s unit of randomization – whether it be individuals, classrooms, schools, or districts – holds other assumptions. Randomizing at the individual level assumes that the treatment being tested will not indirectly affect the control group through their shared social interactions with the treatment group. In the context of MBIs for teachers, this assumes that teachers who work in the same school, some randomized to the treatment group and some to the control group, are not going to discuss or share MBI program content with each other. This concept is regarded in the field as a study’s potential for “contamination” (Rhoads, 2011). As contamination increases, the likelihood of finding significant differences between groups at the end of the intervention decreases, thus reducing the likelihood of finding a significant intervention effect. To solve this issue, researchers can adapt the basic RCT design instead employing a cluster-level design. The cluster-level design eliminates contamination by shifting the unit of randomization from the individual to a cluster, such as a school or a district (Bloom, 2005). This allows all individuals within a defined space, and that are assumed to have regular interaction, to be assigned to the same condition – thus eliminating the threat of contamination.

There is also an innate assumption in RCT studies that the randomization process will create experimental groups that are equal on all measured and unmeasured
characteristics (Shadish et al., 2002). However, this assumption too can be violated if a subset of participants share a specific quality that might make an intervention more or less likely to be effective. When characteristics like this are identified, researchers can use several additional techniques, such as block randomization or matched pairs, to further ensure that the randomization processes are successful (Bloom, 2005; Skinner et al., 2019). Both block randomization and matched pairs methods capitalize on identifying potential variables that may systematically bias results at the start of a study. For example, if researchers believe the impacts of an MBI for teachers might be stronger for teachers working in economically disadvantaged schools, they may choose to ‘block’ teachers by school economic status and randomly assign half of the economically disadvantaged schools to the treatment group and the other half to the control group. This ensures that the majority of one subgroup does not end up in one group or another through basic random assignment (Bloom, 2005).

Lastly, the decision of whether or not to include follow-up times of measurement holds its own assumption about the nature of change expected through program participation (Miller, 2007). RCT designs require a minimum of two time points of measurement – pre-test and post-test. However, most interventions applied in community settings are designed to have lasting impacts in the population of interest. For instance, MBIs for teachers have been developed primarily for the purposes of supporting teachers in managing the heavy workload and ongoing stressors they face in their occupation on a regular basis. As such, it is important to not only determine that
MBIs are effective in the short-term, but also to examine whether they are providing lasting benefits for teachers after the intervention is over. Including follow-up time points of measurement within the RCT framework allows researchers to answer this latter very important question.

**Review of Research on Research Design Features.** Studies included in this review were assessed on several features of research design, including (1) study design, (2) the type of comparison or control group that was included, and (3) whether or not any follow-up times of measurement were assessed. Results of the review of research design features can be found in Table 2.5.

**Study Design.** Of the studies examined, eight utilized a basic RCT design with teachers individually randomized to treatment and control groups; three randomized at the teacher level but accounted for schools as natural blocking variables; and one used a matched pairs design.

**Comparison Group.** All studies included in this review utilized either a waitlist control or active waitlist control comparison group. The waitlist-controlled studies (10 of 12), indicate that teachers randomly assigned to the control group were offered access to the mindfulness program being tested after the intervention study was complete. The two remaining studies utilized an active waitlist control group design, indicating that teachers randomly assigned to the control group received some form of alternative treatment (i.e., music relaxation or meditative awareness practice) during the
intervention phase and were offered access to the intervention components that were withheld in their group, after the intervention phase was over.

**Times of Measurement.** Five of the 12 studies utilized a pre-post design. The remaining seven studies included one additional follow-up time of measurement, ranging from two to five months after the intervention concluded.

**Summary.** In general, the studies reviewed here have been designed extremely well, utilizing some form of waitlist control group and frequently following up the basic pre-post design with an additional time of measurement. Choosing a waitlist design, rather than a control group that doesn’t receive access to a program after the study is complete, may benefit the project itself through enhanced enthusiasm and participation of control group participants. This choice of design, despite being more expensive to implement, also has the potential to benefit more of the study’s target population and therefore enact greater change in the social problem being addressed.

Some studies reviewed here also went a step further to adopt an active waitlist control design, which not only allows researchers to provide resources to all study participants, but that is also well-suited to examine whether specific components of the program are driving change. For example, Garner and colleagues (2018) compared a mindful awareness and social-emotional learning curriculum group to a mindful awareness only group. This design allowed the researchers to test explicitly what benefits the SEL component of the program had, above and beyond the basic features of the intervention that allowed teachers intentional time to focus on their wellbeing.
and learn mindful awareness practices. Additionally, the waitlist control aspect of this design maximizes benefits for teachers as the target population by providing the active control group participants access to the social-emotional learning aspect of the intervention that they missed as part of the efficacy study, after the study was complete.

Many of the studies reviewed here used a basic RCT design, randomizing teachers themselves to treatment and control groups. The use of the basic single-level RCT design in this way is not problematic in a general sense, rather it depends on the level of contamination that occurs through treatment/control group interactions that is of concern (Rhoads, 2011). As such, it is not possible to judge the use of this design without knowing details of each intervention and how many teachers in each condition have shared work environments. It would be useful for future studies to include more information about these concerns and include a justification as to why the basic RCT is acceptable. Examples of ways researchers could incorporate this are: reporting the number of teachers within each school, testing shared variance among clusters (i.e., assessing the intraclass correlation; Hedges & Hedberg, 2007), or reporting reasons why low levels of contamination are anticipated (Bloom, 2005; Rhoads, 2011). An exemplary report of this information reviewed here was provided by Jennings and colleagues (2017), who randomized at the teacher level and provided a theoretical justification for their hypothesis of low contamination. Still, they also utilized hierarchical linear
modeling to account for variation among schools and to increase their likelihood of detecting an intervention effect.

Table 2.5
Review of Studies – Research Design Features

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Intervention Design</th>
<th>Active WLC</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td>Individual Randomization</td>
<td></td>
<td>5-month</td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>Individual Randomization</td>
<td></td>
<td>2-month</td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td>Individual Randomization</td>
<td></td>
<td>3-month</td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td>Individual Randomization</td>
<td>Music Relaxation</td>
<td>4-month</td>
</tr>
<tr>
<td>Garner et al., 2018</td>
<td>Individual Randomization</td>
<td>Meditative Awareness only (no SEL)</td>
<td></td>
</tr>
<tr>
<td>Harris et al., 2015</td>
<td>Block Randomization (two middle schools)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2011</td>
<td>Block (Suburban/Semi-Rural) Randomized Pilot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2013</td>
<td>Matched Pairs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>Cluster Randomized Trial</td>
<td>(Schools as Natural Blocks)</td>
<td></td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td>Individual Randomization</td>
<td></td>
<td>5-month</td>
</tr>
<tr>
<td>Roeser et al., 2013</td>
<td>Individual Randomization</td>
<td></td>
<td>3-month</td>
</tr>
<tr>
<td>Taylor et al., 2016</td>
<td>Individual Randomization</td>
<td></td>
<td>4-month</td>
</tr>
</tbody>
</table>
Chapter 2: Study 1

Intervention Implementation

Implementation science, otherwise known as fidelity of implementation, is a relatively new branch of science that allows researchers to measure the degree to which an intervention is implemented as designed and whether program participants are present and receptive to the program content (Durlak, 2010; Durlak & DuPre, 2008). It is important for researchers to measure implementation, because understanding the nuances of how an intervention plays out can help to identify potential problems with an intervention design. More importantly, if a significant program effect is not found, measures of implementation help to diagnose whether the lack of program effects is due to a program that truly doesn’t work, or just because the intervention wasn’t conducted as it was originally designed.

Program implementation can be measured in a variety of ways, some focused on ensuring an intervention is achievable prior to examining wide-scale use (e.g., program feasibility, acceptability), whereas others are focused on the processes that take place throughout the administration of any intervention (i.e., fidelity of implementation). Below is a summary of many key features of implementation science that are discussed in the literature.

Program feasibility refers to features of the program design that indicate that a program is able to be successfully implemented as designed; specifically, that program demands are manageable, that the program is possible given the set of resources available (e.g., financial resources for program materials, knowledgeable and available...
trainers), and that the program is designed to align with interests from key stakeholders within the community settings of focus (Jephson, 1992). Program acceptability takes the view of the participant in describing the extent to which participants find the program content worthwhile. The degree of interest a participant has in the program will influence their willingness to “buy in” to the program and be motivated to participate (Nastasi et al., 2000). Defined as such, program feasibility and acceptability represent areas of program implementation that should be considered during program design, to maximize the likelihood a program will be effective. When measured well, these two features of general program implementation help to unconfound the conclusion that a program “doesn’t work” from those that determine that a program is not ready to be implemented or that is misaligned to be effective in the population of interest.

Whereas program feasibility and acceptability refer to implementation science components that are largely defined, refined, and solidified prior to large-scale program implementation, fidelity of implementation can be applied to any size study and refers to the degree in which a program is actually implemented as was intended when being tested for effectiveness (Durlak & DuPre, 2008). Generally speaking, fidelity of implementation can be broken into two categories: implementation team efforts to deliver the program to participants, and participant receptivity and engagement in the program. Implementation team components include: program adherence, defined as the degree in which program components are implemented as planned; program quality, which describes the way in which program components are implemented (i.e.,
knowledge and skill of the implementation team); *program dosage*, which quantifies the “full dose” of the program (i.e., the total number of sessions that comprise the “treatment” by program design); *monitoring*, which describes actions taken to understand experiences that both treatment and comparison groups are having that may be related to program content; and *adaptation*, which requires researchers to document changes or additions to program content that may occur during implementation (Berkel et al., 2011; Durlak & DuPre, 2008).

Participant receptivity is usually measured by *participant responsiveness*, which documents the actual dose a participant receives (via attendance to program components), as well as a participant’s engagement and interest in the program. Program acceptability, described above, can be considered one aspect of participant responsiveness, because it measures how useful and important participants believe the program to be. On another level, fidelity of implementation also seeks to compare the quality of similar studies on a broader scale through *program differentiation*, which seeks to distinguish active program components and the program theory of change from similar programs that are also available (Berkel et al., 2011; Durlak & DuPre, 2008).

**Review of Research on Implementation Science.** Table 2.6 summarizes the extent to which fidelity of implementation was reported in all studies included in this review. Implementation measures were broken into several components: (1) program format, which is a descriptive measure of program differentiation; (2) program dose, broken further into formal program sessions, expectations for home practice, and actual
daily practice amounts reported; and (3) other implementation measures that were reported in each study, including program feasibility, acceptability, adherence, and quality.

**Intervention Format.** Interventions assessed vary between four and 16 weeks, with six studies designed to be conducted between four and six weeks, three studies matching the traditional MBSR format of an eight-week duration (Kabat-Zinn, 1990), and three programs extending beyond that recommendation to a maximum of 16 weeks.

**Intervention Dose.** Formal program sessions ranged from nine to 42 hours of program content. Many programs included 90- to 150-minute after school sessions that occurred over the predetermined number of intervention weeks. Some programs (e.g., Benn et al., 2012; Roeser et al., 2013) included several full-day sessions to supplement the after-school sessions, whereas other programs, such as is implemented by Jennings and colleagues (Jennings et al., 2017, 2013; 2011), only include four day-long program sessions supplemented by one-on-one coaching. Still others, such as Harris and colleagues (2016) only require teachers to attend two 20-minute sessions per week that are offered before school.

As is described in the theoretical review, daily practice requirements ranged from 15 to 40 minutes per day for those five studies that reported daily expected practice times. However, this subsample of results indicates that seven studies did not explicitly state what practice requirements were expected. Furthermore, only five studies reported teachers’ actual average amounts of home practice, with two studies
reporting average minutes spent per day (averages of 10 and 16 minutes per day), one study reporting total number of minutes in hours (12.8 hours), and two studies reporting frequency of practice (once per week and 2-5 times per week).

**Other Measures of Implementation.** Researchers also frequently reported on several other measures of implementation: attendance of formal program sessions was reported in seven of the 12 studies (58%), program adherence was measured in four studies (33%), program quality was measured in three studies (25%) and participant responsiveness was measured in seven studies (58%).

**Summary.** Generally speaking, studies of MBIs for teachers demonstrate the importance of measuring fidelity through the inclusion of one or more fidelity measures described above. The descriptive examination of the different types of programs that have been implemented and tested shows a large amount of variation across programs regarding program length and dose. This descriptive look at program differentiation leaves many questions to be answered regarding what dose is sufficient to obtain program impacts at the level desired, but that also maximizes participation by reducing the burden of expectations on teachers. Future work could utilize program dose information, such as attendance to program sessions and participation in home practice to better understand and refine program expectations.

Frequent measurement of program feasibility, acceptability, and participant responsiveness demonstrate that researchers are considering alignment of program components to teachers’ interests and testing whether teachers find program demands
manageable to complete. Some studies even go beyond these general measures and provide information regarding teachers’ perceived benefits from the program, and how likely they are to recommend the program to other teachers (e.g., Roeser et al., 2013). It is encouraging to see researchers using these measures to better understand how these programs are viewed from the perspective of the participant, as understanding whether teachers like and can easily accomplish the goals of the program and further support the use of these programs in teacher populations. However, future work could continue to expand this area of research, by going beyond descriptions of participant responsiveness and explicitly testing participant engagement to outcome relations, as most often, researchers can find program impacts are greater under conditions of higher participation and engagement (Berkel et al., 2011).

Several areas of implementation are less frequently reported, however, such as program quality and adherence; and some important features of fidelity, such as adaptation and monitoring of experiences in both groups, are not discussed in any studies reviewed here. It will be important to consider the measurement of these areas in future work, to fully understand how fidelity of implementation is affecting outcomes for teachers. For instance, it may be particularly important to document how interventions are adapted to fit each cohort of teachers, as changing instructional features to align with participant learning styles and interests may enhance intervention outcomes (Berkel et al., 2011). Additionally, monitoring both treatment and control groups on concurrent experiences they may be engaging in that are related to the
intervention, such as attending community held mindfulness practice sessions, is important for several reasons. Monitoring the control group would help researchers control for intervention-like experiences that might change the control group to be more similar to the treatment group and reduce the likelihood of finding an intervention effect. Whereas monitoring the treatment group would help researchers to disentangle whether program impacts were caused by the intervention itself or other relevant experiences that the treatment group had during the intervention (Berkel et al., 2011).
Table 2.6
*Review of Studies – Intervention Implementation*

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Program Format</th>
<th>Formal Program Sessions</th>
<th>Daily Practice Requirements</th>
<th>Daily Practice Reported</th>
<th>Other Implementation Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td>5 Weeks</td>
<td>9 hours (five 1.5-hour weekly classes + one follow-up session 1-month post-program)</td>
<td>Twice daily for 20 minutes</td>
<td>2-5 times per week</td>
<td>Program Completion Rates/Attendance, Participant Satisfaction, Qualitative reports of Implementor Adherence</td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>5 Weeks</td>
<td>36 hours (nine 2.5-hour sessions + 2 full 6-hour days)</td>
<td>Homework discussed, but not detailed</td>
<td>Average of 10 minutes formal practice</td>
<td></td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td>8 Weeks</td>
<td>36 hours (nine 2.5-hour sessions + two 6-hour days)</td>
<td>Homework discussed, but not detailed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td>10 weeks</td>
<td>15 hours (ten 1.5-hour sessions)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Garner et al., 2018</td>
<td>6 Weeks</td>
<td>12 hours (six 2-hour sessions)</td>
<td></td>
<td></td>
<td>Quality of Intervention Instruction</td>
</tr>
</tbody>
</table>
### Table 2.6
*Continued*

<table>
<thead>
<tr>
<th>Research Study</th>
<th>Program Format</th>
<th>Formal Program Sessions</th>
<th>Program Dose</th>
<th>Other Implementation Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris et al., 2015</td>
<td>16 Weeks</td>
<td>11 hours (32 20-min sessions; &gt;= 2 sessions/week)</td>
<td>Weekly personal practice prompted, but required duration not discussed</td>
<td>Participant Feasibility, Attendance, Adherence measured by implementor self-report and 20% third-party observation of intervention sessions.</td>
</tr>
<tr>
<td>Jennings et al., 2011</td>
<td>4-6 Weeks</td>
<td>24 hours (4 day-long sessions + coaching)</td>
<td></td>
<td>Program Attendance, Program Satisfaction (feasibility, acceptability)</td>
</tr>
<tr>
<td>Jennings et al., 2013</td>
<td>4-6 Weeks</td>
<td>30 hours (4 day-long sessions + coaching and 1-day booster)</td>
<td>Homework discussed, but not detailed</td>
<td>Participant Satisfaction, Adherence (self-report &amp; observed)</td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>4-6 Weeks</td>
<td>30 hours (4 day-long sessions + coaching and 1-day booster)</td>
<td>Home practice discussed, but not detailed</td>
<td>Program Attendance, Adherence, Quality, Participant Satisfaction</td>
</tr>
<tr>
<td>Research Study</td>
<td>Program Format</td>
<td>Formal Program Sessions</td>
<td>Program Dose</td>
<td>Other Implementation Measures</td>
</tr>
<tr>
<td>--------------------</td>
<td>----------------</td>
<td>-------------------------</td>
<td>---------------------------------------</td>
<td>----------------------------------------------------</td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td>8 Weeks</td>
<td>42 hours (4 day-long sessions + 4 evening sessions)</td>
<td>25 minutes per day</td>
<td>Daily and total minutes used in analyses but not reported otherwise</td>
</tr>
<tr>
<td>Roeser et al., 2013</td>
<td>8 Weeks</td>
<td>36 hours</td>
<td>15 minutes per day</td>
<td>Average of 16 minutes per day</td>
</tr>
<tr>
<td>Taylor et al., 2016</td>
<td>9 Weeks</td>
<td>36 hours</td>
<td>16 hours total</td>
<td>Average of 12.8 hours total</td>
</tr>
</tbody>
</table>
Factors that Influence Conclusions of Intervention Effectiveness

There are numerous additional methodological factors, other than the design and implementation features described above, that influence whether or not researchers accurately conclude that an intervention is effective in enacting change. The following supplementary factors focused on here include (1) sample size; (2) attrition and methods for dealing with missing data; and (3) analysis specification. This section is not meant to comprehensively cover all influences in this area, but rather to describe several important factors for intervention research that are less widely known. A more detailed discussion describing each of these three factors and why they are important for intervention research is provided below.

Sample Size. At its most basic level, the sample size of a study has bearing on whether or not intervention effects are detected as statistically significant. Sample size works in tandem with an intervention’s minimum detectable effect size (the smallest true intervention effect that can be detected statistically), level of statistical precision (a predetermined level typically set at .05 for psychological research), statistical power (the study’s likelihood of detecting a significant intervention effect if there truly is one, which is typically held constant at 80%), and the choice in statistical test (one- or two-tailed; Bloom, 2005; Bloom et al., 2007). Typically, the statistical precision, power level, and appropriate statistical test are all held at predetermined acceptable levels within a field of study. Additionally, the predicted minimum effect size of an intervention is typically known from pilot study work on the intervention or estimated using
theoretically relevant benchmarks (Bloom, 2005). This often leaves sample size as the only manipulable factor that researchers can control to instill the most confidence possible in the conclusions they make about their intervention’s effectiveness. Holding all of these other factors constant, as sample size increases, the ability to detect smaller program effects increases. As such, care must be given in determining the number of participants needed for a study, as an insufficient sample size may cause researchers to be unable to demonstrate a statistically significant intervention impact. Additionally, if a small sample size is used, it is important to interpret results based on the contextualized effect size, rather than on pure statistical significance alone, in determining program impacts (Maxwell & Kelley, 2011).

**Attrition and Missing Data Methods.** Attrition can be defined generally as data that are not available at one or more time points for individuals who were originally included in a study (Shadish et al., 2002). Attrition within an RCT is particularly problematic, as the only point at which treatment and control groups are considered to be equal is when the randomization procedure takes place. Anything that occurs after this point, including participants choosing to leave the study before it officially ends, can threaten the strong internal control of the RCT design that is relied on to make causal conclusions regarding intervention impacts (Shadish et al., 1998).

Attrition can bias intervention results in two ways: (1) the general loss of participants reduces the study’s power to detect significant intervention effects, as is described above in the sample size section above, and (2) if the individuals who leave
the study prematurely are systematically different in some way from those who stay, then the conclusions made about the intervention’s effectiveness can be biased to no longer generalize to the entire sample population (Bickman & Reich, 2015). This latter situation may occur most often in intervention studies if the demands of the treatment itself seem unmanageable to a certain subset of individuals. For example, perhaps teachers with the highest levels of stress at the start of a mindfulness intervention find the demands of the program to actually exacerbate their stress rather than reduce it. This appraisal of program demands may increase the likelihood that participants who are the most at-risk to leave the study, do so prematurely. In this scenario, at the end of the study, the test of intervention effectiveness will be positively biased, such that a conclusion could be made that the intervention was more effective than it actually was, had the entire sample’s experiences been represented in the analyses (Enders, 2001).

Fortunately, several methods for dealing with missing data have been developed to assist researchers in recapturing statistical power lost to attrition and mitigating potential bias of intervention effects. Both multiple imputation (MI) and full information maximum likelihood (FIML) represent recommended methods for managing issues of attrition, however, both require certain assumptions about the missing data to be met to be employed appropriately (Graham, 2009). These assumptions about data, otherwise known as missing data mechanisms (Little, 1988; Little & Rubin, 2002), require researchers to demonstrate that the missing data itself is not systematically related to, or caused by, the intervention itself. If this assumption is violated, these
methods can potentially perpetuate the bias associated with the missing data, rather than helping to mitigate it (Graham, 2009).

MI consists of three stages: (1) complete dataset generation, (2) analysis, and (3) pooling. Researchers first generate a predetermined number of complete datasets, typically ranging from five to 20, in which missing values are replaced by plausible values determined by a regression-based estimation method. All of the complete datasets that are generated are then analyzed using familiar complete-case analysis methods, such as listwise deletion. Last, results from each dataset’s analysis phase are pooled together to more accurately represent plausible estimates of an intervention impact, had all participants originally randomized been present in the dataset (Enders, 2010; 2016).

In contrast, FIML is a model-based method of estimation that derives plausible estimates of an intervention impact through an iterative process of comparing the fit of model-based estimates to the sample data on hand. Estimates that provide the best fit are chosen to represent the entire sample of data, had all participants’ data been included in the estimation process (Enders, 2010). Unlike MI, FIML requires the use of sophisticated software and an additional set of underlying assumptions to be met to be used appropriately. Specifically, data need to be normally distributed and the sample size needs to be sufficiently large for this method to derive unbiased estimates (Enders, 2010).

**Analysis Specification.** The choices researchers make in their analysis specification also have the potential for biasing the conclusions they make about
whether or not an intervention was effective (Bickman & Reich, 2015). For instance, all researchers conducting an RCT study are recommended to collect baseline data of participants’ demographic information and baseline levels of the intervention’s outcomes of interest (What Works Clearinghouse, 2017) in order to test for baseline equivalence. Technically speaking, the randomization process should equate groups on all measured and unmeasured variables, however, differences between groups can still occur by chance (Skinner et al., 2019). If these differences are not anticipated and controlled for through design (e.g., by blocking) or otherwise detected and included in the analysis specification of models to adjust for them statistically, results derived from intervention impact analyses may be biased. What Works Clearinghouse (2017) provides general guidelines for dealing with baseline inequivalence, recommending that any variables found to have treatment/control effect size differences between .05 and .25 at baseline should be included as covariates in all subsequent models testing intervention impacts (What Works Clearinghouse, 2017). In contrast, effect sizes less than .05 are small enough not to bias impact estimates, and effects larger than .25 invalidate the baseline equivalence of a study in a manner that reduces the strength of the causal claims that can be made (What Works Clearinghouse, 2017).

If baseline equivalence can be established outright, with all effects at baseline falling under the .05 effect size benchmark, researchers can technically utilize any accepted statistical test that examines differences between groups on outcomes to test for program impacts. This could mean using a test as simple as a t-test, which examines
differences between group means at a single time of measurement (typically post-test) to examine intervention impacts.

However, if researchers have notable differences between groups at baseline, they are recommended to use one of several acceptable methods that allow for statistical adjustment through the inclusion of covariates. Acceptable analysis methods outlined by What Works Clearinghouse (2017) include ordinary least squares regression models adjusted with covariates, hierarchical linear regression models adjusted for covariates, analysis of covariance, and other comparative nonlinear regression approaches. Additionally, other methods are acceptable when pre-test and post-test measures are the same, and it can be demonstrated that they are highly correlated (e.g., simple gain scores).

**Review of Research on Factors that Influence Conclusions of Intervention Effectiveness.** A review of teacher studies regarding the following factors that influence conclusions made about intervention effectiveness can be found in Table 2.7. Each study was examined based on (1) sample size, (2) attrition rates, (3) missing data methods (when applicable), and (4) analysis methods used to examine intervention impacts.

**Sample Size.** Study sample sizes range from 38 to 224. This review purposefully included well designed RCT pilot studies, however, the small sample sizes of several of the studies is reflected here. Additionally, most RCT studies conducted to date have sample sizes of less than 100 participants (9/12 or 75%) and only one study reviewed here had a sample size greater than 200.
Attrition. Three studies reviewed did not report study attrition rates. Of the studies that did report attrition, rates ranged from 0% to 39%. Six of those studies had attrition rates greater than 5%.

Missing Data Methods. Only two studies utilized recommended methods (multiple imputation or full information maximum likelihood) to account for missing data rates higher than 5%. Listwise-deletion (a method of excluding incomplete data from analyses conducted) was the most common method used, in eight of 12 studies reviewed.

Analysis Methods. Ten of 12 studies reviewed explicitly report on baseline equivalence. Even though two studies did not explicitly report baseline equivalence, all studies utilized methods that included covariates, most frequently in the form of the pre-test measure of the outcome and relevant demographic factors. Two pilot studies with small sample sizes adjusted the alpha level from .05 to .10 to assess impacts in the form of “noteworthy trends” that may not have reached statistical significance due to inadequate statistical power caused by the small sample size. All studies reported results in the form of standardized effect sizes.

Summary. The examination of these 12 RCT studies for teachers on factors that can influence conclusions made about intervention effectiveness after an intervention is over elucidates several areas within the field that still need to be addressed. First, generally speaking, the sample sizes examining mindfulness intervention effectiveness for teachers have been small. Inclusion criteria for this review was purposefully inclusive
of small pilot studies with rigorous designs, however, larger studies of efficacy are needed to supplement the general trend of promising results from studies with smaller samples. Maxwell and Kelley (2011) indicate that a sample size of at least 128 is needed for a medium effect size to reach statistical significance when testing differences between groups at the established alpha level of .05. While some studies examined get close to this benchmark, only one study included here had a sample that was greater than 200 participants. Additional studies of this nature are needed to more fully elucidate potential small, but meaningful effects of these interventions on teacher populations.

Second, while several studies had levels of attrition low enough to justify the use of complete-case analysis methods, such as listwise-deletion, most studies had rates of attrition that should have been accounted for by using a recommended missing data method to reduce potential bias on intervention outcomes. For instance, only Kemeny and colleagues (2012) and Jennings and colleagues (2017) utilized maximum likelihood methods to account for attrition of participants in their studies. Additionally, as is described above, maximum likelihood methods have a series of additional methodological assumptions associated with their use that may not be appropriate in small studies, as the consequences of biased statistical tests may increase the likelihood that a researcher incorrectly concludes whether an intervention has been effective (Enders, 2016). It will be important in future work to compare these methods in the
context of intervention research, to better understand which method or methods may be best in producing unbiased estimates of intervention effectiveness.
<table>
<thead>
<tr>
<th>Research Study</th>
<th>Sample Size</th>
<th>Attrition</th>
<th>Missing Data Method</th>
<th>Analysis Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson et al., 1999</td>
<td>91</td>
<td>Not reported</td>
<td>Listwise-deletion</td>
<td>MANCOVA/ANCOVA</td>
</tr>
<tr>
<td>Benn et al., 2012</td>
<td>32 parents,</td>
<td>27 (39%)</td>
<td>Listwise-deletion, attrition</td>
<td>Baseline equivalence, ANCOVA, effect size (d), mediation using bootstrapping</td>
</tr>
<tr>
<td></td>
<td>38 educators</td>
<td></td>
<td>analyses comparing</td>
<td>to estimate indirect effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>completers to non-completers</td>
<td></td>
</tr>
<tr>
<td>Crain et al., 2016</td>
<td>58 Canada,</td>
<td>Not reported</td>
<td>Listwise-deletion</td>
<td>Baseline equivalence, ANCOVA, effect size (d), mediation using Sobel's test and</td>
</tr>
<tr>
<td></td>
<td>55 US</td>
<td></td>
<td></td>
<td>bootstrapping</td>
</tr>
<tr>
<td>Franco et al., 2010</td>
<td>68</td>
<td>None</td>
<td>N/A</td>
<td>Baseline equivalence, between group t-tests, paired samples t-tests, effect size</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(d), percent change pre-post and pre-follow</td>
</tr>
<tr>
<td>Garner et al., 2018</td>
<td>87</td>
<td>None</td>
<td>N/A</td>
<td>Baseline equivalence, RM ANCOVA, effect size (partial eta)</td>
</tr>
<tr>
<td>Research Study</td>
<td>Sample Size</td>
<td>Attrition</td>
<td>Missing Data Method</td>
<td>Analysis Methods</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------</td>
<td>-----------</td>
<td>-------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Harris et al., 2015</td>
<td>42 teachers, 22 other staff</td>
<td>1 (2%)</td>
<td>Listwise-deletion</td>
<td>Baseline equivalence, ANCOVA with p &lt; .10 as benchmark for 'noteworthy trends', effect size (d) and confidence interval</td>
</tr>
<tr>
<td>Jennings et al., 2011</td>
<td>32 pre-service, 11 in-service</td>
<td>4 (9%)</td>
<td>Listwise-deletion</td>
<td>Residualized pre-to-post change t-tests with p &lt; .10 as benchmark for 'noteworthy trends', effect size (d)</td>
</tr>
<tr>
<td>Jennings et al., 2013</td>
<td>53</td>
<td>3 (6%)</td>
<td>Listwise-deletion</td>
<td>Baseline equivalence, ANCOVAs, effect size (d)</td>
</tr>
<tr>
<td>Jennings et al., 2017</td>
<td>224</td>
<td>15 (7%)</td>
<td>FIML, attrition analyses comparing completers to non-completers</td>
<td>Baseline equivalence, HLM, EFA/CFA for measures, effect sizes, and improvement indices</td>
</tr>
<tr>
<td>Kemeny et al., 2012</td>
<td>82</td>
<td>6 (7%)</td>
<td>Restricted ML</td>
<td>Baseline equivalence, RM ANCOVA, effect sizes</td>
</tr>
<tr>
<td>Research Study</td>
<td>Sample Size</td>
<td>Attrition</td>
<td>Missing Data Method</td>
<td>Analysis Methods</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------</td>
<td>-----------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------</td>
</tr>
<tr>
<td>Roeser et al., 2013</td>
<td>113 (58 Canada, 55 US)</td>
<td>6 (5%)</td>
<td>Listwise-deletion</td>
<td>Baseline equivalence, ANCOVA, effect sizes, mediation using Sobel's test and bootstrapping</td>
</tr>
<tr>
<td>Taylor et al., 2016</td>
<td>59</td>
<td>Not reported</td>
<td>Listwise-deletion</td>
<td>Baseline equivalence, ANCOVAs, effect sizes</td>
</tr>
</tbody>
</table>
Discussion

Examining the current evidence base of teacher MBIs using theoretical frameworks put forth by experts in the field of mindfulness in education contexts and current methodological recommendations for RCT studies illuminates several extensions in both categories in need of future work. In terms of theoretical extensions, three areas should be prioritized. First, more work is needed to examine more distal program outcomes in the classroom and for students. While there is a strong empirical evidence base indicating MBIs for teachers are effective in promoting teachers’ mindfulness-related skills and occupational health and wellbeing, few studies have expanded beyond teacher outcomes to examine potential impacts on classrooms and students.

Second, there is a need to empirically test whether teachers engage in mindfulness practices at home during the intervention and whether they continue to use mindfulness practices at home after the intervention ends, since these actions represent the “causally efficacious” processes (Shadish et al., 2002) hypothesized to result from MBIs. And third with the exception of empirical evidence to support the link between mindfulness-related skills and occupational health and wellbeing (Benn et al., 2012; Roeser et al., 2013), more work is needed in testing the mediating pathways or “causal explanations” (Shadish et al., 2002) that link each of these steps together along the theoretical framework, to further understand how MBIs tailored for teachers change various aspects of teachers, classrooms, and students.
In addition to these theoretical extensions, there are several methodological areas within the field that also still need to be addressed to better align with current recommendations for RCT studies. First, while these studies of MBIs for teachers have done an exceptional job meeting basic RCT design standards (e.g., randomization procedures, waitlist control designs, follow-up times of measurement), more work is needed that utilizes larger sample sizes and longer follow-up times of measurement to truly measuring lasting impacts. Second, more studies need to utilize recommended methods for dealing with missing data, rather than listwise deletion, which is not recommended. Lastly, it will be important to expand the types of measures used to assess the fidelity of implementation of MBIs for teachers, to begin to differentiate among these various mindfulness interventions and the unique or common active ingredients they may share. Additionally, monitoring mindfulness practices in both intervention and control groups will also be extremely important, especially as secular mindfulness practices continue to grow in popularity and availability.

In conclusion, the current evidence base examining MBIs for teachers is very promising. Previous research has demonstrated these programs can positively impact teachers in a variety of domains, and may have the potential to extend impacts to classrooms as well (Klingbeil & Renshaw, 2018). Despite this promising start, however, this review highlights several theoretical and methodological extensions that will improve our understanding of the impacts of MBIs for teachers. By examining both theory and methods, this review hopes to inform the direction of future research in
ways that allow for the production of the highest quality research and the most certainty regarding if and how MBIs impart benefits to teachers, their classrooms, and their students.
Chapter 3: Study 2

The Explanatory Role of Mindfulness Practice in the Development of Mindfulness Skills:

Results from Studies of a Mindfulness-Based Intervention with Teachers
Abstract

Objectives: Regular engagement in mindfulness practices is emphasized as key in the development of mindfulness skills and positive wellbeing. Despite this clear theoretical link, however, research examining this relation directly is mixed.

Methods: Using data from three randomized controlled trials testing the impact of an 8-week MBI program for public school teachers (N = 173), this study examined whether the MBI increased teachers’ frequency of use of mindfulness practices during the 8-week program and during the summer months after the MBI ended. This study also utilized structural equation modeling to examine if increases in teachers’ use of these practices explained their improvements in mindfulness and occupational self-compassion, both post-program and four months later.

Results: Both during the intervention (T1.5) and in the summer months after (T2.5), teachers in the MBI group engaged in mindfulness practices significantly more frequently compared to teachers in the waitlist control group ($\beta_{T1.5}=.58, p<.001; \beta_{T2.5}=.41, p<.001$). Mediation analyses indicated teachers’ increases in mindfulness practices explained their long-term improvements in mindfulness ($IDE_{T3}=.13, p<.05$) and short- and long-term improvements in occupational self-compassion ($IDE_{T2}=.17, p<.01; IDE_{T3}=.12, p<.01$).

Conclusions: This study provides additional evidence suggesting frequency of mindfulness practice is causally linked to the development of mindfulness-related skills.
in teachers. Implications related to the timing of these effects and the greater importance of these relations to the fields of mindfulness and education are discussed.

*Keywords*: mindfulness-based intervention, mindfulness practice, mindfulness, self-compassion, teachers
The Explanatory Role of Mindfulness Practice in the Development of Mindfulness Skills: Results from Studies of a Mindfulness-Based Intervention with Teachers

Both Buddhist contemplative traditions and more recent secularized mindfulness-based program adaptations posit that regular engagement in mindfulness practices is essential to the development of mindfulness skills and positive wellbeing (Carmody & Baer, 2008). Despite this clear theoretical assertion, however, research explicitly examining such a connection is mixed. Numerous reasons have been posited to account for these conflicting results, including different operationalizations of mindfulness practices. A general paucity of research directly measuring mindfulness practices or mindfulness-related skills is also a current issue with the field (Davidson & Kaszniak, 2015; Khoury et al., 2015). Additionally, little is known about the time course necessary for mindfulness practices to manifest as changes in mindfulness skills (Davidson, 2010). Using data from three randomized controlled trial mindfulness-based interventions (MBIs) for public school teachers, this study seeks to further investigate mindfulness practice as a mechanism through which an MBI impacts participants’ development of mindfulness-related skills, as well as to expand our understanding of the use of these practices in promoting long-term benefits after the MBI is over.

Mindfulness: What it is and How it Works

Despite recent rapid growth of secular adaptations, the origins of mindfulness and mindfulness practices reside in Buddhist contemplative traditions that are centuries old. Definitions of mindfulness commonly used in scientific literature tend to share
several key features, most predominantly emphasizing non-judgmental acceptance and a present moment focus (Bishop et al., 2004; Kabat-Zinn, 1990). More broadly, mindfulness has also been characterized as an overarching construct that is used to define a series of mental capacities including certain kinds of attention, awareness, memory, and acceptance (van Dam et al., 2018). Additionally, some definitions rooted in Buddhist traditions further describe mindfulness as the development of an understanding of the changing nature of mental and physical states as a means to reduce human suffering (Dreyfus, 2011).

So defined, mindfulness has been characterized as a skill that can be cultivated by engaging in a variety of mindfulness practices, such as: directed attention exercises, which promote attention regulation by focusing attention on a specific target, like the breath or a mental image; body scanning, which enhances an individual’s self-awareness and attention regulation through the process of moving one’s focus piece by piece through the body while practicing non-reactivity towards all physical sensations, thoughts, and feelings that arise; and open-monitoring practices, which emphasize focusing attention broadly to remain aware and attend to any experiences that arise in the body in the present moment (Davidson & Kaszniak, 2015). Although there are apparent differences in the focus of each type of practice, in general, mindfulness practices tend to share the same practice approach: one that emphasizes awareness, curiosity, acceptance, and non-judgement to the present moment experience (Carmody & Baer, 2008; Davidson & Kaszniak, 2015).
Another fundamental element in many mindfulness practices is the cultivation of loving kindness and compassion. This more affectively based mindfulness skill can be described as a desire for the self and others to find happiness and a motivation to alleviate human suffering (Lutz et al., 2008). Similar to mindfulness, compassion is thought to be cultivated through mindfulness practices intended to break down self-centered tendencies as a means of developing kindness and empathy for the self and others. Examples of compassion practices include those that focus on a person’s positive emotional connections with others, and those that involve visualizing the suffering of others as a method of evoking feelings of generosity, empathy, and commonality. In general, compassion-based practices share the goal of reducing an individual’s focus on the self in order to promote happiness and greater altruism of thought and deed in everyday life (Lutz et al., 2008; Neff & Knox, 2017).

This translation of mindfulness practices into enhanced mindfulness-related skills, such as mindfulness and self-compassion, is thought to occur through a process of neuroplasticity, in which regular practice strengthens a series of associated key mental faculties (Hölzel et al., 2011; Tang et al., 2015). Similar to strengthening the body during physical exercise, directing attention intentionally and non-judgmentally has been shown to enhance an individual’s awareness (Carmody & Baer, 2008; Farb et al., 2015; Farb et al., 2013); executive functioning (i.e., attention regulation, working memory, cognitive flexibility) and emotion regulation (Goldin & Gross, 2010; Jha et al., 2007; Tang
et al., 2007; Teper et al., 2013); and kindness and compassion (Lutz et al., 2008; Neff & Knox, 2017).

The development of these mindfulness-related mental faculties is considered to translate into improved functioning in everyday life via several pathways: first, for example, by engaging in mindfulness practices that strengthen non-reactivity and attentional awareness individuals may become better able to take a step back before automatically reacting to a stressful situation (Davidson et al., 2012; Teper et al., 2013). Second, engaging in practices that promote enhanced somatic body awareness may also allow individuals to more effectively detect physiological indicators of stress (e.g., increased heart rate, rise in body temperature) earlier in a stress response; and this enhanced awareness, coupled with improved regulatory capacities, may allow individuals to reduce the overall magnitude and/or duration of their stress reaction (Goldin & Gross, 2010; Williams et al., 2009). Third, through enhanced clarity and decision-making flexibility during moments of stress, individuals may be more likely to choose adaptive emotion regulation/coping strategies, rather than maladaptive strategies (Gross, 1998; Skinner & Beers, 2016). As a last example, activating kindness and compassion during mindfulness practice may increase an individual’s ability to spontaneously evoke these feeling in everyday situations—when they would otherwise react with judgement or negativity—thus reducing subsequent experiences of anxiety, rumination, or stress that may result from negative interactions with others (Macbeth & Gumley, 2012; Neff & Knox, 2017).
Research on Mindfulness Practices

Despite the fact that mindfulness practices are hypothesized to play a central role in the development of mindfulness-related skills, research testing this link directly is mixed. Research that has explicitly studied this connection has predominantly done so in the context of MBIs (but see Soler et al., 2014), which frequently take the form of an 8-week program where participants learn a series of mindfulness exercises that they practice during in-person program sessions and at home between sessions (e.g., Kabat-Zinn, 1990). In a recent meta-analysis of the effects of mindfulness practice during standard 8-week MBI programs, Parsons and colleagues (2017) found a significant small to medium effect of practice on intervention outcomes across twenty-eight studies. However, a number of studies have reported no such relation between amount of practice and outcomes (e.g., Astin, 1997; Davidson et al., 2003; Manuel et al., 2017).

Several other issues are currently prevalent the field as well, including a general lack of measurement of mindfulness practices and mindfulness, and differential operationalizations when they are measured. Specifically, the majority of MBIs fail to report on participants’ frequency of practice (over two hundred MBIs were excluded in Parsons et al.,’s 2017 review for this reason). Many studies have also chosen not to measure mindfulness as an intervention outcome directly (Khoury et al., 2015), rather skipping over this important theoretical link and moving directly to more distal program targets, such as wellbeing. How researchers choose to measure participants’ engagement in mindfulness practices may also contribute to the inconsistent findings
we’ve found so far. Specifically, there are differences in operationalizations of amount of mindfulness practice, with most research on MBIs measuring practice in terms of total minutes spent practicing (e.g., Carmody & Baer, 2008), whereas others examining these relations outside the MBI framework have found significant associations using measures of practice frequency (e.g., days per week; Soler et al., 2014). In a more general sense, there is also a need to further explore the time course over which engagement in mindfulness practices typically manifests as changes in mindfulness-related skills (Davidson & Kaszniak, 2015), as incorrect timing may also explain why the current evidence base is mixed.

**Study Purpose**

Therefore, the purpose of this study was to further clarify the role that mindfulness practices have in the development of mindfulness-related skills, using data from three MBIs for public school teachers. MBIs for teachers have continued to grow in popularity as unique professional development programs useful in buffering teachers against stress and burnout caused by their high work-related demands (Jennings & Greenberg, 2009; Roeser et al., 2012). Similar to MBIs generally, research examining the role of practice in teacher samples is still nascent. In a recent meta-analysis of teacher MBIs, Klingbeil and Renshaw (2018) found that of 16 studies only six reported the amount of time teachers spent engaging in mindfulness practices and only one examined the connection between practice time and mindfulness. Specifically, Kemeny and colleagues (2012) demonstrated that the total number of minutes of mindfulness
practice that teachers engaged in during the MBI significantly predicted increases in mindfulness, reductions in anxiety, and reductions in blood pressure measured during a Trier Social Stress Test.

Furthermore, to date no published studies have utilized randomized controlled MBI methodology to test the effect of practice on the development of mindfulness-related skills. Studying these relations by comparing randomly assigned treatment and control groups, rather than within a treatment group only, will strengthen our ability to make causal inferences about the impact of an MBI on participants’ engagement in mindfulness practices and whether mindfulness practices do in fact represent a causal mechanism through which an MBI influences participants’ development of mindfulness-related skills. Additionally, it is important to examine whether teachers continue to engage in mindfulness practices after the MBI is over, in order to assess the potential role mindfulness practices play in providing teachers with lasting intervention benefits, as evidence of long-term effects is promising in non-teacher samples (Morgan et al., 2014; Solhaug et al., 2019).

Hence, guided by the theoretical model depicted in Figure 3.1, the purposes of this study were twofold: (1) to test the impact of a mindfulness-based intervention on teachers’ frequency of mindfulness practices during and after the MBI, and (2) to examine whether teachers’ frequency of mindfulness practices explain their increases in mindfulness and occupational self-compassion skills, both at post-program and four months after the intervention concluded. Specifically, the following research questions
were tested: (1) Does the MBI impact teachers’ frequency of mindfulness practices during the intervention phase? (2) Does the MBI impact teachers’ frequency of mindfulness practices during the summer months after the intervention is over? (3) Is the impact of the MBI on teachers’ increases in post-program mindfulness and occupational self-compassion skills explained by their frequency of mindfulness practices during the intervention phase? (4) Is the impact of the MBI on teachers’ increases in follow-up mindfulness and occupational self-compassion skills explained by their frequency of mindfulness practices during the summer, after the intervention is over?
Figure 3.1
Mindfulness-Based Intervention (MBI) Theory of Change

Note. Teachers develop mindfulness and occupational self-compassion skills through regular and frequent engagement in mindfulness practices. Short term program impacts on mindfulness and occupational self-compassion occur through the learning and engagement in practices during the intervention phase, whereas long-term program impacts on mindfulness and self-compassion occur through continued practice after the intervention concludes.

Methods

To address these questions, data from three randomized waitlist-controlled trials conducted in the western United States and Canada were combined. Each trial tested the same MBI for teachers and used the same research design and measures. The Canada study took place in 2009, the first U.S. study took place in 2010 (see combined primary intervention results of these first two studies in Roeser et al., 2013), and the second U.S. study took place in two cohorts from 2014 to 2016 (see Roeser et al., in preparation for primary intervention results of this third study). In all three studies,
elementary and/or middle school teachers were recruited using flyers posted in schools and through emails to district leaders. The program was advertised as a voluntary mindfulness-based stress reduction program for teachers. Incentives for teachers included free access to the 8-week mindfulness training and financial compensation for time spent completing study assessments (see impacts papers cited above for additional information about data collection procedures).

Prior to random assignment to study condition, participating teachers provided baseline self-report surveys in the fall of the research school year. Following randomization, MBI group teachers participated in the 8-week mindfulness-based stress reduction (MBSR) program during the winter months, and post-program outcomes were assessed shortly after. Teachers then completed follow-up assessments during the fall of the following school year, 3 to 4 months after the intervention ended. In each trial, after follow-up data collection was complete, waitlist control teachers were offered access to the mindfulness training during the winter months.

**Samples**

Demographic information by intervention condition for each study sample can be found in Table 3.1 (for details, see the respective impacts papers). The combined sample included a total of 173 teachers (MT = 85, WLC = 88). Participants were predominantly female (73%), taught middle school (66%) and had master’s degrees (68%). Of the Canadian teachers, 67% identified as European Canadian and of the U.S. teachers, 88% identified as European American. The age of participants ranged from 25-
64 with an average of 45.13 years (SD = 9.44). Additionally, teacher experience for the combined sample ranged from 1 to 37 years with an average of 13.30 (SD = 8.58) years.
Table 3.1
*Descriptive Statistics by Project Sample and Intervention Condition*

<table>
<thead>
<tr>
<th>Demographic Characteristic</th>
<th>Canada Sample</th>
<th>2010 US Sample</th>
<th>2014-16 US Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MT</td>
<td>WLC</td>
<td>MT</td>
</tr>
<tr>
<td>Number of Teachers (N = 173)</td>
<td>26</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>Sex (% female)</td>
<td>92%</td>
<td>88%</td>
<td>82%</td>
</tr>
<tr>
<td>School level (% middle school)</td>
<td>54%</td>
<td>46%</td>
<td>51%</td>
</tr>
<tr>
<td>Age in years, median (range)</td>
<td>50 (28-59)</td>
<td>46 (29-63)</td>
<td>52 (35-63)</td>
</tr>
<tr>
<td>Years of teaching experience, median (range)</td>
<td>12 (3-35)</td>
<td>10 (4-32)</td>
<td>16 (2-33)</td>
</tr>
</tbody>
</table>
**Intervention**

Teachers who were randomly assigned to the MBI group in each study were offered access to the Mindfulness-Based Emotional Balance (MBEB) program - a fully manualized mindfulness intervention designed for teachers (previously referred to as SMART; see Cullen & Pons, 2015). The 8-week, 36-hour MBEB program comprises approximately 50% MBSR content (Kabat-Zinn, 1990), 30% emotion regulation and theory, and the remaining 20% of the content focuses on compassion and forgiveness. A more detailed description of program contents can be found in Table 3.2. To ensure consistent fidelity of implementation across all projects, the same mindfulness instructor was used in each study. Previous published work with these data has demonstrated the intervention conducted with all three samples was found to be feasible and well accepted by teachers (Roeser et al., 2013; in preparation). Information regarding the feasibility of the intervention and specific effects of the intervention on outcomes per project can be found in the main intervention impacts reports.
Table 3.2
Curricular Components of the Mindfulness Training for Teachers

<table>
<thead>
<tr>
<th>Mindfulness-Based Stress Reduction (50%)</th>
<th>Mindfulness-Based Emotion Skills (30%)</th>
<th>Mindfulness-Based Compassion &amp; Forgiveness (20%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Body scanning to promote somatic awareness</td>
<td>1. Evidence-based information on emotions and expressions</td>
<td>1. Caring for oneself with loving-kindness practices</td>
</tr>
<tr>
<td>2. Breath awareness practices</td>
<td>2. Emotion in connection to reaction/responding</td>
<td>2. Caring for others with loving-kindness practices</td>
</tr>
<tr>
<td>3. Mindful thinking and emotion</td>
<td>3. Interconnectedness of positive/negative emotions and physiology, cognition, and behavior</td>
<td>3. Caring for challenging students with loving-kindness practices</td>
</tr>
<tr>
<td>5. Role play to practice mindfulness in classroom during anger or fear</td>
<td>5. Information on individual differences in emotions (e.g., expression, profiles, triggers)</td>
<td>5. Practices in mindful listening</td>
</tr>
<tr>
<td>6. Role play to practice mindfulness in challenging social interactions</td>
<td>6. How to use mindfulness to detect and reflect on strong emotions; role of refractory period</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7. The development of mindful coping strategies (e.g., reappraisal)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Table adapted from Taylor et al., 2016

Measures

Mindfulness Practice Frequency. To assess teachers’ frequency of mindfulness practices during and after the intervention, teachers were asked whether they had what they considered a “meditation practice” and if so, were asked “how often do you meditate now?” at post-program and follow-up. These questions were recoded into a single variable rated on an 9-point Likert style scale (0 = Never, 1 = Less than once a year, 2 = About once or twice a year, 3 = About once a month, 4 = Nearly every week, 5 = Several times a week, 6 = Nearly every day, 7 = Every day, 8 = More than once per day).

Post-program reports of practice were used to represent teachers’ recollection of their
frequency of mindfulness practices during the intervention phase (T1.5) and follow-up
reports of practice were used to represent teachers’ frequency of practice during the
summer months after the intervention was over (T2.5).

**Mindfulness.** Mindfulness was assessed using the short form of the Five Facet
Mindfulness Questionnaire (FFMQ; Baer et al., 2008). This 39-item scale comprises five
sub scales: *observing sensations, thoughts and feelings* (e.g., ‘I notice visual elements in
art or nature, such as colors, shapes, textures, or patterns of light and shadow’); *acting
with awareness* (e.g., reverse-coded ‘I rush through activities without being really
attentive to them’); *non-judgement* (e.g., reverse-coded ‘I tell myself that I shouldn’t be
feeling the way I’m feeling’); *describing* (e.g., ‘Even when I’m feeling terribly upset, I can
find a way to put it into words’); and *non-reactivity* (e.g., ‘Usually when I have
distressing thoughts or images, I just notice them and let them go’). Items were rated on
a 5-point Likert scale (1 = *Almost never* to 5 = *Almost always*, $\alpha_{T2} = .92$, $\alpha_{T3} = .93$). Due to
a copying error in the Canada sample, the non-reactivity sub scale was not collected.

**Occupational Self-Compassion.** Neff’s (2003) self-compassion scale was
modified to pertain specifically to the occupation of teaching (Roeser et al., 2013). Nine
items were shared among all three studies that capture three subcomponents from this
scale - self-kindness (e.g., “I try to be understanding and patient towards myself when
those aspects of my personality that I don’t like come out in the classroom”), common
humanity (e.g., “When I feel inadequate in my role as a teacher in some way, I try to
remind myself that most teachers experience feelings of inadequacy”), and self-
judgement (e.g., reverse-coded “When times are really difficult at work, I tend to be tough on myself”). All items were rated on a 5-point Likert scale (1 = Not at all true to 5 = Very true; $\alpha_{T2} = .89$, $\alpha_{T3} = .90$). Items from the original self-compassion scale measuring mindfulness were intentionally excluded from our measure of occupational self-compassion to avoid construct overlap. Correlations between our omnibus mindfulness and self-compassion scales demonstrated modest correlations at each time point, suggesting they are related but distinct measures of mindfulness-related skills ($r_{T1} = .56$, $r_{T2} = .61$, $r_{T3} = .57$).

**Baseline Equivalence of Study Samples and Intervention Groups**

Baseline equivalence ANOVA tests were conducted to examine whether there were significant differences at baseline between treatment and control groups, or between project samples across the three study sites. Table 3.3 displays means and standard deviations for teachers’ demographic characteristics and baseline levels for each outcome of interest by intervention group, as well as F-values and significance tests of the main effects of intervention condition and project site, and their interaction. Results showed no significant differences between treatment and control groups on teachers’ baseline outcomes or demographic characteristics that would warrant adding additional covariates to the statistical models proposed. Several main effects of project site were found for teachers’ sex, age, and teaching experience. However, no condition by site interactions were found to be statistically significant, and this lack of significance
was used as an empirical justification to combine all three projects into one sample to run all subsequent analyses of interest.
Table 3.3
Baseline Comparisons of Demographics and Outcomes by Study Condition and Research Site

<table>
<thead>
<tr>
<th>Baseline Measure</th>
<th>Combined MBEB groups (N = 85)</th>
<th>Combined WLC groups (N = 88)</th>
<th>F-value: Study Condition</th>
<th>F-value: Research Site</th>
<th>F-value: Condition x Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>.81 (.40)</td>
<td>.83 (.38)</td>
<td>.10</td>
<td>5.33**</td>
<td>.62</td>
</tr>
<tr>
<td>Age</td>
<td>45.51 (8.96)</td>
<td>44.79 (9.89)</td>
<td>.14</td>
<td>9.82***</td>
<td>.74</td>
</tr>
<tr>
<td>Teaching Experience</td>
<td>13.50 (8.52)</td>
<td>13.11 (8.67)</td>
<td>.10</td>
<td>6.88**</td>
<td>3.00</td>
</tr>
<tr>
<td>Previous Mindfulness Experience</td>
<td>.39 (.49)</td>
<td>.28 (.45)</td>
<td>2.18</td>
<td>.09</td>
<td>.94</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>3.41 (.51)</td>
<td>3.28 (.56)</td>
<td>2.15</td>
<td>.13</td>
<td>.70</td>
</tr>
<tr>
<td>Occupational Self-Compassion</td>
<td>2.89 (.73)</td>
<td>2.76 (.69)</td>
<td>1.20</td>
<td>.68</td>
<td>1.04</td>
</tr>
</tbody>
</table>

*Note.* No condition by site interactions were significant at p < .05.

** p < .01; *** p < .001.
Analyses

Ordinary least squares regressions were used to test the impact of the intervention on teachers’ frequency of mindfulness practices reported at post-program (RQ1) and follow-up (RQ2). Structural Equation Modeling (SEM) was used to examine the mediation hypotheses, testing whether mindfulness practice represents a mechanism through which mindfulness training improves teachers’ mindfulness and occupational self-compassion skills at post-program and follow-up (RQ3 & RQ4). All analyses included project site as a covariate, and were conducted in the Lavaan package (Oberski, 2014) in R (R Core Team, 2017). As a combined sample, on average, missing data rates were low to modest across all three studies. Missing data rates ranged from 6.43-8.84% at post-program and 15.20-21.24% at follow-up across outcomes of interest. Little’s MCAR test determined that the missing data can be assumed to be random, and as such, full information maximum likelihood was utilized to account for missing data and increase power to detect significant effects.

SEM is a valuable statistical method due to its ability to model theoretical constructs of interest as latent factors. Latent factors are created through the process of specifying the measurement component of the SEM model, in which scale items are seen as manifestations of their underlying latent factor, comprised of both true factor variance and measurement error. Specifying the measurement component of the model partitions these two aspects of the observed variables and isolates the true factor variance to be used in the structural component, or hypothesis driven statistical causal
path analysis aspect of the SEM model (Bollen, 2002). Isolating the true factor variance increases statistical precision and power to detect significant path effects. Additionally, SEM models are particularly useful in tests of causal mediation, as the estimation of all regression paths simultaneously lowers the type I error rate of erroneously rejecting the null hypothesis when in fact it’s true – which occurs more often when many statistical tests are conducted to confirm a single hypothesis (Hayes, 2013).

**Preliminary Confirmatory Factor Analysis.** Prior to testing the mediation hypotheses, confirmatory factor analysis (CFA) was used to assess the acceptability of the measurement model for mindfulness and occupational self-compassion outcomes as latent factors at post-program and follow-up. Final CFA model selection was determined based on overall model fit indices (CFI > .90, RMSEA <= .06, SRMR <= .08; Hu & Bentler, 1999). Both mindfulness and occupational self-compassion outcomes were modeled as second-order latent factors, with the subscales described in the measurement section defining the first-order latent factors. Measurement models reached satisfactory overall model fit with the addition of 1 to 4 correlated indicator residuals. Correlations among indicators were chosen based on modification indices and theoretical inspection for item similarity. Correlated indicator residuals are not ideal, as they represent systematic similarity between two items that is not shared by other items and therefore not captured in the overarching latent factor. However, allowing several errors to correlate was chosen as the best statistical choice available, given that these studies required considerable time and resources to conduct and are unlikely to be replicated to improve
measurement of these variables in these samples (Landis, Edwards, & Cortina, 2009). In all of the final models, factor loadings were statistically significant and ranged from .62 to .90 for the mindfulness latent factor and .58 to .95 for the occupational self-compassion latent factor.

**Mediation Analyses**

Once CFA specification was confirmed, mediation analyses were conducted using intervention condition (coded as Treatment = 1, Waitlist Control = 0) as the independent variable, frequency of mindfulness practice as the mediator, latent mindfulness and occupational self-compassion as the outcomes, and project site as a covariate. Research question three tested the impact of the intervention (T1) on teachers’ frequency of mindfulness practices during the intervention phase (T1.5) on post-program mindfulness and occupational self-compassion (T2). Research question four tested the impact of the intervention (T1) on teachers’ frequency of mindfulness practices during the summer after the intervention (T2.5) on follow-up mindfulness and occupational self-compassion, measured in the fall of the following school year (T3). Each structural equation model was estimated using robust maximum likelihood estimation (maximum likelihood estimation with Satorra-Bentler scaled chi-square and standard errors), to correct for kurtosis in the data (Satorra & Bentler, 1994).
Results

Preliminary Results

The impacts of this intervention on teachers’ mindfulness and occupational self-compassion within each of the study samples has been reported elsewhere; Table 3.4 provides means and standard deviations of teachers’ mindfulness, occupational self-compassion, and mindfulness practice frequency at post-program and follow-up. Using Cohen’s (1988) effect size benchmarks of .20, .50, and .80 to indicate small, medium, and large effects, effect sizes reported here reaffirm previous reports with these data demonstrating large effects on teachers’ post-program and follow-up mindfulness and occupational self-compassion in the two earlier studies (Roeser et al., 2013) and small to medium effects on these outcomes in the most recent U.S. sample (Roeser et al., in preparation).

Despite these more modest results on outcomes in the most recent sample, results show effects that are in the expected direction. Additionally, in general it is extremely important to investigate hypothesized mediating variables in intervention research, as it allows researchers to examine whether their hypothesized theory of change has empirical support and if not, helps diagnose the point in which hypotheses and evidence diverge. For instance, if an effect on the mediator is not found, it may indicate the intervention is ineffective at changing key targets of interest or that the mediator is not an essential causal component of the intervention. Or, if an intervention is found to significantly impact mediating variables but a direct program effect on
targeted outcomes is not found, it may indicate the hypothesized timeline for skill transfer is incorrect (MacKinnon et al., 2007). As such, these results, coupled with a priori hypotheses designating teachers’ engagement in mindfulness practices as key in the development of these target intervention outcomes, were used as empirical and theoretical justifications to move forward in testing the proposed mediation hypotheses.
Table 3.4  
*Effect Sizes by Project Sample and Intervention Condition*

<table>
<thead>
<tr>
<th></th>
<th>Canada Sample</th>
<th></th>
<th>2010 US Sample</th>
<th></th>
<th>2014-16 US Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MT M (SD)</td>
<td>WLC M (SD)</td>
<td>MT M (SD)</td>
<td>WLC M (SD)</td>
<td>MT M (SD)</td>
</tr>
<tr>
<td>Post-Program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindfulness Practice (T1.5)</td>
<td>4.88 (.26)</td>
<td>2.03 (1.81)</td>
<td>5.00 (1.36)</td>
<td>2.19 (1.71)</td>
<td>3.68 (2.44)</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>3.62 (.49)</td>
<td>3.25 (.58)</td>
<td>3.63 (.35)</td>
<td>3.24 (.55)</td>
<td>3.47 (.48)</td>
</tr>
<tr>
<td>Occupational Self-Compassion</td>
<td>3.49 (.56)</td>
<td>2.89 (.93)</td>
<td>3.25 (.58)</td>
<td>2.67 (.56)</td>
<td>3.10 (.61)</td>
</tr>
<tr>
<td>Follow-Up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mindfulness Practice (T2.5)</td>
<td>3.43 (2.09)</td>
<td>1.48 (2.41)</td>
<td>4.61 (1.37)</td>
<td>2.33 (1.59)</td>
<td>3.27 (2.21)</td>
</tr>
<tr>
<td>Mindfulness</td>
<td>3.69 (.53)</td>
<td>3.34 (.58)</td>
<td>3.81 (.58)</td>
<td>3.15 (.65)</td>
<td>3.56 (.49)</td>
</tr>
<tr>
<td>Occupational Self-Compassion</td>
<td>3.37 (.63)</td>
<td>3.05 (.92)</td>
<td>3.40 (.54)</td>
<td>2.83 (.39)</td>
<td>3.09 (.74)</td>
</tr>
</tbody>
</table>
**Intervention Impact on Teachers’ Frequency of Mindfulness Practices**

RQ1 and RQ2 examined the impact of the intervention on teachers’ retrospectively reported frequency of mindfulness practice during the intervention and over the summer months after the intervention was over. OLS regression results indicate a significant effect of the intervention on teachers’ frequency of practice at post-program and follow-up ($\beta_{T1.5} = .58, p < .001; \beta_{T2.5} = .41, p < .001$). Specifically, during the intervention phase teachers in the MBI group reported engaging in mindfulness practices nearly every week to several times a week on average ($M = 4.56, SD = 1.80$), whereas teachers in the control condition reported practicing less than once or twice a year on average ($M = 1.91, SD = 1.94$). Furthermore, comparison of teachers’ frequency of practice over the summer months after the intervention was over indicated treatment teachers were still engaging in mindfulness practices several times a month on average ($M = 3.78, SD = 1.98$), whereas control group teachers remained stable in practicing less than once or twice a year on average ($M = 1.90, SD = 2.26$). Additional follow-up analyses examining potential intervention condition by project site interactions were not significant, indicating that significant increases in teachers’ frequency of practice were found across all three project sites (see Table 3.4 for specific site means, standard deviations, and effect sizes).

**Mediation Results**

**Post-Program Impacts.** RQ3 examined whether teachers’ frequency of mindfulness practices during the intervention served as a mechanism through which
mindfulness training increased teachers’ mindfulness and occupational self-compassion skills post-program (See Figures 3.2 & 3.3.). Results testing mediation on post-program mindfulness skills indicated relatively good overall model fit ($\chi^2 (312) = 470.51, p < .001, CFI = .92, RMSEA = .06 CI[.05 - .07], SRMR = .08$). Path estimates indicated a statistically significant overall total effect of the intervention on teachers’ post-program mindfulness as a combined sample ($\beta = .37, p < .001$) and a replication of the impact of mindfulness training on teachers’ frequency of mindfulness practices ($\beta = .58, p < .001$). However, the effect of mindfulness practice on teachers’ post-program mindfulness was not statistically significant, suggesting teachers’ frequency of mindfulness practices during the intervention did not explain teachers’ increases in mindfulness immediately following the intervention.

SEM results testing mediation of mindful practices on teachers’ occupational self-compassion at post-program showed good overall model fit ($\chi^2 (47) = 54.08, p = .22, CFI = .99, RMSEA = .03 CI[.00 - .06], SRMR = .04$). The effect of the intervention on teachers’ mindfulness practices remained statistically significant, as in the previous model, and frequency of mindfulness practice during the intervention significantly predicted teachers’ occupational-self-compassion at post-program ($\beta = .30, p < .01$). The combined mindfulness practice mediation pathway was also significant (IDE = .17, $p < .01$). Additionally, the direct effect of the intervention on teachers’ post-program occupational self-compassion was no longer statistically significant after accounting for
the mediation pathway, indicating that mindfulness practice fully mediated the effect of the intervention on teachers’ post-program occupational self-compassion.

**Figure 3.2**
*Intervention Impact on Teachers’ Post-Program Mindfulness Mediated by Frequency of Mindfulness Practice During the Intervention.*

![Diagram of mediation pathway](image)

*Note. Structural Equation Modeling with MLR estimation (ML with robust standard errors) including project site as a covariate. χ² = 470.51 (df = 312, p < .001), CFI = .92, RMSEA = .06 [.05 -.07], SRMR = .08. IDE = the indirect effect, c = the total effect of the intervention on the outcome, and c' = the direct effect of the intervention on the outcome after accounting for the mediation pathway. ** p < .01; *** p < .001.*
Figure 3.3

*Intervention Impact on Teachers’ Post-Program Occupational Self-Compassion Mediated by Frequency of Mindfulness Practice During the Intervention*

![Diagram](image)

*Note.* Structural Equation Modeling with MLR estimation (ML with robust standard errors) including project site as a covariate. $\chi^2 = 54.08$ (df = 47, $p = .22$), CFI = .99, RMSEA = .03 [.00 - .06], SRMR = .04. IDE = the indirect effect, $c =$ the total effect of the intervention on the outcome, and $c' =$ the direct effect of the intervention on the outcome after accounting for the mediation pathway.

† $p < .07$; ** $p < .01$; *** $p < .001$.

**Follow-Up Impacts.** RQ4 similarly assessed whether teachers’ frequency of mindfulness practice during the summer months after the intervention was over explained treatment teachers’ sustained increases in mindfulness and occupational self-compassion over waitlist controls (see Figures 3.4 & 3.5). Results testing the effect of summertime frequency of mindfulness practice as a mediator of follow-up mindfulness indicated good overall model fit ($\chi^2(313) = 467.89$, $p < .001$, CFI = .92, RMSEA = .05 CI [.04 - .06], SRMR = .09). Path estimates indicated that the effect of the intervention on teachers’ summertime frequency of mindfulness practice remained statistically significant ($\beta = .40$, $p < .001$) and the effect of mindfulness practice on teachers’ mindfulness at follow-up was also statistically significant ($\beta = .33$, $p < .01$), as was the
total indirect effect of the mindfulness practice mediation pathway (IDE = .13, p < .05).

These results indicated that although mindfulness practice was not a significant mechanism through which mindfulness training increased mindfulness in teachers at post-program, it was a significant mediator of teacher mindfulness at follow-up. Additionally, statistical significance of the direct effect of the intervention on teacher mindfulness at follow-up after accounting for the mediation pathway, indicated that teachers’ frequency of mindfulness practice represents one pathway through which the intervention impacts teachers’ mindfulness at follow-up.

Lastly, the model assessing teachers’ frequency of mindfulness practice during the summer months on teachers’ occupational self-compassion at follow-up also had good overall fit ($\chi^2(46) = 73.41, p < .01, \text{CFI} = .96, \text{RMSEA} = .06 \text{ CI}[.03 - .08], \text{SRMR} = .05$). The effect of the intervention on teachers’ frequency of summertime mindfulness practice remained statistically significant. The effect of mindfulness practice on teachers’ occupational self-compassion was also significant ($\beta = .31, p < .01$), as was the total mindfulness practice mediation pathway (IDE = .12, p < .01). Additionally, the direct effect of the intervention on teachers’ follow-up occupational self-compassion was no longer significant after accounting for the mediation pathway, indicating mindfulness practice continued to fully mediate the effect of the intervention on teachers’ occupational self-compassion at follow-up.
Figure 3.4

*Intervention Impact on Teachers’ Follow-Up Mindfulness Mediated by Frequency of Mindfulness Practice After the Intervention*

Note. Structural Equation Modeling with MLR estimation (ML with robust standard errors) including project site as a covariate. $\chi^2 = 467.89$ (df = 313, $p < .001$), CFI = .92, RMSEA = .05 [.04 -.06], SRMR = .09. IDE = the indirect effect, $c$ = the total effect of the intervention on the outcome, and $c’$ = the direct effect of the intervention on the outcome after accounting for the mediation pathway.

* $p < .05$; ** $p < .01$; *** $p < .001$. 
**Figure 3.5**

*Intervention Impact on Teachers’ Follow-Up Occupational Self-Compassion Mediated by Frequency of Mindfulness Practice After the Intervention*

![Diagram showing the relationship between Intervention, Mindfulness Practice, and Self-Compassion](image)

*Note.* Structural Equation Modeling with MLR estimation (ML with robust standard errors) including project site as a covariate. $\chi^2 = 73.41 \text{ (df } = 46, p < .01)$, CFI = .96, RMSEA = .06 [.03 -.08], SRMR = .05. IDE = the indirect effect, $c$ = the total effect of the intervention on the outcome, and $c'$ = the direct effect of the intervention on the outcome after accounting for the mediation pathway.

† $p < .07$; ** $p < .01$; *** $p < .001$.

**Additional Model Specifications**

A set of additional mediation models were tested to examine potential intervention by condition interactions. Across both outcomes and time points mediation path estimates did not meaningfully change in these models. The only noticeable exception to this was that with the inclusion of the interaction term, the main effect of the intervention on mindfulness at post program and follow-up became marginal at $p < .07$. This result was expected and reflects the smaller effect of mindfulness found in the most recent U.S. sample. These additional results indicated that while there may be variation across project sites with regard to program outcome effects, the processes through which those effects occur seemed to be the same. As such, the models
described above that include project site as a covariate were maintained as the final models.

Discussion

The purpose of this project was to examine whether randomization to an MBI increased teachers’ use of mindfulness practices and if so, whether frequency of mindfulness practice served as a mechanism through which the intervention increased teacher mindfulness and occupational self-compassion skills at post-program and follow-up. Overall, results of this study demonstrated significant increases in the frequency in which teachers engage in mindfulness practices both during the intervention and in the summer months after the intervention was over. At its most basic level, these results demonstrate that participants assigned to the MBI group successfully learn key skills that are taught during the program and adopt them into their own repertoire of healthy habits in a relatively lasting manner. This link is important to establish in its own right, as it provides empirical evidence demonstrating the intervention is effective at teaching participants the skills theoretically defined as key to the cultivation of mindfulness-related skills.

Second, these results indicate that teachers’ frequency of mindfulness practices explain increases in their long-term mindfulness and short- and long-term occupational self-compassion skills. These tests of mediation help identify one way in which mindfulness training imparts benefits to participants by illuminating mindfulness practices as key intervention features that lead participants to develop greater
mindfulness and occupational self-compassion. The full mediation of mindfulness practices on occupational self-compassion found at post-intervention and follow-up additionally indicates that mindfulness practices fully explain how teachers improve their occupational self-compassion skills. In contrast, the lack of mediation of mindfulness at post-intervention and the partial mediation of mindfulness at follow-up indicate there may be other important components of the intervention that are driving these changes. It will be important to continue examining other intervention components to understand the aspects that foster positive change in teachers, particularly those that may explain short-term program impacts on mindfulness.

These results also underscore the importance of measuring frequency of practice, in addition to total minutes spent engaging in mindfulness practices, in MBI studies generally. While our measure of frequency of practice is relatively general in nature, its more general focus may actually help explain why previous studies examining dose-effect relations have been mixed. For instance, Soler and colleagues (2014) demonstrated that frequency of practice was significantly associated with more facets of mindfulness than was average number of minutes spent practicing. It could be that brief frequent practices help individuals maintain balance through regular reminders of mindfulness on a continued basis. Similar to physical fitness, consistent regular practice may be key in strengthening the many skills that comprise mindfulness, including executive function, attention, non-judgment, and compassion. Frequency of practice, rather than adherence to the total suggested minutes of practice, may also help explain
how MBI participants gain a variety of benefits despite the fact that most do not adhere to the “full dose” requirements of practice often assigned to them (Berghoff et al., 2017). It will be important to incorporate measures of both frequency and time spent practicing, examined in both treatment and control groups, in future work to better understand these relations.

It is equally important to note, however, that no support was found for one of the most basic hypotheses in this study. It was surprising to discover that mindfulness practices did not help explain teachers’ development of mindfulness at post-program. Because a significant relationship between these two factors was found a few months later at follow-up, these findings may indicate that a longer timeline is needed for such practices to result in the development of greater mindfulness among participants. This failure to find a short-term impact on mindfulness may also help explain why research in general on this topic is mixed. In the future, it may be fruitful to follow participants of MBIs for greater amounts of time in order to better gauge the timeline it takes for mindfulness practices to manifest as improved mindfulness. Additionally, this lack of short-term impact of practice on mindfulness may indicate that other aspects of the training directly impact teachers’ development of mindfulness. For example, mindfulness practice has been previously described as a method for changing participant perspectives, such that through training participants learn to reperceive aspects of their life (Hölzel et al., 2011). This may be one way in which MBIs directly impact participants’ development of mindfulness and the emergence of this new
perspective may not be a linear function of the frequency of mindfulness practices, but may instead initially reflect a qualitative shift that can then be deepened and consolidated through additional practice.

**Limitations**

Several limitations of this project warrant discussion. First, the mediation analyses conducted in this study were not true tests of mediation, as true tests of mediation require three time points of measurement. This methodological limitation was weighed against the benefit of understanding frequency of practice to outcome relations within a closer time frame. Using retrospective reports from both time points allowed us to test specific associations between practice frequency during the intervention in relation to post-program impacts and between practice frequency during the summer months after the intervention in relation to follow-up impacts. This decision turned out to be important in understanding the differences between short- and long-term effects of practice on mindfulness in these samples of teachers, as was demonstrated by the differential effects of practice on mindfulness at each time point. However, future research should consider measuring frequency of practice in a manner that is not retrospectively reported – perhaps through the use of a phone app or digital log – which would allow more accurate concurrent estimates of frequency to be collected. Additionally, as has been discussed elsewhere (e.g., Del Re et al., 2013; Parsons et al., 2017), in addition to quantity, information about practice quality will
likely provide important supplementary information about the effect of practice on mindfulness-related skills.

Second, our study measures have several limitations. The generality of the frequency of practice question at post-program does not distinguish between practices engaged in as part of the MBI (e.g., attending program sessions) and home practice for MBI group participants. In contrast, teachers’ frequency of practice over the summer months is a measure only of practices engaged in outside of the intervention, for all participants. Because of this potential ambiguity, treatment teachers may be differentially reporting their frequency of practice, with some counting required intervention sessions while others do not. In future research on the role of practice in the development of mindfulness, it will be important to distinguish among different types of meditative practices (e.g., formal sitting, mindful movement) and practice type (e.g., intervention sessions, home practice) to distinguish any differential effects of these factors (Vettese et al., 2009).

An additional measurement limitation involves how the outcomes of this study were measured. Due to a copying error in one of the project sites, one of the subscales of mindfulness was not collected. Additionally, the occupational self-compassion items used in this project only represented a subset of the total subscales that conceptually make up previous definitions of self-compassion as a construct. Strategies used in this study, such as examining outcomes as latent factors in the SEM framework and estimating missing data using full information maximum likelihood, do reduce some
possible bias associated with these measurement issues. However, because of these limitations, we decided not to test how practice frequency impacts specific facets of mindfulness and occupational self-compassion. Nevertheless, examining these relations at a more nuanced level is an exciting next step in this line of research. This is particularly true given recent research on mindfulness demonstrating potential links between mindful observing and negative psychological symptoms (Baer et al., 2006), and because certain mindfulness practices place less emphasis on the act of describing experiences (i.e., putting words to thoughts, emotions, or physical sensations; Carmody & Baer, 2008) and as such, may not be significantly related to practice in some cases.

Lastly, several design features could be improved in future work. In order to obtain a sample size (N > 125 suggested by Fritz & MacKinnon, 2007) sufficient to adequately test mediation effects over time, three separate intervention project sites were combined and analyzed simultaneously. ANCOVAs and regression interactions testing for group and project differences indicated no significant differences in outcomes between intervention conditions at baseline, and no meaningful condition by project site interactions on mediation pathways at post-intervention or follow-up; however, future work would be wise to replicate these findings in a single study powered to test mediation explicitly. In addition, for each of these projects, resources only allowed for waitlist control groups, but not active waitlist control groups. Including active control groups in future studies will further elucidate the specific aspects of mindfulness trainings that are key in producing positive outcomes, addressing a serious
limitation in the current field of mindfulness research (Davidson & Kaszniak, 2015; van Dam et al., 2018).

**Conclusion**

This study has demonstrated that an MBI is successful in teaching teachers mindfulness practices and that teachers continue to use these practices after the intervention is over. Additionally, the frequency in which participants engage in mindfulness practices significantly predicts their short- and long-term increases in occupational self-compassion and long-term mindfulness. Increases in mindfulness and occupational self-compassion are important outcomes in their own right, as these skills are associated with a variety of positive outcomes including enhanced attention and emotion regulation, executive functioning, changes in perception of the self, and wellbeing (Hölzel et al., 2011; Neff & Knox, 2017). Additionally, evidence suggests that mindfulness and occupational self-compassion explain teachers’ improvements in occupational health and wellbeing following MBI training (Roeser et al., 2013). Therefore, establishing that mindfulness practices learned during an MBI actually translate into improvements in mindfulness-related skills provides empirical support for the first causal step in the logic model describing how mindfulness interventions benefit participants generally, and teachers specifically in the education context.
Chapter 4: Study 3

Contrasting Missing Data Methods in Applied Experimental Research: Simulated Studies to Inform Best Practices
Abstract

The current study sought to compare modern missing data methods, namely, two types of multiple imputation—combined multiple imputation (CMI) and separate group imputation (SGI)—and full information maximum likelihood (FIML), in producing unbiased estimates of intervention impacts. Simulated data were generated that varied the effect size of the intervention impact, the sample size of the study, and the rate of missing data, to determine which method(s) perform best under common applied intervention research conditions. Collapsing across simulation conditions, several notable patterns of method bias and efficiency emerged. First, few differences were found across all three methods under conditions of a large intervention impact (.80) and little missing data (10%). Second, under conditions of smaller effect sizes (.20 and .50), smaller sample sizes (n = 50 and n = 100), and higher rates of missing data (25% and 50%) CMI consistently produced results that were more biased than results derived using both FIML and SGI methods. Third, it was found that SGI results were more efficient (generated smaller standard errors), as compared to both FIML and CMI, as missing data rates increased. These results indicate that intervention researchers should utilize FIML or SGI, rather than the traditional method of CMI, when analyzing intervention impacts. Implications of this work and future directions for the field are discussed.

Keywords: Missing data; Intervention; Multiple imputation; Maximum likelihood
Contrasting Missing Data Methods in Applied Experimental Research: Simulated Studies to Inform Best Practices

The experimental design—in which participants are randomly assigned to groups, an intervention is offered to and implemented with one group and not the other, and the groups’ outcomes are subsequently compared—is increasingly common in applied social science research to test the impacts that various social programs, policies, and practices have on promoting desired outcomes. Due to the nature of the design, researchers conducting applied experimental studies will likely grapple with missing data caused by measurement non-response and attrition. If not managed appropriately, missing data can be highly problematic and lead to biased parameter estimates, inflated confidence intervals, and reductions in power—all of which contribute to an increased likelihood of committing type 1 and type 2 statistical errors that subsequently produce incorrect conclusions about intervention effectiveness.

Committing an error of either type represents a potential opportunity cost to society at large, if ineffective programs are adopted for wider-scale use or promising programs are discarded. For instance, if a program is truly ineffective, but a study finds statistical support for its effectiveness (a type 1 error) finite public funding and services may be incorrectly allocated to the ineffective program, rather than to an effective one. In contrast, if a researcher concludes a successful program is not efficacious (a type 2 error), at minimum, the cost to society can manifest as resources wasted on program improvement or refinement. On a larger level, discarding a promising program could
have far greater implications for society, by stunting innovation and progress (Leviton, 2011).

These issues are concerning in any research study; however, they are particularly problematic in studies testing the efficacy of interventions, as scientific evidence derived from experimental studies of program impacts is often used by stakeholders to make decisions about which programs should be (or should not be) funded and adopted for wider-scale use. Fortunately, the statistical methods available to help researchers address missing data have advanced in recent years. Maximum likelihood (ML) and multiple imputation (MI) represent two ‘state of the art’ missing data methods that help researchers recapture statistical power lost to attrition, or missing data more generally (Schafer & Graham, 2002). These modern methods have been shown to perform significantly better than older ad hoc methods, such as listwise deletion or single mean imputation, especially as missing data rates increase (Enders & Bandalos, 2001). Indeed, simulation studies have demonstrated that when certain conditions are met (i.e., normally distributed data, sufficient sample size, same model specification) both MI and ML methods generate comparable results (Gelman et al., 2014). Therefore, while specific researchers may prefer to utilize one of these methods over the other based on the nature of a study and proposed statistical analyses (Enders, 2016), both methods are often equally endorsed in the literature when dealing with missing data (Dziura et al., 2013; Schafer & Graham, 2002).
In intervention research specifically, however, additional considerations exist when choosing between methods. Specifically, when using MI to account for missing data, methodologists recommend imputing the treatment and control group data separately (Enders, 2013; Enders & Gottschall, 2011a; Graham, 2009). Despite its recommendation as a best practice, however, this caveat is not well known and even less widely practiced. Separate group imputation (SGI) is substantively the same method as the traditional combined MI method (referred to now as CMI), however, imputing data within each group separately allows the mean and covariance structure of each group to differ (Enders & Gottschall, 2011a). The SGI method better aligns with the assumption underlying the intervention framework, because an intervention is meant to change the trajectory of the treatment group and therefore create distinct subpopulations on important outcomes. As such, failing to account for these differences when deriving replacement values from the MI regression-based estimation process could weaken a true intervention effect and subsequently increase a researcher’s likelihood of making a type 2 statistical error.

Despite its theoretically sound logic and stated practical importance, no studies are known at this time to have tested the performance of SGI, in comparison to CMI and FIML, in the context of interventions. Therefore, the primary purpose of this study was to conduct a series of data simulations to compare how well SGI, CMI, and FIML perform in generating unbiased estimates of intervention effects under varying conditions commonly found in applied intervention research.
Missing Data Mechanisms

Before each missing data technique is described in more detail, it is first important to review the possible underlying mechanisms that create patterns of missingness, as these assumptions largely determine how well each technique will perform under different conditions (Enders & Gottschall, 2011b). Using Rubin and colleagues’ classification system (Little & Rubin, 2002; Rubin, 1976), missing data mechanisms can be categorized as missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR).

In order to be used appropriately, ad hoc missing data methods, such as listwise deletion, require the strictest assumption of MCAR to be met. Data meet the MCAR assumption when missing values for a given variable are unrelated to the variable with missingness itself and all other observed variables (i.e., missingness is truly random). To provide a relevant example, suppose an educational researcher is collecting post-intervention social-emotional learning (SEL) scores in the spring of an academic school year and several students are absent because they have moved to another school district. If the reasons these students moved were truly random, the MCAR assumption would be met and it would not bias intervention impact estimates to use a complete-case analysis method (e.g., listwise deletion; Enders, 2001). In this scenario, excluded cases represent a random subsample of the population and the only consequence of using a complete-case method would be loss of statistical power (Graham, 2009).
MI and FIML are preferred methods for addressing missing data, because they only require the (less strict) assumption of MAR to be met. When data are missing at random, the missing data may relate to other observed variables, but are not related to the variable with missingness itself (Schafer & Graham, 2002). In the previous example, this would occur if student mobility was related to family socioeconomic status (SES) and students’ SES was related to their SEL scores. In this case, the cause of missingness, student mobility, is only indirectly related to the missing SEL scores, through its relation to family SES. If a complete-case analysis method was used in this example, it would bias intervention estimates to the extent that information about a specific subgroup of students with similar SES and SEL scores are no longer accounted for in the sample distribution. To the extent that lower SES is correlated with lower SEL scores, failing to account for this subgroup of individuals who left the program could truncate the lower half of the sample distribution and cause researchers to conclude the intervention is more effective than it actually is. In this situation, using MAR-based methods, such as MI or FIML, allows researchers to utilize all available data from the subsample of participants with missing data when estimating sample statistics, therefore eliminating bias that could occur if they were excluded (Sinharay et al., 2001).

Although tangential to this paper’s focus, scores can also be considered MNAR, which is determined if the missing data are directly related to the scores that are missing. In this scenario, using MI, FIML, or any other method that requires data to be assumed missing at random, are no longer appropriate for use. Extending the above
education intervention example, data could be considered MNAR if the school hosting the SEL intervention expelled students with severe behavioral problems during the intervention timeframe. In this case, students’ low SEL skills would help explain why their post-program SEL assessment scores were missing. Using MAR-based methods to account for missing data in this situation, when the data are MNAR, may not be appropriate because the missing data may be biased in a manner that cannot be fully un-systematized – even by including variables in the analysis specification that reduce non-response bias by helping to distinguish participants with complete and incomplete data (i.e., auxiliary variables; Enders, 2016).

Practically speaking, however, it is impossible to determine whether data are MAR or MNAR, because doing so requires knowledge of the missing values themselves. Thus, while an understanding of these assumptions is important for theoretical reasons, the more practical goal for researchers is to reduce as much potential bias as possible when dealing with missing data, and MAR-based methods are known to perform best in the widest range of scenarios – even in some cases where data have been determined to be MNAR (Enders, 2001; Enders & Bandalos, 2001). Readers interested in a more in-depth description of missing data mechanisms generally are referred to Enders (2010), and those interested in how missing data mechanisms can differentially impact intervention results specifically, should see Dziura and colleagues (2013).
Missing Data Methods

Full Information Maximum Likelihood

The FIML method has often been described as an ‘implicit imputation’ process (Widaman, 2006) because rather than imputing the data directly, all available information is utilized to obtain estimates that best summarize the data. In other words, FIML is conducted on an analysis-by-analysis basis with the goal being ‘to identify the population parameter values that have the highest probability of producing a particular sample of data’ (Enders, 2010, p. 59). Obtaining the most likely population mean and variance requires raw data as the input, which allows the method to use all available data (even what’s partially missing) to generate a set of model-implied mean and covariance matrices. The model-implied matrices that most closely fit with the observed data are then used, instead of the raw data, to generate parameter estimates. These model-implied parameter estimates better reflect what the observed data likely would have generated, had there been no missing data (Enders, 2010; OpenMX Development Team, 2014).

FIML may be a preferred method by many researchers due to its simplicity. For instance, researchers only need to properly specify their analysis specification within a FIML-supported statistical package to use it (Enders, 2016). Availability and use of sophisticated statistical packages continues to expand, and although designed for more complex modeling, these software packages can typically run any models within the general linear model framework (e.g., multiple regression; Graham, 2009). In contrast,
Enders (2016) describes how FIML estimation may be disadvantageous to use in certain scenarios that include complex models with both categorical and continuous variables. To estimate properly, FIML requires a series of distributional assumptions to be met, one of which is multivariate normal data. If researchers include a series of categorical variable dummy codes with missing values in the estimation process, software packages will assume these dichotomous variables are normally distributed. For intervention research specifically, this categorical issue may or may not be of concern, as the categorical intervention condition variable is, by definition, complete in an intent-to-treat analysis. However, it is important to be aware of this limitation, because in many cases, intervention researchers will want to include other important categorical variables in their analysis specification, such as race or gender, that may have missing values.

**Multiple Imputation**

MI is a simulation-based estimation technique that replaces missing data with a set of $m > 1$ possible values, thus creating $m > 1$ complete-case data files. MI improves and extends single-value imputation methods by purposefully reflecting the uncertainty of the imputed values through variation in estimates across the $m$ datasets. MI consists of three steps: imputation, analysis, and pooling. In simplified terms, the imputation process generates plausible values to replace missing values using regression equations based on the mean vector and covariance matrix of the sample data. Additional datasets with different missing data replacement values are then generated using the updated
mean and variance structures from the previous step. This process continues in an iterative fashion until the desired number of datasets is reached (Enders, 2010). After imputation, analyses are conducted on each of the m datasets using complete-case analysis methods. Then, estimates derived from each analysis are pooled to obtain average estimates across all imputed datasets (Enders, 2016).

Unlike FIML, MI datasets are often generated once and then used for many subsequent analyses (Graham & Schafer, 1999). Imputing once for a series of analyses is one advantage of MI over FIML. MI is also advantageous for other reasons. First, while it is possible to include auxiliary variables (those used in the imputation phase but excluded from the analytic specification) when using FIML (see Collins et al., 2001), it is arguably much easier to do in MI. Second, MI may be particularly useful in scenarios where single items on a scale are missing (i.e., item non-response). This is because items can be imputed prior to aggregating scores, thus allowing researchers to use the available inter-item correlations to estimate the missing items more accurately. In contrast, if any items are missing and FIML is used, it is often recommended that researchers leave the entire scale score as missing, thus ignoring the item-level information available to them (Enders, 2016). Lastly, MI is arguably an easier method for researchers to implement, because after data are imputed, analyses can be conducted with familiar complete-case analysis methods.

**Separate Group Imputation.** SGI is substantively the same process of estimation as MI is known in its traditional form, however, it allows the mean and covariance
structure of each group to be estimated separately. In intervention studies, conducting MI on each group separately allows group differences to be preserved in all variables with missing data, merely by separating each group’s data prior to imputing (Enders & Gottschall, 2011a). SGI also provides intervention researchers with greater model specification flexibility, such that intervention interaction terms can be calculated after the imputation process is complete, rather than having to be prespecified if data are imputed as a combined sample (Enders & Gottschall, 2011a). It is important to note, however, that data must meet several requirements for SGI to be utilized appropriately. First, it requires subgroups to be mutually exclusive (Enders & Gottschall, 2011a). This is of no concern for intervention researchers, as participants cannot belong to more than one group in the intent-to-treat framework. Second, SGI requires a sufficient number of participants within each subgroup being tested (e.g., n = 50; Graham & Schafer, 1999), otherwise researchers increase the likelihood of facing model convergence issues, particularly as model complexity increases (Enders, 2010).

Despite its clear value and recommended use, SGI is not well known and even less widely utilized in intervention research. In fact, information about SGI has been left out of several missing data method accounts in clinical research (e.g., Armijo-Olivo et al., 2009; Dziura et al., 2013; Jakobsen et al., 2017), whereas others mention both SGI and CMI as acceptable MI methods (e.g., What Works Clearinghouse, 2017; but see Dong & Peng, 2013; Enders, 2016). Additionally, while SGI has been recommended for use by experts in the field, no studies are known at this time to have empirically examined how
SGI performs in comparison to other recommended methods in intervention research. Therefore, the purpose of this study was to investigate the performance of SGI, relative to FIML and the traditional MI method that combines groups (CMI), when dealing with missing data in intervention research contexts. Specifically, the research question examined is: how do the methods of CMI, SGI, and FIML compare in producing unbiased and efficient estimates of intervention effectiveness? To answer this question, a series of data simulations were conducted to compare these methods and how they differ with respect to three features of an intervention under study—(1) effect size of the intervention impact, (2) percentage of missing data, and (3) sample size.

**Method**

A series of simple simulation studies were conducted to test how well FIML, CMI, and SGI perform when dealing with missing data in the intervention framework. The purpose of these simulations was not to be comprehensive in nature, but instead to illustrate the potential impact missing data may have on conclusions made about intervention effectiveness under common (and more extreme) intervention conditions. Therefore, three simulation studies were performed that modeled .20, .50, and .80 intervention effect sizes. Within each of these studies, four independent variables were manipulated: sample size (50, 100, 250), missing data rate (10%, 25%, 50%), and missing data method (FIML, CMI, SGI).

The chosen effect sizes were determined based on Cohen’s effect size benchmarks indicating small, medium, and large effects of mean differences between
groups (Cohen, 1988). Similarly, sample sizes were chosen to represent a range of common sample sizes found in the social sciences (Kisbu-Sakarya et al., 2013), beginning with a small, underpowered sample size (n = 50) that may be seen in a well-designed pilot, and increasing to a large sample size for an applied experimental study (n = 250). Missing data rates were chosen to cover a range of possible levels of missing data in applied research, starting with 10%, as missing data rates of 5% and 10% have commonly been indicated as benchmark rates in which missingness is non-ignorable (Bennett, 2001; Graham, 2009) and increasing to 50% which is reaching the upper limit of missingness that is acceptable in intervention research (What Works Clearinghouse, 2017). In sum, 27 design cells (3 sample sizes x 3 missing data rates x 3 missing data methods) were examined for each of the three intervention effect size studies. For these method comparisons, the number of imputations were fixed at 5 imputations for both CMI and SGI, as general guidelines indicate 5 imputations are sufficient (e.g., What Works Clearinghouse, 2017).

*Simulation Specification*

R software (R Core Team, 2017), and the rnorm function within the software base package specifically, was used to generate 1,000 normally distributed data samples within each design cell described above. Data samples were generated to reflect standardized estimates (z-scores) of a non-specific intervention outcome of interest, with treatment and control group data generated separately. Control group data samples were generated to model no intervention effect, such that they have a mean of
zero and a standard deviation of one. For every control group sample generated, three specification-matching treatment group samples were also generated, with mean scores corresponding to the pre-specified effect sizes (.20, .50, .80) and standard deviations equal to one. After generating each sample data set, data were randomly removed according to the pre-specified missingness rates. A random seed was specified at the start of each simulation in order to allow for replication of the simulated results exactly, for monitoring purposes.

For each simulated sample, missingness was modeled as occurring completely at random (MCAR) to isolate bias as it relates to the proportion of missing data only, rather than any systematic form of missingness. Additionally, values were only removed on outcome variables, such that the variable representing intervention condition remained complete to represent the intent-to-treat analysis framework. Rates of missing data were also equally distributed across treatment and control groups, to model equal, rather than differential attrition. The same generated datasets were then used in all subsequent analyses testing the performance of each of the three missing data methods. Using the same datasets to test all three missing data methods is comparable to a matched pairs design, such that it removes the within sample variability of the samples, thus allowing any differences among the methods being tested to be detected (Burton et al., 2006).
**Imputation Specification**

The MICE package (van Buuren & Groothuis-Oudshoorn, 2011) in R was used to impute the data for CMI and SGI. Imputations were conducted using the predictive mean matching technique (PMM). PMM is a common regression imputation method that allows missing values to be replaced with an observed value from another respondent whose regression-predicted score is closest to that of the individual’s whose value is missing. This method is considered to be an improvement over mean imputation and many other regression estimation methods because it imputes values that mimic true values in the data and as such, is easy to implement when data are either discrete or continuous (Landerman et al., 1997; Little, 1988). To follow general guidelines for CMI, treatment and control group samples were imputed as a combined data file that includes an intervention condition variable and the outcome variable. To conduct all separate group imputations, treatment and control group data files were imputed separately based on intervention condition assignment and then combined after the imputation process was complete.

**Analysis Specification**

Intervention impact analyses were conducted within the regression framework, with intervention condition as the predictor and the non-specified standardized outcome as the dependent variable. The regression framework was chosen to compare all three methods for several reasons. First, conceptually speaking, the intervention impact unstandardized beta in the regression framework represents the mean (z-score)
difference between treatment and control groups on the outcome and therefore the difference between model derived estimates and the prespecified population effect size can provide a measure of method bias. Second, although the models used here do not include covariates, this framework can be used to generalize to a variety of more complex models commonly found in intervention research that examine unique effects of the intervention, controlling for other variables. Third, using the FIML method requires use of an available SEM package and thus, using the regression framework allows the analytic method to remain consistent across all missing data methods. Analyses were conducted on each sample and estimates from each sample’s results (i.e. intervention impact estimate, estimate standard deviation, and p-value) were saved.

To test the performance of the FIML method, analyses were conducted using the Lavaan package (Rosseel, 2012) in R. For CMI and SGI, regression estimates from all imputations were pooled within each sample and these average estimates were subsequently saved.

**Performance Measures to Compare Missing Data Methods.** The primary purpose of these simulations was to examine the impact of the proposed missing data methods on intervention impact parameter estimates. Specifically, the following performance measures were examined: (1) raw bias, calculated as the difference between the average simulated sample estimate per design cell and the population estimate (prespecified effect size), where negative bias indicates the samples underestimated the population estimate on average, and positive bias indicates the
samples overestimated the population estimate on average; (2) standard error (efficiency), calculated as the average sample estimate standard deviation, where smaller standard errors indicate greater efficiency and larger standard errors indicate lower efficiency; and (3) standardized bias, calculated by dividing the raw bias by its associated standard error, which creates a form of bias that is in standardized units. The entire simulation code used to generate and analyze these data can be found in the Appendix for reference.

**Results**

Results comparing FIML, CMI, and SGI in reproducing intervention impact estimates can be found in Tables 4.1-4.3. Table 4.1 compares each of the methods in reproducing a small (.20) intervention impact estimate, Table 4.2 a medium (.50) intervention impact estimate, and Table 4.3 a large (.80) intervention impact estimate. The results discussed below represent general conclusions and trends in method performance across all three effect size simulation studies, based on the performance measures described above. Because the missing data were modeled as MCAR, and each of the methods being tested represent the best available to date, high rates of bias were not anticipated. However, a certain amount of bias can occur from missing data alone, and as such, simulation conditions that generated standardized bias equal to or exceeding .10 (1/10 of a standard deviation) were bolded to represent notable amounts of bias, given the conditions of these simulations.
Raw Bias

Collapsing across all three intervention effect size studies, overarching trends of bias were as expected—each missing data method demonstrated increasing bias as missing data increased. Additionally, there was little difference in raw bias across all three methods when missingness was low (10%). Comparing each of the methods across all impact effect size studies and simulation conditions indicates levels of raw bias for FIML and SGI remained low even with 25% and 50% missing data, with each never exceeding ±.02 raw bias (see Tables 4.1-4.3). However, as can be seen in Table 4.2, raw bias for the CMI method was considerably larger in magnitude, particularly in the .50 effect size condition—reaching -.08 and -.06 in the small and medium sample size conditions with 50% missing data. Additionally, it is important to note that both FIML and SGI methods generated results that both overestimated (positive bias) and underestimated (negative bias) the true intervention impact to a small degree, whereas CMI consistently underestimated the true intervention impact.

Efficiency

In general, efficiency increased across all methods as sample size increased, and decreased as missingness rates increased. A comparison of the relative efficiency of FIML, CMI, and SGI indicates similar efficiency across all three methods, especially when missing data rates are low (see the first three rows each within Tables 4.1-4.3). Furthermore, the relative efficiency of SGI, compared to the other methods, increased as the rate of missing data increased. Specifically, as can be seen in the 50% missing
data condition within the final three rows of Tables 4.1-4.3, SGI had noticeably smaller standard errors compared to the other two methods. For example, in the large intervention impact study presented in Table 4.3, in the n = 50 and 50% missing data condition, SGI’s efficiency was .32, whereas both FIML and CMI efficiency was .37. This indicates that while all three methods’ efficiency was impacted by increasing missing data (i.e., the precision of estimates decreased as missing data increased), SGI’s precision was less affected compared to FIML and CMI.

**Standardized Bias**

Trends in standardized bias across all three intervention effect size studies matched the pattern of raw bias and increased as missing data increased, however, the standardization reflects this pattern to a greater magnitude and in standard deviation units. Comparing across methods, results indicate FIML produced the least amount of standardized bias, with a maximum of .04 found in Table 4.3 in the n = 50 and 50% missing data condition. In most cases, SGI produced similarly low levels of standardized bias, generally producing levels equivalent to ± .05 or less. However, this generally low trend was violated in one condition—specifically, SGI standardized bias was -.10, or 1/10th of a standard deviation unit, within the medium effect size study (Table 4.2) in the n = 250 and 50% missing data condition. CMI standardized bias was frequently much larger than both FIML and SGI, especially in the .20 and .50 effect size studies and higher rates of missing data. Because CMI generated similar efficiency to the other methods, but consistently had more raw bias, CMI standardized bias equaled or exceeded -.10 in
some conditions, and even exceeded -.20 in several conditions, across the effect size studies.
### Table 4.1
Average Estimates from Small Effect Size (.20) Intervention Impact Simulations

<table>
<thead>
<tr>
<th></th>
<th>FIML</th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
</tr>
<tr>
<td>10% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.20</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.20</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td>25% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.32</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.20</td>
<td>.23</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.20</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>50% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.37</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.19</td>
<td>.28</td>
<td>.01</td>
</tr>
<tr>
<td>N = 250</td>
<td>.20</td>
<td>.18</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.20) and the average sample derived estimate, and standardized bias is raw bias reported in standard deviation units. Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. FIML = full information maximum likelihood; CMI = combined multiple imputation; SGI = separate group imputation. Estimates for CMI and SGI generated using 5 imputations.
### Table 4.2

*Average Estimates from Medium Effect Size (.50) Intervention Impact Simulations*

<table>
<thead>
<tr>
<th>Conditions</th>
<th>FIML</th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td>10% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.50</td>
<td>.49</td>
<td>.50</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>25% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.46</td>
<td>.49</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>50% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.42</td>
<td>.49</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.44</td>
<td>.50</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.49</td>
<td>.50</td>
</tr>
</tbody>
</table>

**Note.** Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.50) and the average sample derived estimate, and standardized bias is raw bias reported in standard deviation units. Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. FIIML = full information maximum likelihood; CMI = combined multiple imputation; SGI = separate group imputation. Estimates for CMI and SGI generated using 5 imputations.
### Table 4.3

*Average Estimates from Large Effect Size (.80) Intervention Impact Simulations*

<table>
<thead>
<tr>
<th></th>
<th>FIML</th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
</tr>
<tr>
<td>10% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td>25% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.31</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.23</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>50% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.82</td>
<td>.37</td>
<td>.02</td>
</tr>
<tr>
<td>N = 100</td>
<td>.81</td>
<td>.28</td>
<td>.01</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.18</td>
<td>.00</td>
</tr>
</tbody>
</table>

**Note.** Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, and standardized bias is raw bias reported in standard deviation units. Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. FIML = full information maximum likelihood; CMI = combined multiple imputation; SGI = separate group imputation. Estimates for CMI and SGI generated using 5 imputations.
Discussion

This study sought to compare the performance of SGI, CMI, and FIML in the intervention framework to determine whether differences existed across methods in generating estimates of intervention impacts. Comparing each of these methods across small, medium, and large intervention effect sizes elucidated several patterns of results that have important implications for researchers addressing missing data. First, these simulation studies indicate that FIML, CMI, and SGI typically reproduce the true intervention impact similarly under conditions of little missing data (10%) and large intervention effects (.80). Second, under other conditions, that are often more likely in applied research (i.e., intervention impact effect sizes in the .20 to .50 range and higher missing data rates), CMI often produced results that were more biased than both FIML and SGI, such that CMI consistently underestimated the true intervention impact—sometimes by over 1/5th of a standard deviation unit. These results provide empirical evidence that intervention researchers should utilize FIML or SGI, rather than CMI, to examine intervention impacts in samples with missing data in their future research (Enders, 2016; Graham, 2009).

The fact that the CMI method consistently underestimated the true intervention impact demonstrates that researchers can lose precision in detecting between group differences simply by imputing treatment and control group data as a combined sample. The consequences of this decision can potentially increase the risk of type 2 error and the likelihood that an intervention will fail to find statistical support for the program’s
effectiveness, even when data are missing completely at random. Additionally, the results found in this study are in contention with current recommendations in the field that still include CMI as an acceptable missing data method for analyzing intervention data. For example, What Works Clearinghouse (2017) states that researchers can use either SGI or CMI, just so long as if CMI is used, researchers include the treatment/control variable in the imputation specification. Results of this study suggest that even when the treatment/control grouping variable is included, the results of CMI can still be demonstrably more biased than those generated using SGI or FIML. As such, the field’s current recommendations of acceptable missing data methods should be updated, to encourage researchers to utilize SGI or FIML, rather than CMI, when examining intervention impacts. Specifically, Table 4.4 translates the results of these simulations into a guide that researchers may find useful in determining which methods may perform best under conditions that most closely align with their own studies.
Table 4.4
Method Recommendations for Best Practice by Study Condition

<table>
<thead>
<tr>
<th>Study Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIML</td>
</tr>
<tr>
<td>.20</td>
</tr>
<tr>
<td>10% Missing</td>
</tr>
<tr>
<td>N = 50</td>
</tr>
<tr>
<td>N = 100</td>
</tr>
<tr>
<td>N = 250</td>
</tr>
<tr>
<td>25% Missing</td>
</tr>
<tr>
<td>N = 50</td>
</tr>
<tr>
<td>N = 100</td>
</tr>
<tr>
<td>N = 250</td>
</tr>
<tr>
<td>50% Missing</td>
</tr>
<tr>
<td>N = 50</td>
</tr>
<tr>
<td>N = 100</td>
</tr>
<tr>
<td>N = 250</td>
</tr>
</tbody>
</table>

Note. Recommendations based on standardized bias. Columns represent effect sizes of intervention impacts. Methods are not recommended for use if standardized bias reached or exceeded ±.10, or 1/10th of a standard deviation unit. • = recommended and x = not recommended under specified conditions.

These simulations have also raised a new question about the use of SGI in the context of intervention research. We found that SGI efficiency, relative to FIML and CMI efficiency, increased as missing data increased in all effect size studies and under all sample size conditions. These results indicate that the precision of estimates derived using the SGI method may be less affected by missing data compared to the other methods examined. If this is a true, improved precision in the face of missing data could be added to the list of benefits of using SGI instead of FIML. However, this relative improvement in efficiency can also lead to greater amounts of standardized bias, even when raw bias is minimal. This occurred in our study within the medium effect size simulation (see Table 4.2) in the 50% missing data and n = 250 sample size condition,
where despite having only -.02 raw bias, SGI standardized bias reached a notable level at -.10.

The increased relative efficiency of SGI as missing data increased could also potentially be explained by features of this study’s design. The current study used the predictive mean matching (PMM) method to impute missing values. PMM replaces missing values with another observed score in the data, so that imputed values mimic plausible observed values (Landerman et al., 1997; Little, 1988). However, a consequence of “borrowing” observed scores is that the imputed data will be forced to remain within the same range of values as the observed data (Allison, 2015). Given this consideration, as missingness increased in our simulations, the overall variation within each sample would likely have decreased—thus reducing the overall variance of each subsample when data were regenerated through imputation. This same issue would likely not have impacted efficiency using the CMI method to the same degree, given that the combined sample size was twice as large as was used in each SGI imputation.

Extending the current study by increasing the complexity of the data generated to include relevant covariates in the imputation process that do not have missing data, as well as comparing these findings with results derived using other imputation methods, would help to clarify SGI efficiency in the context of intervention research.

Lastly, even though FIML and MI methods have been available for use for quite some time now (Graham & Schafer, 1999; Little & Rubin, 2002), meta-analytic results from a clinical RCT review published in 2014 indicates that only 8% of clinical
researchers utilized MI methods to deal with missing data and only 19% used model-based methods (Bell et al., 2014). These review results suggest that there may be a general lack of application of these methods in intervention research contexts (and psychological research more generally; Peugh & Enders, 2004). As such, this discussion of modern missing data methods, and recommendations for their use generally, seeks to communicate the importance of using any of these methods to account for missing data over ad hoc methods such as listwise deletion, particularly when examining intervention effectiveness.

**Limitations and Future Directions**

All simulation studies, particularly those comparing methods, require a certain degree of caution when generalizing their results to situations outside of the scope of their specifications (Enders, 2001). As such, there are several limitations of this study that can be addressed in future work. In order to isolate the effect of each method in producing intervention estimates, data were simulated under ideal conditions that removed other potential factors that could influence bias. Specifically, data were simulated as MCAR and the proportion of missing data was distributed equally across treatment and control groups. Despite the importance of beginning with these ideal conditions to compare methods, these scenarios may not be as likely to manifest in reality. Additional simulation studies that test how these various methods perform under MAR and MNAR assumptions may be warranted, as these mechanisms are more
realistic in the intervention context, given that missing data is often out of an intervention researcher’s control (Schafer & Graham, 2002).

It will also be important to expand this simulated design in future work and apply it to different model specifications and scenarios. Again, to test model performance under ideal conditions, this study utilized a simplified regression framework that specifically examined the effect of a dichotomous intervention condition predictor on a non-specified continuous outcome within a single-level research design. Future work should compare these methods when incorporating dichotomous variables with missing data (as their non-normal distribution violates an assumption of FIML; Enders, 2010) and examine how the inclusion of various auxiliary variables and covariates may influence method performance in this context (Graham & Collins, 2012).

It is also important to note that the conditions under which all of these methods performed least well—in very small sample sizes and with 50% missing data—are conditions under which any claims about intervention effectiveness and generalizability must be made with a great deal of caution. The point of testing recommended method performance under these more extreme conditions was to understand specific scenarios in which these methods may not be optimally suited. However, using any of these methods, even under extreme conditions, is better than deferring to listwise deletion or another ad hoc method.

In conclusion, while the results of this study aim to inform future intervention work by demonstrating which missing data methods are most effective in producing
efficient and unbiased impact estimates, it is important to note that no method can completely recapture what information is lost when data are missing (Allison, 2002). Researchers should take extra care in the planning and implementation phases of their study, to avoid missing data in the first place, especially by considering participant response burden and program feasibility (Enders & Gottschall, 2011b). When missing data methods are needed, we believe this study can help to inform intervention researchers of the available methods that are most successful in producing unbiased estimates under various common intervention conditions.
Chapter 5: Discussion

This dissertation aimed to contribute to the field of mindfulness-based interventions (MBIs) for teachers in three ways: (1) by providing a summary of the current evidence base of randomized controlled trial (RCT) MBI studies with teachers in a manner that elucidates both theoretical and methodological extensions that should be addressed in future work; (2) by extending the field theoretically by examining the role of mindfulness practice in the development of mindfulness-related skills for teachers in an MBI program; and (3) by extending the broader field of intervention research methodologically, by empirically comparing modern missing data methods in producing unbiased and precise estimates of intervention effectiveness under various conditions commonly found in applied intervention research. The following discussion begins with a summary of each study’s findings and study-specific implications and future directions. Then, I transition to a larger discussion of future directions that I would like to address in my own work, that builds upon the work that I have presented here. Lastly, I discuss additional areas of future research that should be addressed in the field in general, and close with the broader implications of this work for teachers and students in schools.

Summary of Current Studies

Study 1 – Review of Mindfulness Interventions for Teachers

Study 1 reviews 12 RCT studies of MBIs conducted with pre-service and in-service teachers who teach kindergarten through high school grades. Studies were examined first for theoretical contributions using a theoretical framework put forth by
experts in the field (Jennings & Greenberg, 2009; Roeser et al., 2012) that describes how MBIs translate into positive outcomes for teachers, classrooms, and students. Results of the theoretical review suggest the field’s current strengths focus on understanding the impact of MBIs on teacher mindfulness-related skills, occupational health and wellbeing, and the causal mediating links between these proximal and more distal outcomes for teachers. Additionally, within the occupational health and wellbeing domain, researchers have predominantly measured outcomes focused on distress reduction.

These results suggest a paucity of empirical evidence in the remaining theoretically proposed domains—including a lack of research on teachers’ engagement in mindfulness practices, and distal program impacts on classroom and student outcomes. More work is also needed that explicitly tests the causal linkages connecting proximal and distal outcomes, other than those that connect mindfulness-related skills and occupational health and wellbeing (see Benn et al., 2012; Roeser et al., 2013). These results indicate that future work should focus on measuring mindfulness practices, positive aspects of occupational health and wellbeing, classroom outcomes, and student outcomes, and use mediation analyses.

Studies were also reviewed from a methodological perspective in order to determine how well they aligned with recommended best practices for executing RCTs testing the impacts of an intervention. Specifically, each study was examined based on factors within three methodological categories: (1) research design features (e.g., study design, level of randomization), (2) intervention implementation, and (3) other factors
that influence conclusions of intervention effectiveness (e.g., sample size, analysis specification). Results of this methodological review indicate that the reviewed MBI studies have exceeded expectations regarding basic RCT design standards by using waitlist-controlled designs (instead of no treatment control designs) and follow-up times of measurement. However, future work is still needed to examine the impacts of these programs in studies with larger sample and that use even longer follow-up times of measurement. More work is also needed that utilizes recommended missing data methods in connection with study attrition, and the field needs to focus to a greater extent on measuring a variety of domains within the field of fidelity of implementation.

**Implications.** The purpose of elucidating areas in the field that require additional work is to continue moving the field of MBIs for teachers in the direction of producing high quality studies that cover a wide range of interesting questions. However, just as important is continuing to improve our confidence in the answers we gain from those questions and using the best methods available to increase our certainty that our results reflect the truth. This is of the utmost importance in applied experimental research, because despite the fact that RCT methodology is particularly well-suited to answer questions of causality, this method too, falls short of perfection. As such, continuing to evaluate the decisions we make as researchers using this design is critical; this is because the motivation for implementing and assessing the effectiveness of intervention programs is, and should always be, to improve the community and individuals in which they are designed to serve.
Future Directions. This review provides a general guide for future work, both theoretically and methodologically, for the field of MBIs for teachers. However, looking beyond the explicit future directions elucidated, this review only included studies that utilized the RCT methodology. While this serves the beneficial purpose of allowing all reviewed studies to be examined at the same level of causal inference and for the same best practices, it is important to note that using an RCT design is not the only way to produce a high-quality study. In fact, there is much to be gained in using diverse methodologies as a means of balancing out the strengths and weaknesses of each method in the process of triangulating in on the truth (Skinner et al., 2019). Therefore, while we continue to improve the use of RCTs for the purpose of understanding MBIs for teachers, it is also important to supplement this work with other well-designed studies that serve to balance the specific weaknesses of the RCT methodology. This process of triangulation is discussed further in the future directions for the field section below.

Study 2 – The Role of Mindfulness Practice in the Development of Mindfulness-Related Skills

Study 2 addresses one theoretical gap highlighted in Study 1. Specifically, Study 2 empirically tests whether teachers’ frequency of engagement in mindfulness practices during and after an MBI serve as a mechanism through which teachers develop mindfulness-related skills. Examining this question in an intent-to-treat framework (i.e., comparing MBI group teachers to control group teachers) allows for causal attributions
to be made regarding the two main study conclusions. First, teachers randomly assigned to the MBI group learn and engage in mindfulness practices both during the intervention and after the intervention is over; and second, that MBI group teachers’ frequency of practice serves as a process through which teachers develop mindfulness and occupational self-compassion skills.

**Implications.** Overall, examining the impact of an MBI on teachers’ frequency of mindfulness practices within the intervention framework provides empirical support establishing practice as an essential active ingredient of the MBI tested. Directly examining hypothesized “active ingredients” or processes within interventions is extremely important, as it allows researchers to empirically test whether their intervention components are truly mechanisms of change. Finding the hypothesized processes are supported, as was the case in this study, helps to reaffirm the importance of those specific intervention components. However, finding the hypothesized processes are not supported is equally important. This is because a lack of support for either (a) a direct intervention impact on the active ingredients, or (b) a direct effect of the active ingredients on key targets, may provide researchers with crucial information that can be used to redesign aspects of the program to be more effective in the future (MacKinnon et al., 2007).

For instance, the fact that this study did not determine mindfulness practices to significantly predict teachers’ development of mindfulness immediately following the intervention may suggest that the timeline for skill transfer is longer for the outcome of
mindfulness as compared to occupational self-compassion (MacKinnon et al., 2007). Determining these differential processes exist may help to guide future work in studying other potential MBI components that help explain the positive short-term changes in teachers’ mindfulness. At the same time, empirical evidence that mindfulness practices play an important role in teachers’ development of both mindfulness (long-term) and occupational self-compassion (short- and long-term) reaffirms the use of these practices in future interventions as positive drivers of change for teachers’ mindfulness-related skills and occupational health and wellbeing. In sum, examining the processes through which MBIs promote teachers’ key program outcomes is essential in understanding how a program works and informing the ways in which improvements can be made to maximize positive change through the intervention in the future.

**Future Directions.** The positive implications of this work described above do not discount the fact that several improvements can be made in examining mindfulness practices in future research. First, it is innately challenging to measure participants’ frequency of practice—researchers must weigh the benefits of obtaining frequent and objective measurement of practice with the increasing possibility of missing data as response burden increases. Additionally, researchers have yet to agree on how mindfulness practices are best measured, whether it be how frequently participants practice (in terms of sessions), the duration of practice (in minutes or in total across an intervention), or the total amount of time a participant has practiced in their lifetime (cumulative practice; Berghoff et al., 2017; Del Re et al., 2013). Similarly, researchers are
also grappling with the best way to capture practice quality, alongside practice quantity, as another important factor likely to influence intervention outcomes (Del Re et al., 2013). Future work in all of these areas will help the field overall to better understand the role mindfulness practices play in producing positive intervention outcomes.

In addition to these future directions for the greater field of MBIs, it is important to consider specifically whether teachers continue to use mindfulness practices to support them long-term in buffering against the high emotional demands and occupational stressors associated with their unique profession. This study has illuminated that teachers continue to use these practices in the summer months after the intervention is over, however, what is most important is understanding whether teachers continue to use these practices to support them once a new school year begins. These programs are meant to impart lasting skills and strategies unto teachers that they can draw on when faced with everyday challenges. However, if continued practice is necessary for some of these improvements to maintain, it is important to understand if teachers continue engaging regularly in these practices after the supports and structure of the intervention are withdrawn and teachers face the increased time demands of a new school year once again. Answering questions of this sort will help researchers to understand if a single MBI is sufficient to achieve the overarching goal of supporting teachers, or if additional “booster” interventions are required to help teachers maintain the benefits they gained initially, in a more lasting manner.
Study 3 addresses a methodological gap elucidated in Study 1 by empirically comparing modern missing data methods—namely, full information maximum likelihood (FIML), combined multiple imputation (CMI), and separate group imputation (SGI)—in producing unbiased estimates of intervention effectiveness. Study 3 addresses this general methodological question in the context of intervention research using simulated data and examines which of these methods may be best for intervention researchers to use under various common applied intervention study conditions. Specifically, results indicate that the FIML and SGI methods both produce intervention results that are less biased than those derived using the CMI under conditions of higher rates of missing data (25% and 50%) and smaller intervention effects (.20 and .50). Additionally, SGI efficiency (precision of standard errors) was relatively better than the efficiency of FIML and CMI, as missing data rates increased. This latter finding raises new questions about the use of SGI that will need to be addressed in future work. Some of the ways in which researchers could examine this question are discussed further in the future directions section below.

Implications. In a general sense, this study sought to provide intervention researchers with information about the possible impacts that attrition, and missing data more generally, can have on biasing conclusions they make about whether an intervention is found to be effective or not. Previous results from a meta-analysis of RCT studies indicates that intervention researchers are often not using any of these
recommended methods, instead relying on listwise deletion or other ad hoc methods (Bell et al., 2014). As such, this study aims to contribute to the general field of intervention research by communicating the importance of using these methods, and by providing practical information to help guide researchers in finding which of the recommended methods may be best for them to use, based on the unique conditions of their study.

Promoting the use of these methods is particularly important for researchers conducting RCTs, because results of these studies are often utilized when making high stakes policy decisions about whether an intervention should be expanded and implemented on a larger scale, or instead discarded. Given the high stakes associated with conducting interventions, it is particularly important to take great care in the planning, implementation, and analysis phases of these studies. Decisions made during each of these phases contributes to increasing the quality of these studies and subsequently reducing the likelihood of making a type 1 or type 2 statistical error. Using the recommended missing data methods examined here may help to minimize bias in intervention tests of effectiveness, thus taking us a step further in increasing our confidence that the programs that are funded to be implemented within communities will be the ones that are best suited to the problem at hand.

**Future Directions.** There are several important future directions of Study 3 that I plan to examine as next steps in this line of research. As is described in the discussion section of Study 3, the enhanced efficiency of SGI, relative to the other methods tested,
raises several new questions about the use of this method in the context of intervention research. It’s possible to interpret this finding as a positive feature of the method (that it is resistant to the typical reduction in precision that occurs with increasing missing data), or as a negative feature of the method (i.e., that imputing at the subgroup level limits its ability to retain the range of variance of the original data). In order to help determine which explanation is true, I plan to design a simple follow-up simulation study that compares (1) the average variance of the complete samples, (2) the average variance of the samples once missing data has been removed, and (3) the average variance of the samples once the data have been imputed (as is reported in Study 3). Comparing the variance at each of these stages of the simulation will help to determine if the SGI imputation process is retaining the precision of the data as missing data increases or if the precision is caused by reduced variance in the samples. In a similar vein, comparing the imputation results derived using the predictive mean matching imputation technique (as is used in Study 3) to other acceptable imputation methods, could illuminate if the increased SGI efficiency is an artifact of the study design and the imputation method used.

Another future direction I plan to take extends the work I originally proposed in this dissertation comparing FIML, CMI, and SGI in determining an intervention’s effectiveness using percent rejection rate. Percent rejection rate represents the percentage of samples that found a statistically significant intervention impact and was an additional performance measure I originally hypothesized would differ among each
of these three methods. When analyzing this aspect of my proposal, results were contrary to what was expected and indicated that SGI’s percent rejection rate increased as missing data increased, particularly in the small and medium intervention impact effect size and small sample size conditions. Interested readers can find the expanded Study 3 results that include percent rejection rate as an additional performance measure in Appendix B, Tables B.1-B.3. Percent rejection rate as a performance measure was removed from Study 3 until it can be studied further in the context of intervention simulations for two reasons. First, because to the best of my knowledge, no researchers have utilized it in this context. In the past, it has been used to compare structural equation modeling simulations comparing model fit (e.g., Enders & Bandalos, 2001). Second, because this finding could actually encourage researchers to lose participants as a means of increasing the likelihood of finding a significant intervention impact using the SGI method. The future directions outlined above regarding SGI efficiency may also help me to elucidate why the SGI percent rejection rate increases as missing data increases.

A second line of inquiry I originally proposed sought to determine whether increasing the number of imputations for SGI and CMI demonstrably changed the power and precision of detecting significant intervention impacts. This secondary exploratory research question was primarily derived from previous work that has demonstrated that up to 20 imputations may be needed for imputed estimates to match the power and precision of estimates derived using FIML (Bodner, 2008; Graham et al., 2007).
Examining the results of this second research question yielded little differences across estimates using either CMI or SGI when increasing from 5 imputations to 10 and 20 imputations. See Tables B.4-B.12 in Appendix B for these specific results by intervention effect size, missing data rate, and sample size. Even though no demonstrable differences were found in the current study, future work should continue to examine whether and under what conditions the number of imputations may matter in the context of intervention impacts.

An additional future direction of this research should focus on taking the information derived from this study to generate a formal guide for researchers to use to help them determine which of these recommended methods is best suited to their study. To provide an example of how a guide would be useful, Table 5.1 presents de-identified intervention impacts on four outcomes (variables labeled X1 through X4) from an RCT study with 58 participants. Each of the four outcomes were examined using all three missing data methods tested in Study 3, namely, full information maximum likelihood (FIML), combined multiple imputation (CMI), and separate group imputation (SGI). Each standardized regression estimate in the table represents the intervention impact on the outcome, controlling for pre-test, and two dichotomous predictors. This study had approximately 20% missing data, which requires the use of a missing data method to align with best practice.

Table 5.1 reveals differences in results derived using each of these missing data methods. Given these differences, how should researchers decide which method to
use? Results of Study 3 suggest that researchers should not utilize CMI, as it has been found to result in higher levels of bias as compared to SGI and FIML. In this example, this decision is actually in the researcher’s favor, as CMI found the least statistically significant intervention impacts. However, before researchers are tempted to conduct analyses using both FIML and SGI methods and then pick the method that depicts the intervention most favorably (i.e., FIML in this case), it is important to remember that using dichotomous predictors with missing data violates an assumption of FIML. As such, it would be best for researchers to conduct these analyses in practice using SGI and to still discuss impacts with some degree of caution, given the small sample size of the study.

It is important to reiterate that this example utilized all methods to illustrate that results can differ in real data, just as they do in the data simulated and reported in Study 3, and that researchers are not expected nor encouraged to analyze their data using more than one of these methods. Developing a guide that builds on the recommendation table presented in Study 3 (see Table 4.4) – to include the assumptions and more nuanced recommendations of these methods to a greater extent – may help researchers determine the method that best fits their analysis plan and study conditions prior to analyzing their data. Additionally, researchers should remember when using all of these methods, that statistical significance is only one facet of intervention evaluation and that significance tests should always be supplemented
with an effect size that contextualizes the effect (the standardized regression estimates in Table 5.1 also serve this purpose; Hubert & Wainer, 2011).

Table 5.1

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th>SGI</th>
<th>FIML</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>.10</td>
<td>.18†</td>
<td>.18†</td>
</tr>
<tr>
<td>X2</td>
<td>.19</td>
<td>.23†</td>
<td>.25*</td>
</tr>
<tr>
<td>X3</td>
<td>.24*</td>
<td>.26*</td>
<td>.21*</td>
</tr>
<tr>
<td>X4</td>
<td>.24†</td>
<td>.32*</td>
<td>.35**</td>
</tr>
</tbody>
</table>

†p < .10; *p < .05; **p < .01.

Note. Estimates are standardized regression betas of the intervention impact. All models include pre-test and two dichotomous covariates. CMI = Combined Multiple Imputation, SGI = Separate Group Imputation; and FIML = Full Information Maximum Likelihood.

Future Directions

There are numerous ways in which the work of this dissertation and the field of MBIs for teachers generally can be improved in future research. Given the broad scope of this dissertation, the future directions section below has been broken into two sections: (1) next steps that I would like to take in my own work, and (2) additional future directions for the field.

Personal Future Directions

Beyond the future directions discussed specifically for each study, there are several additional areas of research within this field that I would like to address in my future work. Specifically, three additional extensions that I would like to focus on next are: (1) to examine possible MBI impacts on teachers’ classrooms using the calm, clear,
kind mindful teacher conceptualization; (2) to combine the theory and methods from Studies 2 and 3 and utilize missing data methods to address the difficulty in obtaining mindfulness practice reports from MBI participants; and (3) to examine moderated impacts of MBI interventions to better understand differential impacts of these programs on specific subgroups of teachers.

Examining Classroom Outcomes. An additional theoretical extension that requires future work in the field of MBIs for teachers is examining whether intervention impacts extend into the classroom context. As is described in Study 1, several studies have examined the more distal impacts of MBIs on teachers’ classrooms using the CLASS observation tool. Specifically, Jennings et al., (2017, 2011) and Flook et al., (2013) have begun examining changes in classrooms due to MBI training for teachers, however, no studies are known to have examined classrooms with any other observational measures of classrooms. As such, future work is still needed in this area, to examine if and how MBIs for teachers translate into positive changes in the classroom context.

Given the lack of research in this area, it is important to continue examining possible program impacts on teachers’ classrooms and to develop observational tools designed to measure teachers’ mindfulness in the classroom. One way in which this gap in the literature is being addressed is through current work being conducted by colleagues of mine. Over the last several years my colleagues have developed the Calm, Clear, Kind theoretical framework, which seeks to define and measure observable manifestations of teachers’ mindful behaviors in the classroom (see Rickert et al., 2019;
Taylor et al., 2020). Specifically, this framework defines a mindful teacher by three categories of behaviors (and their antitheses) in the classroom, namely being calm (and not reactive) during stressful moments in the classroom; clear-minded (and not distracted) with an ability to stay focused in the face of classroom disorganization or disruption; and kind (and not critical) through supportive and compassionate words and actions towards students (Rickert et al., 2019).

This measure has been validated (see Rickert et al., 2019) and conceptualized (see Taylor et al., 2020) in previous work and shows great promise in terms of contributing to this next exciting step in the field of MBIs for teachers. However, no research has yet examined possible intervention impacts using this framework and measure. As such, my goal is to examine possible intervention impacts on teachers’ calm, clear, and kind behaviors, using a sample of 58 middle school teachers that participated in a recent MBI RCT. This study will help to inform the field by (1) adding to the empirical literature examining whether MBI impacts extend into the classroom context and (2) utilizing a new and well-defined multi-informant methodology to examine these possible impacts.

**Missing Data Methods and Mindfulness Practices.** Another area of research I would like to contribute to in my future work bridges the work I’ve done examining missing data methods with work I’ve done studying mindfulness practices in MBIs for teachers. One issue being discussed regarding the study of mindfulness practices in the current field is how to deal with large amounts of missing data associated with practice
frequency logs that are never handed in by participants. This issue is raised by Parsons and colleagues’ (2017) in their meta-analytic review of mindfulness practice in MBI studies. In particular, MBI studies suffer several limitations relating to missing data of mindfulness practice reports, namely, underreporting of low levels of practice (where participants who do not adhere to practice requirements are the least likely to turn in reports) and an inability to obtain reports from participants who drop out of the study prior to completion (where researchers cannot compare practice frequency reports of “completers” versus “non-completers” of the program; Parsons et al., 2017). These issues can cause researchers to have an incomplete picture of the role of practice, and depending on how the issue is managed, can bias analyses examining practice-to-outcome relations. For instance, if researchers utilize listwise deletion and only test practice-to-outcome relations using participants who handed in their practice frequency logs, researchers are likely to overestimate how often the entire sample of participants engaged in mindfulness practices (i.e., underreporting of low practice frequency). In contrast, if researchers attempt to include all participants who provide any practice data (albeit incomplete) by filling in the missing data with “no practice” values (a method utilized by Carmody & Baer, 2008), researchers are likely to underestimate how often this larger sub-sample engaged in mindfulness practices.

Both of these measurement issues relate to missing data and could possibly benefit from missing data theory and practice. First, it would be useful to set up a protocol and recommendations for researchers describing how to determine the
underlying missing data mechanism (e.g., missing at random or missing not at random) associated with their missing data from participants who do not hand in practice frequency logs. While it is impossible to truly distinguish between missing at random and missing completely at random mechanisms, researchers can determine whether their data are not missing at random. Specifically, researchers can create a missing or not missing (1/0) dummy-coded variable in their dataset and then use regression methods to examine if the missing data is associated to any baseline or demographic variables of participants (Enders, 2010). Second, knowing the underlying missing data mechanism would allow researchers to proceed forward in using a recommended missing data method, and their practice-to-outcome analyses, using their entire sample and in a manner that is least likely to perpetuate biased information about mindfulness practice frequency. Conducting analyses in this way, coupled with improvements in the ways in which we measure mindfulness practice, may lend themselves well to addressing some of the current limitations in this field.

**Moderated Intervention Impacts.** Another important next step in the field of MBIs for teachers is examining moderated intervention impacts, or questions regarding the types (subgroups) of teachers that may benefit most from the program. Understanding what makes a program most effective for specific individuals is a powerful tool that can be utilized to help refine programs and tailor them to specific subgroups’ needs (Mark & Lenz-Watson, 2011). In the field of MBIs for teachers, this area of inquiry could take many forms. For instance, differences in MBI benefits found
comparing early and later career teachers may indicate that teachers at different stages of their career need differential supports emphasized in their MBI. For example, early career teachers may benefit more from a program that emphasizes building mindfulness-related skills and classroom management strategies rooted in mindfulness techniques. In contrast, later career teachers, who may already have a strong foundation of skills and strategies that they have gained through years of experience, may benefit more from a program that focuses on recapturing the joy of teaching or cultivating compassion for themselves and their students.

While the above examples represent ways in which moderation analyses can help inform how MBI programs can be made more effective, it is equally true that moderation tests can be used to inform possible negative side effects of these programs for select subgroups as well. Mark and Lenz-Watson (2011) highlight this less-emphasized use of moderation, describing the importance of first measuring possible negative outcomes and then also examining whether certain subgroups are disproportionately impacted on those outcomes. Understanding the possible negative side effects of program participation would help inform revisions and improvements to programs in the future and provide evidence to support policy decisions focused on terminating programs that have severe unintended consequences, that would otherwise have gone unnoticed (Bickman & Reich, 2015).

I plan to use moderation in my own work as a next step of Study 2, examining the role of teachers’ frequency of mindfulness practice in the development of
mindfulness-related skills. Moderation questions will help me to better understand whether a certain demographic of teachers (e.g., early career, later career) or teachers with specific qualities (e.g., those with high stress or previous experience with mindfulness) are most likely to engage in mindfulness practices regularly throughout the program. Additionally, I also think it will be extremely important to examine which type(s) of teachers continue to use mindfulness practices after the intervention is over, to better understand the long-term benefits of these programs. Understanding program participation at this more nuanced level will help to uncover what motivates teachers to engage in mindfulness practices regularly and subsequently what can be adjusted in these programs to enhance motivation and participation, to maximize benefits for teachers.

**Additional Future Directions for the Field**

Outside the scope of my own current work and next steps, there are several other future directions for the field that should be mentioned. Although this is not comprehensive, two key areas of extension that are important to discuss are: the measurement of fidelity of implementation, and the triangulation of methods to supplement evidence found through RCT studies.

**Fidelity of Implementation.** As is described in Study 1, more implementation science work is needed in the field of MBIs for teachers generally. This includes increasing the measures used in implementation areas such as program quality, adherence, and dose, as well as participant responsiveness (see Berkel et al., 2011).
However, I think an even greater emphasis should be given to the adaptation, monitoring, and differentiation areas of fidelity of implementation in this field. To review, adaptation refers any additions or changes made to the program during implementation; monitoring refers to the documentation of experiences and services provided to both treatment and control groups that is distinct from but related to intervention content; and differentiation describes the uniqueness of the intervention being tested in comparison to other similar programs (Berkel et al., 2011). These areas may be particularly important to study in the context of mindfulness interventions for several reasons. First, adaptation, although referring to potential changes that are made throughout a study, may be a valuable tool for program improvement. Specifically, documentation of the ways in which implementors tailor program content to enhance the relevance of the program to participants may inform progressive changes to the program in the future (Berkel et al., 2011). In many ways, documenting adaptation in this way could be an intriguing and creative way to enhance the relevance and effectiveness of MBI programs for teachers. And, even if changes are not synonymous with high quality and attuned facilitation, negative adaptations that are documented can be used to contextualized null or negative impacts of the intervention as well.

Monitoring both treatment and control groups’ experiences is another method that can help researchers to understand their intervention and potential unexpected study findings. For instance, participants who are randomized to the control group may not be satisfied that they will not be able to participate in the program until the
research study is over, and so they may seek out experiences in the community that are similar to the intervention being tested. If treatment contamination occurs in this way, and researchers have not asked questions throughout the study about these experiences, researchers may be more likely to make a type 2 statistical error and conclude their intervention doesn’t work. This may be particularly true for mindfulness research, given its continued rise in popularity and participants’ assumed interest in mindfulness due to the voluntary nature of these studies.

Monitoring the treatment group, although less often done, is equally important (Berkel et al., 2011). If an intervention is found to be effective, but participants in the treatment group have had experiences similar to the intervention being tested during the intervention phase, researchers may not know which experiences were the active ingredients responsible for causing the impacts of their study. For instance, if teachers in the treatment group of an MBI study were simultaneously provided with a new mindfulness curriculum to implement in their classrooms, the effect of mindfulness practice as part of the intervention and the effect of the new curriculum would be confounded. Measuring relevant “intervention-like” experiences can help researchers to maintain internal control of their study, by producing variables that allow researchers to statistically control for these outside factors, or to at least descriptively contextualize possible confounds of their impacts.

Additionally, given the rise in popularity of mindfulness, it will be incredibly important for the field of MBIs to begin understanding how each program being tested
differs from the others. What defines a mindfulness intervention can vary extensively from study to study. For instance, MBIs can range in duration from three days to three months and can be administered in a traditional “retreat” format, completed online, or completed using a phone application (Creswell, 2017). This significant variation in program type and content leads to various additional questions that may be best answered from studying the fidelity of implementation of these programs. For instance, is three days of mindfulness practice a sufficient dose for participants to see positive outcomes? Is a three-month retreat only beneficial for a select type of participant who is willing to undertake such a time-demanding process?

Many of these questions regarding program differentiation can be answered as researchers increasingly measure the other components of fidelity of implementation. For instance, requiring researchers to describe their intervention components in more detail in publications will allow programs to begin to be differentiated regarding each program’s definition of mindfulness, program dose, and the core components that are conceptualized as the active ingredients (e.g., which mindfulness practices are being taught). Additionally, increasing the measurement of other facets of fidelity of implementation, such as program quality, adaptation, and participant responsiveness, will be important in helping researchers to distinguish between low- and high-quality mindfulness interventions. In sum, studying fidelity of implementation in MBIs in the future can shift the field forward in categorizing MBIs at a more nuanced level, rather than as a catch-all phrase encompassing a variety of programs with varying degrees of
effectiveness. This work will be an extremely useful next step in understanding what types of mindfulness programs work, for whom, and under what conditions.

Lastly, a larger consideration still needs to be made regarding where (in terms of location) the fidelity of implementation information of an RCT study should be discussed. Should implementation information be required in RCT impacts write-ups for quality control purposes and function similarly to baseline equivalence tests? In contrast, should entire supplementary articles be published that focus specifically on a study’s implementation? In many ways, a study’s quality of implementation will determine the results of their study, however, there is no clear practice in place as to whether these different fields of research should be presented together or separately. If these interrelated aspects of a study remain separate, it may be useful to develop a system that links the separate published articles together to help readers obtain both components of a study. For example, in my own systematic review of MBIs for teachers reported in Study 1, I discuss a need for additional fidelity of implementation work in the field. This conclusion may be incorrect, if fidelity data is merely underreported in impacts papers in this field and instead published in a separate article. Unfortunately, as long as implementation science remains a separate field, this limitation remains in assessing this component of quality within RCT studies.

**Triangulation of Methods.** All research designs have strengths and weaknesses. The goal of research should always be to use a variety of methodologies, with strengths that address the weaknesses of others, in order to triangulate in on the true phenomena
in which we seek to study. Similar to all research designs, RCTs suffer from a variety of limitations that are innate to the design itself. For instance, the strong internal control of the design has the unintended consequence associated with it that reduces its generalizability to situations outside of the intervention. Similarly, while being the optimal choice for answering the ‘does it work’ question, the RCT design often falls short in answering process-oriented questions and contextualized questions that are often more interesting and rich—both theoretically and practically. Additionally, in the attempt to maintain causal inference, RCT designs require researchers to ignore the experiences participants actually have (i.e., whether they actually attend the program doesn’t matter), instead basing all analyses at the point in which participants were randomized to groups (Skinner et al., 2019). Given these limitations, and despite the fact that much of this dissertation has focused on improving the quality of studies using the RCT design, future work should also focus to a greater extent on understanding the effects of MBI programs for teachers using other methodological designs.

Several ways in which this process of triangulation could be enhanced in future work could be to supplement MBI RCT studies with (a) quasi-experimental studies that are process-oriented, (b) creative RCT studies that focus on examining unique effects of program components within the intent-to-treat framework, and (c) qualitative studies that are focused on generating and refining the theory of MBI programs. One type of quasi-experimental study that may complement RCT studies particularly well is a longitudinal design that has been referred to as “confirmatory program evaluation”
(Reynolds, 1998) or “theory-based evaluation” (Shadish et al., 2002). This design seeks to illuminate the causal processes or explanatory mechanisms of a program by clearly delineating a program’s theory of change prior to program implementation. Similar to the theoretical framework presented in Study 1 examining MBIs for teachers, this design would test each link within the theoretical framework longitudinally in a single study that only follows a treatment group (no counterpart is used as a control). This design is thought to increase in its ability to claim causal impacts as linkages along the theory of change are supported empirically (Reynolds, 1998).

There is no reason this theory driven evaluation method cannot be conducted using the RCT framework as well (Shadish et al., 2002), despite the fact that adding a comparison group to this longitudinal design would require a substantial increase in the number of resources needed to implement the study. Given these practical limitations, it may be best to utilize this “theory-driven design” quasi-experimentally, and to focus on other more feasible ways to creatively utilize the intent-to-treat framework that gets closer to answering questions about causal mechanisms. For instance, several recent studies have been conducted in the greater field of MBIs that creatively decompose MBI trainings to examine unique effects of specific active program components. For example, Hunt and colleagues (2018) sought to test the unique effects of various components of a mindfulness intervention by randomly assigning participants to a mindfulness-only condition, a yoga-only condition, and a combined yoga and mindfulness condition; and compared these three intervention conditions to an active
control group and an inactive control group. Assigning participants to a “full dose” condition (i.e., the combined yoga and mindfulness condition), several “half dose” conditions, and different control groups, allows researchers to better understand what these different features of MBIs (namely, mindfulness practices, movement practices, and intentional time to de-stress generally) have on program outcomes.

Similarly, Kropp and Sedlmeier (2019) designed a study that examined the unique effects of specific mindfulness practices on program outcomes by randomly assigning participants with no prior meditation experience to (1) a breathing meditation only condition, (2) a body scan only meditation condition or, (3) a loving-kindness only mediation condition. Creatively designed studies such as these have the potential to enhance our understanding of how mindfulness interventions work without sacrificing the high internal control that RCT designs offer.

Lastly, quantitatively focused quasi-experimental and experimental studies should both be supplemented with qualitative studies that serve to inform our theoretical understanding of mindfulness interventions for teachers. Qualitative methods, particularly those such as semi-structured interviews and focus groups, could be used for program improvement purposes to refine MBI programs to better fit the needs of teachers. Used another way, qualitative methods could be employed a priori with teachers in order to develop “bottom-up” MBI programs that are developed specifically aligned to teachers’ interests. Generally speaking, qualitative methods can provide researchers with unique theory-generating insights into their phenomena of
interest and allow researchers to collect rich data that celebrates the perspective of the participant in a manner that cannot be captured using quantitative methods (Hesse-Biber, 2017).

**Conclusion**

This dissertation sought to contribute to the current field of MBIs for teachers in a manner that extends our understanding of both the theory and methods. The overarching goal of this dissertation was to critically examine the current use of RCT studies in this field in order to better understand the ways in which we can continue moving towards producing the highest quality applied research studies serving to support teachers and students. However, the focus on RCT studies, and MBIs for teachers generally, is but one “bottom-up” approach to helping teachers better manage the many occupational demands associated with their jobs. This approach, while promising and well deserving of research, does not solve the “top-down” burden teachers face that gives rise to the need to support them in this way. Additional future work should therefore also focus on generating institutional-level changes that lift some of the burden off of teachers entirely, rather than focusing on providing supports to help them better manage unrealistic demands. If this could be done, MBIs for teachers could still function as unique professional development opportunities for teachers. Instead of focusing on reducing negative occupational experiences, MBI programs could shift to focusing predominantly on cultivating positive change in schools, as community spaces where teachers, administrators, and students collectively can all thrive.
References


https://doi.org/10.1037/a0018480

Davidson, R. J., Dunne, J., Eccles, J. S., Engle, A., Greenberg, M., Jennings, P., Jha, A.,


Hargreaves, A. (2000). Mixed emotions: Teachers’ perceptions of their interactions with
students. *Teaching and Teacher Education*, 16(8), 811–826.


https://doi.org/10.1016/j.tate.2017.01.015


https://doi.org/10.1037/edu0000187


https://doi.org/10.1037/a0026118


https://doi.org/10.1001/archpsyc.59.10.877


References


Neff, K. D. (2003). The development and validation of a scale to measure self-
References


teachers’ professional development: An emerging area of research and practice.

*Child Development Perspectives, 6*(2), 167–173.


https://doi.org/10.1007/s10902-008-9100-y


https://doi.org/10.1002/jclp.20237


https://doi.org/10.1177/1745691617709589


Appendix A – Study 3 Simulation Code

# Choles Dissertation Missing Data Paper Syntax
# The below syntax is complete for producing 5 imputations.
# If syntax for 10 or 20 imputations is desired, please contact the first author

# load libraries
#====================
library(mice)
library(miceadds)
library(lavaan)
library(knitr)

# set seed for replication purposes
#=====================
set.seed(1313)

# function to generate data
#=================================
exact_sample = function(n, mu=0, sigma=1, missing=F, mprop=m){
y = rnorm(n)  # generate size of sample wanted
# Obtain a true z score
y = y-mean(y)
y=y/sd(y)
# transform to desired distribution
y = y*sigma  # scales the sd
y = y+mu    # centers at new value
# handle missingness by hand
if(missing){
  if (mprop >1){
    mprop = mprop/100 # makes missing data a percentage
  }
  # proportion of missingness modeled
  msize = round(length(y)*mprop)
  # randomly select that number of values to remove
  midx = sample(1:length(y), msize)
  # remove those values
  y[midx] = NA
}
return(y)
}
# create vectors for desired conditions to iterate through
#===========================================================
samp_ns = c(25, 50, 125) # sample sizes - represent half numbers of total sample (T/C)
samp_ds = c(.2, .5, .8) # effect sizes
samp_miss = c(10, 25, 50) # missingness rates
# variable for number of replications
num_reps = 1000

# Generate the data!
#===============================
for(i in 1:num_reps) { # loop this 1000 times 1:1000
  for(n in samp_ns) { # calling in vector of sample sizes
    for(m in samp_miss) { # calling in vector of missing data rates
      ctrl_fname = paste('REP_',i,'_CTRL_N',n,'_M',m,'.csv',sep='') # file name
      ctrl = cbind('condition'=0, 'outcome'=exact_sample(n, missing=T, mprop=m)) # create the sample
      write.csv(ctrl, file=ctrl_fname, row.names=F) # save file
      for(d in samp_ds){ # effect size stuff
        effsize = d*10
        trt_fname = paste('REP_',i,'_TRT_N',n,'_M',m,'_D',effsize,'.csv',sep='')
        trt = cbind('condition'=1, 'outcome'=exact_sample(n=n, mu=d, missing=T, mprop=m)) # trt sample
        write.csv(trt, file=trt_fname, row.names=F) # save files
      }
    }
  }
}

# function for combining T and C samples
#=======================================================
combine_datasets = function(fctrl,ftrt){
  # read in control file
  ctrl_dat = read.csv(fctrl)
  for(t in ftrt){ # loop over treatments
    temp_dat = read.csv(t) # read in treatment data
    combo = rbind(ctrl_dat, temp_dat) # bind them!!!
    # save under new name
    combo_name = sub('TRT','COMBO',t)
    # write out combined data
    write.csv(combo, file=combo_name, row.names=F) # save files
  }
}

# function for combining T and C samples
#==============================================

# combine T and C samples
#=======================================
# grab control file
# read parameter specs (n, d, miss)
# match to treatments over effect size ranges
ctrl_files = list.files(pattern='_CTRL_')
for(f in ctrl_files)
  # load the control file as a data.frame using read.csv()
  # manipulate the file name to get the treatment files as csv()
  new_name = sub('CTRL','TRT',strsplit(f,'\.[0-9]+')[[1]][1]) # swap CTRL for TRT
  trt_files = list.files(pattern=new_name) # grab corresponding treatment sims
  combine_datasets(f,trt_files)

# SGI Imputation
#=======================================
# relevant libraries needed
#=======================================
library(mice)
library(miceadds)

# imputing control group files
#=======================================
ctrl_files = list.files(pattern = '_CTRL')   # find all control files
for(f in ctrl_files){
  # load control file as a data.frame using .csv()
  ctrl = read.csv(f)
  # modify control file name using paste()
  SGI_name = paste('SGI_5_', f, sep='')
  # impute the data using mice
  SGI = mice(data=ctrl, m = 5, method = 'pmm')
  # isolate imputed data files
  cSGI = complete(SGI, action = 'long', include = TRUE)
  # save file
  write.csv(cSGI, file=SGI_name, row.names=F)
}
# imputing treatment group files
===================================================================
trt_files = list.files(pattern = '_TRT')  # find all treatment files
for(t in trt_files){
  # load treatment file as a data.frame using .csv()
  trt = read.csv(t)
  # modify treatment file name using paste()
  SGI_name = paste('SGI_5_', t, sep='')
  # impute the data using mice
  SGI = mice(data=trt, m = 5, method = 'pmm')
  # isolate imputed data files
  # updated our data organization here. From wide to long.
  # keeps the .imp and .id columns in the imputed set for analysis later.
  tSGI = complete(SGI, action = 'long', include = TRUE)
  # save file  ### figure out how to save a mids object
  write.csv(tSGI, file=SGI_name, row.names=F)
}

# Combining SGI data files
===================================================================

# subset to isolate SGI files

# use function for combining data files used above

# grab control file
# read parameter spects (n, d, miss)
# match to treatments over effect size ranges
ctrl_files = list.files(pattern='SGI_\d+_REP_|\d+_CTRL_')  # regexp!!!
for(f in ctrl_files){
  # load the control file as a data.frame using read.csv()
  # manipulate the file name to get the treatment files as csv()
  new_name = sub('CTRL','TRT',strsplit(f,'.')[[1]][1])  # swap CTRL for TRT
  trt_files = list.files(pattern=new_name)  # grab corresponding treatment sims
  combine_datasets(f,trt_files)
}

# Traditional MI Imputation
===================================================================
 ======
# grab combo file (all combined datafiles)
combo_files = list.files (pattern = '^REP_\d+_COMBO') # more than just one type of COMBO

for(c in combo_files){
    # load the file as a data.frame using .csv()
    combo = read.csv(c)
    # modify combo name using paste()
    mi_name = paste('MI_5_', c, sep = '')
    # impute treatment/control combined data files
    MI = mice(combo, m = 5, method = 'pmm')
    # isolate imputed data files
    cMI = complete(MI, action = 'long', include = TRUE)
    # save file
    write.csv(cMI, file = mi_name, row.names=F)
}

# MI Analysis

mi_files = list.files (pattern = 'MI_5') # specify N, Miss, ES (run 1000 samps per cell)
# preallocate a matrix to store results
mi_analysis_results.df = data.frame('file'=mi_files,
    'cell'=rep(NA, length(mi_files)),
    'est'=rep(NA,length(mi_files)),
    'se'=rep(NA,length(mi_files)),
    'tval'=rep(NA,length(mi_files)),
    'pval'=rep(NA,length(mi_files)),
    'rej'=rep(NA, length(mi_files)))
i = 1 # r starts counting at 1
for(f in mi_files){
    # load combined data files as mids object
    mi = as.mids(read.csv(f))
    # analyze data
    res <- with(data=mi, exp=lm(outcome~condition))
    # save analysis results
    sum_res <- summary(pool(res))
    # Adding p-value rejection column
    sum_res$rej = ifelse(sum_res$p.value < .05, 1, 0)
    # extract relevant pooled estimates
    est = as.numeric(sum_res$estimate[2])
    se = as.numeric(sum_res$std.error[2])
tval = as.numeric(sum_res$statistic[2])
pval = as.numeric(sum_res$p.value[2])
rej = as.numeric(sum_res$rej[2])

# table each cell's results, create column for average results per cell
mi_analysis_results.df[i, 'cell'] = strsplit(f,'COMBO_|\.')[[1]][2]
mi_analysis_results.df[i, 'est'] = est
mi_analysis_results.df[i, 'se'] = se
mi_analysis_results.df[i, 'tval'] = tval
mi_analysis_results.df[i, 'pval'] = pval
mi_analysis_results.df[i, 'rej'] = rej
i = i+1 # iterate
}

# SGI Analysis
#==================================
sgi_files = list.files (pattern = 'SGI_\d+_REP_\d+_COMBO') # specify N, Miss, ES (run 1000 samps per cell)
# preallocate a matrix to store results
sgi_analysis_results.df = data.frame('file'=sgi_files,
  'cell'=rep(NA, length(sgi_files)),
  'est'=rep(NA,length(sgi_files)),
  'se'=rep(NA,length(sgi_files)),
  'tval'=rep(NA,length(sgi_files)),
  'pval'=rep(NA,length(sgi_files)),
  'rej'= rep(NA, length(sgi_files)))

i = 1 # r starts counting at 1
for(f in sgi_files){
  # load combined data files
  # new attempt
  # crufty fix.
  dat = read.csv(f)
  dat = dat[order(dat$.imp,dat$.id),c('.imp','condition','outcome')]
  # sgi = as.mids(read.csv(f)) # load data as mids object
  sgi = as.mids(dat) # load data as mids object
  # analyze data
  res <- with(data=sgi, exp=lm(outcome~condition))
  # save analysis results
  sum_res <- summary(pool(res))
  # Adding p-value rejection column
  sum_res$rej = ifelse(sum_res$p.value < .05, 1, 0)
  # extract relevant pooled estimates
est = as.numeric(sum_res$estimate[2])
se = as.numeric(sum_res$std.error[2])
tval = as.numeric(sum_res$statistic[2])
pval = as.numeric(sum_res$p.value[2])
rej = as.numeric(sum_res$rej[2])

# table each cell's results, create column for average results per cell
sgi_analysis_results.df[i, 'cell'] = strsplit(f,'COMBO_\.|')[1][2]
sgi_analysis_results.df[i, 'est'] = est
sgi_analysis_results.df[i, 'se'] = se
sgi_analysis_results.df[i, 'tval'] = tval
sgi_analysis_results.df[i, 'pval'] = pval
sgi_analysis_results.df[i, 'rej'] = rej
i = i+1 # iterate

# FIML Analyses
#=======================================================================
==
fiml_files = list.files(pattern = '^REP_\d+_COMBO')

library(lavaan)

# preallocate a matrix to store results
fiml_analysis_results.df = data.frame('file'=fiml_files,
    'cell'=rep(NA, length(fiml_files)),
    'est'=rep(NA, length(fiml_files)),
    'se'=rep(NA, length(fiml_files)),
    'tval'=rep(NA, length(fiml_files)),
    'pval'=rep(NA, length(fiml_files)),
    'rej' = rep(NA, length(fiml_files)))

i = 1 # r starts counting at 1
for(f in fiml_files){
    # load combined data files
dat = read.csv(f)
    # analyze data
    f1 = 'outcome ~ condition'
lavf1 = sem(f1, data = dat, missing = 'ml')
    # save analysis results
    f1res <- parameterestimates(lavf1, se = TRUE)
    # Adding p-value rejection column
    f1res$rej = ifelse(f1res$pvalue < .05, 1, 0)
# extract relevant pooled estimates
est = f1res$est[1]
se = f1res$se[1]
tval = f1res$z[1]
pval = f1res$pvalue[1]
rej = f1res$rej[1]

# table each cell's results, create column for average results per cell
fiml_analysis_results.df[i, 'cell'] = strsplit(f, 'COMBO_' | '')[[1]][2]
fiml_analysis_results.df[i, 'est'] = est
fiml_analysis_results.df[i, 'se'] = se
fiml_analysis_results.df[i, 'tval'] = tval
fiml_analysis_results.df[i, 'pval'] = pval
fiml_analysis_results.df[i, 'rej'] = rej

i = i+1 # iterate

# Aggregation of Results for each missing data method
#==================================================================================

# MI
MI_Table = with(mi_analysis_results.df,
    aggregate(cbind(est, se, tval, pval, rej) ~ cell,
               FUN=mean))

# SGI
SGI_Table = with(sgi_analysis_results.df,
                 aggregate(cbind(est, se, tval, pval, rej) ~ cell,
                            FUN=mean))

# FIML
FIML_Table = with(fiml_analysis_results.df,
                  aggregate(cbind(est, se, tval, pval, rej) ~ cell,
                             FUN=mean))

#Adding an effect size column to each method's dataframe
#==================================================================================

# MI
MI_Table$ES = as.numeric(rep(c(.2,.5,.8),9))

# SGI
SGI_Table$ES = as.numeric(rep(c(.2,.5,.8),9))
# FIML
FIML_Table$ES = as.numeric(rep(c(.2,.5,.8),9))

# Calculating Raw Bias
#==============================================

# MI
MI_Table$rawbias = (MI_Table$est - MI_Table$ES)

# SGI
SGI_Table$rawbias = (SGI_Table$est - SGI_Table$ES)

# FIML
FIML_Table$rawbias = (FIML_Table$est - FIML_Table$ES)

# Calculating Standardized Bias
#==============================================

# MI
MI_Table$stdbias = (MI_Table$rawbias / MI_Table$se)

# SGI
SGI_Table$stdbias = (SGI_Table$rawbias / SGI_Table$se)

# FIML
FIML_Table$stdbias = (FIML_Table$rawbias / FIML_Table$se)

# save tabled results
#==============================================
write.csv(MI_Table, file="IMP5_MI_Table.csv", row.names = F)
write.csv(SGI_Table, "IMP5_SGI_Table.csv", row.names = F)
write.csv(FIML_Table, "IMP5_FIML_Table.csv", row.names=F)
Table B.1

Average Estimates from Small Effect Size (.20) Intervention Impact Simulations

<table>
<thead>
<tr>
<th></th>
<th>FIML</th>
<th></th>
<th></th>
<th></th>
<th>CMI</th>
<th></th>
<th></th>
<th></th>
<th>SGI</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
<td>Std. Bias</td>
<td>%</td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
<td>Std. Bias</td>
<td>%</td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>10% Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.29</td>
<td>.00</td>
<td>.01</td>
<td>0</td>
<td>.19</td>
<td>.29</td>
<td>-.01</td>
<td>-.02</td>
<td>0</td>
<td>.21</td>
<td>.29</td>
</tr>
<tr>
<td>N = 100</td>
<td>.20</td>
<td>.21</td>
<td>.00</td>
<td>.01</td>
<td>0</td>
<td>.19</td>
<td>.21</td>
<td>-.01</td>
<td>-.04</td>
<td>0</td>
<td>.20</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.20</td>
<td>.13</td>
<td>.00</td>
<td>.00</td>
<td>8</td>
<td>.19</td>
<td>.13</td>
<td>-.01</td>
<td>-.04</td>
<td>8</td>
<td>.20</td>
<td>.13</td>
</tr>
<tr>
<td>25% Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.32</td>
<td>.00</td>
<td>-.02</td>
<td>1</td>
<td>.17</td>
<td>.32</td>
<td>-.03</td>
<td>-.10</td>
<td>1</td>
<td>.20</td>
<td>.31</td>
</tr>
<tr>
<td>N = 100</td>
<td>.20</td>
<td>.23</td>
<td>.00</td>
<td>.00</td>
<td>2</td>
<td>.18</td>
<td>.23</td>
<td>-.02</td>
<td>-.10</td>
<td>2</td>
<td>.20</td>
<td>.22</td>
</tr>
<tr>
<td>N = 250</td>
<td>.20</td>
<td>.15</td>
<td>.00</td>
<td>.00</td>
<td>14</td>
<td>.19</td>
<td>.15</td>
<td>-.01</td>
<td>-.10</td>
<td>12</td>
<td>.20</td>
<td>.14</td>
</tr>
<tr>
<td>50% Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.37</td>
<td>.00</td>
<td>.00</td>
<td>2</td>
<td>.16</td>
<td>.35</td>
<td>-.04</td>
<td>-.13</td>
<td>2</td>
<td>.19</td>
<td>.32</td>
</tr>
<tr>
<td>N = 100</td>
<td>.19</td>
<td>.28</td>
<td>-.01</td>
<td>-.02</td>
<td>5</td>
<td>.16</td>
<td>.26</td>
<td>-.04</td>
<td>-.17</td>
<td>4</td>
<td>.19</td>
<td>.23</td>
</tr>
<tr>
<td>N = 250</td>
<td>.20</td>
<td>.18</td>
<td>.00</td>
<td>.00</td>
<td>13</td>
<td>.17</td>
<td>.17</td>
<td>-.03</td>
<td>-.19</td>
<td>12</td>
<td>.20</td>
<td>.15</td>
</tr>
</tbody>
</table>

Notes. Est. is the average sample derived estimate of the intervention, SE is the standard error of the estimate, raw bias is the difference between the population parameter (.20) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. FIML = full information maximum likelihood; CMI = combined multiple imputation; SGI = separate group imputation. Estimates for CMI and SGI generated using 5 imputations.
## Table B.2

*Average Estimates from Medium Effect Size (.50) Intervention Impact Simulations*

<table>
<thead>
<tr>
<th></th>
<th>FIML</th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
</tr>
<tr>
<td>10% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.50</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td>25% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.32</td>
<td>-.01</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.23</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>50% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.37</td>
<td>-.01</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.28</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.18</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.50) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. FIML = full information maximum likelihood; CMI = combined multiple imputation; SGI = separate group imputation. Estimates for CMI and SGI generated using 5 imputations.
## Table B.3
**Average Estimates from Large Effect Size (.80) Intervention Impact Simulations**

<table>
<thead>
<tr>
<th></th>
<th>FIML</th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
</tr>
<tr>
<td>10% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td>25% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.31</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.23</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>50% Missing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.82</td>
<td>.37</td>
<td>.02</td>
</tr>
<tr>
<td>N = 100</td>
<td>.81</td>
<td>.28</td>
<td>.01</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.18</td>
<td>.00</td>
</tr>
</tbody>
</table>

Note. Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. FIML = full information maximum likelihood; CMI = combined multiple imputation; SGI = separate group imputation. Estimates for CMI and SGI generated using 5 imputations.
### Table B.4
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .20, Missing Data = 10%

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>5 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.19</td>
<td>.29</td>
</tr>
<tr>
<td>N = 100</td>
<td>.19</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.19</td>
<td>.13</td>
</tr>
<tr>
<td>10 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.20</td>
<td>.29</td>
</tr>
<tr>
<td>N = 100</td>
<td>.19</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.19</td>
<td>.13</td>
</tr>
<tr>
<td>20 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.19</td>
<td>.29</td>
</tr>
<tr>
<td>N = 100</td>
<td>.19</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.19</td>
<td>.13</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
Table B.5
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .20, Missing Data = 25%

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>5 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.17</td>
<td>.32</td>
</tr>
<tr>
<td>N = 100</td>
<td>.18</td>
<td>.23</td>
</tr>
<tr>
<td>N = 250</td>
<td>.19</td>
<td>.15</td>
</tr>
<tr>
<td>10 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.17</td>
<td>.32</td>
</tr>
<tr>
<td>N = 100</td>
<td>.18</td>
<td>.23</td>
</tr>
<tr>
<td>N = 250</td>
<td>.19</td>
<td>.14</td>
</tr>
<tr>
<td>20 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.17</td>
<td>.31</td>
</tr>
<tr>
<td>N = 100</td>
<td>.18</td>
<td>.22</td>
</tr>
<tr>
<td>N = 250</td>
<td>.19</td>
<td>.14</td>
</tr>
</tbody>
</table>

Note. Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
### Table B.6
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .20, Missing Data = 50%

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>5 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.16</td>
<td>.35</td>
</tr>
<tr>
<td>N = 100</td>
<td>.16</td>
<td>.26</td>
</tr>
<tr>
<td>N = 250</td>
<td>.17</td>
<td>.17</td>
</tr>
<tr>
<td>10 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.16</td>
<td>.35</td>
</tr>
<tr>
<td>N = 100</td>
<td>.16</td>
<td>.25</td>
</tr>
<tr>
<td>N = 250</td>
<td>.17</td>
<td>.17</td>
</tr>
<tr>
<td>20 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.16</td>
<td>.34</td>
</tr>
<tr>
<td>N = 100</td>
<td>.15</td>
<td>.25</td>
</tr>
<tr>
<td>N = 250</td>
<td>.17</td>
<td>.16</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
### Table B.7
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .50, Missing Data = 10%

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td><strong>5 Imputations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.30</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.13</td>
</tr>
<tr>
<td><strong>10 Imputations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.29</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.13</td>
</tr>
<tr>
<td><strong>20 Imputations</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.49</td>
<td>.29</td>
</tr>
<tr>
<td>N = 100</td>
<td>.50</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.50</td>
<td>.13</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
### Table B.8
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .50, Missing Data = 25%

<table>
<thead>
<tr>
<th>Imputations</th>
<th>N = 50</th>
<th>N = 100</th>
<th>N = 250</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>.46</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>SE</td>
<td>.32</td>
<td>.23</td>
<td>.14</td>
</tr>
<tr>
<td>Raw Bias</td>
<td>.04</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Std. Bias</td>
<td>.12</td>
<td>.08</td>
<td>.00</td>
</tr>
<tr>
<td>% Reject</td>
<td>15</td>
<td>56</td>
<td>99</td>
</tr>
<tr>
<td>SGI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>.49</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>SE</td>
<td>.31</td>
<td>.22</td>
<td>.14</td>
</tr>
<tr>
<td>Raw Bias</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Std. Bias</td>
<td>.04</td>
<td>-.02</td>
<td>-.02</td>
</tr>
<tr>
<td>% Reject</td>
<td>27</td>
<td>62</td>
<td>90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Imputations</th>
<th>N = 50</th>
<th>N = 100</th>
<th>N = 250</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>.46</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>SE</td>
<td>.32</td>
<td>.23</td>
<td>.14</td>
</tr>
<tr>
<td>Raw Bias</td>
<td>.04</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Std. Bias</td>
<td>.12</td>
<td>.07</td>
<td>.00</td>
</tr>
<tr>
<td>% Reject</td>
<td>16</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>SGI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>.49</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>SE</td>
<td>.31</td>
<td>.22</td>
<td>.14</td>
</tr>
<tr>
<td>Raw Bias</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Std. Bias</td>
<td>.04</td>
<td>-.02</td>
<td>-.01</td>
</tr>
<tr>
<td>% Reject</td>
<td>27</td>
<td>64</td>
<td>91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Imputations</th>
<th>N = 50</th>
<th>N = 100</th>
<th>N = 250</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>.46</td>
<td>.48</td>
<td>.50</td>
</tr>
<tr>
<td>SE</td>
<td>.32</td>
<td>.23</td>
<td>.14</td>
</tr>
<tr>
<td>Raw Bias</td>
<td>.04</td>
<td>.02</td>
<td>.00</td>
</tr>
<tr>
<td>Std. Bias</td>
<td>.12</td>
<td>.07</td>
<td>.00</td>
</tr>
<tr>
<td>% Reject</td>
<td>15</td>
<td>62</td>
<td>100</td>
</tr>
<tr>
<td>SGI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Est.</td>
<td>.49</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>SE</td>
<td>.31</td>
<td>.22</td>
<td>.14</td>
</tr>
<tr>
<td>Raw Bias</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Std. Bias</td>
<td>.04</td>
<td>-.02</td>
<td>-.01</td>
</tr>
<tr>
<td>% Reject</td>
<td>27</td>
<td>64</td>
<td>90</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
Table B.9
Comparing CMI and SGI Estimates Across Imputation Levels (Effect Size = .50, Missing = 50%)

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th></th>
<th>SGI</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
<td>Std. Bias</td>
</tr>
<tr>
<td>5 Imputations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.42</td>
<td>.36</td>
<td>.08</td>
<td>.23</td>
</tr>
<tr>
<td>N = 100</td>
<td>.44</td>
<td>.27</td>
<td>.06</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.49</td>
<td>.16</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>10 Imputations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.42</td>
<td>.35</td>
<td>.08</td>
<td>.22</td>
</tr>
<tr>
<td>N = 100</td>
<td>.45</td>
<td>.26</td>
<td>.05</td>
<td>.19</td>
</tr>
<tr>
<td>N = 250</td>
<td>.49</td>
<td>.16</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>20 Imputations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.42</td>
<td>.35</td>
<td>.08</td>
<td>.23</td>
</tr>
<tr>
<td>N = 100</td>
<td>.45</td>
<td>.26</td>
<td>.05</td>
<td>.21</td>
</tr>
<tr>
<td>N = 250</td>
<td>.49</td>
<td>.16</td>
<td>.01</td>
<td>.07</td>
</tr>
</tbody>
</table>

Note. Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
## Table B.10

*Comparing CMI and SGI Estimates Across Imputation Levels — Effect Size = .80, Missing Data = 10%*

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th></th>
<th></th>
<th></th>
<th>SGI</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
<td>Std. Bias</td>
<td>% Reject</td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
</tr>
<tr>
<td><strong>5 Imputations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
<td>.00</td>
<td>99</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
<td>-.01</td>
<td>100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
<td>-.01</td>
<td>100</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td><strong>10 Imputations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
<td>.00</td>
<td>100</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
<td>.00</td>
<td>100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
<td>-.01</td>
<td>100</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
</tr>
<tr>
<td><strong>20 Imputations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.80</td>
<td>.29</td>
<td>.00</td>
<td>.00</td>
<td>99</td>
<td>.81</td>
<td>.29</td>
<td>-.01</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
<td>-.01</td>
<td>100</td>
<td>.80</td>
<td>.21</td>
<td>.00</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
<td>-.01</td>
<td>100</td>
<td>.80</td>
<td>.13</td>
<td>.00</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
### Table B.11
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .80, Missing Data = 25%

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th></th>
<th>SGI</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Raw Bias</td>
<td>Std. Bias</td>
</tr>
<tr>
<td>5 Imputations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.79</td>
<td>.32</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.23</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.14</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>10 Imputations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.79</td>
<td>.32</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.22</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.14</td>
<td>.00</td>
<td>.01</td>
</tr>
<tr>
<td>20 Imputations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.79</td>
<td>.32</td>
<td>.01</td>
<td>.04</td>
</tr>
<tr>
<td>N = 100</td>
<td>.80</td>
<td>.22</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.14</td>
<td>.00</td>
<td>.02</td>
</tr>
</tbody>
</table>

**Note.** Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.
Table B.12
Comparing CMI and SGI Estimates Across Imputation Levels – Effect Size = .80, Missing Data = 50%

<table>
<thead>
<tr>
<th></th>
<th>CMI</th>
<th>SGI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.76</td>
<td>.37</td>
</tr>
<tr>
<td>N = 100</td>
<td>.79</td>
<td>.26</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.16</td>
</tr>
<tr>
<td>10 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.76</td>
<td>.36</td>
</tr>
<tr>
<td>N = 100</td>
<td>.78</td>
<td>.26</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.16</td>
</tr>
<tr>
<td>20 Imputations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 50</td>
<td>.76</td>
<td>.36</td>
</tr>
<tr>
<td>N = 100</td>
<td>.78</td>
<td>.26</td>
</tr>
<tr>
<td>N = 250</td>
<td>.80</td>
<td>.16</td>
</tr>
</tbody>
</table>

*Note.* Est. is the average sample derived estimate of the intervention, SE is the standard error of that estimate, raw bias is the difference between the population parameter (.80) and the average sample derived estimate, standardized bias is raw bias reported in standard deviation units, and % reject is the percentage of samples that found a significant intervention effect (p < .05). Conditions with bias that exceeded .10 (1/10 of a standard deviation unit) are shown in bold type. CMI = combined multiple imputation; SGI = separate group imputation.