Electronic Theses and Dissertations, 2004-2019

2017

Barriers to Adoption of Wellness Programs: A Worked Example of an Augmented Best-Fit Framework Synthesis

Dalton Oliver
University of Central Florida

BARRIERS TO ADOPTION OF WELLNESS PROGRAMS: A WORKED EXAMPLE OF AN
AUGMENTED BEST-FIT FRAMEWORK SYNTHESIS

by

DALTON OLIVER
B.S. Auburn University, 2007
M.S. University of Central Florida, 2012

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Education
in the College of Education and Human Performance
at the University of Central Florida
Orlando, Florida

Spring Term
2017

Major Professor: Thomas Fisher

ABSTRACT

Background: Qualitative syntheses have the potential to offer a great deal of insight into complex problems of practice. However, their methods often appear unclear and warrant ongoing scrutiny by the research community.

Aim: This study introduces a novel combination of methods for synthesizing qualitative literature and explores the utility of these methods through a worked example of a real-world problem of practice.

Methods: Qualitative studies that investigated barriers to adoption of wellness programs through the perspectives of key informants were systematically collected for synthesis. Key informants were identified as decision makers at small- to medium-sized businesses. The primary method used in this study was the Best-Fit Framework Synthesis (BFS). The BFS was augmented with Alignment Scores, CERQual Analysis, and a novel Saturation of Inquisition Test. Dedoose software was used to support data analysis.

Results: The systematic search returned 4 studies that met the inclusion criteria. Diffusion Theory was systematically selected to develop a framework for analyzing qualitative findings. The synthesis generated four analytical themes and led to the development of a contextually rich conceptual framework. Analytical themes deeply informed the research questions while the framework offered a broader view of the overall problem. CERQual Analysis provided an added dimension of ranking amongst findings based on their level of confidence. The Saturation of Inquisition Test identified gaps in current research and validated decisions made during the synthesis. Alignment Scores identified specific points of misalignment and supported decision-making during the synthesis.

Conclusion: The augmented BFS was a valuable method for synthesizing qualitative findings in a manner that informs practitioners and builds on relevant theory. The additional methods integrated seamlessly with the original BFS while enhancing transparency, reliability, and practical value of the synthesis. Further replication and critical evaluation of the overall methodology and its individual components is warranted.

ACKKNOWLEDGMENTS

I wish to thank Sabrina Gordon, the external reviewer noted throughout this paper. Her timely

assistance was critical to this work. I'd also like to acknowledge the influential role Dr. Thomas

Fisher has played on the past 3 years of my professional career. I would not have achieved this

milestone without his mentorship. Lastly, I'd like to thank the members of my committee, Dr.

David Boote, Dr. Bonnie Swan, and Dr. Anna Valdes, for their support and critical feedback that

was integral to this paper.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF COMMON ACRONYMS or ABBREVIATIONS

ACA   Affordable Care Act

ADA   Americans with Disabilities Act

AMA   American Medical Association

BFS   Best-Fit Framework Synthesis

CDC   Center for Disease Control

CPR   Cardiopulmonary Resuscitation

DM   Decision Maker

DT   Diffusion Theory

EAP   Employee Assistance Program

EEOC   Equal Employment Opportunity Commission

ERISA   Employee Retirement Income Security Act

GINA   Genetic Information Nondiscrimination Act

HIPAA   Health Insurance Portability and Accountability Act

HP   Health Promotion

HR   Human Resources

HRA   Health Risk Assessment

SMB   Small- to Medium-Sized Business

WP   Wellness Program

CHAPTER ONE: INTRODUCTION

<u>Background</u>

A unique strength of qualitative research is its ability to analyze decision-making processes through the end-user's perspective. This perspective becomes valuable to researchers when investigating the rejection of evidence-based health and fitness strategies. This seemingly illogical rejection of such strategies becomes more apparent every year as preventable diseases continue to climb and cripple America's healthcare system. Qualitative research and syntheses thereof offer potentially valuable insight into this user-centric problem of practice. However, qualitative approaches are not without their limitations. Such work is often vulnerable to misinterpretation and may yield less than meaningful results if conducted or reported in a less than thorough manner. Therefore, the process of qualitative research—and especially syntheses thereof—warrants the utmost scrutiny, ongoing evaluation, and continuous refinement.

This paper pilot tests a novel combination of qualitative methods through a worked example of a synthesis of qualitative literature. Namely, this paper refers to this set of methods as an Augmented Best-Fit Framework Synthesis (A-BFS). The methods were selected based on their complementary effects and potential utility in health and fitness research.

The pilot test is set within the context of a highly relevant and timely topic in the health and fitness industry—wellness program (WP) adoption. The aim of the A-BFS is to uncover potential barriers to WP adoption amongst small- to medium-sized businesses (SMBs). The information to be analyzed and synthesized is the perceptions of decision makers (DMs).

The problem is thoroughly described throughout Chapters 1 and 2. The overview provided in these chapters offers a valuable point of reference when discussing the relatively

complex methodology in Chapters 3, 4, and 5. The overview also offers readers from other areas of research a better position from which to evaluate the value of any outcomes from the A-BFS.

Problem Statement

There is a discrepancy in WP adoption between smaller and larger businesses. Smaller businesses are less likely to have a WP in place than their larger counterparts (Claxton, Rae, Panchal, Whitmore, Damico, Kenward, & Long, 2015; Linnan, Bowling, Childress, Lindsay, Blakey, Pronk, & Royall, 2008; McCoy, Stinson, Scott, Tenney, & Newman, 2014; Reeves, 2015). Smaller businesses that do offer WPs tend to offer less comprehensive services than their larger counterparts (Linnan et al. 2008). This latter point may be equally as important as the former, considering that comprehensive WPs present a much higher likelihood of success than less-comprehensive versions thereof (Hersey, Williams-Piehota, Sparling, Alexander, Hill, Isenberg, Rooks, & Dunet, 2008; Merrill, Aldana, Vyhlidal, Howe, Anderson, & Whitmer, 2011; Pronk, 2014; Terry, Seaverson, Grossmeier, & Anderson, 2008; Zula, 2014). Relevant highlights of this discrepancy in WP adoption and service offerings are provided in APPENDIX A.

This discrepancy in WP adoption is exacerbated by the potential benefits of WPs. WPs have demonstrated potential as a viable solution to the preventable diseases that continues to plague America's healthcare system (Arena et al. 2013; Goetzel & Ozmlnkowski, 2008; Goetzel et al. 2014; Ye et al. 2013). Declining health and general physical capacity in older workers is also leading to a growing number of unplanned retirements (Pitt-Catsouphes, James, & Matz-Costa, 2015). Unplanned retirements, especially of senior employees, may have a greater effect on smaller companies due to the lower number of personnel to absorb the impact.

This discrepancy is also occurring despite repeated government-led initiatives in funding and promotion (Fielding, Kumanyika, & Manderscheid, 2013; Linnan et al. 2008). Some evidence points to international interests in policy reform that would support health across borders (Labonte, 2014). Federal- and state-allocated grants and tax incentives also exist to support WPs in businesses of all sizes (Anderko et al. 2012; Swords, 2014; Tjoa, Ling, Bender, Brittenham, & Jha, 2012). Such grants may be important to smaller companies that lack the advantage of economy of scale. Their smaller size may very well raise costs and limit their return on investment (ROI) decreasing their likelihood of adopting such programs (Baicker, Cutler, & Song, 2010). This is only one barrier among many that are likely playing a role in this discrepancy between smaller and larger businesses.

Despite such barriers, SMBs still demonstrate interest in such programs (Hughes, Patrick, Hannon, Harris, & Ghosh, 2011; Witt, Olsen, & Ablah, 2013). Furthermore, SMBs show significant promise at reaping the benefits from WPs (Hersey et al. 2008; Merrill et al. 2011; Newman et al. 2015). Small businesses also account for more than 50% of the private workforce in the U.S. (McDowell, 2010). With this potential for success and broad level of exposure, increasing adoption of WPs amongst SMBs could positively affect the health and economy of the country overall (McCoy, Stinson, Scott, Tenney, & Newman, 2014; Shepherd, 2016).

Simply put, the American population is becoming more likely to develop chronic disease (Roger et al. 2012; Smith et al. 2012), work later into the lifespan (Toosi, 2012), and pass potentially debilitating health costs on to the employer than ever before (Schultz, Chen, & Edington, 2009; Swords, 2014). Smaller businesses represent half of America's private workforce and are underserved regarding WPs. This discrepancy constitutes a substantial and timely problem warranting investigation.

<p style="text-align: center;"><u>Purpose</u></p>

Aim: This study explores the utility of a novel combination of methods at understanding barriers to adoption of wellness programs (WP)s.

Sample: Qualitative studies that interviewed key informants on their perceptions of WP adoption. Key informants are identified as DMs of SMBs with less than 1,000 employees.

<p style="text-align: center;"><u>Significance</u></p>

The context for this pilot test was prompted by a number of social and political trends. Specifically, government interest and support in wellness initiatives, such as *Healthy People 2020* (U.S. Department of Health and Human Services, 2017; U.S. Department of Labor, 2014); lagging adoption of WPs by smaller businesses (McCoy et al. 2014; Linnan et al. 2008); and growing evidence that supports the value of comprehensive wellness initiatives (Goetzel et al. 2014). Several reviews have investigated potential discrepancies in WP adoption (Kaspin, Gorman, & Miller, 2013; McCoy et al. 2014; Milat, King, Bauman, & Redman, 2012; Newman et al. 2015). However, few have attempted to translate perceived barriers amongst SMB DMs across studies in a manner that deeply and thoroughly informs practitioners (Catford, 2009). To accomplish this, this study aimed to systematically analyze DMs' views on barriers to adoption and test their transferability across contexts. If possible, evidence will be synthesized into themes that produce the deepest insight and most actionable solutions to date.

Directly investigating this phenomenon on a review scale is prudent and timely. The current body of evidence appears to be approaching a need for re-alignment; warrants a deeper probe into specific aspects of the problem; and would benefit from elimination of redundancy in the research. The surface questions facing adoption have been addressed, but analysis and

synthesis of those findings may better direct deeper inquisition. This *deeper dive* may supply

more actionable solutions for practitioners while testing the viability of a relatively novel

methodology. This latter outcome is the overarching goal of the paper; to produce a worked

example of an A-BFS for critical evaluation by qualitative researchers and methodologists.

Research Question

The overarching purpose of this paper was to explore the utility of a novel combination

of qualitative methods for synthesizing qualitative literature. To accomplish this in a thorough

and practical manner, two research questions were proposed to test the utility of the

methodology. Indeed, any methods employed are only as valuable as their propensity to solve a

given research question. The questions for this worked example are as follows:

- *What root barriers are hindering adoption of WPs amongst SMBs?*

- *Are some perceived barriers more universal, or substantial, than others?*

Definitions

Defining Small- to Medium-Sized Business

Business size is classified by employee number or revenue depending on the industry

sector. It's important to note that there is often a slight discrepancy in the literature regarding

how small, medium, and large businesses are defined. Some studies include companies of up to

1,000 employees when analyzing small businesses while others stop at 199. This review defines

SMBs as companies employing less than 1,000 people.

Conceptual Framework

A conceptual framework will be developed simultaneously with, but independently from, the literature search. This method will allow for optimal fit between framework and findings while further supporting the objective nature of the analysis. The *a priori* framework will also be augmented as a result of the synthesis, leading to a contextually rich *posteriori* framework. This Method is well-defined by Carroll, Booth, Leaviss, & Rick, (2013) as the Best-Fit Framework Synthesis (BFS). The BFS is the foundational methodology used in this paper. All other methods are intended to augment the highly valuable methods offered by Carroll et al (2013).

Assumptions

A few assumptions underlie this study. These assumptions are noted where relevant and thoroughly described in the Limitations section of this paper. Generally speaking, this paper assumes the following:

- the databases used to garner literature adequately reflects breadth of scientific literature,

- the literature accrued adequately reflects the scientific discoveries, and

- the reader is informed as to the basic intentions of a Dissertation in Practice.

Organization

This study opens with a literature review to deepen the reader's understanding of the problem being investigated by the A-BFS. While health and fitness professionals may find the content in Chapter 2 valuable, the literature review was written for an audience of researchers and methodologists. The purpose of Chapter 2 is to allow readers from various fields ample opportunity to identify with the given problem. In doing so, the reader may better appreciate the

value of any outcomes produced. The A-BFS methodology is then presented in a transparent and

thorough manner throughout Chapter 3. Results are then provided in Chapter 4 that offer

findings accrued during the study while explicitly noting the roles various methods played during

the synthesis. While health and fitness professionals may find these results useful, they are

intended to act as a worked example of the expected outcome from the A-BFS. Chapter 5 revisits

the process of the A-BFS and explores its underlying rationale. The discussion in Chapter 5

considers specific methods used in isolation as well as combination with one another.

Limitations are then offered to better describe parameters of the various methods and their

outcomes.

CHAPTER TWO: LITERATURE REVIEW

An initial review of literature was performed to provide contextual breadth and depth to the investigation. The literature review was predominantly guided by the researcher's personal and professional experience in the health and fitness industry. The literature review was organized according to the influential work of Beile and Boote (2005).

Specifically, the aim of this chapter is to a) state what is currently known about barriers to adoption of WPs; b) provide a thorough description of WPs and SMBs; and c) situate the problem in a historical and cultural context. An overarching aim is to provide a reference point for discussion on the utility of the A-BFS. The intended result is a more informed reader that can critically evaluate the relevance of any outcomes and the utility of this paper's methodology. The importance of this informed perspective on the part of the reader cannot be overstated, especially when referencing these contextual elements to describe the utility of the complex set of methods employed in Chapter 3.

Historical Context of Wellness Programs

WPs have proliferated and evolved over the past few decades as the result of cultural, political, and economic factors. More recent proliferation is likely due in part to the rising prevalence of chronic disease and disability that is attributed to the Obesity epidemic. For example, Type 2 Diabetes has increased substantially throughout the U.S. across all ages, ethnicities, and genders since 1988 (Menke, Casagrande, Geiss, & Cowie, 2015). The trend is relatively curvilinear, showing a gradual progression in the relative percentage of people affected. When considering the impact on individuals, families, and the American health care system, the trend becomes an important problem of practical significance. Furthermore, this

trend is occurring despite the often preventable nature of Type 2 Diabetes and is mirrored by

many other preventable diseases that arise from poor lifestyle and fitness behaviors.

<center>*Re-Branding of Employee Assistance Programs*</center>

Haaz, Maynard, Petricia, and Williams (2003) date organized and accredited employee

assistance programs (EAPs) as far back as the 1940s. From their remarks, it appears that such

programs likely placed more emphasis on alcoholism and mental health than fitness and physical

activity. The time-period itself may be a key indicator of the changing norms, denoting different

cultural ideals, problems, and/or opportunities. Such programs also likely varied in their

offerings based on their own users' needs and organizational goals. In any case, programs that

support employee wellness have likely existed at least as long as Haaz et al. (2003) stated—if not

longer'—and reflect a combination of the needs of the target audience and cultural influences.

Well after EAPs were officially established the ideologies and aims of such programs

appear to have morphed. Some evidence suggests that around 1980 EAPs began prioritizing

physical fitness (Howard & Mikalachki, 1979; Shepard, 1981; Shephard, 1983). Interestingly,

the goals of these WPs appear somewhat performance-based as opposed to health-based. In the

earlier research reviewed, health care does not appear to be a primary factor until more recently.

This makes superficial sense when comparing this potential shift in priorities to trends in health

care costs. With more than a third of Americans now suffering from Obesity and their employers

often paying the bill, it is logical that corporations integrate some method of mediation. Of note,

Obesity is now known to cause declines in health on its own as well as spurring various other

diseases and disorders. As such, Obesity was recently classified as a disease warranting medical

<center>9</center>

attention (Flegal, Carroll, Kit, & Ogden, 2012; Ogden, Carroll, Kit, & Flegal, 2012; Stoner & Cornwall, 2014).

Causes of Obesity are multifaceted, but predominantly linked to lifestyle choices that are readily modifiable. While some people may be more prone to the development of Obesity than others, most can manage or completely prevent it with proper nutrition and physical activity. This aspect becomes important because it implies that solutions exist yet these solutions are not diffusing amongst in-need populations. Together these complex problems point to a need for more user-centric, actionable research. Such research may improve the diffusion of preventive and corrective strategies across systems and populations.

<center>Defining "Wellness Program"</center>

"Wellness Program" is defined for this paper as *a formal program that supports health and wellness of a target organization's personnel*. However, it is important to note that WPs are far from standardized. Processes, offerings, and even definitions vary from system to system. This leads to significant potential of confounding nuances when attempting to measure, compare, or even standardize WPs. This leaves practitioners, researchers, and DMs with complex—and often conflicting—information. Such ambiguity does not lend itself well to simple decisions or actionable solutions. WP adoption is then left to the best guesses of DMs with limited objective information from which to base their decisions. This creates a context that warrants further investigation from many different realms and perspectives.

Evidence-based best practice is a key area of investigation amongst WPs. Such research is indeed vital to the effective standardization of WPs. Standardization of WPs would then streamline evaluation and, potentially, knowledge dissemination. However, the conceptualization

of best practices remains difficult for its own underlying reasons. For one, we see large amounts of data coming from surveys and studies that sample volunteers from given populations. Such studies in the health and fitness setting are notorious for types of selection bias. This appears so commonplace in the literature that the concept has earned its own line of contextually-based research (Li & Sung, 1999). This seems to stem from greater participation by healthy individuals than their less-healthy counterparts. In other words, health-related surveys are more likely to be completed by people who are already healthy and/or aim to support the implied cause of the survey. As such, these studies should not be interpreted as representative of the broader population. Misguided interpretation of such data may also skew true representation of the perceived barriers to WP adoption. This was an important concern when reviewing literature on barriers to adoption because it may imply that WP adopters are more likely to participate in research than non-adopters. Moreover, information on barriers from non-adopters may be more valuable than adopters since adopters are less likely to be affected by, or perceive, certain barriers.

Understanding of what a WP entails obviously affect people's perceptions of them, and these perceptions matter during the adoption-decision process. Ill-defined and unstandardized WPs are inherently difficult to evaluate which leaves them equally as difficult to predict. This, in turn, makes their adoption an inherently riskier decision. Therefore, to enhance the adoption of WPs on a broader scale they must first be thoroughly defined and standardized to some degree.

Establishing a Standard for Wellness Programs

Defining a gold standard for WPs appears difficult at best and potentially impossible at worst. Nevertheless, much work has been done to elucidate best practices (Das et al. 2014;

Hersey et al. 2008; Reavley, Livingston, Buchbinder, Bennell, Stecki, & Osborne, 2010; Terry et al. 2008; Terry, Grossmeier, Mangen, & Gingerich, 2013; Zula, 2014). These practices, noted next under the heading *Evidence-Based Components* act as this paper's description of what a standardized WP would ideally entail. It bears noting that papers reviewed may not have considered this same description. As such, these descriptions function here solely for the sake of this review.

There are a few other caveats to note before moving forward. First, the components that comprise this paper's view of WPs were predominantly proposed for, and/or identified from, larger settings. Many of these components, and especially all of them together, may not be feasible for SMBs. This notion is indirectly supported by evidence of smaller entities remaining significantly less likely to offer all components of a comprehensive program despite interest in doing so (Taylor, Pilkington, Montgomerie, & Feist, 2016).

Second, standardized procedures likely vary between organizations due to different needs, values, and/or organizational interests. Cultural and organizational processes, whether formal or informal, indirectly affect the efficiency and effectiveness of WPs. Such confounding factors often influence outcomes and blur potential standards or benchmarks.

Third, WPs are likely influenced by temporal aspects of the systems in which they reside. For example, the reviewed definitions of *wellness* have evolved over time. A paper by Kirkland (2014) highlights this difference in definitions of wellness over the past few decades as well as between cultures and fields of operation. However, it does appear that common elements exist between scholars' definitions of wellness. These differences and commonalities become clear when comparing these elements between industries, geographical locations, and generations.

Some elements of wellness (social, mental, spiritual, etc.) may also be more important to one entity than another. As such, these elements may require different prioritization amongst WPs. A WP that prioritizes social wellness is likely to employ—or at least emphasize—very different strategies than one that focuses on physical wellness. As noted earlier, this makes comparison between WPs difficult and generalizable benchmarks and practices nearly impossible to ascertain at the programmatic level.

Nayer et al. (2010) notes that some innovative employers who are readily involved in wellness programming differ on definitions of what WPs entail, but agree somewhat on the intended outcome. Specifically, Nayer proposes the following general goals in order of importance amongst informants: 1) reduced healthcare expenditures; 2) improved productivity; and 3) reduced absenteeism. Improved health status, retention, and engagement were other noteworthy responses. Notably, Nayer et al. (2010) also suggested that WP vendors aim for program effectiveness, as opposed to cost efficiency, during the initial implementation phase. This was an interesting and potentially debatable perspective when considering the common theme of "cost" throughout the literature (McCoy et al. 2013). However, the topic often being implied by the mention of "cost" could be "the cost equation" or "cost/benefit ratio". If this is the case, increasing the benefit would indeed justify a greater cost. Either way, the study by Nayer et al. (2010) offers an example of how the value of outcomes, prioritization of strategies, and underlying definition of WPs might be perceived in different ways between organizations.

Evidence-Based Components of Wellness Programs

Five seemingly well-established components of WPs were noted in a widely-cited landmark study by Linnan et al. (2008). These components appear to originate from Healthy

People 2010 (a government-led initiative to improve the health of the U.S.). The components are noted as fundamental to a comprehensive WP. In other words, a program was deemed "non-comprehensive" if it lacked on or more of these 5 components.

While other elements have been integrated into WP models, these 5 components have provided a valuable reference point for researchers when discussing foundational elements of WPs (Goetzel et al. 2014; Linnan et al. 2008; McCoy et al. 2013). These components are still reflected in Healthy People 2020 (the latest revision of the government-sponsored initiative). These components are also directly in line with other common elements proposed in the literature and highlight the importance of individual, environmental, and organizational aspects of WPs (Hersey et al. 2008; McCoy et al. 2013; Terry et al. 2008; Terry et al. 2013; Zula, 2014). More specific information on Healthy People 2020 and evidence-based elements of WPs can be found at the Center for Disease Control's website, CDC.gov. Namely, the 5 components of comprehensive WPs included:

- health education,

- supportive social and physical environments,

- integration of programs into organizational structure,

- linkages with existing programs, and

- screening programs.

The components listed above act as a general standard for comprehensive programs, but other evidence-based elements warrant noting. Terry et al. (2008) describes best-practice criteria in a list of nine quality components. Programs that employed WP strategies in a manner congruent with Terry et al. (2008)'s descriptions were significantly more successful in

participation rates, completion rates, and health risk reduction rates than programs that did not meet the criteria. Unfortunately, programs being compared were less than standardized which led to weakened baseline of comparison. However, the study did provide a few valuable insights nonetheless. First, the study supported the effectiveness of well-defined comprehensive programs over those that are not. Second, the authors note relatively specific criteria that WPs should aim to meet to maximize performance. Programs that met this "best-practice criteria" performed significantly better at participation rates and program completion. Third, they offered a benchmark for practitioners which appears to be a rare find amongst the present literature. The benchmarks offered were a 70% participation rate in health risk assessments and 48% participation rate in coaching among high-risk candidates. Importantly, these benchmarks were the averages amongst participants of the best-practice programs.

Other investigators have proposed similar descriptions of their own best-practice criteria. Most recently, Zula (2014) proposed their own five strategic factors for success that built upon previous literature. Those strategic factors were: effective and efficient communication; leadership engagement and commitment; relationships and partnerships to leverage resources; accessible and involved employees; and relevance and continuous improvement.

Terry et al. (2013) expanded on the potential relationship between demographic factors and various types of incentives. In this study, there appeared to be a relationship between the response to different incentives (monetary versus non-monetary), age, and gender. This directly added to the lead author's previous work on the topic. Where Terry et al. (2008) notes that financial incentives should be used to maximize participation, this later research by the same lead author (Terry et al. 2013) noted that financial incentives may undermine interest and decrease participation in older males. An alternative was non-monetary incentives, such as duffel

15

bags, for this demographic. Importantly, financial incentives still enhanced participation amongst females. Again, this work was not a direct contradiction of the previous findings as much as a refinement thereof. Older males tended to prefer non-monetary incentives over monetary versions; this was the opposite effect seen in their female and younger male counterparts.

There seems to be no shortage of recommendations for specific strategies and program components. However, the theme that one program does not suit all organizations should not be overlooked or undermined in lieu of grand strategies. A great deal of attention to the needs analysis portion of programming is warranted while deferring to evidence-based components when in doubt. Terry et al. (2013) explicitly noted the importance of designing interventions through specific and relatively subjective analyses first before looking to objective, generalizable, best-practice methods. In any case, the prevalence of seemingly effective strategies throughout the literature further substantiates the need to examine barriers to their adoption.

<center>Barriers to Adoption of Wellness Programs</center>

To further establish the reader's perspective, common barriers to adoption were collected from previous work (Linnan et al. 2008; McCoy et al. 2014). These barriers are described below. Namely, the most common barriers found included:

- cost,
- time,
- employee interest,
- expertise, and
- legal concerns.

<center>16</center>

Overview

Generic barriers to WP adoption have been noted throughout the literature, often irrespective of company size. More specific to small businesses, McCoy et al. (2014) conducted a systematic review of experimental and quasi-experimental studies that assessed factors influencing the adoption and effectiveness of WPs in settings of less than 500 employees. "Cost", both direct and indirect, were noted as crucial considerations when deciding which components to adopt. This was supported by other literature (Hughes et al. 2011; Linnan et al. 2008; Taylor et al. 2016; Witt et al. 2013) and appeared to be a potentially universal theme. Other themes were noted by various studies, the most common of which are further described below.

Cost

Cost was a common barrier to adoption of WPs noted throughout the literature (Hughes et al. 2011; McCoy et al. 2014; Taylor et al. 2016; Witt et al. 2013). As such, establishing an ROI may be a valuable communication tool for WP vendors. Likewise, direct comparisons between WP cost and savings in healthcare expenditures appears frequently throughout the literature as a method of justifying adoption of WPs (Goetzel et al. 2014; Hubley & Dutram, 2011; Liu, Weinberger, Serxner, Mattke, & Exum, 2013; Mukhopadhyay & Wendel, 2013). However, this may be an incomplete method for evaluating WPs due its potential for undermining other benefits (Bishop & Yardley, 2010; Mukhopadhyay & Wendel, 2013). This is especially true for new programs, due to the time lag in cultivating certain elements of the workplace culture and attitudes of the staff that are important to the success of wellness initiatives (Mukhopadhyay & Wendel, 2013).

There is evidence that supports the notion of focusing on less tangible benefits rather than decreasing costs as a way to gain the greatest value from a program (Kaspin et al. 2013). Comparing WP costs to potential these potential intangible values also appears in the literature (Connor, 2016; Ozminkowski, Serxner, Marlo, Kichlu, Ratelis, & Van de Meulebroecke, 2016). Enhanced engagement, for example, is another potential metric for measuring value of WPs. Surveying employees for participation rates, job satisfaction, and other metrics related to engagement may offer a degree of perspective when evaluating the otherwise less-tangible values of such programs. Recruiting advantages and staff retention are highly viable outcomes to consider when adopting WPs as well. Such elements may also carry greater magnitude in smaller settings. For example, the gain or loss of one skilled employee has a larger relative impact on small business operations than larger counterparts. That one employee gained or retained may quickly outweigh the costs of a WP for that organization.

Furthermore, the literature has not been consistent regarding the generalizability of potential ROI of WPs (Mukhopadhyay & Wendel, 2013; Newman et al. 2015; Zula, 2014). The diverse nature of smaller organizations exacerbates the flaws of generalizing ROI. For example, one organization of 30 predominantly obese people may be in dire need of weight loss programs while another small business of 20 healthy individuals may not.

Costs also vary per service, length of program, number of employees, and so forth. Zula (2014) notes the large range in price from below $300 per employee to more than $500 per employee. This was along with 29% of their surveyed organizations (n=9) stating that they were unsure of the exact amount spent on wellness initiatives. This was not central to their investigation, but it does demonstrate a lack of standardized pricing which seems to coincide

with a lack of standardized processes. While costs are often a general concern of doing business, the lack of certainty behind ROI and program costs may exacerbate this concern.

Time

Time costs and scheduling conflicts are noted as a potential barrier to adoption of WPs (McCoy et al. 2014). This makes superficial sense, considering that time constraints are often a concern of business in general. However, this may be an issue of priorities as much as time itself. Qualitative interviews repeatedly note the importance of aligning tasks—and therefore, time— with company goals (Audrey et al. 2015; Williams & Snow, 2012). Many company goals are not necessarily exclusive from WP activities; some company goals may even be supported by WP activities or their outcomes. More *root barriers*, such as inadequate personnel or inefficient processes, may present themselves as "time" without DMs being aware. This would be especially true from the perspective of once-removed positions that DMs commonly hold. In any case, "time" appears to be a common barrier in a generic sense at the very least.

Employee Interest

A key factor that may influence the adoption of WPs is employee interest. A survey by Linnan et al. (2008) showed that 63.5% of respondents noted "lack of employee interest" as a barrier to effectiveness and adoption of WPs. Moreover, many participants from the same study (48.2% of all respondents) noted the lack of participation on the part of high-risk employees, specifically, as a key challenge of such programs. This is a well-founded concern, considering that those at risk of chronic disease would likely have the most to gain from such programs.

While maintaining the health of healthy workers is an important goal of many WPs, such "maintenance outcomes" are arguably difficult to gauge when compared to positive changes in health indicators and medical expenditures of unhealthy workers. As noted earlier, these changes in expenditures are often the metric of choice for DMs. This may be due in part to their ease of comparison. For example, losing 20lbs to attain a healthy bodyweight is an easily measured benefit whereas maintaining a healthy body image is less so. While both are beneficial, one is simply more easily evaluated against standard health-risk data. Either way, WPs would likely benefit from maximizing participation by all employees, healthy or not, by employing methods to create and sustain behaviors that support health and prevent disease. Any lack of interest in the program, especially by large percentages of high-risk individuals, should draw concern for WP vendors and DMs alike.

Expertise

Lack of expertise is mentioned as a barrier to adoption and implementation (McCoy et al. 2014). With regard to adoption, this lack of expertise would increase uncertainty of a given program's outcome. Inadequate expertise may also reflect a hindrance of the DMs ability to evaluate WPs, leading to a barrier of ignorance. Expertise is likely a chief concern amongst DMs and may hinder adoption even in those who are otherwise eager to invest in WPs.

However, expertise as a general concept is not so straightforward. The concept is too broad to specifically answer when addressed at a programmatic level. Expertise of specific roles would be, but evidence is far from clear in this particular area. For example, health coaches are often a fundamental facilitator of WPs, but there was no description of credentials or education for this role listed in the studies reviewed. This omission by researchers may lead to ambiguity

regarding the role and a discrepancy in the quality of health coaching. At best, this discrepancy

may lead to widely varied results from otherwise similar programs. At worst, this may lead to

failure of otherwise well-designed programs. Either way, it provides just one example of the

ambiguity that may undermine the adoption of WPs.

Legal Concerns

  Legal concerns may be a potential barrier to WP adoption, and for good reason. Legal

policies and regulations are deceptively vague and present their own host of unique risks (Plump

and Ketchen, 2013). The extent of these risks varies from organization to organization due to a

number of factors. The ability to navigate legal policies, for example, may be easy or difficult for

different DMs based on their own expertise and/or access to professional counsel. SMBs may be

at a disadvantage with regards to personnel who can navigate legal systems. Another example

could be an organization's vulnerability to negative health consequences. Employees who exhibit

greater health risks may have more to gain from expert programs, but also pose a bigger threat of

negative effects if mishandled in some way.

  Plump and Ketchen (2013) offer a few general strategies for basic WP purposes

regarding legal concerns. For one, WPs should be steadfast against any form of potential

discrimination and aim for equal treatment. Individuals who feel they have been discriminated

against may decide to pursue legal action against the discriminator. There are specific policies in

place stating what is and is not discrimination under different contexts. However, personal law

suits may be incurred despite adherence to these policies.

  WPs are also directed to limit the collection of medical information (Plump & Ketchen,

2013). Collection, use, and dissemination of health information appears to be highly regulated.

Legal consequences could be incurred if such regulations are violated. Importantly, legal consequences may be incurred regardless of any negative consequences to the individual whose information was violated.

Plump and Ketchen (2013) also propose that WPs maximize the voluntary nature of program enrollment and clearly differentiate between wellness activities and workplace initiatives. This latter strategy is somewhat debatable, and regulations that would imply any given strategy are not quite cut and dry. However, the general notion of equal opportunities and voluntary involvement seem to be productive directions when developing and implementing WPs. Once again, this notes the ambiguity that underlies WP adoption and, therefore, barriers thereto.

<u>Summary</u>

Many barriers to adoption of WPs likely exist amongst businesses of all sizes, but some of these barriers may carry more weight in smaller settings. These barriers also vary in significance based on the host organizations unique elements. The literature review offered here aggregates some of the more commonly perceived barriers that vendors and researchers may encounter. Just as importantly, this review offers perspective as to the ambiguity of the problem and lack of actionable knowledge collected to date. Qualitative inquiry into the perceptions of those making the decisions to adopt or reject WPs may offer deeper insight. Such insight garnered directly from those making the adoption-decision may provide more actionable knowledge.

CHAPTER THREE: METHODS

The aim of this study was to explore the utility of a novel combination of qualitative synthesis methods for informing a real-world problem of practice. This chapter describes the implementation process of these methods. The BFS was the foundational method employed during the study. Supplemental methods used to augment the BFS are noted as such throughout this paper. Namely, the set of supplemental devices and methods used to augment the BFS consisted of:

- the addition of Alignment Scores to quality appraisals;

- Confidence of Findings (CERQual) scores; and

- a Saturation of Inquisition Test.

The A-BFS is described first from a macro perspective. A more detailed description of the individual methods used is then provided in an instructional manner for the sake of replication and transparency. These methods are then discussed regarding their interaction and influence over the outcome throughout Chapters 4 and 5. The overall methodology, referred to as the A-BFS, aimed to answer the following questions:

- *What root barriers are hindering adoption of WPs amongst SMBs?*

- *Are some perceived barriers more universal, or substantial, than others?*

<u>Overview of the Best-Fit Framework Synthesis Method</u>

Step 1 of the BFS consisted of two independent searches: one for theoretical framework and another for empirical literature. The framework and the literature were selected based on

their potential to inform the problem of practice. Framework and literature searches occurred simultaneously and independently of one another. Specific search processes are outlined later in this chapter and illustrated descriptions are provided in APPENDIX B.

Step 2 consisted of two independent processes: data extraction and *a priori* framework development. Framework development and data extraction also occurred simultaneously and independent of one another. The framework was used to define codes and analyze findings. Data extraction targeted the findings from the Results sections of selected studies. Other information from studies was used for quality appraisal purposes, but was not coded for synthesis.

Step 3 involved the coding and constant comparison of findings (i.e. data analysis). Empirical findings were coded using the *a priori* framework as a guide. Findings were compared for similarities and differences across studies. Findings that were not readily matched with definitions provided by the framework were identified for Step 4. An external reviewer duplicated this process for reliability.

Step 4 involved the identification of descriptive themes and emergent themes. Descriptive themes were intended to reflect the evidence through the perspective of the *a priori* framework. Themes that did not fit the framework are identified as *emergent themes.* Emergent themes were defined to reflect the empirical evidence and identified as new codes to be compared across studies. Emergent themes should be seen as a sub-type of descriptive theme. An external reviewer critically examined and verified the outcome for reliability.

Step 5 involved integrating emergent themes into the *a priori* framework as new codes. This process was somewhat cyclical, and required comparison between all codes, evidence, and the conceptual framework. A *sensitivity analysis* was used to justify the addition of any given codes into the conceptual framework. Any *a priori* codes that were not used or explicitly rejected

by the empirical evidence were considered for removal. A Saturation of Inquisition Test was used to justify the removal of any codes that were not used during analysis.

Step 6 involved refinement of the *posteriori* framework. Any overlap or redundancy between codes and definitions thereof was addressed as needed. This was repeated until all codes were aligned and confirmed to represent the empirical findings as viewed through the selected theory.

Importantly, a Saturation of Inquisition Test was added to Step 5. This was intended to add a level of analysis to the original methods without detracting from them. This test warrants further evaluation and consideration by researchers. An illustrated overview of this part of the A-BFS process is provided in Figure 1.

| Systematically identify theoretical framework elements. | Systematically identify empirical research. |

⬇ ⬇

| Aggregate key theoretical framework elements into an a priori framework. | Extract data from the empirical literature. |

⬇ ⬇

Code empirical evidence against the framework while using the constant comparison method between findings. Record findings that are not adequately captured by the framework.

⬇

Compare findings that are not adequately captured by the framework across studies to generate novel, emergent themes.

⬇

Create a new framework that better informs the problem of practice. This is done by 1) operationalizing emergent themes as codes to fill the "gaps" in the a priori framework, 2) refining a priori codes to accommodate emergent themes if necessary, and 3) removing any a priori codes that were rejected by the findings.

⬇ ⬆

Evaluate and refine the new framework to minimize issues of redundancy or dissonance.

Figure 1. Overview of steps in the original Best-Fit Framework Synthesis

<u>Detailed Components of the A-BFS Methodology</u>

The specific methods and processes of the A-BFS methodology are described below in detail. For the most part, these components are described in the order they were used. However, the macro perspective just provided demonstrates how some methods are used simultaneously and/or interdependently. Care should be taken to consider this when employing this set of methods.

Systematic Search Process

The University of Central Florida's OneSearch tool was initially used for the search processes. This specific tool searches multiple databases simultaneously to provide the greatest breadth to the search. However, Dr. David Boote at the University of Central Florida was consulted and found the search tool inadequate to support a systematic review without a follow-up of specific databases. For this reason, the EBSCO interface was used to search the following databases to ensure comprehensive results: Academic Search Premier, Alt Health, Business Source Premier, and Medline.

The search process used here included a two-tailed search approach: one search to find a most relevant theory; and one search to find most relevant findings. Specifically, the SPIDER approach (Cooke, Smith, & Booth, 2012) was used to search for relevant literature while the BeHEMoTh approach (Booth & Carroll, 2015) was used to search for a theoretical framework. Readers should view both search strategies as independent from one another rather than sequenced or interdependent.

Framework-Specific Search: BeHEMoTh

Relevant theories and models were searched and evaluated according to their ability to thoroughly inform this study's problem of practice. If a theory was not found that comprehensively informed the phenomenon under study, elements from more than one theory would be synthesized into a novel framework. If multiple theories were found yet no justification could be made to exclude more than one theory, the relevant theories were to be synthesized into a meta-framework. Specifically, the theories and/or theoretical elements were selected based on their ability to thoroughly analyze barriers to adoption of WPs.

The BeHEMoTh approach (Booth & Carroll, 2015) was used to identify a most relevant theory or elements thereof in a systematic manner. However, exploring barriers to WP adoption is an inherently multidisciplinary task. The participants (DMs), context (SMBs), and phenomenon (perceived barriers to program adoption) all present elements best viewed through organizational theory, but the nature of WPs denotes strong health promotion and behavior change components. Considering this, the following key words were used to search for potential models and theories: *organizational theor**; *wellness program* OR *health promotion*; and *adoption*. These key terms were further justified after a comparison to the initial literature review. This list of key terms was intended to be precise enough to generate most relevant results while broad enough not to exclude potentially viable theories and models. A summary table of the BeHEMoTh criteria and a screenshot of the search is included in APPENDIX B.

Literature-Specific Search: SPIDER

The SPIDER approach (Cooke, Smith, & Booth, 2012) was used to systematically search for findings that may directly inform this study's problem of practice. For the sake of relevance

and reuse in future research that may build upon this study, the following key terms were searched: *wellness program* OR *health promotion*; *barriers* AND *adoption*; and *qualitative*. These key terms were deemed appropriate based on the research question and necessary modifications to the SPIDER process (discussed in Chapter 4 and 5). A screen shot and summary table of the SPIDER search criteria is included in APPENDIX B.

To collect the most relevant studies in a comprehensive manner, studies that met the following criteria were included:

- English language,

- published after 2010,

- explicit investigations of barriers and/or facilitators to adoption of WPs,

- explicit targeting of DMs, and

- investigations that included SMBs (defined here as 999 employees or less).

Studies were excluded if they:

- pooled evidence from sources other than SMB DMs,

- used findings accrued before 2005,

- used exclusively quantitative methods,

- included their own inclusion or exclusion criteria that was too confining for generalization to DMs of other SMBs, or

- were conducted outside the U.S.

Quality Appraisal with Alignment Scores

A COREQ (Tong, Sainsbury, & Craig, 2007) checklist was used to appraise quality of reporting. Alignment Scores were added as a reference point during the analysis and synthesis portion of this review. Alignment Scores are intended to identify potential inconsistencies between an aspect of the primary study and that of this study. Aspects of a given study were identified as *well-aligned* or *not well-aligned*. *Well-aligned* indicated that the given aspect of the primary study directly addressed or reflected this study's line of inquisition. *Not well-aligned* indicated that the given aspect of the primary study did not directly address or reflect this study's line of inquisition. Quality of Reporting and Alignment Scores are provided in APPENDIX C and further discussed in Chapter 5.

Data Extraction

Data extraction was two-tailed in a similar manner as the literature search. The framework specific process involved aggregating conceptual elements from the chosen theory into a framework to operationally define *a priori* codes. This step was carried out by defining the code to be used, the code's key elements, and its potential mediating factors. All codes were based on the same unit of analysis (i.e. *potential barriers to WP adoption*). The literature specific process of data extraction included the extraction of all content in the Results sections of the primary studies. While each study was comprehensively appraised, only the content in the results section was used for analysis and synthesis.

Data Analysis and Synthesis

After the *a priori* framework was developed and empirical findings were extracted, content analysis was used to code empirical findings against the *a priori* framework. Content analysis was carried out by the author and an external reviewer in a cyclical fashion until full agreement on coding was reached. Any discrepancies in coding that could not be resolved were removed from the synthesis and recorded as *unresolved codes* in Chapter 4. Coding was done using Dedoose software.

Constant comparison was carried out simultaneously with the coding process to unearth emergent themes and test the transferability of all themes across contexts. Emergent themes are defined here as a sub-type of descriptive themes that are not readily absorbed by the *a priori* framework. Emergent themes were then defined to best reflect their empirical source and compared across studies for transferability. Emergent themes were integrated into the *a priori* framework unless rejected by the sensitivity analysis (discussed next).

Refinement to the *a priori* codes was then considered if there was overlap or discrepancies in findings. Refinements were only made if necessary, and care was given to ensure all codes aligned with one another while accurately reflecting the empirical findings. Analytical themes were generated based on frequency, distribution, co-occurrence, and transferability of codes across studies. The general process of data analysis and synthesis is illustrated in Figure 2.
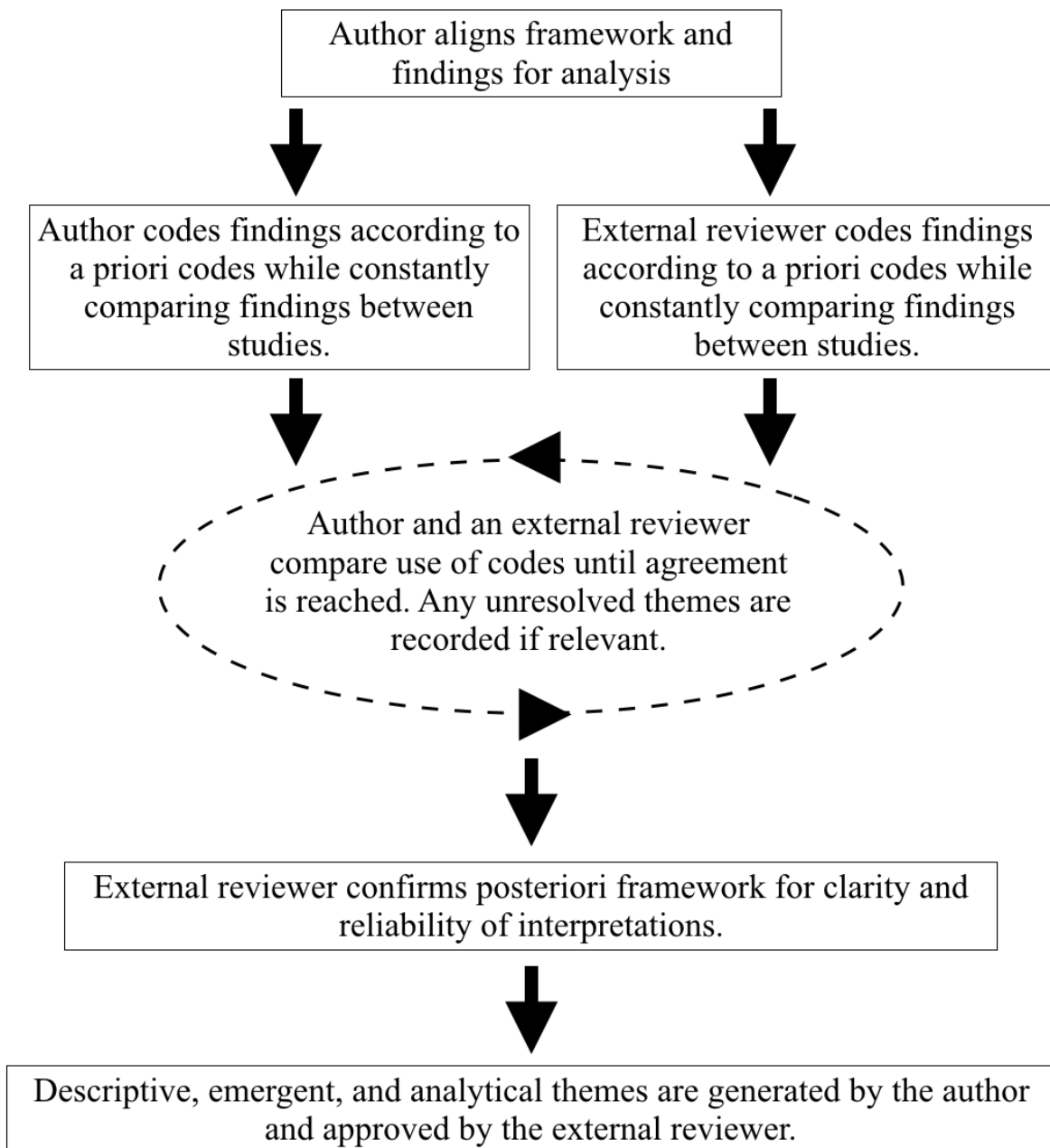
Figure 2. Steps in data analysis.

Sensitivity Analysis

A sensitivity analysis was carried out to test for the influence of lower quality or weakly-aligned studies on the findings. Any findings that were predominantly based on lower quality or weakly-aligned studies were then identified for further evaluation. If the finding was then found to be irrelevant to this study's line of inquisition or context it was considered for removal.

Saturation of Inquisition

A novel Saturation of Inquisition Test was used to justify deletion of unsupported codes as well as identify gaps in research. The test accomplishes this by comparing interview questions from the primary studies to the *a priori* framework. This identified domains within the *a priori* framework that were inadequately examined by researchers. Any domains (i.e. codes) that were not directly assessed by the original interviews were retained in the framework for future use and identified as gaps in inquisition. The rationale underlying this test is discussed in Chapter 5.

CERQual Analysis

Confidence of Findings was then assessed as outlined by Lewin et al. (2015). This allowed for ordinal ranking of individual findings based on the likelihood that they represented the phenomenon under investigation. More specifically, this method focused on rating well-defined domains of each study that contributed to a given finding. This allowed for a ranking of findings based on the studies that supported them. The domains of interest during the CERQual Analysis are described in light detail below.

- *Methodological Limitations* were assessed using relevant elements from the COREQ checklist.

- *Relevance* was assessed via the directness and comprehensiveness in which the original study informed all core domains of this study's research questions. Core domains were identified as *population* (SMB), *participant perspective* (DM), *phenomenon of interest* (barrier to adoption), and *independent variable* (WP).

- *Coherence* was assessed according to variances in patterns that led to the finding.

- *Adequacy of Data* was assessed in terms of the number of informants supporting the finding and richness in descriptions of findings.

Each component of measurement was identified as presenting minor, moderate, or substantial concerns. Again, this was done on an ordinal scale using all other studies as the reference point. This process is similar to constant comparison in this manner; it gains its point of reference from the other studies being reviewed. Initial outcomes were then compared between studies to ensure consistency of rating. Findings were then identified as high, moderate, low, and very low based on this comparison between scores. If any scores were considered "very low" at any point they were considered for further evaluation. Further description of the rubric employed during this process is offered in APPENDIX D and APPENDIX E.

Self-Evaluation

This paper was designed to meet or exceed standards proposed by a number of seminal works. Carroll et al. (2013), Sandelowski (2007), and Lewin et al. (2015) supplied guidance for much of the methods used regarding quality appraisal, data extraction, analysis, and synthesis.

The ENTREQ Statement (Tong, Flemming, McInnes, Oliver, & Craig, 2012) was used as a checklist to ensure comprehensive reporting of all relevant criteria in this study as well as those analyzed (see Table 2 and Table 3 in Chapter 4). The initial literature review was organized to best satisfy criteria proposed by Beile and Boote (2005). While other studies and books were consulted, these works were used on a regular basis as benchmarks, guides, and rubrics of sorts to evaluate the selection and use of methods throughout this paper.

<div align="center">Basic Strategies to Support Methodological Rigor</div>

Qualitative inquiry, and especially synthesis thereof, appears to fall victim to somewhat prejudiced criticism by the scientific community (Barbour, 2003). This has been noted to possibly stem from the imposition of quantitative paradigms and methodological rules unto this very different form of inquisition and reporting (Barbour, 2003; Weed, 2006). In any case, qualitative syntheses may indeed require an additional degree of discipline and scrutiny on the part of the researcher and reader alike. For instance, a set of quantifiable measurements can be segregated and transferred externally between contexts relatively easily when the unit of measurement is consistent. Qualitative literature is not so easily dissected. The unit of measurement (better stated as unit of analysis) may be defined specifically for the particular research project. The codes supplied during the study in this paper are an example. Codes are most often operationally defined units that may vary based on the framework or context. The complexity of these contextually-based concepts and units of analyses becomes even more apparent during syntheses of qualitative evidence. Metaphorically speaking, a qualitative work may be described in this regard as a puzzle of inter-dependent pieces rather than a set of blocks that can be independently re-aligned. As such, qualitative works must be evaluated as a whole

before attempting to transfer their outcomes across contexts. Quality of Reporting, Alignment Scores, review of underlying frameworks (when present), comparison of study characteristics, and other analytical devices aided in this effort. However, the reader is now tasked with their own interpretation of this work. As such, they should remain critical when transferring any findings or acting on any conclusions drawn.

For these reasons, the reporting process of qualitative syntheses requires a significant degree of detail to support the repeatability of the process (Atkins et al. 2008; Tricco et al. 2016) and parameters for the work's outcomes (Weed, 2006). This quickly becomes apparent when reviewing qualitative literature; even the most rigorous qualitative methods rely heavily on rich descriptions to maintain their validity and reliability. For this reason, elements of methodological rigor are addressed next in thorough detail.

Upholding Validity

In accordance with Sandelowski and Barroso (2007), the iterative nature of certain forms of qualitative syntheses warrants a high degree of transparency by the author and scrutiny by the reader. This is especially true for author-generated third-order themes that arise from secondary analysis. Sandelowski and Barroso (2007) also state the value of explicitly addressing potential threats and describing any strategies used to minimize potential weaknesses. Below is this paper's attempt to fulfill such a standard.

To support *face validity*, a sensitivity analysis was performed. This was deemed appropriate due to the inclusion of a study that was relatively low in quality and alignment of the given sample and phenomenon being examined. More specifically, Nelson et al. (2015) provided a methodologically rigorous study, but assessed perceived barriers to adoption and integration of

programs (i.e. wellness programs and occupational health and safety programs). The investigation was likely directed in part toward the concept of occupational health programs and/or the phenomenon of integration. How well the study aligned with the investigation of this review would have been better understood if interview guides were provided. However, reporting of these important elements was less than thorough leading to a lower quality score. For this reason, a post hoc sensitivity analysis was conducted by removing the evidence provided by Nelson et al. (2015) and noting any change in outcome.

To support *external validity* of findings, a negative case analysis was performed. This was done by directly searching the findings for evidence that opposed or limited the initial results in some way. Such cases would then influence definitions, findings, and limitations as appropriate. If cases were found that could not be reconciled they would be recorded. A rich description of any such cases would be provided in the results and/or discussion sections.

Theoretical validity is admittedly a difficult aspect of validity to defend. The concept is defined specifically for this review as *accurate operationalization of theoretical elements*. This paper attempted to uphold this aspect of validity through the well-defined tables and figures that illustrate the conceptual processes involved; macro and micro views of the processes and framework; thoroughly defined and aligned operational definitions; and identification of most likely mediators for codes and themes.

Descriptive validity is defined here as *the factual accuracy of findings*. This would be upheld by directly restating claims and contextual factors directly influencing those claims. Care was given to accurately depict relevant characteristics of the reviewed studies as well as factors that appeared most relevant to the discovery of those findings.

Interpretive validity is defined here as *fair representations of points of view*. This is inherently conceptual and difficult to control for without violating descriptive validity. For instance, a participant may state that "cost" is their most prevalent barrier to adopting WPs while implying a lack of "cost-effectiveness". The informant could mean the absolute cost exceeds their available resources, but they could also mean that the relative cost is too high compared to other alternatives. To uphold interpretive as well as descriptive validity, coding was kept somewhat conservative toward factual reporting while implications and interpretations were noted when relevant. The discussion portion of the study would have been used to explore potential differences in interpretations.

Pragmatic validity is defined here as *the utility of findings*. This is primarily upheld through a direct connection of findings to the research questions. For example, the research questions are re-stated throughout the work and answered in Chapter 4. The utility of findings to WP vendors and researchers is also redundantly noted throughout this paper.

Upholding Transparency and Reliability

Direct attention was made to the referential adequacy (Lincoln & Guba, 1985) of findings reported in this paper. To achieve this while minimizing redundancy, information was separated into that which was analyzed and that which is archived simply for reference. While archived findings have little influence on the outcome they are still a valuable component of this paper's audit trail. Readers may find archived material valuable when testing the rationale of conclusions, transferring findings to future projects of their own, or simply adding to their perspective. Ample use was also made of the Appendix to provide any and all information that may be deemed relevant when examining the decisions made, processes used, and rationale

behind conclusions in this paper. Any redundancy between the body and the Appendix should be seen as a productive err toward the transparent side of the dissemination continuum.

Attention was also paid to the reliability of data analysis. A doctoral candidate in an unrelated field was consulted as an external reviewer during coding. The external reviewer played an important part in supporting reliability of findings.

Ethical Considerations

No ethical violations were apparent during this review. This does not mean that recommendations made are free of ethical scrutiny, especially if implemented negligently or outside of the context presented here. Care should be taken by the reader to ensure actions based on any recommendations do not potentially violate policy, rules, regulations, or rights of workers. Liability issues, for example, may arise when adopting, implementing, or evaluating programs intended to affect participants' health and wellness.

CHAPTER FOUR: RESULTS

The aim of this study was to explore the utility of a novel combination of qualitative synthesis methods at informing a real-world problem of practice. These methods were explored based on their ability to thoroughly inform practitioners on a relevant problem of practice. Specifically, the collection of methods was employed to answer the following questions:

- *What root barriers are hindering adoption of WPs amongst SMBs?*

- *Are some perceived barriers more universal, or substantial, than others?*

This chapter briefly explains the results from the A-BFS. These results act as a valuable reference point when examining the utility of the methodology. Notably, the methods were selected and designed to leverage the many strengths of qualitative research while simultaneously accounting for its inherent weaknesses. Results from the A-BFS included:

- the *a priori* framework;

- study selection results;

- Quality of Reporting and Alignment Scores;

- outcomes from the sensitivity analysis;

- descriptive, emergent, and co-occuring themes;

- analytical themes;

- *posteriori* framework;

- Confidence of Findings; and

- gaps in inquisition.

A Priori Framework

The BeHEMoTh approach was used as a standardized method to find the most relevant model or theory for this study's problem of practice. Theoretical elements that were most relevant were operationally defined and aligned with one another to create the *a priori* framework. Results of the search, selection, and development of *a priori* framework are described below.

BeHEMoTh Search Results

The BeHEMoTh search strategy was used to discover potentially relevant theories and models. This approach was explicitly recommended by Carroll (2013) for use in BFS. The search initially identified 246 articles, of which 23 abstracts were screened for potentially relevant theories. Upon screening, 11 papers were rejected and 12 full-text articles were downloaded for brief review and comparison of theoretical frameworks. The results of the search are summarized in Figure 3.
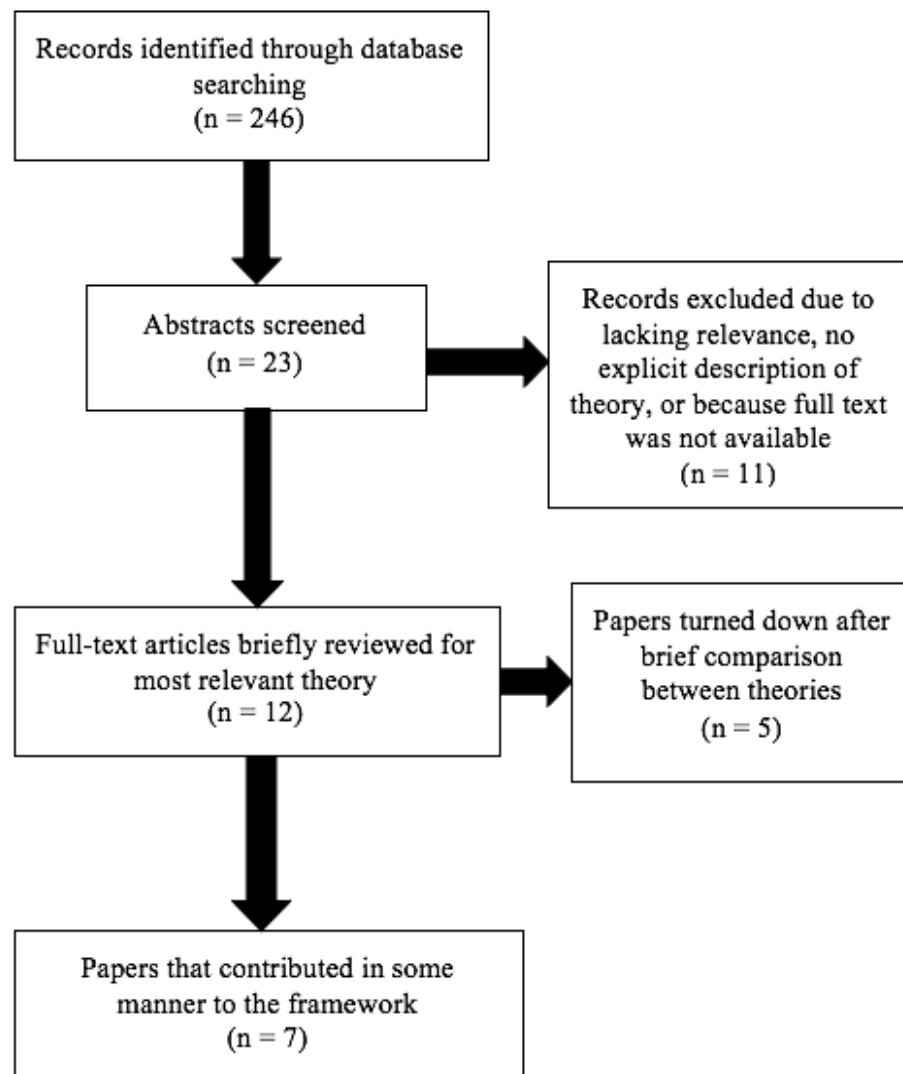
Figure 3. BeHEMoTh search results.

Abstracts were reviewed if they appeared to fit the problem of practice. Full-texts were

downloaded for further review if the abstract noted the use of a potentially relevant theory and

evaluated the phenomenon of program adoption. Diffusion Theory (DT) was selected due to its

overwhelmingly frequent mention as well as its direct applicability to this paper's problem of

practice. No other elements were used to augment the framework. While the argument can be made that other theories provided relevant insight, any changes in the DT framework also presented a potential threat. Therefore, it was deemed appropriate for the overarching goal of this paper to err on the conservative side of generating theory if possible in order to maximize reliability. The DT-based framework is outlined in Table 1.

Table 1. *A Priori Framework*

Potential Barrier	Key Elements	Operational Definition	Potential Mediating Factor
System Characteristics			
Champion Presence is Inadequate	Personnel (Internal)	An advocate (current or potential) is not present within the system.	The level of influence (formal or informal) and general ability of the champion.
Opinion Leader Support is Inadequate or Rejects the WP	Personnel (Internal or External)	A member of the social system who has influence over acceptance of new ideas is not adequately supporting adoption or is opposing it to some degree.	The extent of influence on DMs and employees.
Social Norms Do Not Support Adoption	Social System Influence	Patterns of behavior within the systems (internal and external) that DMs operate do not support adoption of WPs or aspects of their operation.	Innovativeness of the adopter; and influence from Opinion Leaders.
Organizational Slack is Inadequate	Organizational Resources	The organization does not have adequate resources to devote to the WP.	Shifts in organization revenue; and potential for capacity building.
DM Characteristics			
Need is Perceived as Inadequate	Knowledge of Problem	The DM does not perceive a problem to exist, or does not perceive the problem to be a worthy issue to correct.	The extent of how actively the DM seeks relevant knowledge; knowledge of problem; knowledge of the severity of the problem; and underlying beliefs and attitudes toward the problem.
Awareness is Inadequate	Knowledge of Solution	The DM does not perceive a solution to exist, or does not perceive the solution as worthy to implement.	General knowledge of the WP; knowledge of the WP viability; knowledge of WP accessibility; knowledge of the WP usability; knowledge of the underlying principles of the WP; and underlying beliefs and attitudes toward the WP.
Communication Channels between Adopters and Vendors			
Communication Channels are Not Leveraged Adequately	Inter-System or Intra-System Interactions	Either party does not actively communicate outside of their social system, or communicates poorly within their social system.	Availability of interactions between systems; and alignment of communication strategies employed within systems.
Heterophilous Traits Hinder Communication	Personality Traits	The culture, social status, or other background elements inhibits empathy and communication.	The extent of cultural and social differences; and presence of common values or interests that may bridge gaps.

Potential Barrier	Key Elements	Operational Definition	Potential Mediating Factor
WP Characteristics			
Relative Advantage is Inadequate	Observable Benefits	The advantages to adopting the WP do not outweigh alternative decisions.	Perceived effectiveness of the WP over current solutions; presence of alternatives (direct or indirect); and perceived effectiveness of alternatives.
Compatibility is Inadequate	Integration with Values, Beliefs, and Daily Processes	The WP does not appear to integrate well within the system.	Perceived processes involved in the WP; and perceived underlying principles of the WP.
Ease of Use is Inadequate	Complexity	The WP is perceived as too complex to employ effectively.	Background knowledge of the DM; and the WP agent's ability to educate the DM.
Trialability is Inadequate	Observable Benefits	The WP is not available to try.	Previously observed effects; and third-party subjective notions regarding the WP.
Observability is Inadequate	Observable Benefits	The WP is not perceived as having shown demonstrable benefits. Note: This goes beyond effectiveness toward a more perceptual "observed effectiveness".	Extent of trust in alternative verification of the benefits; and third-party subjective notions regarding the WP.
Potential for Re-Invention is Inadequate	Integration with Values, Beliefs, and Daily Processes	The WP is not perceived as customizable to the changing needs of the organization.	Anticipated future needs of the organization.
Vendor Characteristics			
Change Agent Power is Inadequate	Personnel (Internal)	The Change Agent's abilities to gain decision to adopt is low.	Heterophilous background with DM; access to Opinion Leaders; and incompetence as a facilitator.
Aide Effectiveness is Inadequate	Personnel (Internal or External)	The Change Agent's Aides are not able to build strong enough informal connections to the Opinion Leaders or DMs.	Extent of heterophilous background with adopter personnel.

<u>Literature Search</u>

   The SPIDER approach was used to search for relevant literature to inform this review's research questions. Relevant studies were then evaluated based on inclusion and exclusion criteria noted in Chapter 3. A description of the results from this process is provided below.

SPIDER Search Results

   Of the 234 articles returned during the search, 16 abstracts appeared to meet the inclusion criteria. Abstracts were then screened for relevance. Eight studies were excluded due to reliance on primarily quantitative methods. Full-text articles were then retrieved and assessed. Three of these articles were excluded for being conducted outside the U.S., and 1 was excluded due to business size. This left 4 studies to be included in the synthesis. Results from the search are illustrated below in Figure 4.
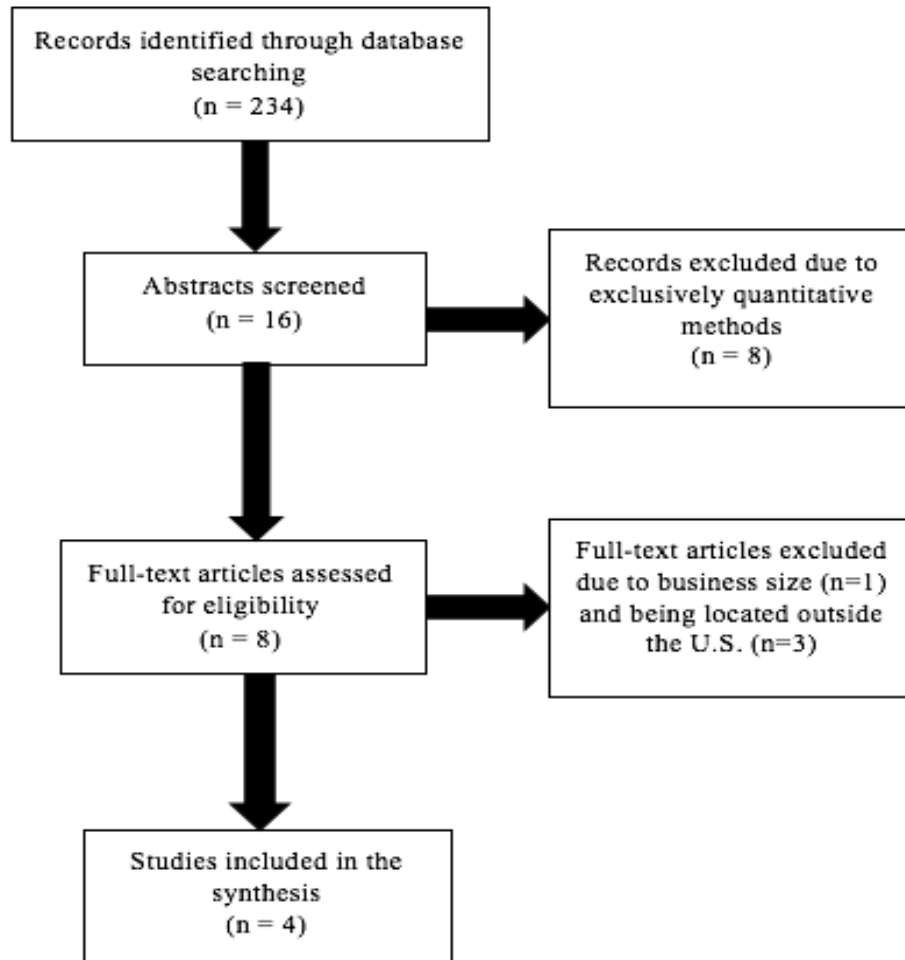
Figure 4. SPIDER search results.

Study Characteristics

  Relevant study characteristics were collected and compared. This comparison across studies allowed for a richer understanding of the context in which any themes may exist. A summary of relevant study characteristics is provided below in Table 2 and Table 3. Reporting of characteristics was intended to exceed the standards set by the ENTREQ Statement (Tong et al. 2012). Characteristics are organized vertically for easy comparison by the reader.

Table 2. *ENTREQ-Based Report of Study Characteristics*

	This Synthesis	Hughes	Witt	Kuehl	Nelson
Aim	*To identify and explore root barriers to wellness program adoption.*	Explore the decision-making processes for adopting and implementing health promotion programs at SMBs.	Identify SMB DMs' criteria, perceived barriers and aids, and values regarding adoption and implementation of WPs.	Identify and evaluate determinants of WP adoption amongst fire departments.	Describe perceptions of acceptability and feasibility of implementing an integrated WP and occupational safety and health program (OSH).
Research Question	*What are the root barriers that hinder the adoption of WPs by SMB DMs? Are some more substantial than others?*	N/A	N/A	N/A	N/A
Intended Audience	*Health promotion researchers, practitioners, and interdisciplinary qualitative methodologists.*	Health promotion practitioners and researchers.	Health promotion practitioners and researchers.	Health promotion practitioners and researchers.	N/A
Underlying Framework	*Generated through Best-Fit Framework Synthesis; Diffusion Theory (Rogers, 2003)*	Organizational Health Environment Model (Golaszewski, Allen, & Edington, 2008)	Hughes et al. 2011	PHLAME, an evidence-based and NIH-funded program.	N/A
Sampling Strategy	*Purposive*	Purposive	Purposive	Purposive	Convenience
Data Collection and Analysis	*Best-Fit Framework Synthesis (BeHEMoTh/SPIDER; and Thematic Analysis); CERQual Analysis; and Saturation of Inquisition Test.*	Qualitative interviews	Qualitative interviews	Qualitative interviews	Qualitative interviews
Software Used	*Dedoose*	ATLAS.ti	N/A	N/A	N'Vivo
Number of Coders	*1 and an external reviewer*	2	2 or more	2	1
Output	*Theoretically-grounded and empirically-tested analytical themes and contextualized framework.*	Themes that underlie the decisions to adopt, implement, and continue WPs.	Themes that underlie the adoption-decision process at SMBs.	Factors that likely mediate the adoption of WPs amongst Fire Departments.	Basic description of potential opportunities and barriers for adopting and/or integrating programs.

Table 3. *Other Study Characteristics*

	Hughes	Witt	Kuehl	Nelson
Year	2011	2013	2013	2015
Geographical Area	Pacific Northwest	Wichita, KS	Oregon and Washington	Greater Minneapolis Area
Industries Represented	Manufacturing	Manufacturing, Schools, Social Services, Financial Services, Public Administration, and Rehabilitative Services	Fire Departments	Manufacturing
Company Sizes	75 – 800 employees	29 – 880 employees	31 – 187 employees	"less than 50" – 500 employees
Number of Participants	24	12	36	14
Previous Adopter	Yes: 3 No: 21	Yes: 11 No: 1	Yes: 12 No: 24	Unclear
Saturation Reached	Yes	Yes	Unclear	Unclear
COREQ Reporting Score	21	18	16	16
Key Findings	Vendors and practitioners should: 1) appeal to company financial success; 2) leverage insurers and benefits brokers as channels for communication and enhanced facilitation; 3) target information to senior management and human resource personnel; and 4) study and demonstrate the effectiveness of WPs as they relate to potential SMB adopters.	The concept of "engaged versus unengaged in WPs" lies on a continuum rather than lending itself to dichotomous categories. *Cost* was the predominant factor behind WPs, with employee well-being noted as secondary. Employee buy-in was a concern for many DMs when considering WPs. Key influencers on the adoption-decision process appeared to be distributed throughout various roles in the organization including employees.	Effective mailer connection, champion presence, and a willing fire chief were all noted as factors likely to influence the adoption-decision process.	Factors that influenced the adoption of combined programs were similar to that of individual programs. The same factors that facilitate implementation in some cases may be seen as barriers, albeit for different reasons, in other cases.

Quality Appraisal and Alignment Scores

Quality of Reporting was appraised using a COREQ Checklist (Tong, Sainsbury, and Craig, 2007). Alignment Scores were added to the checklist to identify points of misalignment between original studies and this study's line of inquisition. The study by Nelson et al. (2015) scored considerably low in reporting as well as alignment. As such, findings that relied on this study were analyzed through a sensitivity analysis (see Table 6). The completed quality appraisals with Alignment Scores are offered in APPENDIX C.

Data Analysis, Evaluation, and Synthesis

Content analysis led to the identification of several descriptive and emergent themes. The unit of analysis for coding was *potential barrier to adoption* as based on definitions provided by the *a priori* framework. Emergent themes were operationalized as codes and integrated into the framework as relevant. All codes reflect the basis of themes and are identified in the framework as *Potential Barriers*. The code application is illustrated in Figure 5.
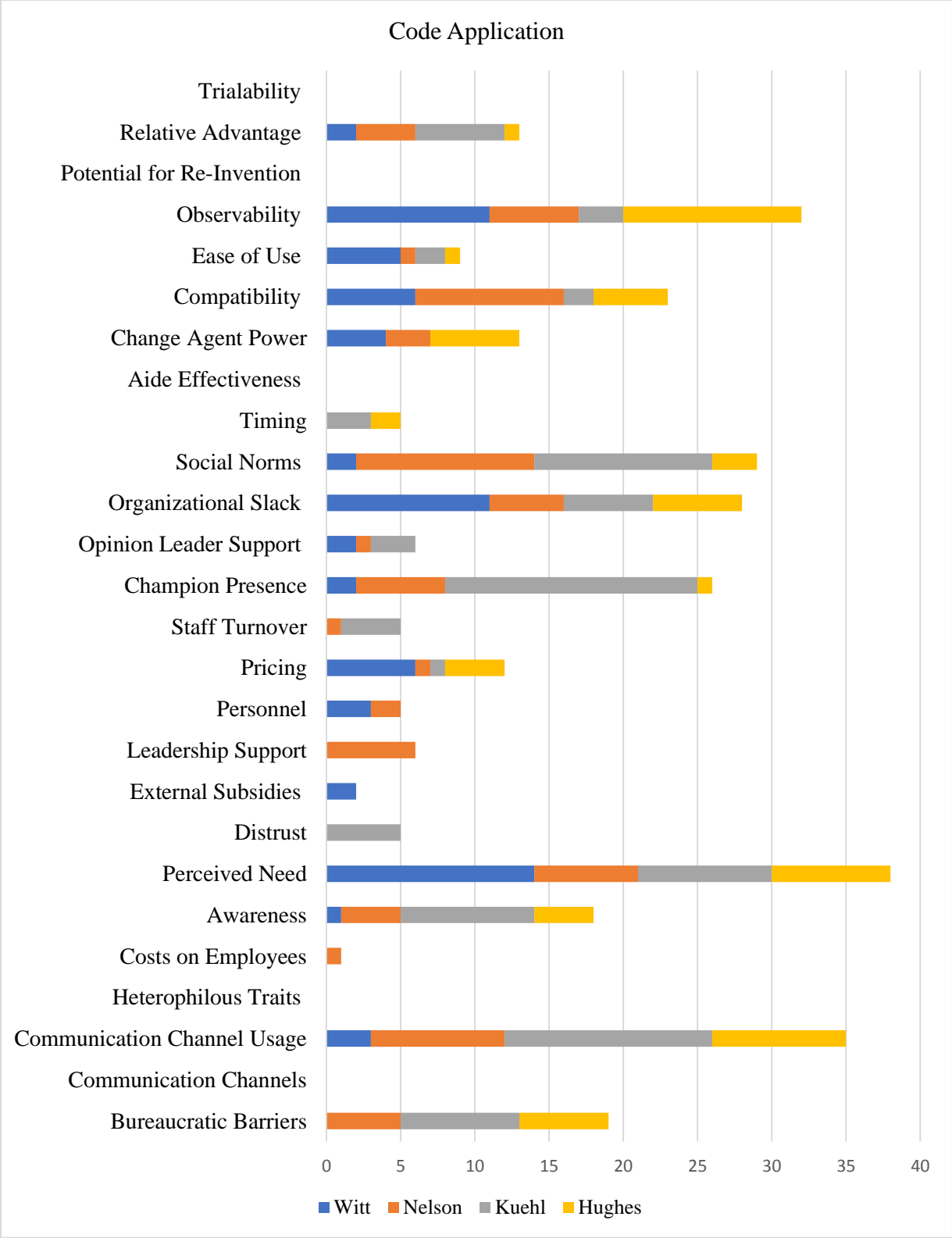
Figure 5. Code application across studies.

Figure 5 illustrates total code application across studies. *Perceived need* was the most frequently used code and appeared to transfer well across studies. Code application led to the generation of several descriptive, emergent, and analytical themes.

Descriptive Themes

Communication Channels (n=35), *Organizational Slack* (n=28), *Perceived Need* (n=38), *Observability* (n=32), *Champion Presence* (n=26), *Social Norms* (n=29), and *Compatibility* (n=23) were the most frequently mentioned themes regarding barriers to adoption of WPs. All were readily informed by DT and appeared to transfer well across studies. As such, these potential barriers warrant a degree of consideration amongst practitioners and researchers alike. However, this should not undermine less frequently noted barriers when considering individual client needs. For example, *Staff Turnover* (n=5) was noted far less frequently, but appeared to have a detrimental effect on the adoption process (Kuehl, Mabry, Elliot, Kuehl, & Favorite, 2013). This less common yet highly significant barrier may easily be overlooked if not conscientiously addressed.

Emergent Themes

Data extraction and analysis led to the identification of themes that were not directly aligned with the *a priori* framework. These important themes emerged directly from the empirical evidence and were defined to reflect their source of support. Definitions of these themes were compared to the existing definitions of codes supplied by the *a priori* framework.

This allowed new codes to be defined in a manner that complimented the original codes and enhanced the overall framework. These resulting *emergent themes* are described below.

Bureaucratic Barriers (n=19) and *Pricing* (n=12) were common emergent themes. These themes also directly complimented other concepts supplied by DT. For example, *pricing* can be directly compared to *organizational slack* and *observability* to gain a more comprehensive as well as precise view of the nuances that underlie the cost equation. These themes provided new, contextually-rich knowledge that built on the well-established theory of DT. *Bureaucratic Barriers* offered a more precise element to its broader counterpart, "communication channels". This more precise look at underlying root barriers provides much deeper and actionable information than the generic concepts often offered in the literature.

Emergent themes were a high value outcome from the investigation. Accordingly, *Bureaucratic Barriers* and *Pricing* were added to the framework and appeared to influence at least one analytical theme. Other emergent themes that influenced the framework and/or analytical themes were *Staff Turnover* (n=5), *Distrust* (n=5), *Timing* (n=5), *Personnel* (n=5), and *External Subsidies* (n=2).

However, not all emergent themes were added to the framework or incorporated into analytical themes. For example, *Leadership Support* (n=6) and *Cost to Employees* (n=1) was difficult to define within the given context. Sensitivity analysis showed that support for these themes came exclusively from one study that was also rated lowest in alignment. This lack of alignment may have been the cause of the ambiguity of the finding. A review of the original source validated this concern, leading to rejection of the codes.

Co-Occurrences

It warrants noting that some codes occurred together at relatively high rates. This is likely

due to one of two reasons: 1) there was overlap in the operational definition of codes; or 2) the

concepts are related in practice somehow. Either would warrant further examination of the codes

to determine the nature of the interaction. Importantly, any relationship between codes may have

an influence on other findings. Likewise, these relationships may provide practical implications

for practitioners. For these reasons, co-occurrences were a key element of this investigation. An

illustrated view of the code co-occurrence output from Dedoose is provided below in Figure 6

and Figure 7.

Codes	Bureaucratic Barriers	Communication Channels	Communication Channel	Heterophilous Traits	Costs on Employees	DM Characteristics	Awareness	Perceived Need	Distrust	Exemplar Quotes	External Subsidies	Leadership Support	Personnel	Pricing	Staff Turnover	System Characteristics	Champion Presence	Opinion Leader Support	Organizational Slack	Social Norms	Timing	Vendor Characteristics	Aide Effectiveness	Change Agent Power	WP Characteristics	Compatibility	Ease of Use	Observability	Potential for Re-Invention	Relative Advantage	Trialability	Totals
Bureaucratic Barriers			5									2					3		1	1						2		1				15
Communication Channels																																
Communication Channel	5						10	5	2			1		2			1	2	2					9		4	1	3		1		48
Heterophilous Traits																																
Costs on Employees																																
DM Characteristics																																
Awareness			10					4						1	2				1					5				3		3		29
Perceived Need			5				4					1		2			1		1	8				5		3		7		5		42
Distrust			2																1													3
Exemplar Quotes																																
External Subsidies														1					1													2
Leadership Support	2		1					1									3									3						10
Personnel																	3		3							1	3					10
Pricing							1	2			1								6							2		1		1		14
Staff Turnover			2				2												1													5
System Characteristics																																
Champion Presence	3		1					1				3						1	2	1						1	3	1		1		18
Opinion Leader Support			2														1											2				5
Organizational Slack	1							1			1		3	6			2	1		2				1		2	2	2				24
Social Norms	1		2				1	8	1			3		1			1		1					1		13				5		38

Figure 6. Screenshot of code co-occurrences: group one.

Codes	Bureaucratic Barriers	Communication Channels	Communication Channel	Heterophilous Traits	Costs on Employees	DM Characteristics	Awareness	Perceived Need	Distrust	Exemplar Quotes	External Subsidies	Leadership Support	Personnel	Pricing	Staff Turnover	System Characteristics	Champion Presence	Opinion Leader Support	Organizational Slack	Social Norms	Timing	Vendor Characteristics	Aide Effectiveness	Change Agent Power	WP Characteristics	Compatibility	Ease of Use	Observability	Potential for Re-Invention	Relative Advantage	Trialability	Totals
Timing																			2													2
Vendor Characteristics																																
Aide Effectiveness																																
Change Agent Power			9				5	5						2					1	1						1		5		1		30
WP Characteristics																																
Compatibility	2		4					3				3	1				1		2	13				1			4	6		1		41
Ease of Use			1										3				3		2							4		1		1		15
Observability	1		3				3	7						1			1	2	2					5		6	1					32
Potential for Re-Invention																																
Relative Advantage			1				3	5						1			1			5				1		1	1					19
Trialability																																

Figure 7. Screenshot of code co-occurrences: group two.

Social Norms and *Compatibility* occurred together the most frequently (n=13). After a re-examination of these *a priori* framework and empirical sources, this appeared to be due to the influence culture has on establishing formal processes and/or vice versa. A proposed mediator of *Compatibility* was its congruence with formal and informal processes which are included in the definition of *Social Norms*. This would support the notion that a WP's congruence with workplace culture will have a substantial influence on the adoption of WPs. This leads to a strong support for one analytical theme and the recommendation that practitioners thoroughly assess such norms well before promoting WPs to the potential adopter.

Communication Channels and *Awareness* occurred together at the second highest rate (n=10). The two codes also appeared to act in a sequential fashion; those who did not leverage communication channels were less likely to be aware of WP availability and processes. As such, targeting the former may solve for the latter.

Pricing and *Organizational Slack* occurred together at a relatively high rate (n=6). This seemed natural, since the two can be directly compared to assess program feasibility. Of note, "Cost" was noted throughout the initial review of literature, but led to shallow insight until it was differentiated into more *root* barriers. For example, some informants may use *cost* to indicate the price difference between programs and alternatives while others may mean *cost* as it compares to what they have to spend. Others may imply cost-benefit interaction, as in the cost must not outweigh the expected benefit. For this reason, more discrete variables were developed for the *a priori* framework. Such discrete concepts may aid communication between researchers and informants and allow for a more precise and actionable solutions. *Organizational Slack*, *Observability*, and *Relative Advantage* all offered precise definitions that better informed the

generic concept of cost. *Pricing*, however, was an emergent theme that further informed the concept of "cost". As such, this was a highly valuable finding.

Most practically, *Observability* occurred with *Perceived Need* (n=7), *Change Agent Power* (n=5), and *Compatibility* (n=6) relatively frequently. This was deemed practical because of its actionable and correctable nature as WP and vendor characteristics (predominantly). For example, *Social Norms* are highly relevant yet difficult for vendors to influence. However, *Observability* can be directly addressed by vendors through demonstrations, documented case studies, referrals from exemplar clients, and citing relevant literature—all of which may also enhance *Change Agent Power*.

Change Agent Power and *Perceived Need* occurred together 5 times. Indeed, a substantial portion of the Change Agent's role may be to illuminate a need for their proposed innovation. However, a need is not always present—or at least perceived as such. In either case, this warrants consideration on the part of the vendor and Change Agent to be sure the agent is well-prepared to diagnose otherwise hidden needs within potential adopting organizations.

Codes that appeared to affect the broadest range of other codes were *Communication Channels* (co-occurring with 14 other codes); *Social Norms* (co-occurring with 12 other codes); *Organizational Slack* (co-occurring with 12 other codes); *Champion Presence* (co-occurring with 11 other codes); *Perceived Need* (co-occurring with 11 other codes); and *Compatibility* (co-occurring with 12 other codes). As such, relationships between and amongst these highly transferable codes were prioritized during the generation of analytical themes.

Sensitivity Analysis

One study (Nelson, Allen, McLellan, Pronk, & Davis, 2015) demonstrated relatively low Quality of Reporting and Alignment Scores. As such, it was removed from the analysis to test its effect on findings. After removal, the relative frequency and transferability of most codes across studies was still apparent with the exception of two emergent codes. These emergent codes were then re-assessed based on the evidence from the Nelson study before integration into the *posteriori* framework or analytical themes. Codes from which 34.1% (1 standard deviation of a normal curve) or more occurred in the Nelson et al. (2015) study were also identified for further evaluation. The output of the sensitivity analysis is provided below in Table 4.

Table 4. *Sensitivity Analysis*

Code	Nelson et al.	All Studies Combined	Percentage Coming from Nelson's Findings
Bureaucratic Barriers	5	19	26.32%
Communication Channels	6	35	17.14%
Costs to Employees	*1*	*1*	*100.00%*
Awareness	4	18	22.22%
Perceived Need	7	38	18.42%
Distrust	0	5	0%
External Subsidies	0	2	0%
Leadership Support	*6*	*6*	*100.00%*
Pricing	1	12	8.33%
Staff Turnover	1	5	20.00%
Champion Presence	6	26	23.07%
Opinion Leader Support	1	6	16.67%
Organizational Slack	5	28	17.85%
Social Norms	*12*	*29*	*41.37%*
Timing	0	5	0%
Compatibility	*10*	*23*	*43.47%*
Ease of Use	1	9	11.11%
Observability	6	32	18.75%
Relative Advantage	4	13	30.76%
Trialability	0	0	0%

Note. Italics identify codes that were predominantly based on findings by Nelson et al (2015).

Further review of the original source showed two codes, *leadership support* and *cost to employees*, were emergent themes that did not align well with this study. Specific points of misalignment can be found in the augmented COREQ appraisals in APPENDIX C. In any case, these two themes were not added to the new framework, but noted here for consideration by the reader.

Results also demonstrated that *Social Norms* and *Compatibility* relied heavily, but not exclusively, on Nelson et al (2015) for support. After removing the Nelson study from the synthesis these codes were still apparent in other studies and of high practical value to this study's line of inquisition. Therefore, *Social Norms* and *Compatibility* were kept in the *posteriori* framework and still influenced at least one analytical theme.

Answering the Research Question: Analytical Themes

Throughout the analysis, a number of themes tended to heavily influence one another or the adoption-decision itself. Certain codes also occurred somewhat universally, presenting themselves in multiple contexts across all studies with no contradictory evidence. Such codes were used to form the analytical themes by synthesizing their operational definitions with their empirical evidence and considering their relationships to the phenomenon. Analytical themes are the primary output of the methods used and provide the answer to the research questions. Namely, analytical themes provide the answers to the following questions:

- *What root barriers are hindering adoption of WPs amongst SMBs?*

- *Are some perceived barriers more universal, or substantial, than others?*

Analytical Theme: Perceived Need Mediates the Magnitude of Barriers to WP Adoption

This was well-supported by *Perceived Need* occurring frequently across all studies and in

conjunction with all other key barriers. Even when this code was not explicitly stated, it was

often implied. The definition of the code also lends itself directly to the nature of the

phenomenon, potentially mediating the effect of other codes listed in the framework.

Analytical Theme: Ineffective Communication of Program Availability and Effectiveness is Hindering WP Adoption

This was supported by the frequently occurring codes *Communication Channels*,

Awareness, and *Observability*. Ignorance likely hindered informants from offering *Awareness* as

a perceived barrier without being prompted, making this theme potentially more substantial than

what is observed during interviews. Either way, this theme was well-established despite this

potential for understatement. Therefore, it is plausible that awareness of availability,

effectiveness, and need are all undermined by ignorance. Furthermore, it is plausible that this can

often be corrected through improved use of communication channels.

Analytical Theme: Misalignment between Program Price Points and SMBs' Available Resources is Hindering WP Adoption

This was supported by the transferable theme of *Organizational Slack* and substantiated

by the emergent theme of *Pricing*. Both codes occurred across all contexts, and were well-

supported in higher-quality studies. Further, the two codes offer complimenting perspectives on

the more generic barrier of *cost*.

Analytical Theme: Inefficient Alignment of WPs with Organizational Processes and Roles is Hindering Adoption

This was supported by the recurring themes of *Champion Presence*, *Compatibility*, *Social Norms*, and *Personnel*. These transferable and relevant themes directly informed this study's line of inquisition in a highly practical manner. However, some concern was raised during the CERQual Analysis. As a result, this theme was downgraded in confidence.

<center>*Confidence in Primary Findings*</center>

Confidence of analytical themes was assessed via the CERQual Analysis (Lewin et al. 2015). This offered an added dimension of relative order to the findings. Results are illustrated below in Table 5 along with brief explanations of the ordinal ranking. While confidence scores are not considered a primary finding, they are still interesting to consider by both researchers and practitioners. At the very least, these confidence scores offer a point of reference for further discussion. A more thorough description of the CERQual scoring rationale is offered in APPENDIX D and APPENDIX E.

Table 5. *CERQual Summary: Analytical Themes with Respective Confidence Scores*

Objective: To identify, appraise, and synthesize qualitative research on small- to medium-sized business decision-makers' perceived barriers to adoption of wellness programs.				
Perspective: Perceptions of small- to medium-size businesses.				
Included Informants: Decision-makers at small- to medium-sized businesses.				
Review Finding: Analytical Themes	Confidence in the Evidence	Explanation of Confidence	Studies Contributing to the Finding	Codes that Captured the Theme
Perceived need mediates the magnitude of barriers to WP adoption.	[*]*High*	This was consistently supported and appears to be a universal theme.	Hughes et al. 2011; Witt et al. 2013; Kuehl et al. 2013; Nelson et al. 2015	Perceived Need; Observability
Ineffective communication of program availability and effectiveness is hindering adoption.	*Moderate*	This was consistently supported and appears to be a universal theme. Minor concern regarding quality of data captured by the Kuehl study.	Hughes et al. 2011; Witt et al. 2013; Kuehl et al. 2013; Nelson et al. 2015	Communication Channels; Observability; Awareness; Relative Advantage
Misalignment between program price points and SMBs' available resources is hindering adoption.	*Moderate*	Minor concern over quality of data captured by *pricing* codes. This is due to potential interpretive validity of the cost versus price construct.	Hughes et al. 2011; Witt et al. 2013; Kuehl et al. 2013; Nelson et al. 2015	Organizational Slack; Pricing
Inefficient alignment of organizational processes and roles is hindering adoption.	*Low*	Minor concern due to adequacy of data from sources other than Kuehl et al. 2013. The homogenous sample from Kuehl (firefighters) is limited in transferability of findings. Restraint is warranted due to the reliance on potentially indirect relevance of findings from the Nelson study.	Hughes et al. 2011; Witt et al. 2013; Kuehl et al. 2013; Nelson et al. 2015	Champion Presence; Compatibility; Social Norms

[*]It is highly likely that the review finding is a reasonable representation of the phenomenon of interest.

<u>Saturation of Inquisition</u>

Interview prompts were analyzed to determine the extent to which domains of the *a priori* framework were investigated. Results showed that certain areas of the a priori framework were not directly investigated at all, while others were thoroughly examined. This was valuable information for a few reasons. For one, domains that were not assessed may provide important knowledge yet to be uncovered. Second, removal of unsupported codes is warranted during a BFS, but if these codes were not investigated they may still be highly relevant. The fact that these codes were informed by a well-established theory further supports this position. Therefore, removing DT-supported concepts from the framework was deemed inappropriate if these concepts were not adequately addressed. Namely, areas that were not addressed included: *Trialability; Potential for Re-Invention;* and *Aide Effectiveness. A priori* codes and the interview items that probed them are provided below in Table 6.

It is important to note before moving forward that no interview items from Keuhl et al. (2013) were found and items from Nelson et al. (2015) were incomplete. It could be the case that these interviews probed the domains. However, it was still deemed inappropriate to remove these otherwise-established concepts without explicit justification.

Table 6. *Saturation of Inquisition: A Priori Codes Addressed by Interview Items*

Code	Hughes/Witt	Nelson	Kuehl
Champion Presence	--	Item 10	n/a
Opinion Leader Support	35, 36, 39	n/a	n/a
Social Norms	Items 3, 4, 5, 6, 8, 16, 17, 18, 28, 31, 40	Item 4, 8	n/a
Organizational Slack	Item 7, 8, 9	Item 7	n/a
Extent of Perceived Need	Item 6, 7, 8, 9, 21	n/a	n/a
Awareness	Item 6, 7, 8, 9, 14	Item 5	n/a
Communication Channels	Item 12, 13, 14, 15, 16, 17	Item 2	n/a
Extent of Heterophilous Traits	Item 42	n/a	n/a
Relative Advantage	Item 7, 8, 9	n/a	n/a
Compatibility	Item 26, 27, 29, 32	Item 3, 4, 6, 8	n/a
Ease of Use	Item 23, 24	n/a	n/a
[a]Trialability	--	n/a	n/a
Observability	Item 5	Item 9	n/a
[a]Potential for Re-Invention	--	n/a	n/a
Extent of Change Agent Power	Item 7, 9	n/a	n/a
[a]Aide Effectiveness	--	n/a	n/a

[a]These items indicate gaps in research.

Posteriori Framework

Themes were aggregated, analyzed, and compared to the *a priori* framework. Emergent themes were added to the framework while aligning the new operational definitions with pre-existing concepts. Emergent themes are indicated via right-alignment of the code. The result was a more contextualized framework that thoroughly and directly informed the given problem of practice. This *posteriori* framework is offered below in Table 7.

Table 7. *Posteriori Framework*

Potential Barrier	Key Elements	Operational Definition	Potential Mediating Factor
System Characteristics			
Champion Presence is Inadequate	Personnel (Internal)	An advocate (current or potential) who aides in the facilitation of WPs (paid or unpaid) is not present within the system or is generally ineffective.	The level of influence (formal or informal) and general abilities of the champion.
Opinion Leader Support is Inadequate or Rejects the WP	Personnel (Internal or External)	A member of the social system who has influence over acceptance of new ideas is not adequately supporting adoption or is opposing it to some degree.	The extent of influence the Opinion Leader has on DMs and employees.
Social Norms Do Not Support Adoption	Social System	Patterns of behavior within the systems (internal and external) which DMs operate do not support adoption of WPs or aspects of their operation.	Innovativeness of the adopter; influence from Opinion Leaders; and employee characteristics.
Organizational Slack is Inadequate	Organizational Resources	The organization does not have adequate resources devoted to WPs.	Shifts in organizational revenue; and potential for capacity building.
External Subsidies are Inadequate	External Resources	External mechanisms for capacity building are not present, inadequate, or blocked.	External processes or policies that facilitate capacity building or the use of external resources.
DM Characteristics			
Need is Perceived as Inadequate	Knowledge of Problem	The DM does not perceive a problem to exist, or does not perceive the problem to be a worthy issue to correct.	The extent of how actively the DM seeks relevant knowledge; knowledge of presence of the problem; knowledge of the severity of the problem; and underlying beliefs and attitudes toward the problem.
Awareness is Inadequate	Knowledge of Solution	The DM does not perceive a solution to exist, or does not perceive the solution as worthy to implement.	Knowledge of WPs; access to external communication channels; and use of inter- and intra-system communication channels.

Potential Barrier	Key Elements	Operational Definition	Potential Mediating Factor
Communication Channels (Inter- and Intra-System)			
Communication Channels are Not Leveraged Adequately	Inter- or Intra-System Interactions	Either party does not actively and/or effectively communicate within or outside their social system.	Availability of interactions between systems; and alignment of communication strategies employed within and between systems.
Bureaucratic Processes Hinder Adoption	Intra-system Communication Processes	Bureaucratic processes may hinder timeliness, completeness, and authenticity of information exchange.	The number of intra-organizational levels to the communication processes; and number of entry and exit points in the communication system.
Staff Turnover Hinders Adoption	Staff Turnover	Staff turnover results in dramatic loss of information from the system.	Presence and effectiveness of strategies (formal or informal) to retain personnel, or their knowledge, in the system.
Heterophilous Traits Reduce Empathy	Personality Traits	The culture, social status, or other background elements inhibits empathy and communication.	The extent of cultural and social differences.
Distrust of Proposed Innovation or Innovator	Past Experience	DMs have already formed negative perceptions of a vendor or product from previous experiences that involve similar processes, materials, or actors.	Interactions and outcomes with previous information or vendors in the market.

Potential Barrier	Key Elements	Operational Definition	Potential Mediating Factor
WP Characteristics			
Relative Advantage is Inadequate	Observable Benefits	The advantages to adopting the WP do not outweigh alternative decisions.	Perceived effectiveness of the WP over current solutions; presence of alternatives (direct or indirect); and perceived effectiveness of alternatives.
Pricing is Not Considered Competitive	Market Position	Price point does not match the organizational slack or effective position in the market.	Flux in the market or organizational slack.
Compatibility is Inadequate	Integration with Values, Beliefs, and Daily Processes	The WP does not appear to integrate well within the system.	Perceived processes involved in the WP; perceived underlying principles of the WP; and expected acceptance amongst users.
Ease of Use is Inadequate	Complexity	The WP is perceived as too complex to employ efficiently and effectively.	Background knowledge of the DM; and the WP agent's ability to educate the DM.
Personnel is Inadequate	Personnel	Availability of potential, formally compensated, internal facilitators is inadequate.	Extent of resources needed versus resources available to supply adequate personnel.
Trialability is Inadequate	Observable Benefits	The WP is not available to try.	Previously observed effects; and third-party subjective notions regarding the WP.
Observability is Inadequate	Observable Benefits	The WP is not perceived as having shown demonstrable benefits.	Extent of knowledge necessary to evaluate the WP; demonstrable processes and effects; and tangible, as opposed to abstract, nature of WP outcomes.
Potential for Re-Invention is Inadequate	Integration with Values, Beliefs, and Daily Processes	The WP is not perceived as customizable to the changing needs of the organization.	Anticipated future needs of the organization.

Potential Barrier	Key Elements	Operational Definition	Potential Mediating Factor
Vendor Characteristics			
Change Agent Power is Inadequate	Personnel (Internal)	The Change Agent's abilities to gain decision to adopt is low.	Ability to negotiate; heterophilous background with DM; and access to important members of the system.
Aide Effectiveness is Inadequate	Personnel (Internal or External)	The Change Agent's Aides are not able to build strong enough informal connections to the Opinion Leaders or DMs.	Congruence of cultural and personal background with adopter personnel.
Timing is Inadequate	Temporal Aspects of Solicitation and Delivery	Timing of sales strategies are not aligned with systematic opportunities or are employed in the presence of temporal barriers.	External and internal fluctuations in the system that mediate barriers, opportunities, and resources.

CHAPTER FIVE: DISCUSSION

The overarching goal of this paper was to explore the utility of a novel combination of methods in the context of a real-world problem of practice. These methods were used to synthesize qualitative research findings through a most-relevant theoretical lens to answer two specific research questions. This chapter discusses the rationale behind the methodology and examines the utility of specific methods in isolation as well as in combination with one another.

Augmented Best-Fit Framework Methodology

The BFS method was chosen over other forms of qualitative synthesis for its pragmatic capacity to thoroughly inform this paper's problem of practice. It accomplished this by objectively analyzing evidence through the lens of a most relevant theory. "Most relevant theory" is used thoughtfully here, because the theory was chosen specifically for the problem of practice in a systematic manner as part of the BFS method. The selection and use of this theory is arguably a predominant strength of BFS over other forms of qualitative synthesis.

Some evidence has shown that a comparison of empirical findings to well-established theoretical concepts may lead to much richer analysis than translating observations across contexts alone (Lorenc, Pearson, Jamal, Cooper, & Garside, 2012). The final synthesis in the BFS also directly built upon the chosen theory to provide a contextually rich version thereof. This *posteriori* framework becomes a valuable tool that can then be replicated or further refined by future researchers.

The systematic selection of the given theory allowed a degree and transparency for the reader to better understand why and how the theory was chosen and operationalized. The use of this theory as a coding framework added a level of objectivity to secondary analysis by

71

supporting consistency during coding. This framework also acts well as a reference point for discussion amongst researchers, or as a means to evaluate any discrepancies during the coding process. The framework also offers the reader clear evidence for the justification of themes generated during the synthesis portion of the study.

This paper proposed supplemental methods to bolster findings and enhance the outcome, reporting, and rigor of the BFS process. It is important to note that many of the supplemental tools only support the BFS and do not stand alone. These additional methods also added to the study rather than directly influencing it. If any method or procedure is later deemed inappropriate, the additional method and its results can be omitted from the review and the main findings from the BFS would remain predominantly intact. This was deemed important due to the novelty of certain analytical tools and methods.

SPIDER: Searching the Evidence

SPIDER was initially chosen for its standardized approach. However, certain modifications were made to the SPIDER search strategy to accomplish the task of the study. Importantly, these modifications were only made after the strategy was deemed inadequate. For example, the criteria indicated by the original SPIDER approach was *small- to medium-sized business* AND *decision makers*. This search was conducted and repeatedly delimited, but ultimately returned no viable evidence. Therefore, *wellness program* OR *health promotion* was substituted (see APPENDIX B). This is an important modification for the reader to note. Equally so, it is important to note that such modifications appear to be in line with the systematic review process so long as they are justified and clearly stated. This transparency is a fundamental aspect of systematic reviews as it allows for reproducibility by the reader.

The SPIDER approach could be a valuable asset to the scientific community when selecting from large quantities of studies. For example, its universal use would lead various scientists investigating a given phenomenon to find similar results. In theory, this would expedite the scientific consensus as to what is known or not known on a topic. However, while this approach may be ideal in theory, it was not shown to be practical for this specific research question or problem of practice.

Other standardized search approaches are offered, but the variety of approaches seems to contradict why they were created in the first place. If the goal was universal strategies to enhance reliability, this goal becomes less and less attainable with every strategy proposed. Either way, researchers may be inclined to employ standardized strategies first and making modifications only as necessary. After all, explicitly identifying and reporting the search strategy is arguably the more important aspect and a fundamental component of a systematic review's audit trail.

Including vs Excluding Grey Literature

Various forms of inquiry may search through grey literature such as reports, working or unpublished papers, and government documents. This would likely benefit some research methods that aim for further artifacts to inform their conclusions. However, it was explicitly stated by Carroll et al. (2013) that findings for BFS should be limited to primary evidence. This seems debatable as a more comprehensive search may uncover novel and relevant information. However, an implied aim of the A-BFS was to discover what is known *as a result of* the scientific and qualitative inquiry. This was deemed most appropriate so that future investigators may better understand gaps in the current research. Put another way, an implied goal of the

73

synthesis was to enhance future *scientific* inquiry by providing evidence of gaps in inquisition; it would be misleading to fill such gaps with grey literature.

The synthesis also sought out the perceptions of DMs to gain an unadulterated view of the barriers they perceived as most important. In doing so, this would better illuminate most relevant barriers for WP vendors to address. The nuances that are inherent to perception offer important insights for practitioners and detecting such nuances is a potential strength of qualitative methods. However, the very nature of seeking *root barriers* warranted the utmost scrutiny when collecting evidence to synthesize. Secondary analysis of perceptions noted throughout grey literature would have lacked such scrutiny due to the ambiguous manner in which the primary evidence could have been accrued. A qualitative synthesis, after all, can only be as valid as the original evidence it synthesizes. In this respect, grey literature may have offered more misinformation than legitimate insight.

It warrants reiterating here that *perceived* barriers may be as significant as, if not more than, their more tangible counterparts. While many potential obstacles to the adoption of WPs exist, a targeted focus on the perceptions of DMs may offer more direct insight as to what is actually limiting WP adoption. Targeting these specific perceived barriers of DMs may move practitioners away from the "data rich, decision poor" paradigm and toward a more direct, efficient, and effective selection of strategies.

Quality of Reporting and Alignment Scores

In line with the nature of systematic reviews and the resources consulted on qualitative synthesis (Carroll et al. 2013; Sandelowski & Barroso, 2007; Thomas & Harden, 2008), primary studies were appraised for quality as it related to thoroughness of reporting. While findings were

not excluded based on quality, a low Quality of Reporting score was considered when drawing interpretations from the related evidence. For example, if a given theme was based predominantly on an ill-reported study, the original source supporting the theme was re-evaluated. If warranted, the original study would then go through a sensitivity analysis to test for its influence on this study's findings. If findings were kept hesitation was still warranted and findings may have been downgraded in confidence depending on what was missing from the reports.

Alignment Scores were added to the COREQ quality appraisals to gauge how well a primary study aligned with the synthesis's line of inquisition. This appeared to complement the Quality of Reporting checklist and add another dimension of quality to the appraisal. For instance, a paper may have reported the given elements of a study quite well, but certain elements may have been misaligned with the synthesis's aim. Such a case would warrant a high Quality of Reporting score, but a lower score for alignment. Studies that scored relatively low in alignment were also analyzed through a sensitivity analysis. An example of a misalignment is provided below in Figure 8.

COREQ CHECKLIST: DOMAIN 3 (NELSON ET AL. 2015)

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 3: Analysis and Findings				
Data analysis				
Number of data coders	24	How many data coders coded the data?	0	0
Description of the coding tree	25	Did authors provide a description of the coding tree?	0	0
Derivation of themes	26	Were themes identified in advance or derived from the data?	1	1
Software	27	What software, if applicable, was used to manage the data?	1	--
Participant checking	28	Did participants provide feedback on the findings?	0	0
Reporting				
Quotations presented	29	a.Were participant quotations presented to illustrate the themes/findings? b.Was each quotation identified? e.g. participant number	a. 1 b. 1	a. 1 b. 1
Data and findings consistent	30	Was there consistency between the data presented and the findings?	1	1
Clarity of major themes	31	Were major themes clearly presented in the findings?	1	0
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	1	1
		Domain Total	7	5
		Report Total	16	12

Figure 8. *Screenshot of an example of a misalignment.*

The red star in Figure 8 highlights a misalignment that occurred in the study by Nelson et al. (2015). While the major themes were clearly reported, the themes and/or manner in which they were investigated did not directly align with this study's line of inquisition. Importantly, this is a relatively subjective judgment that does not necessarily indicate anything on its own.

However, it provided a valuable reference point for further investigation and offered another level of transparency for the reader.

The value of this indicator will likely become progressively more apparent as more information is synthesized. This would be especially true in studies with multiple authors. For example, this indicator may support communication and reduce the cognitive load that comes with rigorous analysis of dozens of studies between multiple authors. This analytical device of sorts may also offer a new and simple tool to novice researchers. Full Quality of Reporting and Alignment Scores are provided in APPENDIX C.

Excluding vs Including Low Quality Studies

As noted earlier, inclusion of lower quality studies is contrary to the methods used in systematic reviews of quantitative research, but consistent with the methods of qualitative syntheses and resources used for this review (Carroll et al. 2013; Cooper, 1998; Sandelowski & Barroso, 2007; Thomas & Harden, 2008). The rationale for including lower quality studies appears to be based on the decision to err on the side of inclusion rather than exclusion of findings based on *a priori* criteria. Considering the exploratory nature of this review, and qualitative research in general, any *a priori* judgments on quality are likely incomplete by default. The researcher becomes more informed as the research is conducted, supporting the notion that *posteriori* decisions may offer less bias and a richer analysis of the problem.

Further, assessments of methodological rigor and quality as it relates to anything other than thoroughness of reporting could only be considered subjective at best—or ignorant at worst—from the once-removed position of secondary analysis. The quality of such elements is also a key focus of the peer-review process. Although that process itself is not infallible, it does

indirectly justify the somewhat necessary assumption of methodological quality to some extent. "Somewhat necessary" is used here thoughtfully due to 1) again, the once removed perspective of secondary analysis, and 2) the skill set of this paper's author as a doctoral candidate and novice researcher.

Considering these potential caveats, the quality of a study's reporting—not methodology—is still critically judged in this paper. However, this is analyzed more for formative purposes rather than summative. For example, any limitations that may stem from inadequate reporting would be integrated into the finding itself rather than completely thrown out. Put another way, Quality of Reporting is seen here as a variable to be considered during the synthesis rather than a measurement from which to base exclusion of peer-reviewed, published findings.

Notably, Quality of Reporting scores are solely intended to measure positive support via the accumulation of specific descriptive elements. In other words, the scores support the extent that something is likely true. Conversely, scores are not intended to evaluate findings for the sake of rejection, nor should scores reduce the meaningfulness of findings. In other words, the scores do not support the extent that something is likely false. This is an important distinction—scores measure support as opposed to rejection. After all, highly valuable studies may still receive lower quality scores due to potentially irrelevant factors such as word limits of journals or the author's own foresight regarding reporting.

BeHEMoTh: The A Priori Framework

A search was conducted for the most relevant theory or model that informed the problem of practice. Of note, one paper summarized many highly relevant theories and acted as a valuable

reference point during the selection process (Batras, Duff, & Smith, 2014). The initial review of literature and this study's research questions directed the selection. Namely, key elements that were sought included:

- the extent of established literature on the theory;

- the theory's potential to offer practical implications; and

- the theory's ability to distinguish the phenomenon of program adoption from other processes.

After papers were reviewed, DT was selected. *Diffusion of Innovations* (Rogers, 2003) was then consulted for a more direct interpretation of the theory. Concepts and elements that were relevant to potential barriers of WP adoption were operationally defined for use as a coding (*a priori*) framework. This process—performing the initial review and then locating a theory that best informed it—was an important part of developing the best possible set of codes. Codes were labeled in the framework as *Potential Barriers*. Few modifications were needed, and no further theories were used to create the framework.

Articles that were relevant to the selection process are offered in APPENDIX B. Articles that did not directly contribute to the framework may still provide value to other, similarly positioned reviews. Readers may also be interested in alternative lenses through which to view the findings; articles that were discovered during the BeHEMoTh search may offer such a perspective.

Analysis, Evaluation, and Synthesis

Despite the low number of studies directly informing this study's problem of practice, recurring themes became present during analysis. Per the methods of BFS, themes that were not readily absorbed by the *a priori* framework were coded as *emergent themes*. Such themes were defined to reflect their empirical source and fit the framework while minimizing redundancy. These themes were further operationalized as codes to be used in the framework and constantly compared across studies. This process continued in a cyclical manner to ensure all codes were applied to the evidence. Coding was considered complete when no new codes were generated.

Emergent themes were highly valuable because they represented potential knowledge that was missing from the theoretical framework. This alone is a valuable outcome of the BFS methodology. Descriptive themes were noted as such because they described the phenomenon. Both types of themes, however, only described the phenomenon in a purely superficial manner. Hence the call for synthesis across studies and the generation of analytical themes.

Elements of themes that were unanimously supported across contexts without contradiction were synthesized and reported as analytical themes. Analytical themes are the primary outcome of the study. These themes represent a logical connection between a given element and outcome. For example, analytical themes predict or rationalize a given factor's influence on adoption or interaction with other codes. Such themes reflect the phenomenon in a manner similar to that of their descriptive counterparts while also informing readers on a slightly deeper, analytical level.

Sensitivity Analysis

The impact of lower quality studies on findings is evaluated in the original BFS method through a post-hoc sensitivity analysis. The exact manner that this was carried out by Carroll et al. (2013) was unclear, but most likely it was done in a similar manner as seen here. This analysis was carried out by removing the lower quality or misaligned studies to test for impact on individual findings. Indeed, restraint should be used when attempting to generate themes from codes that come exclusively, or even somewhat predominantly, from one study; this is especially true of ill-reported or misaligned sources.

It is noteworthy that while Confidence of Findings reflects the outcome of the sensitivity analysis, both were run separately and do not directly affect one another. However, congruence between them does seem to support their reliability. This could be valuable in several ways. For one, this could provide a degree of intra-rater reliability to single author projects, such as dissertations. Also, the outcome of each is less powerful on its own yet very powerful in conjunction. While we cannot quantify power in qualitative research, we can see the role that ordinal ranking may play in the implications of a given study. This "power" ranking is the outcome of the CERQual Analysis, and might indicate the need for justification of such a ranking. The sensitivity analysis does exactly that; it provides further justification for otherwise subjective decisions on the Confidence of Findings.

It should also be noted that the metric used to identify codes during the sensitivity analysis (34.1%) would violate certain statistical assumptions, and should not be considered a scientifically sound method of rejecting a given theme (i.e. hypothesis). This was included in this paper simply for practical purposes, such as perspective of comparison to other codes and a reference point for constructive dialogue. As with many analytical tools used throughout this

methodology, the value of individual methods and devices may appear meaningless when judged in isolation yet become quite apparent when used in strategic combinations.

Primary Outcome: Posteriori Framework and Analytical Themes

Key outputs of the A-BFS were the *posteriori* framework and analytical themes. The framework offered a more inclusive description of *possible* barriers and their conceptual components while the analytical themes point to more *probable* barriers and complex interactions thereof. Metaphorically speaking, the framework may be seen as a map while the analytical themes represent a compass. Each is valuable on its own, but the two complement one another exceedingly well.

Analytical themes appear as transferable across studies reviewed, but should not be considered generalizable to other contexts. An example of this potential "fault" in generalizing across industries (or systems) is the lacking concern for cost in the study by Kuehl et al. (2014). While the initial review of literature noted cost as a key barrier, this was not seen in the study on firefighters for a few potential reasons. It may also be the case that cost is simply to generic of a concept. As such it can be miscommunicated between participants, researchers, and practitioners. This was a valuable strength of the *posteriori* framework—it provided a more precise lens to distinguish between the nuances of root barriers. Meanwhile analytical themes provided a deeper look at what barriers were likely to exist, how they were likely to interact with other variables, and which solutions may be most relevant.

Practical relevance was the primary concern when adding or deleting concepts from the *a priori* framework. Again, while the analytical themes reflect the most prevalent and distinguishable barriers, the *posteriori* framework is more exhaustive; it covers potential barriers

in the same detail as those that are most-likely. The *posteriori* framework sacrifices depth for breadth in this manner, with no distinction between possible and probable barriers.

Of note, no codes were dropped from the framework. While a few codes were not supported by the findings, the original interviews did not probe all of the *a priori* constructs in a manner that justified exclusion of DT-based codes (see Table 8). A lack of evidence under the scope of inadequate measurement does not speak to a lack of existence. Nevertheless, it warrants consideration that the evidence does currently support certain barriers more than others.

To summarize, the value of the framework and analytical themes are best demonstrated when used together. Analytical themes direct the practitioner's perspective while the framework broadens it. This combination, which is the primary output of the original BFS method, is a powerful combination of knowledge for practitioners and researchers alike.

Secondary Outcome: Confidence of Findings

A CERQual Analysis (Lewin et al. 2015) was used to communicate the level of confidence in each analytical theme. This level of confidence can be viewed as the extent to which a given finding from this review reflects the phenomenon being investigated. Specifically, the process involves assessing the methodology, relevance, coherence, and adequacy of the primary studies that support the given finding against pre-defined criteria. By evaluating the review finding in this manner the reader may have a clearer understanding of the extent to which the finding likely explains the problem of practice.

A few caveats should be considered when using the CERQual confidence scores. First, components of confidence (methodological quality, coherence, etc.) are assumed to be equally weighted. This is not always the case since the weight would likely vary between components

83

from case to case with any given finding. For example, relevance may matter more to one finding than adequacy of data or vice versa. However, both are treated relatively equally in all cases when using this approach.

This was deemed appropriate since low ratings were only used for formative purposes. For example, this ordinal metric simply presents a potential hierarchy within the context reviewed. Confidence scores also reflect support—not rejection. Low ratings simply indicate the need for more evidence as opposed to rejection of the finding. All findings should also be considered reliable and valid within the given context. Confidence scores are only meant to further inform the reader as to the differing degrees of empirical support between findings (i.e. analytical themes). Confidence scores can then be compared to gain a relative sense of hierarchy amongst findings. Again, confidence scores offered in this paper are not intended to act as absolute metrics that can be compared externally.

It is important to note that the assessment of confidence is not a primary outcome of the study, nor does it affect the primary outcomes of the study. Again, all findings noted should be viewed as valid and reliable representations of the phenomenon even if they are ranked relatively low. The ongoing examination of this relatively new method (CERQual Analysis) is also warranted. If it is found invalid or rendered inappropriate for any reason, this aspect of the study can be omitted and the primary findings would remain intact.

Description of CERQual Components

As a reference point, Lewin et al. (2015) briefly compared CERQual components to their GRADE (Balshem et al. 2011) counterparts. GRADE is a method of assessing confidence of evidence in intervention analyses (mixed and quantitative). As an example, "precision" is offered

by GRADE whereas this same concept is viewed through the CERQual lens as "Adequacy of Data". Some readers may be familiar with GRADE and not CERQual. For this reason, a description of the CERQual levels of measurement is offered in APPENDIX D and a description of the CERQual components of measurement is offered in APPENDIX E.

Secondary Outcome: Gaps in Inquisition

A Saturation of Inquisition Test was used to identify areas of the *a priori* framework that were not directly probed by the primary studies. This was an important test for two key reasons. First, the justification for removing any *a priori* codes from a well-established framework must be based on evidence—not a lack thereof. If the researchers did not directly probe an aspect of the framework it may still be highly relevant. Stated another way, if the aspect was not directly assessed it cannot be judged as irrelevant. As noted earlier, a lack of support resulting from inadequate measurement is by no means ground for rejection of theoretical constructs that are otherwise supported through diffusion research. However, aspects of the *a priori* framework that were directly assessed through the original interview prompts were considered for removal if they were explicitly rejected. It is worth noting here that codes that were not supported by the literature did not influence any outputs of the study. Analytical themes, for example, were based heavily on empirical support.

Second, any gaps in inquisition may hold high value as directions for research. For example, areas that have not been directly probed may hold critical findings yet to be discovered. Therefore, this Saturation of Inquisition Test is a logical step for researchers to take when refining future research projects. While similar approaches may exist to some degree, no

versions of the approach were found in the literature reviewed. This paper provides a formal approach to the process for the sake of transparency and constructive dialogue.

Of note, this type of testing can be done in a systematic way to hone the precision of future research questions. Per this study, comparing the interview prompts to the DT framework provides a clear indication of what has not been investigated. Likewise, this approach can be applied to other elements of a research study to gain a better perspective of aspects that have not been addressed. The same method can be used to identify other aspects of samples, research instruments, timelines, etc. that may provide high value for research.

As an example, Figure 9 compares a list of industries to those explicitly investigated by the studies reviewed. Gaps in inquisition are circled in red. As such, investigations in these industries may yield valuable knowledge that is yet to be uncovered. Furthermore, replications of studies in these industries may provide knowledge that can be directly compared across studies.

NAICS SECTORS REPRESENTED IN THIS REVIEW

Sector	Description	Studies	Descriptor
11	Agriculture, Forestry, Fishing, and Hunting		
21	Mining, Quarrying, and Oil and Gas Extraction		
22	Utilities		
23	Construction		
31-33	Manufacturing	Hughes; Witt; Nelson	Mechanical Products, Plastics Products, Furniture and Related Product, and Imaging Systems
42	Wholesale Trade		
44-45	Retail Trade	Hughes	Car Sales
48-49	Transportation and Warehousing		
51	Information		
52	Finance and Insurance	Hughes	Financial Firms
53	Real Estate		
54	Professional, Scientific, and Technical Services	Hughes	Legal Services
55	Management of Companies and Enterprises		
56	Administrative, Support, Waste Management, and Remediation Services		
61	Educational Services	Witt	Schools
62	Health Care and Social Assistance	Hughes; Witt	Rehabilitation Services, Social Service Organizations
71	Arts, Entertainment, and Recreation		
72	Accommodation and Food Services		
81	Other Services		
92	Public Administration	Witt; Kuehl	Municipalities, Chambers of Commerce, Fire Departments

Figure 9. Screenshot of results from a post-hoc Saturation of Inquisition Test.

Gaps in inquisition are likely to hold the highest potential value to researchers and are often a prerequisite for research projects. However, these gaps may be difficult to uncover without a systematic process. The method identified here may expedite this process while offering a level of objectivity and precision needed to justify otherwise vague research ideas.

Conclusion

The A-BFS offered a viable set of methods and tools that thoroughly answered the research questions and accrued a substantial amount of valuable information. This was done despite relatively limited evidence, which highlights the ability of the A-BFS to maximize synthesis output. The additional tools and methods offered in this paper enhanced the overall process and outcome of the BFS method offered by Carroll et al. (2013). Outcomes also directly built on a well-established theory as well as the empirical literature. Specific outputs from the A-BFS included: analytical themes; contextually rich posteriori framework; gaps in inquisition; Confidence of Findings; and a rich audit trail that supports reliability.

The potential value of the A-BFS will likely be more evident in larger research projects. Alignment Scores, for example, offer an expedited means of tracking and communicating important elements of findings. Saturation of Inquisition Tests allow straightforward and precise identification of key elements as well. This paper offers a worked example of these two novel analytical devices, along with other relatively new methods. However, further scrutiny of these methods in combination as well as isolation is warranted.

<u>Limitations</u>

This section discusses limitations, potential threats, and weaknesses. DT is then discussed in detail due to its influence over the generative portion of the synthesis and overall findings. Care was taken to ensure the most transparent audit trail throughout this paper, but readers should remain steadfast on aspects that may be missing or ill-reported.

The small number of studies that directly investigated this study's problem of practice presented a few limitations. First, not all industries were represented by the findings. This creates concern when translating findings into industries that were not directly assessed. Second, informants' professional roles varied between studies and organizations. This could have proven useful if roles were identified in congruence with findings. However, roles were not readily identifiable and findings were pooled across these roles despite few representatives from each.

Incomplete Saturation of Inquisition left gaps of knowledge in the framework. For example, no descriptions of DM personal traits and behaviors were offered, leaving little evidence to speak to the hetero- versus homophilous nature of relationships between adopters and vendors. Hetero- versus Homophily was noted throughout DT as a fundamental factor in the adoption decision process. Gaps of inquisition into domains such as this leave a substantial deficit in our understanding of the phenomenon. Such missing information may have influenced other barriers or affected adoption-decisions directly.

Quite a bit of work is needed regarding all elements of WPs. This variable (WPs) is difficult to conceptually define in a generalizable manner and, therefore, may hinder communication. This ambiguity leaves perceptions vulnerable to change in the presence of relatively small amounts of new knowledge.

Qualitative interviews, and syntheses thereof, may be particularly vulnerable to a number of threats and bias. One threat noted during the initial literature review was selection bias. It may be the case that people who engage in WP surveys and interviews do so out of prior interest in them. This may lead to a biased perspective, especially when noting whether someone would adopt a service or not. However, this threat was well-controlled for in the studies reviewed through a comparison to non-adopters. For instance, two studies (Hughes et al. 2011; Kuehl et al. 2013) garnered more evidence from non-adopters than adopters. This was valuable for two reasons: 1) adopters may view WPs more favorably, limiting their perceived significance of barriers to adoption; and 2) non-adopters are more likely to have firsthand accounts of barriers that were significant enough to blockade adoption of WPs.

Content analysis is susceptible to various threats. Namely, researcher bias and reliability was a concern during the content analysis/synthesis process. This was addressed through the use of an external reviewer. After comparison between codes, discrepancies were discussed until agreement on coding was reached. A thorough description of codes used are provided in the framework; justifications for interpretations are provided throughout Chapter 4 and Chapter 5. Operational definitions of codes were established by the author and reviewed by the external reviewer as well. Any ambiguity of definitions offered in the framework were addressed until agreement was reached between the author and external reviewer. However, the recreation of DT into a framework for coding is itself a deductive and inductive process. The reader should consult the original source (Rogers, 2003) rather than their own interpretations of the framework for a better understanding of concepts that appear most relevant.

Qualitative syntheses are vulnerable to various threats of validity. While the interpretive nature of this process provides many strengths, the reader must decide if the information can be

generalized to their own context or problem of practice. For this reason, the reader is compelled to consider the material as a whole rather than looking exclusively to isolated recommendations offered by the author.

While the methods used in this review present their own strengths and weaknesses, the overarching methodology was intended to harness the former while correcting for the latter. Moreover, the methodology was formed specifically, but not exclusively, for qualitative syntheses in health and fitness research. To the author's knowledge, this combination of methods has not been used before—especially in this context. For this reason, replication and critical evaluation by other researchers will hold the utmost value regarding the trustworthiness of the approaches used throughout this paper. Until the work is examined by other researchers, readers should treat this work as a unique example before considering it a validated one.

Critical Views of Diffusion Theory Research

While the value of DT has been demonstrated over decades of research, it is not without its limitations. Indeed, the use of any given theory to analyze empirical evidence should be accompanied by a thorough description of that theory's inherent assumptions and potential weaknesses. What follows is a description of these assumptions, biases, and potential weaknesses that may arise when superimposing this inherently broad, conceptual paradigm (DT) onto discrete findings. Rogers himself supports such critical appraisal by referencing a time-period that lacked such critiques: "Such absence of critical viewpoints may have indeed been the greatest weakness of diffusion research" (Rogers, 2003).

Assumptions and Biases of Diffusion Theory

Assumptions are inherent, and arguably necessary, to every research endeavor. When developing lines of inquisition these assumptions allow for deeper and more meaningful discoveries, lest the researcher be hindered by describing more surface level factors in a more redundant than constructive manner. Indeed, these assumptions allow research to move the scientific conversation forward, as opposed to superfluously "studying" the same phenomena. This is an inherent submission scholars must make as the dialogue moves toward deeper and distinct lines of inquisition.

DT assumes that innovations denote progress. As such, adoption of such innovations may be considered the standard to reach. This is noted by Rogers (2003) as "pro-innovation bias". This becomes increasingly apparent throughout Roger's text. For example, DT offers its lens to evaluate the adoption process while assuming the effectiveness of the given innovation. This is a substantial assumption that should be considered by researchers and practitioners alike. If a WP doesn't work, its adoption may be hindered despite accounting for all barriers listed in this paper.

Publication bias likely affects the assumed rules of DT as well. Rogers (2003) notes that this may be a significant threat to the growth of DT. Indeed, publication bias may limit a reviewer's selection to positive and significant studies whereas studies that show no trend, or even *negative* trends, may not receive as much attention. Study selection may magnify this threat by focusing on instances of rapid adoption rather than rejection, slower adoption, or simply random change with no identifiable causation or correlation. Moreover, studies that directly focus on rejection of productive innovations may be a highly productive quest when informing barriers to adoption. Such studies were sought and discovered for this review, but limited in breadth and depth. In this paper, it is assumed that this threat is accounted for by the reader. This

is a necessary assumption to move the conversation forward. Other papers are readily available that offer more direct discussion on publication bias.

Source bias, individual-blame bias, weak connections to causality, and equal weight amongst different innovations were also noted as potential weaknesses of DT (Rogers, 2003). Many sponsors of early diffusion research were likely most interested in increasing adoption of their given product. This may have led the theory on the path to pro-innovation bias, but such a claim is difficult to validate or reject. As such, this is seen here simply as another limitation that the reader is compelled to consider.

According to Rogers (2003), a significant portion of diffusion research has also been correlational at best, further weakening any conclusions of causality. While certain elements may predict likelihood of adoption and rate thereof, this does not necessarily indicate the "why" behind the given phenomenon. As such, these predictions would be susceptible to a host of confounding factors when transferred across contexts.

Summary of Limitations

By explicitly noting these limitations the reader may have a better vantage point from which to view any outcomes of this study. Many of these limitations should not detract from findings as much as add to the ongoing conversation. Leveraging any given theory is likely to include such limitations and should be seen by the reader as an integral part of the research process. Indeed, limitations should be explicitly stated by the researcher and thoroughly understood by the reader for constructive knowledge transfer to take place. By understanding these limitations, threats, and underlying assumptions, readers may stay vigilant and critical without sacrificing potentially meaningful and practical findings.

93

APPENDIX A: DISCREPANCIES IN WELLNESS OFFERINGS
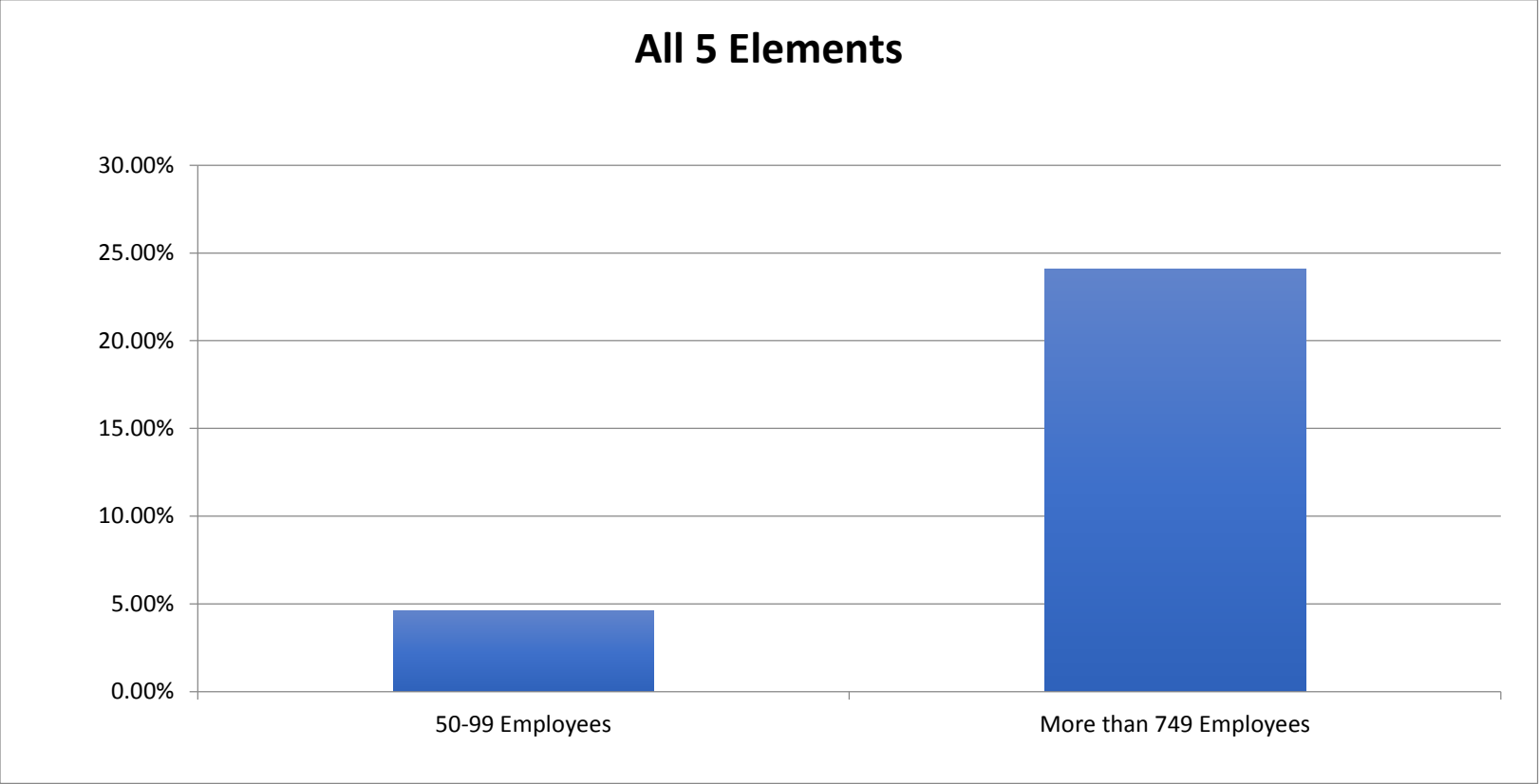
All 5 Elements

Note. This graph shows the difference in percentage (4.6% and 24.1%, respectively) of companies with 50-99 employees (n=179) and more than 749 employees (n=111) that offered all 5 elements of an evidence-based comprehensive WP in a 2004 survey (Linnan et al. 2008).
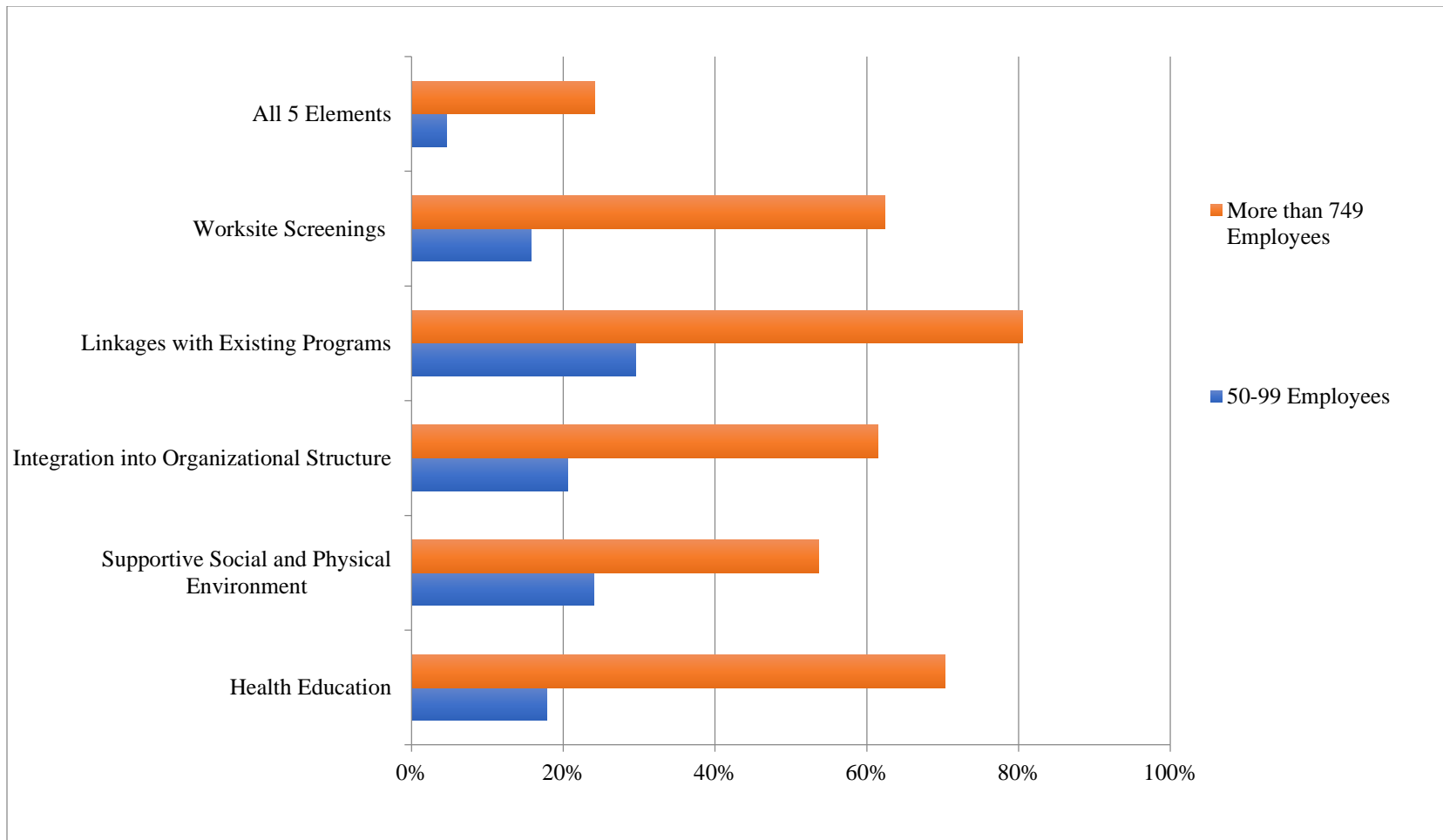
Note. This graph compares the prevalence of 5 key elements of wellness programs according to business size (Linnan et al. 2008).
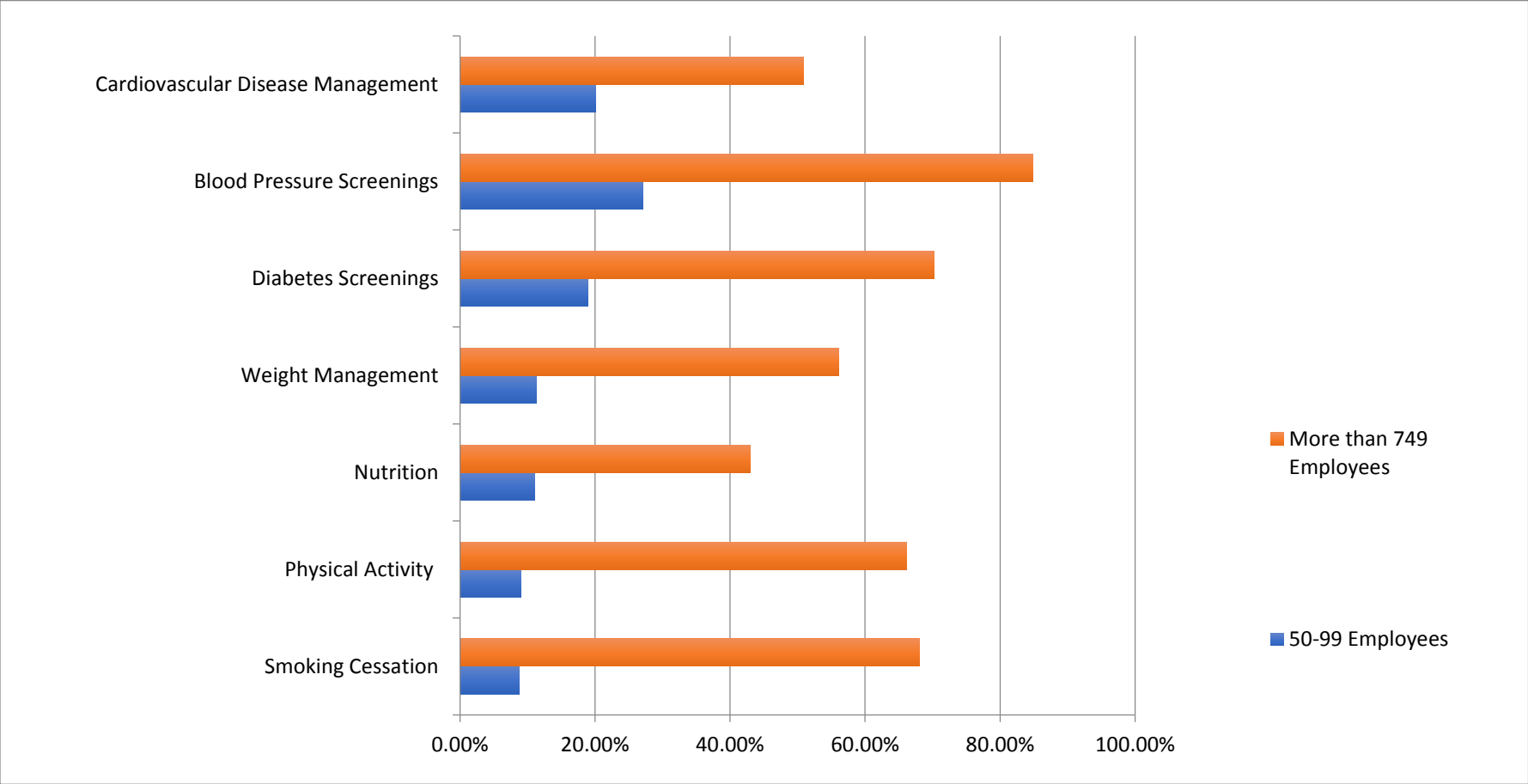
Note. This graph compares components of wellness programs offered amongst companies of different sizes that were surveyed by Linnan et al. (2008).

APPENDIX B: SEARCH METHODS AND RELEVANT FINDINGS

Table B1. *BeHEMoTh Search Criteria*

Concept	Criteria Used
Behavior of Interest	*Adoption*
Health Context	*Wellness program* OR *Health Promotion*
Exclusions	Exclusively used in a field other than health promotion.
Model or Theory	*Organizational theor**

Quick **Search**

organizational theor*	Select a Field (option... ▼	**Search**
AND ▼	wellness program OR health prom	Select a Field (option... ▼
AND ▼	adoption	Select a Field (option... ▼	⊕ ⊖

Basic Search   Advanced Search   Search History ▶

Results   Relevance ▼   Page Optic

earch   ⌄   **Search Results: 1 - 30 of 246**

Table B2. *SPIDER Search Criteria*

Concept	Terms
Sample (Substituted for the Independent Variable)	*Wellness Program* OR *Health Promotion*
Phenomenon of Interest	*Barriers;* AND *Adoption*
Design	(left blank to err toward inclusion)
Evaluation	(left blank to err toward inclusion)
Research Type	*Qualitative*

Quick **Search**

wellness program OR health promotion	Select a Field (option... ▾

AND ▾	barriers AND adoption	Select a Field (option... ▾

AND ▾	qualitative	Select a Field (option... ▾

Basic Search   Advanced Search   Search History ▶

sults

Search Results: 1 - 30 of 234

Table B3. *BeHEMoTh Search Results (Summary Table)*

Abstract Reviewed	Full Text Reviewed	Contributed to the Framework	Reason for Exclusion (as Compared to the Selected Theory)
Michaels & Greene, 2013	Yes	Directly	--
Batras et al. 2014	Yes	Directly	--
Downey & Sharp, 2007	--	--	Deemed less relevant to barriers.
Herzog et al. 2016	--	--	Deemed less relevant overall.
Griffin-Blake & DeJoy, 2006	Yes	--	Deemed less relevant overall.
Milat et al. 2012	Yes	--	Deemed less relevant to adoption.
Kreps, 2009	--	--	Deemed less relevant to adoption.
Van Nassau et al. 2016	--	--	No explicit mention of theory.
Deschesnes et al. 2010	--	--	No explicit mention of theory.
Naaidenberg et al. 2009	--	--	Deemed less relevant to adoption.
Antikainen & Ellis, 2011	Yes	--	Deemed less relevant overall.
Wolfe et al. 1993	--	--	No explicit mention of theory.
Steckler et al. 1992	--	--	Full text was not available.
Tarlov, 1999	--	--	Deemed less relevant.
Blackman et al. 2013	Yes	--	Deemed less relevant.
Miller & Shinn, 2005	--	--	No explicit mention of theory.
Yancey, 2009	--	--	Full text was not available.
Little et al. 2015	Yes	Indirectly	--
Dearing et al. 2006	Yes	Indirectly	--
Dunn et al. 2012	Yes	Indirectly	--
De Civita & Dasgupta, 2007	Yes	Indirectly	--
Fleury & Lee, 2006	Yes	--	Deemed less relevant overall.
Atun et al. 2010	Yes	Indirectly	--

APPENDIX C: AUGMENTED QUALITY APPRAISAL

Table C1. *COREQ Checklist: Domain 1 (Hughes et al. 2011)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 1: Research Team and Reflexivity				
Personal Characteristics				
Interviewer/Facilitator	1	Which author/s conducted the interview or focus group?	0	--
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	1	1
Occupation	3	What was their occupation at the time of the study?	1	1
Gender	4	Was the researcher male or female?	0	--
Experience and training	5	What experience or training did the researcher have?	0	0
Relationship with participants				
Relationship established	6	Was a relationship established prior to study commencement?	0	0
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	0	0
Interviewer characteristics	8	What characteristics were reported about the interviewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	0	0
		Domain Total	2	2

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C2. *COREQ Checklist: Domain 2 (Hughes et al. 2011)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 2: Study Design				
Theoretical framework				
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study?	1	1
Participant selection				
Sampling	10	How were participants selected?	1	1
Method of approach	11	How were participants approached?	1	1
Sample size	12	How many participants were in the study?	1	1
Non-participation	13	How many people refused to participate or dropped out? Reasons?	1	1
Setting				
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	1	1
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	0	0
Description of sample	16	What are the important characteristics of the sample?	1	1
Data collection				
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	1	1
Repeat interviews	18	Were repeat interviews carried out? If yes, how many?	0	0
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	1	1
Field notes	20	Were field notes made during and/or after the interview or focus group?	1	1
Duration	21	What was the duration of the interviews or focus group?	1	--
Data saturation	22	Was data saturation discussed?	1	1
Transcripts returned	23	Were transcripts returned to participants for comment and/or correction?	0	0
		Domain Total	12	11

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C3. *COREQ Checklist: Domain 3 (Hughes et al. 2011)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 3: Analysis and Findings				
Data analysis				
Number of data coders	24	How many data coders coded the data?	1	1
Description of the coding tree	25	Did authors provide a description of the coding tree?	0	0
Derivation of themes	26	Were themes identified in advance or derived from the data?	1	1
Software	27	What software, if applicable, was used to manage the data?	1	--
Participant checking	28	Did participants provide feedback on the findings?	0	0
Reporting				
Quotations presented	29	a.Were participant quotations presented to illustrate the themes/findings? b.Was each quotation identified? e.g. participant number	a. 1 b. 0	a. 1 b. 0
Data and findings consistent	30	Was there consistency between the data presented and the findings?	1	1
Clarity of major themes	31	Were major themes clearly presented in the findings?	1	1
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	1	1
		Domain Total	7	7
		Report Total	21	20

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C4. *COREQ Checklist: Domain 1 (Witt et al. 2013)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 1: Research Team and Reflexivity				
Personal Characteristics				
Interviewer/Facilitator	1	Which author/s conducted the interview or focus group?	0	--
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	1	1
Occupation	3	What was their occupation at the time of the study?	0	0
Gender	4	Was the researcher male or female?	0	--
Experience and training	5	What experience or training did the researcher have?	0	0
Relationship with participants				
Relationship established	6	Was a relationship established prior to study commencement?	0	0
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	0	0
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	0	0
		Domain Total	1	1

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C5. *COREQ Checklist: Domain 2 (Witt et al. 2013)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 2: Study Design				
Theoretical framework				
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study?	1	1
Participant selection				
Sampling	10	How were participants selected?	1	1
Method of approach	11	How were participants approached?	1	1
Sample size	12	How many participants were in the study?	1	1
Non-participation	13	How many people refused to participate or dropped out? Reasons?	1	1
Setting				
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	1	1
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	0	0
Description of sample	16	What are the important characteristics of the sample?	1	1
Data collection				
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	1	1
Repeat interviews	18	Were repeat interviews carried out? If yes, how many?	0	0
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	1	1
Field notes	20	Were field notes made during and/or after the inter view or focus group?	1	1
Duration	21	What was the duration of the interviews or focus group?	1	--
Data saturation	22	Was data saturation discussed?	1	1
Transcripts returned	23	Were transcripts returned to participants for comment and/or correction?	0	0
		Domain Total	12	11

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C6. *COREQ Checklist: Domain 3 (Witt et al. 2013)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 3: Analysis and Findings				
Data analysis				
Number of data coders	24	How many data coders coded the data?	0	0
Description of the coding tree	25	Did authors provide a description of the coding tree?	0	0
Derivation of themes	26	Were themes identified in advance or derived from the data?	1	1
Software	27	What software, if applicable, was used to manage the data?	0	--
Participant checking	28	Did participants provide feedback on the findings?	0	0
Reporting				
Quotations presented	29	a.Were participant quotations presented to illustrate the themes/findings? b.Was each quotation identified? e.g. participant number	a. 1 b. 0	a. 1 b. 0
Data and findings consistent	30	Was there consistency between the data presented and the findings?	1	1
Clarity of major themes	31	Were major themes clearly presented in the findings?	1	1
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	1	1
		Domain Total	5	5
		Report Total	18	17

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C7. *COREQ Checklist: Domain 1 (Nelson et al. 2015)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 1: Research Team and Reflexivity				
Personal Characteristics				
Interviewer/Facilitator	1	Which author/s conducted the interview or focus group?	0	--
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	0	0
Occupation	3	What was their occupation at the time of the study?	0	0
Gender	4	Was the researcher male or female?	0	--
Experience and training	5	What experience or training did the researcher have?	0	0
Relationship with participants				
Relationship established	6	Was a relationship established prior to study commencement?	0	0
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	0	0
Interviewer characteristics	8	What characteristics were reported about the inter viewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	0	0
		Domain Total	0	0

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C8. *COREQ Checklist: Domain 2 (Nelson et al. 2015)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 2: Study Design				
Theoretical framework				
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study?	0	0
Participant selection				
Sampling	10	How were participants selected?	1	1
Method of approach	11	How were participants approached?	1	1
Sample size	12	How many participants were in the study?	1	1
Non-participation	13	How many people refused to participate or dropped out? Reasons?	1	1
Setting				
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	0	0
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	0	0
Description of sample	16	What are the important characteristics of the sample? e.g. demographic data, date	1	1
Data collection				
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	1	0
Repeat interviews	18	Were repeat interviews carried out? If yes, how many?	0	0
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	1	1
Field notes	20	Were field notes made during and/or after the inter view or focus group?	1	1
Duration	21	What was the duration of the inter views or focus group?	1	--
Data saturation	22	Was data saturation discussed?	0	0
Transcripts returned	23	Were transcripts returned to participants for comment and/or correction?	0	0
		Domain Total	9	7

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C9. *COREQ Checklist: Domain 3 (Nelson et al. 2015)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 3: Analysis and Findings				
Data analysis				
Number of data coders	24	How many data coders coded the data?	0	0
Description of the coding tree	25	Did authors provide a description of the coding tree?	0	0
Derivation of themes	26	Were themes identified in advance or derived from the data?	1	1
Software	27	What software, if applicable, was used to manage the data?	1	--
Participant checking	28	Did participants provide feedback on the findings?	0	0
Reporting				
Quotations presented	29	a.Were participant quotations presented to illustrate the themes/findings? b.Was each quotation identified? e.g. participant number	a. 1 b. 1	a. 1 b. 1
Data and findings consistent	30	Was there consistency between the data presented and the findings?	1	1
Clarity of major themes	31	Were major themes clearly presented in the findings?	1	0
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	1	1
		Domain Total	7	5
		Report Total	16	12

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C10. *COREQ Checklist: Domain 1 (Kuehl et al. 2013)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 1: Research Team and Reflexivity				
Personal Characteristics				
Interviewer/Facilitator	1	Which author/s conducted the interview or focus group?	0	--
Credentials	2	What were the researcher's credentials? E.g. PhD, MD	1	1
Occupation	3	What was their occupation at the time of the study?	0	0
Gender	4	Was the researcher male or female?	0	--
Experience and training	5	What experience or training did the researcher have?	0	0
Relationship with participants				
Relationship established	6	Was a relationship established prior to study commencement?	0	0
Participant knowledge of the interviewer	7	What did the participants know about the researcher? e.g. personal goals, reasons for doing the research	0	0
Interviewer characteristics	8	What characteristics were reported about the interviewer/facilitator? e.g. Bias, assumptions, reasons and interests in the research topic	0	0
		Domain Total	1	1

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C11. *COREQ Checklist: Domain 2 (Kuehl et al. 2013)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 2: Study Design				
Theoretical framework				
Methodological orientation and Theory	9	What methodological orientation was stated to underpin the study?	1	1
Participant selection				
Sampling	10	How were participants selected?	1	1
Method of approach	11	How were participants approached?	1	1
Sample size	12	How many participants were in the study?	1	1
Non-participation	13	How many people refused to participate or dropped out? Reasons?	1	1
Setting				
Setting of data collection	14	Where was the data collected? e.g. home, clinic, workplace	0	0
Presence of non-participants	15	Was anyone else present besides the participants and researchers?	0	0
Description of sample	16	What are the important characteristics of the sample?	1	1
Data collection				
Interview guide	17	Were questions, prompts, guides provided by the authors? Was it pilot tested?	0	0
Repeat interviews	18	Were repeat interviews carried out? If yes, how many?	0	0
Audio/visual recording	19	Did the research use audio or visual recording to collect the data?	1	1
Field notes	20	Were field notes made during and/or after the interview or focus group?	1	1
Duration	21	What was the duration of the interviews or focus group?	0	--
Data saturation	22	Was data saturation discussed?	0	0
Transcripts returned	23	Were transcripts returned to participants for comment and/or correction?	0	0
		Domain Total	8	8

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

Table C12. *COREQ Checklist: Domain 3 (Kuehl et al. 2013)*

Topic	Item No.	Guide Question/Prompt	Quality of Reporting Score	Alignment Score
Domain 3: Analysis and Findings				
Data analysis				
Number of data coders	24	How many data coders coded the data?	1	1
Description of the coding tree	25	Did authors provide a description of the coding tree?	0	0
Derivation of themes	26	Were themes identified in advance or derived from the data?	1	1
Software	27	What software, if applicable, was used to manage the data?	0	--
Participant checking	28	Did participants provide feedback on the findings?	0	0
Reporting				
Quotations presented	29	a.Were participant quotations presented to illustrate the themes/findings? b.Was each quotation identified? e.g. participant number	a. 1 b. 1	a. 1 b. 1
Data and findings consistent	30	Was there consistency between the data presented and the findings?	1	1
Clarity of major themes	31	Were major themes clearly presented in the findings?	1	1
Clarity of minor themes	32	Is there a description of diverse cases or discussion of minor themes?	1	1
		Domain Total	7	7
		Report Total	16	16

Note. Quality of reporting was scored as (1) adequate or (0) inadequate/non-existent. Alignment was scored as (1) well-aligned or (0) not well-aligned. Adapted with permission from Tong et al. (2007).

APPENDIX D: CERQUAL LEVELS OF CONFIDENCE

Table D. *CERQual Levels of Confidence*

Level	Definition
High Confidence	It is highly likely that the review finding is a reasonable representation of the phenomenon of interest.
Moderate Confidence	It is likely that the review finding is a reasonable representation of the phenomenon of interest.
Low Confidence	It is possible that the review finding is a reasonable representation of the phenomenon of interest.
Very Low Confidence	It is not clear whether the review finding is a reasonable representation of the phenomenon of interest.

Note. Reproduced from Lewin et al. (2015) under the terms of the Creative Commons Attribution License.

APPENDIX E: CERQUAL COMPONENTS OF MEASUREMENT

Table E. *CERQual Components of Measurement*

[a]Component	[a]Definition	[b]Determinants of Confidence
Methodological Limitations	The extent to which there are problems in the design or conduct of the primary studies that contributed evidence to a review finding.	Core domain #1: *Measurement tool* adequately investigates the problem and directly resulted in the finding in the manner it is presented. Core domain #2: *Data collection* is performed in an adequate manner. Core domain #3: *Number of original coders* supports objectivity. Core domain #4: Reporting score from the *COREQ appraisal* is adequate.
Relevance	The extent to which the body of evidence from the primary studies supporting a review finding is applicable to the context (perspective or population, phenomenon of interest, setting) specified in the review question.	*Direct:* All domains are comprehensively addressed. *Indirect:* One domain of the research question has been substituted for another. *Partial:* All domains of the research question are not comprehensively addressed. *Uncertain:* There are deficiencies in the reporting of details from the domains being investigated. Note: The Alignment Scores on the COREQ appraisal acted as valuable points of reference here.
Coherence	The extent to which the review finding is well grounded in data from the contributing primary studies and provides a convincing explanation for the patterns found in these data.	Coherence is defined here as the extent to which *contextually based* data or *conceptually based* data can explain the pattern recognized by the finding. More simply, how well the individual finding aligns with general patterns seen across studies.
Adequacy of Data	An overall determination of the degree of richness and quality of data supporting a review finding.	Adequacy is a function of *richness* and *quantity* of data. Richness is measured here as the extent to which data describes details of the phenomenon. Quantity is (hesitantly) measured here as a function of saturation.

[a]Adapted from Lewin et al. (2015) under the terms of the Creative Commons Attribution License.
[b]Developed specifically for this review as directed by Lewin et al. (2015).

LIST OF REFERENCES

Anderko, L., Roffenbender, J. S., Goetzel, R. Z., Howard, J., Millard, F., Wildenhaus, K., DeSantis, C., & Novelli, W. (2012). Promoting prevention through the Affordable Care Act: Workplace wellness. *Prev Chronic Dis,* 9:120092. doi: http://dx.doi.org/10.5888/pcd9.120092

Arena, R., Guazzi, M., Briggs, P. D., Cahalin, L. P., Myers, J., Kaminsky, L. A., Forman, D., Cipriano, G., Borghi-Silva, A., Babu, A., & Lavie, C. J. (2013). Review: Promoting health and wellness in the workplace: A unique opportunity to establish primary and extended secondary cardiovascular risk reduction programs. *Mayo Clinic Proceedings*, *88,* 605-617. doi:10.1016/j.mayocp.2013.03.002

Atkins, S., Lewin, S., Smith, H., Engel, M., Fretheim, A., & Volmink, J. (2008). Conducting a meta-ethnography of qualitative literature: Lessons learnt. *BMC Medical Research Methodology, 8,* 21. doi:10.1186/1471-2288-8-21

Baicker, K., Cutler, D., & Song, Z. (2010). Workplace wellness programs can generate savings. *Health Affairs*, *29*(2), 304-311. doi:10.1377/hlthaff.2009.0626

Balshem, H., Helfand, M., Schunemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. (2011) GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol 64,* 401-406.

Barbour, R. S. (2003). The newfound credibility of qualitative research? Tales of technical essentialism and co-option. *Qualitative Health Research, 13*(7), 1019-1027. doi: 10.1177/1049732303253331

Batras, D., Duff, C., & Smith, B. (2014). Organizational change theory: implications for health

 promotion practice. *Health Promotion International, 31*(1), 231-241. doi:

 10.1093/heapro/dau098.

Beile, P., & Boote, D. (2005). Scholars before Researchers: On the Centrality of the Dissertation

 Literature Review in Research Preparation. *Educational Researcher, 34*(6), 3-15.

Bishop, F., & Yardley, L. (2010). The development and initial validation of a new measure of lay

 definitions of health: The wellness beliefs scale. *Psychology & Health*, *25*(3), 271-287.

Carroll, C., Booth, A., Leaviss, J., & Rick, J. (2013). "Best fit" framework synthesis: refining the

 method. *BMC Medical Research Methodology, 13*(37). doi:10.1186/1471-2288-13-37.

Catford, J. (2009). Advancing the 'science of delivery' of health promotion: Not just the 'science

 of discovery'. *Health Promotion International, 24*(1), 1-5. doi: 10.1093/heapro/dap003

Claxton, G., Rae, M., Panchal, N., Whitmore, H., Damico, A., Kenward, K., & Long, M. (2015).

 Health benefits in 2015: Stable trends in the employer market. *Health Affairs*, *34*(10),

 1779-1788. doi:10.1377/hlthaff.2015.0885

Connor, M. (2016). A broader value proposition for reporting wellness results. *Employee Benefit*

 Plan Review, 70(7), 8-11.

Cooke, A., Smith, D., & Booth, A. (2012). Beyond PICO: The SPIDER tool for qualitative

 evidence synthesis. *Qualitative Health Research, 22*(10), 1435-1443.

Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Thousand

 Oaks, CA: Sage.

Das, B., Petruzzello, S., & Ryan, K. (2014). Development of a logic model for a physical

 activity-based employee wellness program for mass transit workers. *Preventing Chronic*

 Disease, 11. doi: 10.5888/pcd11.140124

Fielding, J. E., Kumanyika, S., & Manderscheid, R. W. (2013). A perspective on the

    development of the Healthy People 2020 framework for improving U.S. population

    health. *Public Health Reviews (2107-6952)*, *35*(1), 1-24.

Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and

    trends in the distribution of body mass index among US adults, 1999–2010. *Journal of

    the American Medical Association*, *307*(5), 491–97.

Goetzel, R., & Ozmlnkowski, R. (2008). The health and cost benefits of work site health-

    promotion programs. *Annual Review Of Public Health*, *29*303-323.

Goetzel, R., Tabrizi, M., Henke, R., Benevent, R., Brockbank, C., Stinson, K., Trotter, M., &

    Newman, L. (2014). Estimating the return on investment from a health risk management

    program offered to small Colorado-based employers. *Journal Of Occupational And

    Environmental Medicine*, *56*(5), 554-560.

Haaz, E. J., Maynard, J., Petricia, S. C., & Williams, C. E. (2003). Employee assistance program

    accreditation: History and outlook. *Employee Assistance Quarterly, 19*(1), 1-26.

    doi:10.1300/J022v19n01_01

Hersey, J., Williams-Piehota, P., Sparling, P. B., Alexander, J., Hill, M. D., Isenberg, K. B.,

    Rooks, A., & Dunet, D. O. (2008). Promising practices in promotion of healthy weight at

    small and medium-sized US worksites. *Preventing Chronic Disease*, *5*(4), A122.

Howard, J. & Mikalachki, A. (1979). Fitness and employee productivity. *Can J Appl Sport Sci,

    4*(3), 191-8.

Hubley, T. & Dutram, K. (2011). Crafting a foundation for evaluating a worksite wellness

    program. *Journal of Multidisciplinary Evaluation, 7*(16), 148-161.

Hughes, M. C., Patrick, D. L., Hannon, P. A., Harris, J. R., & Ghosh, D. L. (2011).

>Understanding the decision-making process for health promotion programming at small
>to midsized businesses. *Health Promotion Practice*, *12*(4), 512-521.
>doi:10.1177/1524839909349162

Kaspin, L. C., Gorman, K. M., & Miller, R. M. (2013). Systematic review of employer-

>sponsored wellness strategies and their economic and health-related outcomes.
>*Population Health Management, 16*(1), 14-21. doi:10.1089/pop.2012.0006

Kirkland, A. (2014). What is wellness now?. *Journal Of Health Politics, Policy & Law*, *39*(5),

>957. doi:10.1215/03616878-2813647

Kuehl, H., Mabry, L., Elliot, D. L., Kuehl, K. S., & Favorite, K. C. (2013). Factors in adoption of

>a fire department wellness program: Champ-and-chief model. *Journal Of Occupational*
>*And Environmental Medicine*, *55*(4), 424-429. doi:10.1097/JOM.0b013e31827dba3f

Labonte, R. (2014). Health in all (foreign) policy: Challenges in achieving coherence. *Health*

>*Promotion International*, *29*48-58.

Lewin, S., Glenton, C., Munthe-Kaas, H., Carlsen, B., Colvin, C.J., Gülmezoglu, M., Noyes, J.,

>Booth, A., Garside, R., & Rashidian, A. (2015) Using qualitative evidence in decision
>making for health and social interventions: An approach to assess confidence in findings
>from qualitative evidence syntheses (GRADE-CERQual). *PLoS Med 12*(10), e1001895.
>doi:10.1371/journal. pmed.1001895

Li, C. Y., & Sung, F. C. (1999). A review of the healthy worker effect in occupational

>epidemiology. *Occup Med (Lond), 49*(4), 225-229. doi: 10.1093/occmed/49.4.225

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry.* Newbury Park, CA: Sage

>Publications.

Linnan, L., Bowling, M., Childress, J., Lindsay, G., Blakey, C., Pronk, S., & Royall, P. (2008). Results of the 2004 National Worksite Health Promotion Survey. *American Journal Of Public Health*, *98*(8), 1503-1509 7p. doi:10.2105/AJPH.2006.100313.

Liu, H., Harris, K., Weinberger, S., Serxner, S., Mattke, S., & Exum, E. (2013). Effect of an employer-sponsored health and wellness program on medical cost and utilization. *Population Health Management, 16*(1), 1-6. doi:10.1089/pop.2011.0108.

Lorenc, T., Pearson, M., Jamal, F., Cooper, C., & Garside, R. (2012). The role of systematic reviews of qualitative evidence in evaluating interventions: A case study. *Res Synth Meth, 3,* 1-10. 10.1002/jrsm.1036.

McCoy, K., Stinson, K., Scott, K., Tenney, L., & Newman, L. (2014). Health promotion in small business: A systematic review of factors influencing adoption and effectiveness of worksite wellness programs. *Journal Of Occupational And Environmental Medicine*, *56*(6), 579-587.

McDowell, I. (2010). Measures of self-perceived well-being. *Journal of Psychosomatic Research, 69*(1), 69-79. doi:10.1016/j.jpsychores.2009.07.002

Menke, A., Casagrande, S., Geiss, L., & Cowie, C. C. (2015). Prevalence of, and trends in, Diabetes among adults in the United States, 1988-2012. *JAMA, 314*(10), 1021-1029. doi:10.1001/jama.2015.10029

Merrill, R. M., Aldana, S. G., Vyhlidal, T. P., Howe, G., Anderson, D. R., & Whitmer, R. W. (2011). The impact of worksite wellness in a small business setting. *Journal Of Occupational & Environmental Medicine*, *53*(2), 127-131 5p. doi:10.1097/JOM.0b013e318209e18b

Milat, A., King, L., Bauman, A., & Redman, S. (2012). The concept of scalability: Increasing the

    scale and potential adoption of health promotion interventions into policy and practice.

    Health Promotion International, 28(3), 285-298. doi:10.1093/heapro/dar097.

Mukhopadhyay, S., & Wendel, J. (2013). Evaluating an employee wellness program.

    International Journal Of Health Care Finance & Economics, *13*(3-4), 173-199.

Nayer, C., Berger, J., & Mahoney, J. (2010). Wellness, hard to define, reduces trend up to 4

    percent. *Population Health Management*, *13*(2), 83-89. doi:10.1089/pop.2009.0014

Nelson, C. C., Allen, J. D., McLellan, D., Pronk, N., & Davis, K. L. (2015). Integrating health

    promotion and occupational safety and health in manufacturing worksites: Perspectives

    of leaders in small-to-medium sized businesses. *Work*, *52*(1), 169-176.

    doi:10.3233/WOR-152038

Newman, L., Stinson, K., Metcalf, D., Fang, H., Brockbank, C., Jinnett, K., Reynolds, S.,

    Trotter, M., Witter, R., Tenney, L., Atherly, A., & Goetzel, R. (2015). Implementation of

    a worksite wellness program targeting small businesses: The Pinnacol Assurance Health

    Risk Management Study. *Journal Of Occupational And Environmental Medicine*, *57*(1),

    14-21.

Ogden, C. L., Carroll, M. D., Kit, B. K., & Flegal, K. M. (2012). Prevalence of obesity and

    trends in body mass index among US children and adolescents, 1999–2010. *Journal of*

    the American Medical Association, 307(5), 483–90.

Ozminkowski, R., Serxner, S., Marlo, K., Kichlu, R., Ratelis, E., & Van de Meulebroecke, J.

    (2016). Beyond ROI: Using value of investment to measure employee health and

    wellness. *Population Health Management, 19*(4), 227. doi:10.1089/pop.2015.0160

Pitt-Catsouphes, M., James, J. B., & Matz-Costa, C. (2015). Workplace-based health and

wellness programs: The intersection of aging, work, and health. *The Gerontologist*, (2),

262. doi:10.1093/geront/gnu114

Plump, C. M., & Ketchen, D. J. (2013). Paving a road to well? How the legal pitfalls of wellness

programs can harm organizational performance. *Business Horizons, 56*(3), 261-269.

Pronk, N. (2014). Best practice design principles of worksite health and wellness programs.

Acsms Health & Fitness Journal, *18*(1), 42-46.

Reavley, N., Livingston, J., Buchbinder, R., Bennell, K., Stecki, C., & Osborne, R.H. (2010). A

systematic grounded approach to the development of complex interventions: The

Australian WorkHealth Program – Arthritis as a case study. *Social Science & Medicine,*

70(3), 342-350. doi:10.1016/j.socsimed.2009.10.006

Reeves, J. (2015). *The 2015 UBA Health Plan Survey* (Executive Summary). Indianapolis, IN:

United Benefits Advisors, LLC. Available to download at:

http://content.ubabenefits.com/thank-you-page-level-2?submissionGuid=71dd89e1-801f-

4b75-b199-0d454e160b6d

Roger, V., Go, A., Lloyd-Jones, D., Benjamin, E., Berry, J., Borden, W., Bravata, D., Dai, S.,

Ford, E., Fox, C., Fullerton, H., Gillespie, C., Hailpern, S., Heit, J., Howard, V., Kissela,

B., Kittner, S., Lackland, D., Lichtman, J., Lisabeth, L., Makuc, D., Marcus, G., Marelli,

A., Matchar, D., Moy, C., Mozaffarian, D., Mussolino, M., Nichol, G., Paynter, N.,

Soliman, E., Sorlie, P., Sotoodehnia, N., Turan, T., Virani, S., Wong, N., Woo, D., &

Turner, M. (2012). Executive summary: Heart disease and stroke statistics--2012 update:

A report from the American Heart Association. *Circulation*, *125*(1), 188-197 10p.

doi:10.1161/CIR.0b013e3182456d46

Rogers, E. M. (2003) *Diffusion of innovations*, 5th edition. New York, NY: Free Press.

Sandelowski, M, & Barroso, J. (2007). *Handbook for synthesizing qualitative research.* New York, NY: Springer Publishing Company.

Schultz, A. B., Chen, C., & Edington, D. W. (2009). The cost and impact of health conditions on presenteeism to employers: A review of the literature. *Pharmacoeconomics*, (5), 365.

Shephard, R. J. (1983). Employee health and fitness: The state of the art. *Prev Med, 12*(5), 644-53.

Shepard, R. J. (1981). Fitness program participation: Its effect on worker performance. *Journal of Occupational Medicine, 23*, 359-363.

Shepherd, H. R. (2016). It saves to be healthy: Using the tax code to incentivize employer-provided wellness benefits. *Indiana Law Journal*, *91*(2), 597-616.

Smith, J. C., Collins, A., Ferrari, R., Holmes, J. R., Logstrup, S., McGhie, D. V., Ralston, J., Sacco, R., Stam, H., Taubert, K., Wood, D., & Zoghbi, W. A. (2012). WHF/AHA/ACCF/EHN/ESC presidential advisory: Our time: A call to save preventable death from cardiovascular disease (heart disease and stroke). *Global Heart*, 7297-305. doi:10.1016/j.gheart.2012.08.002

Stoner, L., & Cornwall, J. (2014). Did the American Medical Association make the correct decision classifying obesity as a disease? *The Australasian Medical Journal*, *7*(11), 462–464. http://doi.org/10.4066/AMJ.2014.2281

Swords, S. (2014). Don't let them eat cake: Implementing a government-sponsored wellness program that utilizes tax credits to promote healthy behaviors. *Journal Of Legal Medicine*, *35*(4), 579-598.

Taylor, A. W., Pilkington, R., Montgomerie, A., & Feist, H. (2016). The role of business size in assessing the uptake of health promoting workplace initiatives in Australia. *BMC Public Health*, *16*1-8. doi:10.1186/s12889-016-3011-3

Terry, P., Seaverson, E., Grossmeier, J., & Anderson, D. (2008). Association between nine quality components and superior worksite health management program results. *Journal Of Occupational And Environmental Medicine*, *50*(6), 633-641.

Terry, P. E., Grossmeier, J., Mangen, D. J., & Gingerich, S. B. (2013). Analyzing best practices in employee health management: How age, sex, and program components relate to employee engagement and health outcomes. *Journal Of Occupational And Environmental Medicine*, *55*(4), 378-392.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *Bmc Medical Research Methodology*, *8.*

Tjoa, A., Ling, V., Bender, C., Brittenham, M., & Jha, S. (2012). Special issue: Health insurance and the affordable care act: Wellness programs. *Journal Of The American College Of Radiology*, *9*(Health Insurance and the Affordable Care Act), 894-899. doi:10.1016/j.jacr.2012.09.011

Tong, A., Sainsbury, P., & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): A 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care, 19*(6), 349-357. doi: intqhc/mzm042

Tong, A., Flemming, K., McInnes, E., Oliver, S., & Craig, J. (2012). Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Medical Research Methodology, 12*, 181.

Tricco, A. C., Antony, J., Soobiah, C., Kastner, M., MacDonald, H., Cogo, E., Lillie, E., Tran, J.,
      & Straus, S. E. (2016). Knowledge synthesis methods for integrating qualitative and
      quantitative data: A scoping review reveals poor operationalization of the methodological
      steps. *Journal of Clinical Epidemiology, 73*, 29-35. doi:10.1016/j.jelinepi.2015.12.011

U.S. Department of Health and Human Services (2017). Retrieved from:
      https://www.healthypeople.gov/2020/About-Healthy-People

U.S. Department of Labor. (2014). Fact Sheet. Retrieved from:
      https://search.usa.gov/search?utf8=✓&affiliate=ebsa&query=wellness

Weed, M. (2006). Interpretive qualitative synthesis in the sport & exercise sciences: The meta-
      interpretation approach. *European Journal of Sport Science, 6*(2), 127-139. doi:
      10.1080/17461390500528576

Williams, S. J., & Snow, D. M. (2012). Promoting health in small and medium-sized enterprises.
      Journal Of Small Business & Enterprise Development, *19*(4), 729.
      doi:10.1108/14626001211277497

Witt, L. B., Olsen, D., & Ablah, E. (2013). Motivating factors for small and midsized businesses
      to implement worksite health promotion. *Health Promotion Practice*, *14*(6), 876-884.
      doi:10.1177/1524839912472504

Ye, C., Foster, G., Kaczorowski, J., Chambers, L. W., Angeles, R., Marzanek-Lefebvre, F.,
      Laryea, S., Thabane, L., & Dolovich, L. (2013). The impact of a cardiovascular health
      awareness program (CHAP) on reducing blood pressure: A prospective cohort study.
      BMC Public Health, doi:10.1186/1471-2458-13-1230

Zula, K. (2014). Workplace wellness programs: A comparison between best practice guidelines
      and implementation. *Journal Of Applied Business Research*, *30*(3), 783-791.