EXAMINING PRODUCTIVE FAILURE INSTRUCTION IN DENTAL ETHICS

by

Karen K. Tam

Copyright © Karen K. Tam 2017

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017
As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Karen Tam, titled Examining Productive Failure Instruction as an Effective Teaching Method in Dental Ethics and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Date: 4/13/17
Ronald Marx

Date: 4/13/17
Francesca Lopez

Date: 4/13/17
Heidi Legg-Burross

Final approval and acceptance of this dissertation is contingent upon the candidate’s submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Date: 4/13/17
Dissertation Director: Ronald Marx
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Karen K. Tam
ACKNOWLEDGEMENTS

My deepest gratitude goes to all those who supported and encouraged me along this meaningful path. Without them, this would not have been possible. I would especially like to thank Dr. Ron Marx, the chairman of my committee who provided me extensive personal and professional guidance to put together this manuscript. I am privileged and forever grateful to have you as my mentor who through good conversations, humor, availability, and understanding provided me with unfailingly good advice.

Special thanks to Dr. Francesca Lopez and Dr. Heidi Burross for their valuable comments and suggestions to this project. Dr. Francesca Lopez shared her knowledge and expertise which provided insight and clarity into my data. Dr. Heidi Burross provided endless support from the very inception of this project. I am grateful for the opportunity to have worked with both of you.

Heartfelt thanks to Toni Sollars who kept me firmly on this path. I know I would not be here if it was not for your kindness and guidance. I also wish to thank Karen Sesslers for continuing the support for which I greatly appreciate.

And to my family members, eternal thanks to my parents, Dr. C.S. and Theresa Tam, Margaret Junker, and my sisters, Melinda and Mary for cheering me on throughout this process. More importantly, I wish to thank my husband, E.J. for his love and patience to help me see this through when I needed it the most.

Finally, I wish to express my sincere thanks to Holly Houck for her assistance in this study and the dental hygiene faculty, students, and their institutions who unselfishly gave me their time and support so that I may pursue my work.
DEDICATION

To Jasmine and Aidan, my wonderful children, who remind me every day that nothing is more important than patience when we fail, persistence when we struggle, and most of all, love for support when we strive to be the best that we can be.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>10</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>13</td>
</tr>
<tr>
<td>a. Productive Failure</td>
<td>14</td>
</tr>
<tr>
<td>b. Research Questions</td>
<td>17</td>
</tr>
<tr>
<td>II. PRODUCTIVE FAILURE: INSTRUCTIONAL DESIGN</td>
<td>20</td>
</tr>
<tr>
<td>a. Abstract</td>
<td>21</td>
</tr>
<tr>
<td>b. Principles of Productive Failure</td>
<td>24</td>
</tr>
<tr>
<td>c. Assessments of Productive Failure</td>
<td>30</td>
</tr>
<tr>
<td>d. Role of the Teacher</td>
<td>32</td>
</tr>
<tr>
<td>e. Scaling Productive Failure Learning Environment</td>
<td>33</td>
</tr>
<tr>
<td>f. Research in Productive Failure</td>
<td>34</td>
</tr>
<tr>
<td>g. Conclusion</td>
<td>38</td>
</tr>
<tr>
<td>III. A PILOT STUDY EXAMINING PRODUCTIVE FAILURE AS AN EFFECTIVE TEACHING METHOD IN DENTAL ETHICS</td>
<td>39</td>
</tr>
<tr>
<td>a. Abstract</td>
<td>40</td>
</tr>
<tr>
<td>b. Definition of Key Concepts</td>
<td>44</td>
</tr>
<tr>
<td>c. Literature Review</td>
<td>45</td>
</tr>
<tr>
<td>d. Methods</td>
<td>55</td>
</tr>
<tr>
<td>e. Results</td>
<td>67</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Patterns of Shifts in Moral Schemas from N2 Pretest to Posttest</td>
<td>115</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLE

3.1 Characteristics of the Sample ................................................................. 56
3.2 Productive Failure Design Compared to Lecture and Practice

   Instructional Design ..................................................................................... 63
3.3 Ethical Dilemma Problem Design .............................................................. 65
3.4 Comparison of Pretest and Posttest of DIT2 N2 Scores ......................... 69
4.1 School and Instructional Characteristics .................................................... 91
4.2 Ethical Dilemma Problem Design .............................................................. 99
4.3 Demographic Characteristics of Participants ............................................ 109
4.4 N2 Pretest-Posttest Scores Means ............................................................. 111
4.5 Regression of the N2 Posttest Scores on Pretest Scores and Teaching

   Methods ........................................................................................................ 112
4.6 Shift of Distributions Between Moral Schemas Using Mean

   Differences................................................................................................. 113
4.7 Descriptive Statistics for Process Analysis on Practice Problem ............. 116
4.8 Posttest Scores of Written Responses ...................................................... 117
4.9 Regression of the Well-structured Problem Posttest Scores on Teaching

   Methods ...................................................................................................... 118
4.10 Mean Difference of Pretest and Posttest scores on Moral Dilemma

   Problems.................................................................................................... 119
Abstract

These three papers examine “productive failure” as a viable learning design to improve problem solving skills using open-ended problems. Productive failure, a teaching method, is based on the premise of unsuccessful learning performance in solving for complex problems with little to no support while yielding productive learning for subsequent problems. Kapur (2008) argues that hidden efficacies of learning exist in failure in which learners potentially learn through experimentation from their exploration and struggle in solving complex problems in a way that learners must first try and solve complex, novel problems on their own, but ultimately will fail to reach a solution.

We have limited understanding if this type of design would be effective on complex problems with multiple solutions since previous studies on productive failure focused on problems with a canonical solution. In the three papers, I examine the extent to which students learn how to solve moral dilemmas in productive failure (PF) compared to lecture and practice (LP) and to what extent instruction in PF helps students learn skills in transferable problem solving.

One paper describes a pilot study that was conducted with 21 second-year dental hygiene students. In the randomized-controlled study, analysis did not show significant differences on moral reasoning ($p = .06$) and transfer of knowledge ($p = .58$) between PF and LP instructional method. However, the effect size on students’ posttest scores was high ($d = .76$) which as a result of the educational intervention, suggests that PF students demonstrated acquisition of new thinking and approached the complex problem in a more sophisticated moral way of thinking. To replicate these findings, the results from the
pilot study were used to make adjustments in instructional and research design for a full-scale study.

The second study on 77 second-year dental hygiene students from four dental hygiene programs further shows that PF students gained a deeper conceptual understanding and were better prepared for subsequent problems. PF students, I found, demonstrated greater shifts from simplistic thinking to post conventional thinking compared to LP students. Although PF students performed similarly when compared to LP students on their posttest scores in the moral responses, LP students scored lower than their pretest problem and the difference between pretest and posttest scores in LP School had a moderate effect in a negative direction ($d = -.64$).

Findings in both studies suggest that productive failure design has the potential to help students reach a deeper conceptual understanding when they 1) analyze their own failure; 2) use the learned concept to build upon their own prior knowledge; and 3) repair existing mental models to successfully solve complex problems. As such, continued exploration of various instructional approaches like productive failure is still needed as alternatives to lecture and practice for developing problem solving skills.
CHAPTER ONE: INTRODUCTION

Developing problem solving skills, specifically moral reasoning skills, in dental and dental hygiene students is essential to their professional success. Dental hygiene students must now employ interpersonal relations, self-management skills, judgment and knowledge of dental disease processes, as well as demonstrate competency in addressing the physical, psychological, and social well-being of the patient (Chambers, 1987; Khatami & MacEntee, 2011). In response to changes in professional responsibilities, the Commission on Dental Accreditation’s (CODA; 2011) standards documents on dental hygiene programs asserts that there is a clear need for developing problem solving and moral reasoning skills among dental hygiene students. Under the heading Ethics and Professionalism, Standard 2–19 states that “graduates must be competent in applying the principles of ethical reasoning, ethical decision making and professional responsibility as they pertain to the academic environment, research, patient care and practice management” (p. 27). CODA (2003) states that to prepare student dental hygienists for a profession, learning environments and instruction need to be designed to “evaluate student experiences which promotes ethics, ethical reasoning and professionalism” (p. 27). Under the heading Critical Thinking, Standard 2–23 states that “graduates must be competent in problem solving strategies related to comprehensive patient care and management of patients” (p. 28). Problem solving as defined by the dental profession is “the process of answering a question or achieving a goal when the path or answer is not immediately obvious, using an acceptable heuristic or strategy such as the scientific method” (ADEA, 2009, p. 30). The implications of these standards on dental hygiene
curriculum are that students must be provided opportunities to demonstrate moral reasoning skills and the ability to solve problems or moral dilemmas.

In this dissertation, I take up this aim by examining “productive failure” as a viable learning design that is structured to improve problem solving skills. Additionally, I compare productive failure with lecture and practice, a teaching method commonly used by dental hygiene faculty even though trends in education stress using active learning over lecture (Turner, Prihoda, English, Chismark & Jacks, 2016). Turner et al. (2016) described lecture as a format where teachers’ provide students information in a structured face-to-face environment. Moreover, lecture and practice focuses on learning goals, follows a deliberate implementation of demonstration, guided practice, and then independent practice using structured and explicit instruction (McMullen & Madelaine, 2017). From this point forward, I conceptually define lecture and practice as an instructional design that is teacher-led to provide structured and explicit instruction followed by demonstration and guided practice used to meet student learning goals. In what follows, I briefly describe productive failure as an instructional design that, unlike lecture practice, focuses on developing problem solving skills instead of meeting learning goals and the learning environment as a flexible space rather than structured. Also, I discuss the contributions of these three papers and how the findings will add to the body of research of productive failure.

**Productive Failure**

Productive failure is described as “a learning design that affords students opportunities to generate representations and solutions to a novel problem that targets a concept that they have not learned yet, followed by consolidation and knowledge
assembly where they learn the targeted concept” (Kapur, 2015, p. 52). The concept of productive failure is based on situative and social constructivist learning theories. Collaboration and working in small groups to solve authentic complex problems are central in productive failure, a recommendation that was made by dental educators when considering teaching strategies for dental and dental hygiene students (Hendricson, 2012). The difference lies on the structure of when to implement instruction with the use of external support such as scaffolding and in the construction of activities that lead learners to ultimately fail while maintain the role of persistence to finding a solution. In a study on productive failure among middle-school mathematics students, Manu Kapur (2010) suggests that “delaying the structure received by students from the ill-structures groups (who solved complex, ill-structured problems collaboratively followed by well-structure problems individually) helped them discern how to structure an ill-structured problem, thereby facilitating a spontaneous transfer of problem – solving skills…designing opportunities for learners to generate and develop their own structures-representations, problem –solving methods, conceptions-in the absence of external structure may well lead to performance failure in the shorter term. However, this very process may be germane to learning in the longer term” (p.525).

Productive failure appears to be a part of a larger program of research and development in the area of problem solving under unsuccessful conditions or impasse-driven learning (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003) and manipulations that creates “desirable difficulties” such as varying learning conditions to be less
predictable, practice sessions as learning events, and reducing feedback to the student (Bjork & Kroll, 2015). With that, failure has many operational definitions which are dependent on how each learner and teacher defines failure. Failure could be related to a learner’s underperformance, lack of goal-orientedness, lack of effort, and/or lack of motivation to learning and to a teacher’s lack of preparedness, unclear instruction or weak assessments. But failure also presents with other qualities that may be beneficial to novice learners. Kapur (2008) explains that the challenge of the problem and one that ultimately leads to an impasse or failure require task persistence for novice learners to process and accommodate to the new information and experiences to create and develop multiple problem-solving methods. It is through the learners’ struggle or failure that is needed to change their existing mental presentations into creating new mental presentations.

According to Kapur (2008), the delay of instruction is essential to producing a hidden efficacy. By introducing instruction after students struggle, persist, and fail to find a solution while developing their own mental representations of concepts, a hidden efficacy may exist that helps learners understand why canonical concepts are structured and assembled that way and as such explains to learners why their methods did not work. The hidden efficacy is that students potentially learn through failure, making errors, and experimentation from their exploration and struggle in solving complex problems (Bjork & Kroll, 2015; Schwartz & Martin, 2004). Generating a response even when it is incorrect becomes an active process of attention to the specific concept being taught (Bjork & Kroll, 2015; Kapur 2008, 2009).
Productive failure aims to improve the abilities of dental hygiene students in problem solving skills. However, we have limited understanding about whether this design would be effective if the ill-structured complex problems did not have a canonical solution, but multiple solutions that are typical of open-ended problems, such as moral dilemmas. As the scope of practice for dental hygienists continues to expand and moral responsibility increases, cultivating professional responsibility in the profession is necessary (Blue, 2013).

The main goal of this study is to examine the effects of productive failure on student learning. The two research questions that guided this study were:

1. To what extent do students learn how to solve moral dilemmas in productive failure instruction compared to lecture and practice?

2. To what extent does productive failure help students learn skills in transferable problem solving?

In this dissertation, I discuss how these questions will be addressed and the contributions of these three papers. Chapter 2, I analyze productive failure design and the mechanisms that promote problem solving skills. I review the literature to unpack the principles of productive failure to explain how the conceptual components guide and explain student learning. I set out to examine previous studies on productive failure that were used in various learning formats to better understand how to implement the instruction design in a professional education program.

In Chapter 3, I designed a pilot study and conducted quantitative analyses to examine the effectiveness of productive failure as teaching method for dental hygiene students solving moral dilemmas. I also analyzed the extent to which productive failure
supports transfer of knowledge. The analysis drew on data from twenty-one students’ moral schemas and performance on moral dilemmas.

In the analysis presented in Chapter 4, I conducted a full-scale study to understand the development of moral reasoning skills of 77 students across four dental hygiene programs within the same southwestern state. Drawing on findings from Chapter 3, I made modifications to the study design, while examining the same research questions from Chapter 3. I was interested in analyzing potential patterns and shifts of moral schemas of dental hygiene students and identifying conditions with which productive failure predicted a greater degree of problem solving skills.

With increasing changes to the profession such as evolving scope of practice and healthcare needs for a growing multicultural and diverse population, students are expected to uphold ethical principles and high ethical standards of conduct. Developing problem solving skills are necessary to help students navigate through the complexity of dealing with oral health issues, changing skills in the practice of the profession, and encountering moral dilemmas presented in their professional life. Productive failure design is different from other instructional approaches because students are given room to create variations and multiple approaches to address a complex problem. Solving problems together, learning to think by explaining, listening, and asking questions that provoke understanding have been slow in being implemented in dental and dental hygiene education even though these are evidence-based teaching strategies (DePaola, 2012). Our profession still requires several high stake exams for licensure which can add to student stress. However, if we can teach our students to be effective problem solvers and facilitate their learning to encourage persistence to continue in their problem solving
methods, then I believe that students are better prepared to meet the challenges of moral dilemmas in the profession.
CHAPTER TWO

PRODUCTIVE FAILURE: INSTRUCTIONAL DESIGN

Karen Tam
University of Arizona
College of Education
Abstract

Productive failure, a teaching method, is based on the premise of unsuccessful learning performance in solving for complex problems with little to no support while yielding productive learning for subsequent problems. Kapur (2008) argues that hidden efficacies of learning exist in failure in which learners potentially learn through experimentation from their exploration and struggle in solving complex problems in a way that learners must first try and solve complex, novel problems on their own, but ultimately fails to reach a solution. This paper describes the principles of productive failure to explain how invention activities, delayed instruction, and the role of the teacher guide and support student learning. The types of assessments used in productive failure are also discussed. A review of the literature further enhances our understanding on how productive failure was implemented in various teaching settings and whether the hidden efficacies of learning existed.
Productive Failure: Instructional Design

Typically, problems in the real world are by nature complex and ill-structured. Sometimes trying to figure out the solution leads a person to struggle and even fail because the task is beyond one’s skills and ability. However, misconceptions that failure offers little efficacy and meaning in learning have stimulated discussion and research to explore the role of failure in learning (Kapur, 2008, 2009). There is growing body of research that is exploring the notion of failure and learning opportunities in problem solving (Hung, Chen, & Lim, 2009; Kapur, 2008, 2009, 2010; Kapur & Kinzer, 2009; Kennedy-Clark, 2009; Pathak, Kim, Jacobson & Zhang, 2011; Roll, Holmes, Day & Bonn, 2012; Wiedmann, Leach, Rummel & Wiley, 2012; Westermann & Rummel, 2012). Productive failure is a teaching method, based on the premise that unsuccessful learning performance in solving complex problems with little to no support yielding productive learning for subsequent problems.

The goal of productive failure is to enhance conceptual understanding and the ability to solve subsequent and future problems (Kapur 2008, 2009). Research that explores the elements of designing for failure may seem unproductive and inefficient. Why not tell the student the answer or use highly structured scaffolds to direct learning? Kapur (2008) argues that there are hidden efficacies of learning that exist in failure. The hidden efficacy is that learners potentially learn through failure and experimentation from their exploration and struggle in solving complex problems (Schwartz & Martin, 2004), but in the way that learners must first try and solve complex, novel problems on their own. Unlike productive failure, productive success does not allow learners to struggle too much and certainly not to reach failure, which means productive success learners may
not be exposed to the same hidden efficacies of learning as those in productive failure. Environments that will accommodate the goals of productive failure need to be designed to influence student learning and performance.

Currently, the focus of productive failure design research has focused on mathematics, science, and computer supported learning environments with middle or high school students (Kapur, 2008, 2009; Kapur & Kinzer, 2012; Pathak et al., 2011); and university students (Kennedy-Clark, 2009; Roll et al., 2012; Westermann & Rummel, 2012; Wiedmann et al., 2012). This could be due in large part to curriculum standards related to designing ways to problem solve in math and science areas. Interestingly, all the problems that were posed had a canonical solution (Kapur, 2008, 2009; Kapur & Kinzer, 2012; Kennedy-Clark, 2009; Pathak et al., 2011; Roll et al., 2012; Wiedmann et al., 2012; Westermann & Rummel, 2012). Wiedmann et al. (2012) suggest that problems with a canonical solution may be effective for students to understand lessons about new mathematical formulas more conceptually instead of just procedurally and that the age group may be a factor in learning by the activities used in productive failure. If problems have no canonical solutions, younger students may not benefit as much as college students because they lack the skills in dealing with complex problems (Wiedmann et al., 2012). Additionally, knowing how much prior knowledge is a necessary consideration before teachers begin to design complex problems. Therefore, choosing the appropriate target concept and the types of activities related to age group may be essential when using a productive failure approach.

Another consideration related to age group is deciding which type of support systems will keep groups engaged in the task of problem solving and maintain
persistence even if it leads to failure. Implementing support systems that attend to affective behavior may be necessary especially with novice learners who may not be familiar with the features that promote productive group work. Furthermore, scaffolds that are affective-related and not content-related teach novice learners what metacognitive processes adaptive experts use to persist when learning new concepts (Bransford et al. 2006; Roll et al., 2012; Westermann et al., 2012).

**Principles of Productive Failure**

To understand the principles of productive failure one needs to understand the role of failure in learning and problem solving. Examining impasse-driven learning, VanLehn et al. (2003) found that tutored students learned better when they got stuck or made an error compared to when they were told the answer. Coached tutoring, a collaborative problem solving design, keeps students on a correct solution path with the tutor pointing out the error and producing most of the explanations (VanLehn et al., 2003). VanLehn et al. (2003) explains that learning opportunities that end in an impasse provides opportunities for students to solve problems themselves (VanLehn et al., 2003). Students were more likely to keep thinking and talking until they produced an explanation that satisfied the tutor (VanLehn et al., 2003) and thus constructing learning that was more meaningful to them. Impasse-driven learning provides the space for students to explore, discover, and generate diverse solutions to complex problems.

When it comes to solving complex, ill-structured problems in the absence of support structures, there is the belief that there is little or no efficacy in failure (Kapur, 2008; Kapur & Kinzer, 2008). However, past research has suggested there may be a hidden efficacy in failure (Hung et al., 2009; Kapur, 2008; Schwartz and Martin, 2004;
VanLehn et al., 2003). Schwartz and Martin (2004) found that learners demonstrated a hidden efficacy in learner-generated conceptions and understanding even though these activities failed to produce canonical conceptions and solutions during the invention phase. The hidden efficacy is that learners potentially learn through failure and experimentation from their exploration and struggle in solving complex problems (Schwartz & Martin, 2004). Kapur (2008, 2009) found that as long as students were able to persist in the problem-solving processes they would find different ways of developing representations while producing diverse solution methods. Hung et al. (2009) expands on hidden efficacies of learning in that metacognitive processes can also take place in failure. When learners fail they are more likely to reflect upon and think about different possibilities that may change the status of their understanding (Bransford et al., 2006; Hung et al., 2009). Moreover, metacognitive processes can help learners isolate key factors involved in the problem even though the solution results in failure.

The principle of contrasting cases supports the goal of the ability to solve subsequent and future problems. According to Marton (2006), transfer is about people being able to do similar things in different situations by identifying the similarities and differences between those situations. In other words, the contrast in the problem structure helps learners discern between similarities and differences found in complex problems. By recognizing differences or variation in something in a situation, one can notice the similarities between the two situations (Marton, 2006). Marton (2006) further explains that transfer has to include a specific solution to the first problem. The link between contrasting problems is having an “identical element” or key concept factors in place between the two structured problems (Marton, 2006). The “identical element”
thereby influences the learner’s ability to transfer understanding from one problem to the next. The specific solution serves as a resource for learners to draw upon and that the solution has to be distinguished from the specific problem. This can only happen if there are at least two different instances or contrasting cases of the same solution which create experiences to notice or learn the key concepts (Marton, 2006). Consequently, Schwartz and Bransford (1998) support Marton’s (2006) idea of using contrast cases for transfer to future learning. For example, Schwartz and Bransford’s (1998) work on preparation for future learning demonstrated that when students examined patterns of similarities and differences among contrasting cases, it prepared them to derive greater benefit from a subsequent problem on that concept. The attention to these critical features may be what allowed learners to better conceptualize novel concepts (Kapur 2008, 2009).

Finally, the role of collaboration as a design principle is consistent with socioconstructivists theories of learning emphasizing the need to engage learners in the learning task. In productive failure, collaboration helps learners build shared understandings as they engage in the problem with group members. Their discussion and discourse in generating representational solution methods build to create a solution that is exclusive to the group itself. Furthermore, collaboration extends to the teacher with the learners. The teacher works with the learners on the canonical solution to the problem and then works with them to figure out what went wrong with the students’ solution methods. Collaboration is central to productive failure.

These principles guide the kinds of specific activities to be used in an environment of productive failure (Harris et al., 2006). Not only should instructional
activities be engaging to the learner, but it should connect with the goals of the design (Wiggins & McTighe, 2006).

**Invention Activities**

Invention activities such as ill-structured problems and well-structured problems representing a novel concept extract hidden efficacies of learning that lead to enhancing conceptual understanding of the novel concept (Kapur, 2008, 2009). According to Schwartz and Martin (2004), invention activities encourage a form of exploratory behavior. Ill-structured problems map onto the principle of impasse-driven learning (VanLehn, 2003) and hidden efficacies of learning (Hung et al., 2009; Kapur, 2008; Schwartz & Martin, 2004) to support the goal of learner-generated conceptions and understanding of novel concepts. Ill-structured problems are characterized to have many unknowns in the problem parameter of which leads to multiple solution paths. These problems require learners to make assumptions, judgments and express personal opinions and beliefs (Kapur 2008). Ill-structured problems allow learners to explore varying levels of relevant and specific solution methods thereby increasing interactions between individual learner and the problem as well as with group members (Kapur, 2008).

Additionally, structuring problems with many unknown parameters are similar to the problems that experts are more likely to encounter in the real world. Not only are ill-structured problems more authentic, but they are teach novice learners how to deal and persist with novel concepts in complex problems in the way that experts would when they learn new concepts or ideas (Kapur, 2008, 2009, Bransford et al., 2006; VanLehn et al., 2003). Following the principle of adaptive expertise, Bransford et al. (2006) explains that experts are constantly evolving their competencies and expanding their knowledge base.
as the need arises. Learners persist to engage in problem solving to generate as many diverse representations and solution methods at the same time practice skills in critiquing, explaining, and elaborating why their solution methods make sense (Kapur, 2008, 2009). Persisting in this process allow for learners to become flexible and adaptive in problem solving (Bransford et al. 2006; VanLehn et al., 2003).

Equally important are well-structured problems in productive failure since it is the counterpart to ill-structured problems. The principle that maps onto this instructional activity is the design of contrasting cases. The design of well-structured problems is to be restrictive in terms of freedom to explore for solutions (Kapur, 2008). Well-structured problems possess fewer parameters and a limited number of rules and principles that are organized in predictive ways of which reduces the exploration process. This problem design makes the solutions of well-structured problems more comprehensive to the learner and therefore, stops short of critiquing other possible solution paths that may exist. However, well-structured problems are important in that they make the novel concept more visible which is then becomes the link between the ill-structured and well-structured problems (Marton, 2006). The specific solution found in well-structured problems serves as the resource for learners to draw upon to transfer to ill-structured problems (Marton, 2006).

For productive failure, problem structuring was the experimental manipulation between the two groups: ill-structured problem group and well-structured group (Kapur 2008, 2009) Kapur & Kinzer, 2012; Pathak et al., 2011; Roll et al., 2012). Contrasting cases have a retrospective and prospective approach (Marton, 2006). This activity maps onto the principle of contrasting cases to transfer a novel concept in a same and different
situation (Marton, 2006; Schwartz and Bransford, 1998). Students benefit by learning how to structure future ill-structured problem and in recognizing and noticing patterns between the two problems. A hidden efficacy of learning in this design is the ability to solve subsequent future problems and therefore, learners become better solvers of ill-structured problems (Bransford et al., 2006; Kapur, 2009).

**Instructional Interventions of Productive Failure**

There are two phases in productive failure design. The first phase is the generation phase in which group members struggle with ill-structured problems to generate as many solution methods even though the process leads to failure. By delaying instruction and first solving ill-structured problems, learners attempt to assemble key ideas and elements to generate their own solution methods and representations despite failing in their solution methods (Kapur, 2008, 2009). Delaying instruction also allows the learner to reach an impasse (VanLehn, 2003) which offers opportunities and space for learners to come up their own solutions. Kapur (2008, 2009) found that delayed instruction helped learners to discern how to structure an ill-structured problem and later apply it to other ill-structured problem. In this phase, the principle of persistence is vital to keep the learners in the problem solving process of developing diverse representations and multiple solution methods (Kapur 2008). To help learners persist, the teacher encourages the learner using affective support. If learners do not persist there is likelihood that the number of generated and quality of solution methods would be minimal and that the goal of conceptual understanding would not be met.

The second phase is the instruction or consolidation phase. Consolidation is given either in the form of direct instruction or well-structured problem solving lead by
the teacher. Collaboratively, students work in applying the correct solution and reflecting upon why the group’s incorrect solution failed. The consolidation phase assists learners on how to organize and assemble key ideas and concepts as they are shown the correct solution. In this phase, teachers can coach learners either by direct instructions or well-structured problem solving to notice and recognize critical features of the novel concept and then apply them to other problems. Moreover, feedback in the form of comparisons of the canonical solution to their failed solution method, reflection upon why their solution method failed, and practice in applying the canonical solutions to well-structured problems may lead to a deeper understanding of the novel concept. Without the consolidation phase, learners are more likely to flounder around trying to find a solution, which in turn may make the learning opportunity extremely frustrating and unproductive.

Assessments of Productive Failure

Assessments provide evidence of the understanding that is being sought in productive failure (Wiggins & McTighe, 2006). Appropriate assessments to demonstrate performance and conceptual understanding are vital in deciding if the productive failure approach is considered a good way of teaching. In productive failure, groups are asked to produce artifacts of their problem representations and solution methods. These drawings and sketches provide a source of information in terms of what group work looks like, how problems are attended to and possible patterns across all the groups (Kapur 2008, 2009). Much like showing mathematical work on paper, evidence of identifying and linking important elements may be important in how learners discover the novel concept being taught. Informal checks for understanding like generated representations and
attempts provide feedback to the teachers about learners’ understanding and conceptualization of the novel concept.

Group discussions are used as another informal check for evaluating the relationship between the representations produced by the groups related to solution methods that groups discussed and used in their attempt to solve the problem (Kapur, 2008). The group discussions can reveal complexity of the problem which, of course, is necessary as problem structure is a key element in productive failure. Complex problems are expected to produce diverse representations and multiple solution methods because of low group convergence (Kapur, 2008). Group discussions in productive failure are chaotic and divergent, but qualitatively insightful in how the group builds their own original representations based on each other’s questions, critique, and evaluation process. Additionally, teachers can hear if a lack of persistence begins to set in and therefore, step in and provide affective scaffolds to continue in solving the problem.

Finally, post-tests are used to assess the performance of an individual learner’s understanding of the targeted concept. The design of each post-test is important to the goal of productive failure. The first post-test assesses if learners can apply the targeted concept to other well-structured problems. The second post-test assesses if learners are able to solve another ill-structured complex problem with a concept that had not been taught as of yet. Since the goal of productive failure is to enhance conceptual learning and understanding, the outcome of the post-tests would be evidence to demonstrate whether or not learners learned and understood the novel concept. The second post-test in its ill-structured problem form would also provide evidence of goal attainment in that
transfer of the newly learned concept could be applied to another problem with more complexity and a concept that has not been learned (Kapur, 2008, 2009).

**The Role of the Teacher**

The role of the teacher in productive failure is to create the groupings, if necessary, provide affective scaffolding to continue and persist in problem solving, be the expert in designing the structured problems and share the canonical solutions to the targeted concepts with the student. Teachers using this type of instructional design would need to be familiar with the student’s prior knowledge and the learner’s Zone of Proximal Development by Lev Vygotsky as Kapur (2008, 2009) points out when developing structured problems in the invention activities. If the learner does not have enough background knowledge then learning may become even more difficult and frustrating. Consequently, learners with too much knowledge may avoid struggling altogether to find a solution thus hidden efficacies in learning by struggling may not exist for those learners. In groups, if one learner has more knowledge than other group members, then the goal of self-critiquing, exploring and critiquing each other’s solutions would be limited (Kapur 2008). Such problem development also requires teachers to have a solid understanding of the novel concept being posed as well as how to design structured problems that have specific criteria necessary to be considered ill-structured or well-structured. Professional development or even collaboration with other teachers in the same discipline may need to be available for teachers to learn how to design specific structured problems. A problem designed with little thought may impact the effectiveness of the instructional design making the activity disconnected from the goal of conceptual understanding.
Additionally, the role of the teachers in the generation phase of productive failure is to provide affective scaffolding to continue and persist in the problem solving process. This might require some training to learn how to provide affective scaffold without revealing content. Moreover, teachers may have to teach learners various strategies on how to work collaboratively for productive learning. Although collaborative group work is not new in educational teaching settings there may be scaffolding techniques that work better than others in productive failure.

**Scaling Productive Failure Learning Environment**

Productive failure has been used in classroom settings (Kapur 2008, 2009; Roll et al., 2009; Westermann & Rummel, 2012; Wiedmann et al., 2012) and with computer technology (Kapur & Kinzer, 2009; Pathak et al., 2011). In order to scale up, the focus of productive failure has to move into larger contexts of schools and school systems. Since problem solving is a key component in state and national standards, productive failure provides a learning environment that meets those standards. Moreover, productive failure researchers have moved away from the initial studies of productive failure (Kapur 2008, 2009) in middle to high schools to across university school systems (Roll et al., 2009; Westermann & Rummel, 2012; Wiedmann et al., 2012) and professional development courses (Lai, Portolese, & Jacobson, 2016). Minimal to no modifications were made to existing school structures during the implementation of productive failure. In fact, productive failure design stayed within the institutions’ time constricts making the timeframe relatively short for middle school at 50-minutes (Kapur, 2008, 2009) and high school at 50 minutes to 1.5 hours (Kapur & Kinzer, 2009; Pathak et al., 2011); whereas, timeframe at university institutions were close to 1.5 to two hours (Roll et al.,
2012; Westermann & Rummel, 2012; Wiedmann et al., 2012). In terms of technological infrastructure, the researchers incorporated what students used on a daily basis (Kapur & Kinzer, 2009; Pathak et al., 2011). This could be seen as a valid design that can fit within an existing educational structure. If the productive failure hypothesis could be demonstrated and replicated with minimal changes to the school curriculum, teacher training, and technological infrastructure, and within a relatively short timeframe, then it would only speak well of productive failure design’s practical significance (Kapur & Kinzer, 2009). However, if one is looking for examples of the structure and design of productive failure, Kapur’s (2009) study describes the theoretical framework behind productive failure, details the learning goals, lists the intend teaching practices and provides specific criteria of designing low and high structured problems. It would be a good place to start, but would probably need more materials to ensure its success to other school systems.

**Research in Productive Failure**

Productive failure design research is ongoing with the initial study by Kapur in 2008. Since then, a growing body of research is being conducted to analyze the features that support this instructional design (Hung et al., 2009; Kapur, 2008, 2009; Kapur & Kinzer, 2009; Kennedy-Clark, 2009; Lai et al., 2016; Loibl & Rummel, 2015; Pathak et al., 2011; Roll et al., 2012; Song & Kapur, 2017; Westermann & Rummel, 2012; Wiedmann et al., 2012). One set of studies kept to the original structure of productive failure where minimal scaffolding process were focused on helping the learners persist in the generation phase with no structured technique (Kapur 2008, 2009; Lai et al., 2016; Loibl & Rummel, 2015; Song & Kapur, 2017). Two other studies implemented various
affective scaffolding in the form of role-play scripts (Westermann & Rummel, 2012) and
metacognitive scaffolding using Guided Inventions (Roll et al., 2012). These studies
support Kapur’s (2008, 2009) finding that the invention phase helps learners become
more effective problem solvers in the absence of highly structured scaffolding.
Westermann and Rummel’s (2012) aim was to extend the positive effect found for
delaying instruction to different learning situations. The authors implemented a learning
condition known as Think Ask Understand (TAU) to college students. Students were
paired and one of the students took on the role of the thinker who explains the problem or
solution and the other student took the role of the questioner who asks questions about
the student’s explanation as the role-play script. Westermann and Rummel (2012) found
that by adding the role-play script as a collaborative problem solving method, the
students were more productive in their collaboration efforts by asking more questions,
elaborating on explanations, and triggering metacognitive processes used to monitor their
own understanding. As such, the TAU learning condition did not affect the students’
initial interaction with the problem even when the group failed in finding a correct
solution. Results of post-test found that students using the TAU condition in the posttest
problem performed better once they adopted the role-play script as second nature.

Roll et al. (2012) studied metacognitive scaffolding effects in supporting students
in the generation phase and by using contrasting cases for the students to compare.
Guided Inventions are scaffolding prompts to help students engage in the exploratory
analysis when learning a novel concept in physics. The first prompt asked students to
rank and compare and analyze four different contrasting cases. In the second prompt, the
students were asked to provide self-explanations by explicitly asking them to explain
their reasoning and the third prompt was to compare answers to other group members prior to designing their group methods. Students in the Guided Invention groups made almost twice as many high-level comments per students compared with Unguided Invention students who did not receive any prompts ($M=.55$; $F(92, 131) = 4.0, p=.05$) making an average of 0.91 comments per student on target features compared with 0.34 comments per student in Unguided Invention groups. Thirteen percent of the students in the Guided Invention groups presented multiple solutions compared with 3% in the Unguided Invention students. The findings are consistent with Kapur (2008, 2009) that with invention activities, students made more comments on target features to help them better understand canonical solutions. Also, these students used reflective reasoning practices more often when developing original solutions methods. This study corroborates with Kapur (2008, 2009) in that exposing students to the challenges of the domains prior to giving them tools to overcome these challenges is an important and effective design feature of productive failure.

Wiedmann et al.’s (2012) study investigated the effectiveness of invention activities and as mediated by small group composition of math abilities. This study found that groups with mixed ability students produced the most approaches compared to high ability groups and low-ability group. Additionally, group compositions of students in either high ability groups or mixed groups did not differ in post-test performance.

More important, Wiedmann et al.’s (2012) study supports Kapur’s (2008, 2009) findings that groups who produced more different approaches or solution methods performed better on post-test. Wiedmann et al. (2012) used contrasting complex problems similar to Kapur’s (2008) to create learner-generated representation and
solutions methods. The authors found that the number of representations produced by the mixed ability groups during the generation phase significantly affected their performance on conceptual understanding of a novel concept and transferability of the conceptual understanding in the post-test, $F(2,65) = 6.95$, $p=.01$, $\eta^2 = .18$. This is significant in that multiple research studies observed that the more generated representations and attempts the better the uptake and understanding of novel concepts (Kapur 2008, 2009; Roll et al., 2012; Wiedmann et al., 2012). Moreover, groups solving for ill-structured problems presented broader and multiple solution methods compared to well-structured groups (Kapur 2008, 2009; Roll et al., 2012; Wiedmann et al., 2012). Groups in the productive failure design significantly outperformed those in the overly structured design in post-tests (Kapur, 2008, 2009; Wiedmann et al, 2012).

Studies have examined productive failure by infusing technology into teaching and learning, primarily in the context of computer-supported collaborative learning (CSCL) and online courses (Lai et al., 2015; Pathak et al., 2011). Most of the design elements were exactly the same except for the communication format between learners. The CSCL structure used a text chat only application to communicate on the invention activities instead of paper and pencil used in Kapur’s (2008) initial study; whereas the online course did not have an instructor and learners did not collaborate with one another. Instead, an online course unit was developed as the direct instruction with quizzes at the end of each topic. Results were similar to the Kapur’s (2008) in that learners were more engaged in the problems generating representations and multiple solution methods. Pathak et al. (2011) found that overtly scripted activities did not produce authentic practices that experts in the field would engage in and productive failure learning
students were more likely to use reflective reasoning principles. Students in the productive failure group were more random in their exploration and able to articulate different levels of the structure in the problem from domain-specific and domain-general. Therefore, students were more engaged in finding a solution compared to the non-productive failure group. Pathak et al. (2011) also found that the productive failure group used reflective reasoning to understand the findings.

Lai et al. (2015) found that even without an instructor, process engineers gained greater conceptual understanding and were able to transfer their new knowledge to new problems. By merely changing the sequencing of learning activities, the engineers recognized they had pre-existing gaps of knowledge which were then filled in after taking the online course unit. This finding supports existing productive failure studies that prior knowledge is relevant when it comes to recognizing existing gaps of knowledge and repairing existing mental model.

**Conclusion**

Productive failure has many positive benefits and effects in terms of short and long term memory in conceptual understanding and transfer. Productive failure adds to the literature of designing effective learning environments for best teaching practices by looking at failure in terms of developing problem solving strategies, deeper understanding, and skills to transfer learning by applying concepts that have not been introduced to the learner to an ill-structured problem.
CHAPTER THREE

A PILOT STUDY EXAMINING PRODUCTIVE FAILURE AS AN EFFECTIVE TEACHING METHOD IN DENTAL ETHICS

Karen Tam
University of Arizona
College of Education
Abstract

There is an unmet need for instructional methods in teaching dental ethics. This pilot study examines the effectiveness of productive failure, an instructional design compared to lecture and practice as a teaching method for dental hygiene students solving moral dilemmas and the extent to which the productive failure supports transfer of knowledge. Twenty-one second-year dental hygiene students at a community college in the southwestern United States. Analysis did not show significant differences on moral reasoning ($p = .06$) and transfer of knowledge ($p = .58$), but the effect size on students’ posttest scores was high ($d = .76$). These pilot results suggest that productive failure might enhance students’ approach to moral dilemmas and, with adjustments in instructional and research design, a full-scale study is warranted.

*Keywords:* dental hygiene ethics instruction, solving moral dilemmas, productive failure, transfer of knowledge
A Pilot Study Examining Productive Failure as an Effective Teaching Method in Dental Ethics

One of the problems faced by dental educators is that there is an unmet need for instructional methods in teaching dental ethics especially in the use of active learning and self-assessment (Lantz, Bebeau & Zarowski, 2011). With recent indictments levied against colleges and universities concerning their economic value (Bennett & Wilezol 2013), it has become important for researchers to examine college and its impact on student learning (Berk, 2001). Part of the examination process involves scrutinizing the learning purported to take place on college campuses and the mechanisms researchers use for its assessment (Bennett & Wilezol 2013) and professional programs are no different. Accreditation standards are holding colleges and programs accountable in how learning is measured empirically and the extent to which measurements themselves may influence the learning process (Commission on Dental Accreditation (CODA), 2013). Specially, recent CODA standards (2013) were approved in the area of professionalism. The most widely used instructional methods to develop these skills are structured in the form of lectures, formal examinations, small group designs, case-based formats, problem-based learning, and reflective writing (Berk, 2001; Brondani & Rossoff, 2010; Christie, Bowen &Paarmann, 2007; Duly, Fitzpatrick, Zormosa, Lambert & Mitchell, 2009; Lantz et al.,2011).

One approach to teaching ethics in dental hygienic is to engage the learners in solving moral problems or dilemmas. Yet, to be effective, situated problems need to be complex. While these problems encourage critical thinking, they generally use support structures like instructor scaffolding and intentional guiding to keep students from failing. Recent studies have shown that structured support activities may actually interfere in how
students prepare to learn (Kapur, 2008; Kapur & Bielacyzc, 2012; Schwartz & Martin, 2004) by constraining students to a narrow scope of problem solving and limiting the production or generation of other possible solutions (Kapur, 2008; Kennedy-Clark, 2009) and that students can learn even when they reach an impasse solving complex problems (Van Lehn, 2000). Moreover, other studies have shown there are benefits to learning in the process of struggle, negotiation, and persistence to solve for solutions in complex problems (Kapur, 2008, 2009, 2011) and how failure, more specifically short-term failure, and the process of failure may be necessary to the learning process and in transferring knowledge and skills for future novel problems.

Productive Failure (PF), an instructional design, is based on the premise that learning that comes from the experience of struggle and failure in solving complex problems with little to no support from teachers can lead to students solving subsequent complex problems (Kapur 2008, 2009, Kapur & Bielacyzc, 2012; Kennedy-Clark, 2009; Roll, Holmes, Day & Bonn, 2012). Hung, Chin, and Lim (2009) describe learning as a developmental process and if that process does not allow for experimentation and struggle, the hidden efficacy of learning, cognition and metacognition may be inhibited, thus limiting students’ further learning once the outcome has been reached. While failing as an outcome is undesirable for most learning and most learning opportunities prevent students from failing, Kapur and Hung et al. argue that failure may be necessary to develop the metacognition for deeper learning, which can then lead to solving future problems. Kapur (2008) defines productive failure as the student’s engagement and involvement with a complex, ill-structured problem to figure out a solution only to find no solution and ultimately failing. There is growing research exploring the notion of

By conducting an experimental study, I sought to provide data to examine if Kapur’s (2009) productive failure is effective as a teaching method for dental hygiene students solving moral dilemmas with multiple non-canonical solutions. The study examined the effects of productive failure and its outcome in student learning. The research questions that guided this study were:

1. To what extent do students learn how to solve moral dilemmas in Productive failure compared to lecture and practice?
2. To what extent does instruction in productive failure help students learn skills in transferable problem solving?

I hypothesize that learners who receive productive failure will outperform learners in lecture and practice instruction in moral development on open-ended moral dilemmas. Additionally, for learners in the productive failure, I hypothesize that productive failure students will activate their prior knowledge differently to better prepare and transfer problem solving skills to a novel problem. Subsequently, learners will engage in more complex interactional discussions and generate and produce more inventive solutions to
the problem than learners in a lecture and practice instruction design. Understanding learning processes and conditions in which students learn best can help instructors provide purposeful and meaningful instruction and feedback for students who may struggle with problem solving skills.

**Definitions of Key Concepts**

In order to describe the theory of Productive Failure, there is a need to define key concepts that was used to frame this study.

- **Productive Failure Model** (Kapur, 2008) is engaging students in solving complex, ill-structured problems **without** support structures (scaffolding, expert guiding, providing resources, etc.). Invention activities to stimulate prior knowledge precede instruction to deepen understanding of complex problems.

- **Transfer** is a product of deeper learning through which the learner takes what was learned in one situation and applies it to a novel situation, a central topic in education (NRC, 2012).

- **Dilemma** “is a situation in which rights or interests of affected parties conflict. Alternatively, a dilemma can be described as a situation in which the protagonist feels he/ she has conflicting obligation” (Bebeau, 1995, p.3).

- **Moral Judgment** “is a psychological construct that characterizes the process by which people determine that one course of action in a particular situation is morally right and another course of action is wrong. Moral judgment involves defining what the moral issues are, how conflicts among parties are to be settle, and the rational for deciding on a course of action” (Rest, Thoma & Edwards 1997, p.5).
Literature Review

Academic dental institutions introduce students to a “moral community” where students are prepared to exhibit values that are central to the profession and society (Taichman, Parkinson, Nelson, Nordquist, Young, & Thompson, 2012). However, most students do not see the study of professionalism or ethics as challenging enough to maintain real interest and they see ethics as subjective, making it difficult to evaluate someone else’s moral position (Schwitzgebel, 2013). Moreover, researchers argue that ethical training really does not change students’ ethical reasoning (Ritter, 2006; Schwitzgebel, 2013) and that if effects were found, they were so small that any positive effects negated the ethics training may have had on the student (Schwitzgebel, 2013). Furthermore, the attitude of the student about ethical issues, their moral compass, and ability to make ethical judgments significantly influences the ethical decision-making process and outcome (Duley, 2009; Kacerik, Prajer, & Conrad, 2006). However, Berk (2001) reported that teaching approaches in ethics have moved from didactic teaching to more interactive and introspective. Case-based learning, small groups, and problem-based learning are now the most commonly used teaching approaches emphasizing communication and group problem-solving behaviors, self-reflection, and ethical examinations (Berk, 2001). Considerable attention is being paid to transform the learning experiences to help develop students’ ethical reasoning from school to practice by seeking and developing new methods that support and provide students opportunities to develop moral reasoning. For example, in 2013, changes in the Commission on Dental Accreditation, Standard 2–22 now call for dental hygiene students to be able to critique and apply the principles of “ethical reasoning, ethical decision making and professional

With an aging population expected to double in 30 years and chronic disease on the rise, there are huge implications on its effect on oral health (HRSA, 2013). In the future, dental hygienists will be seen as primary healthcare providers in the nation’s healthcare delivery system and not people seen only in a dental office under the supervision of dentists. In fact, the future dental hygienist “would be trained to work independently in a community clinic. The dental hygienist should be able to assess risk and manage disease, be mindful of the needs of special groups, and show cultural competence, as well as knowledge of health serves research and public policy development” (HRSA, 2013, p.7). No longer working as a dependent healthcare provider, the future dental hygienists are working collaboratively as independent interdisciplinary team members with, but not limited to, nurses, pharmacists, physicians, and public health professionals. As such, the current educational model of dental hygiene programs and their curriculum will require an evaluation in how to prepare future dental hygienists to work collaboratively as interdisciplinary team members.

According to the Human Resources and Services Administration (HRSA, 2013) report, there are four core competencies or domains that curriculum should include to promote newer approaches to practice: 1. Roles and responsibilities, 2.Interprofessional education, 3.Teams and teamwork, and 4. Values / ethics. With the responsibility of working independently and the emphasis of ethics and values in the new curriculum, current forms of instruction in ethics used in dental hygiene schools may not be sufficient enough to prepare students for collaborative practice. Although Berk (2001) reported
that trends of teaching ethics are more interactive and introspective, several studies suggest that ethics are generally taught using traditional methods of lectures, discussion, quizzes, and written assignments (Duley, 2009; Lanz et al., 2011). These forms of instruction emphasize the memorization of facts with minimal challenge and are inadequate in fostering critical thinking and problem solving skills that are necessary in developing moral and ethical reasoning (Duley, 2009; Lantz et al., 2011; Taichman et al., 2012). To develop critical thinking and problem solving skills, students must have time and space to reflect, think critically, and make connections between failed attempts and successful endeavors of the target concepts by asking questions that challenge existing ideas, and apply those ideas to new solutions and in novel ways (Kapur, 2008; Taichman et al., 2012). Productive failure is an instructional method that emphasizes self-directed learning by designing learning activities that challenge students to critically think to generate multiple solutions in novel ways (Kapur, 2010).

**Productive failure**

Similar to current teaching methods, the instructional design of productive failure (Kapur, 2008) is centered around situative and social constructivist learning theories and beliefs by working collaboratively in small groups on authentic complex problems. In a series of studies, Kapur (2008, 2009, 2011) argues that designing learning activities that challenge students to be creative can have long-term learning benefits even though students might reach an impasse and fail.

Productive failure (Kapur, 2008), a teaching method, is based on the premise that unsuccessful learning performance in solving for complex problems with little to no support can yield productive learning for subsequent problems. Using Kapur’s (2011)
definition, failure is when the student is unable to solve the problem by themselves or incorrectly solves the problem. Thus, in this proposed study the conceptual definition of productive failure is engaging students in solving complex, ill-structured problems without the provision of support structures even when performance is low and an impasse or failure is reached (Kapur, 2008, 2009). Despite students believing they failed in solving ill-structured problems initially, Kapur (2008, 2009) found that students outperformed traditional lecture and practice instruction during transfer tasks. Although Kapur’s studies (2008, 2009) were on high school students in math, other studies found similar results on university undergraduate students taking courses in math (Weidman, Leach, Rummel & Wiley, 2012; Westermann & Rummel, 2012), physics (Roll 2012), and ecology sustainability (Trueman, 2014).

There are two phases in productive failure design. The first phase is the generation phase in which group members struggle with ill-structured problems to generate as many solution methods even though the process leads to failure. The teacher groups learners in dyads or triads. By delaying instruction and first approaching ill-structured problems, learners attempt to assemble key ideas and elements to generate their own solution methods and representations despite failing in their solution methods (Kapur 2008, 2009). Delaying instruction also allows the learner to reach an impasse (VanLehn, 2003), which offers opportunities and space for learners to come up their own solutions. Kapur (2008, 2009) and others (Loehr, Fyfe, &Rittle-Johnson, 2015; Trueman, 2013) found that delayed instruction helped learners to discern how to structure an ill-structured problem and later apply it to other ill-structured problem. In this phase, the principle of persistence is vital to keep the learners in the problem solving process of
developing diverse representations and multiple solution methods (Kapur, 2008). If learners do not persist there is likelihood that the number and quality of solution methods would be minimal and that the goal of conceptual understanding would not be met.

The second phase is the instruction or consolidation phase. Consolidation is given either in the form of direct instruction or well-structured problem solving lead by the teacher. Collaboratively, students work in applying the correct solution and reflecting upon why the group’s incorrect solution failed. The consolidation phase assists learners on how to organize and assemble key ideas and concepts as they are shown the correct solution. In this phase, teachers can coach learners either by direct instruction or well-structured problem solving to notice and recognize critical features of the novel concept and then apply them to other problems. Moreover, feedback in the form of comparisons of the canonical solution to their failed solution method, reflection upon why their solution method failed, and practice in applying the canonical solutions to well-structured problems may lead to a deeper understanding of the novel concept. Without the consolidation phase, learners are more likely to flounder around trying to find a solution, which in turn may make the learning opportunity extremely frustrating and unproductive.

When it comes to solving complex, ill-structured problems in the absence of support structures, there is the belief that there is little to no efficacy in failure (Kapur, 2008; Kapur & Kinzer, 2008). However, past research suggests there may be a hidden efficacy in failure (Hung et al., 2009; Kapur, 2008; Schwartz & Martin, 2004; VanLehn et al., 2003). The hidden efficacy is that learners potentially learn through failure and experimentation from their exploration and struggle in solving complex problems (Schwartz & Martin, 2004). Schwartz and Martin (2004) found that learners generated
conceptions and understanding even though the activities failed to produce canonical conceptions and solutions during the Invention phase. Kapur (2008, 2009) found that as long as students were able to persist in the problem-solving processes they would find different ways of developing representations while producing diverse solution methods.

Hung et al. (2009) expand on the hidden efficacies of failure in that metacognitive processes can also take place in failure. When learners fail they are more likely to reflect upon and think about different possibilities that may change the status of their understanding (Bransford et al., 2006; Hung et al., 2009). Moreover, metacognitive processes such as reflection can help learners isolate key factors involved in the problem even though the solution results in failure. Other researchers found that adding supports in the form of affective support by the instructor (Kapur, 2009), group collaboration (Westermann & Rummel, 2012) and metacognition scaffolding by group members on tasks and domain-specific inventions (Roll et al., 2012) improved the inquiry process without compromising the fundamental principles of productive failure design because the supports were not content-specific and therefore did not lead students to a canonical solution of the problem.

This study applies the instructional design created by Kapur (2008, 2009; Kapur & Bielaczyc, 2012) and extends the previous work of prior investigators to open-ended problems such as moral dilemmas that have multiple perspectives and possibly multiple canonical solutions.

Productive Failure and Transferable Knowledge

Transfer of knowledge and skills from one situation to a novel situation is argued by many researchers to be a central topic in education (National Research Council, 2012).
However, current forms of instruction in ethics employed by dental hygiene education programs are lectures, discussion, quizzes, and written assignments (Duley, 2009). This type of learning primarily requires learners to recall facts, concepts or procedures, rather than helping learners engage in deeper learning. Many reports have revealed that this form of instruction is inadequate to prepare dental hygiene students in real life experiences (Duley, 2009; Kacerik et al., 2006; Lantz et al., 2011).

In a review of research on instructional strategies for learning and transfer, Bransford et al. (2006) suggests an adaptive expertise approach to help novice learners prepare for future learning and transfer rather than a recall of facts and procedures. The main tenets of adaptive expertise centers around the learner to recognize and notice familiar elements (prior knowledge), work through the problem by inventing new strategies, testing them, and disconfirming them to solve novel problems (Bransford, 2006). Similar to experts solving problems, noticing familiar elements requires the expert to have a good grasp of the underlying principle or concept in order for transfer to a new situation to occur. Consequently, novice learners tend to learn concepts superficially since they are less likely to initially identify key elements (Bransford, 2006; Mayer, 2010). This process of noticing key elements describes deeper learning through which the learner takes what he or she learned in one situation and applies it to a novel situation and thus, transferrable knowledge (NRC, 2012). Much different than the recall of facts disassociated from meaning and context, problem solving using a deeper learning process enhances the retention of moral concepts as well as develops skills that are relevant in the dental field (Lantz et al., 2011).
Invention activities such as ill-structured problems and well-structured problems representing a novel concept extract hidden efficacies of learning that lead to enhancing conceptual understanding of the novel concept (Kapur, 2008, 2009). According to Schwartz and Martin (2004), invention activities encourage a form of exploratory behavior. Ill-structured problems map onto the principle of impasse-driven learning (VanLehn, 2003) and hidden efficacies of learning (Hung et al., 2009; Kapur, 2008; Schwartz & Martin, 2004) to support the goal of learner-generated conceptions and understanding of novel concepts. Ill-structured problems are characterized by having many unknowns in the problem parameters, which leads to multiple solution paths. These problems require learners to make assumptions and judgments and express personal opinions and beliefs (Kapur 2008). Ill-structured problems allow learners to explore varying levels of relevant and specific solution methods thereby increasing interactions between individual learners and the problem as well as with group members (Kapur, 2008). Additionally, structuring problems with many unknown parameters is similar to the problems that experts are more likely to encounter in the real world. Not only are ill-structured problems more authentic, but they teach novice learners how to deal and persist with novel concepts in complex problems in the way that experts would when they learn new concepts or ideas (Bransford et al., 2006; Kapur, 2008, 2009; VanLehn et al, 2003). Following the principle of adaptive expertise, Bransford et al. (2006) explain that experts are constantly evolving their competencies and expanding their knowledge base as the need arises. Learners persist in problem solving to generate as many diverse representations and solution methods, while at the same time they practice skills in critiquing, explaining, and elaborating why their solution methods make
sense (Kapur, 2008, 2009). Persisting in this process allows learners to become flexible and adaptive in problem solving (Bransford et al. 2006; VanLehn et al., 2003).

Equally important in productive failure are well-structured problems because they are the counterpart to ill-structured problems. The principle that maps onto this instructional activity is the design of contrasting cases. The design of well-structured problems restricts freedom to explore solutions (Kapur, 2008). Well-structured problems possess fewer parameters and solutions that require a limited number of rules and principles organized in predictive ways, thus limiting the exploration process. This problem design makes the solutions of well-structured problems more comprehensible to the learner requiring less mental effort and therefore, stops learners short of critiquing other possible solution paths that may exist. However, well-structured problems are important in that they make the novel concept more visible, which becomes the link between the ill-structured and well-structured problems (Marton, 2006). The specific solution found in well-structured problems serves as a resource for learners to draw upon to transfer to ill-structured problems (Marton, 2006).

In productive failure studies, problem structuring was the experimental manipulation between the two groups: an ill-structured problem group and a well-structured group (Kapur 2008, 2009; Kapur & Kinzer, 2012; Pathak et al., 2011; Roll et al., 2012). Contrasting cases such as ill-structured problems to well-structured problems employ a retrospective and prospective approach (Marton, 2006). This activity maps onto the principle of contrasting cases to transfer a novel concept in same and different situations (Marton 2006; Schwartz & Bransford, 1998). Students benefit by learning how to structure future ill-structured problem and in recognizing and noticing patterns
between the two problems. A hidden efficacy of learning in this design is the ability to solve subsequent future problems and therefore, learners become better solvers of ill-structured problems (Bransford et al., 2006; Kapur 2009).

**Ethics Dilemma**

Despite evidence supporting the effectiveness of productive failure, most problems in productive failure studies have been in mathematics (Kapur, 2009; Loehr et al., 2014; Westermann & Rummel, 2012; Weidmann, Leach, Rummel & Wiley, 2012), physics (Kapur, 2008, Roll et al., 2012), computer learning (Pathak, Kim, Jacobson & Zhang, 2011), and science (Kennedy-Clark, 2009; Trueman, 2013). These types of problems are designed to have one main canonical solution. However, most problems in health care, specifically ethical dilemmas in dental hygiene, have multiple perspectives in that there is not one canonical solution (Bebeau, 1995), but many possibilities. Most dental hygiene ethical dilemmas relate to making decisions based on providing the highest level of patient care and that makes the most economic sense (Duley, 2009). Based on Kohlberg’s (1981) theoretical framework for moral reasoning, people are guided by one’s perception of what constitutes an ethical dilemma (cognition) and the interactions with society and situations of what behaviors should be applied (society). Kohlberg’s (1981) theory consists of three levels that describe one’s decision making preferences based on the individual’s concerns, rules of social behaviors, and view-point of what constitutes equality, justice, and human rights (Bergling, 1981). As such, there are multiple pathways and judgments that an individual can make depending on one’s moral schema preference or type all of which can be seen as acceptable points of view (Bebeau, 1995; Bebeau & Thoma, 2003). Although some studies report an increase in
moral judgment as a result of changes made in dental ethic courses, (Bebeau, 2002; Lantz, 2011) the complexity of making ethical decisions is still unclear to dental hygiene students (Duley, 2009) and yet little has changed in the methods of instruction in dental hygiene curriculum (Duley, 2009; Kacerik, Prajer, & Conrad, 2006). In a survey to 147 dental hygiene programs, Kacerik et al. (2006) found that the majority of ethics course were still being taught using class discussion (99%) and lecture (97%). Thus, there is a need for implementing instructional methods that focuses on ethic concepts in real life experiences. Specifically, in the development of moral reasoning Bebeau (1995) explains that a student who learns by repeating the appropriate solution to one problem may have a difficult time transferring the idea to new situations which means that conceptualizing moral problems is inadequate for the student. In this study, I created problems based on ethical dilemmas related to exceeding scope of practice and conflicts in allocating resources for dental hygiene students to solve and generate as many multiple solutions path. Given the potential benefits of productive failure in generating multiple solutions and the transfer of knowledge to novel situations, it is worthy to explore the boundaries of productive failure with problems that do not have one canonical solution, but several possibilities.

**Methods**

**Participants**

Second year dental hygiene students (N = 21) from a community college in the southwestern United States participated in the study. The setting of this study was chosen out of convenience because it is my place of employment. Table 3.1 describes the
demographic makeup of participants. All 21 participants were female with a mean age of 30 (SD =7.23). This is not uncommon since 98% of dental hygiene students across the U.S. are females (ADEA, 2015). In this study, students identified themselves as Asian/Pacific Islander (1), Hispanic (9), Caucasian (8), and Other (3). Fifty-seven percent of study participants (12) completed Junior College, 29% of students (6) completed a vocational or technical college (29%), and 10% or two students went on to complete most of their senior year in college. The majority of study participants (81%) reported that English was their primary language with 19% of participants (4) reported speaking another language.

Table 3.1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N = 21</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>21</td>
<td>100</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Hispanic</td>
<td>9</td>
<td>43</td>
</tr>
<tr>
<td>American Indian/ Other Native American</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Caucasian (other than Hispanic)</td>
<td>8</td>
<td>38</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Education level completed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocational/ Technical School</td>
<td>6</td>
<td>29</td>
</tr>
<tr>
<td>Junior College</td>
<td>12</td>
<td>57</td>
</tr>
<tr>
<td>College Senior</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>English as primary language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>17</td>
<td>81</td>
</tr>
<tr>
<td>No</td>
<td>4</td>
<td>19</td>
</tr>
</tbody>
</table>
Ethical Safeguards

University of Arizona’s Institutional Review Board (IRB) and Pima Community College IRB reviewed and approved my study in order to protect the participants. I obtained approval from the Dean of Health Related Professions and Dental Studies Department Chair in which I conducted my research. After I gained permission, each participant signed a consent form to participate in the study. In this study, informed consent was obtained (see Appendix A) and no reporting of the actual names of the participants occurred. Instead the students’ actual names were replaced by an assigned number that only the researcher had access to. This research did not affect the grade of any student involved in this study. The students in either the PF or LP instructional design received the same content and material. Confidentiality and privacy of students’ actual names, raw data, and codes to the data were stored in a locked file cabinet. At no point in the study were students placed in danger or exposed to any harm.

Design

This experiment had a pretest–intervention-posttest design. The intervention occurred in a classroom that students normally have instruction. The students were randomly assigned to either to the productive failure (PF) \(n = 10\) or lecture and practice (LP) \(n = 11\) conditions.

Measures

The Defining Issues Test-2 (DIT-2) online survey served to measure my first research question on the effectiveness of PF instruction on teaching moral reasoning skills for moral development. This instrument was used as an outcome measure and was chosen because it was created to measure moral and ethical reasoning and has been
shown to be sensitive to interventions designed to promote more advanced reasoning (Bebeau & Thoma, 2003). In 2008, 15 out of 54 U.S. dental schools reported using the DIT as an outcome measurement for assessing ethical competency (Lantz et al., 2011). The DIT-2 is a modified updated version of the validated DIT-1 instrument. Drawing upon Kohlberg’s (1984) model of cognitive-developmental theory of moral reasoning, the DIT-1 follows the basic assumptions of how individuals hold opinions of societal relationships, rules, roles systems, and formal institutions (Rest, Narvaez, Thoma & Bebeau, 1999). Rest et al. (1999) have since revised the DIT-1. The primary changes to the DIT-1 is the DIT-2’s shorter length, updated dilemmas and items, adoption of a new scale to calculate a developmental score (the N2 index) and a revised check method for bogus data. Like the DIT-1, the DIT-2 online is capable of discriminating between different levels of moral reasoning. Construct validity of DIT-2 was found to be robust against the DIT-1 with high correlation between DIT-1 and DIT-2 ($r = .79$) and with high internal consistency (alpha coefficients of .90) in the scenarios (Rest et al., 1999). In this study, the DIT-2 was administered online as a pretest/posttest. The pretest served to ensure equivalency between groups and compared scores posttest scores, measures change as a result of participating in productive failure.

DIT-2 measures students’ ability to identify, articulate, and reflect critically on ethical issues from their own perspective both personally and professionally (DIT-2 Spring, 2009, p.3) which is specific to the purpose of this study. The instrument has five ethical moral dilemmas followed by issue statements that are rated and ranked (Rest, 1979). The first item asks the individual which action to take: *Do something, Can’t decide,* or *Do nothing.* The next 12 items ask the individual to rank the issues as *Great,*
Much, Some, Little, or No. The last four items rank the importance of 4 of the 12 issues as Most important issue, Second most important, Third most important, and Fourth most important. This process continues with the other four dilemmas. Posttest scores were used to detect group differences in moral reasoning development at the end of instruction.

For the purpose of this study, I examined the N2 or New Index Score to measure changes in moral judgment as a result of an intervention. Like the name, N2 is a new index that has generally replaced the P or post conventional score on the DIT-1 to measure moral reasoning. The post conventional score is a developmental index that represents arguments that appeal to moral ideals for resolving complex moral issues which is different from the personal interest index that represents arguments that appeal to personal interests (Bebeau & Thoma, 2003). The N2 index combines how much post conventional moral reasoning is used and how little personal interest is used for the same decision. Specifically, the N2 score represents the degree to which respondents prioritize post conventional items which can be seen as a demonstration of more sophisticated thinking (Bebeau & Thoma, 2003). Higher N2 scores reflect an individual’s increased capacity for reasoning about moral issues based on a system of fairness that serves the public good; lower N2 scores tend to reflect reasoning about moral issues from a self-serving understanding of fairness (Mayhew, 2014). In this study, the targeted concept of resource allocation is based on a system of fairness of how to effectively and efficiently distribute resources that maximize health benefits for the public good at the lowest cost. This supports using the N2 scores to measure changes in students’ moral reasoning ability because of the complex nature of resource allocation. The DIT-2 N2 score (Cronbach’s α = .81) has a stronger internal reliability than DIT-2 P (Cronbach’s α = .74)
as it relates to observed effects of educational interventions (Bebeau & Thoma, 2003). Furthermore, in a meta-analysis with a sample of 503 students comparing Post conventional with N2 scores, average effect size for the N2 index on the sensitivity to educational intervention was reported to be $d = .31$, a moderate size effect (Rest, Thoma, Narvaez, & Bebeau, 1997). Additionally, moral development can be observed as N2 scores increase by educational levels with average scores in the upper 20’s for students in vocational and technological programs, in the lower to mid-30’s for junior college students and college juniors, and in the mid to upper 30’s for college seniors (DIT2 Guide, 2003). This was an important feature of the DIT-2 since I wanted to assess pre and posttest scores to measure if students’ moral schemas change using PF instruction method.

To measure my second research question on problem solving skills and the transfer of knowledge from one problem to another novel problem, I used Bebeau’s (1995) criterion checklist that was created to identify relevant issues and possible arguments of the moral problem (Appendix A). The checklist had four criterion categories to determine an individual’s ethical reasoning process. Ethical reasoning was evaluated by how the individual: 1. Described the ethical issues; 2. Identified affected parties; 3. Described the consequences of acting; and 4. Described the relevant obligations or duties. Each category was rated on how students articulated and developed their responses on an ethical dilemma. Bebeau’s (1995) checklist was used to evaluate students’ written responses to the problem presented and discussed.

Although Bebeau (1995) was aware that there may be other important issues that are not identified and listed in the checklist, she suggested that modifications may be
necessary to attend to as many possible issues and arguments within the problem that may be relevant within the discipline. On the basis of this assumption, two dental hygiene instructors discussed and generated as many plausible arguments and justifiable possibilities, which were then used to create parameters and rating scales for each criterion being measured.

The same two dental hygiene instructors independently assessed and scored 10% of the same pretest and posttest student responses using Bebeau’s checklist. These scores were discussed and negotiated until agreement was made on the measurement. After finalizing the parameters for each criterion, the raters independently scored the remaining student responses. Interrater reliability was assessed using intraclass correlations (ICC), a two-way mixed, absolute agreement measure (Hallgren, 2012). The ICC assessed the degree to which two independent raters provided agreement in their ratings of student’s ability to identify relevant issues and possible arguments of the moral problem across other students. The results showed high agreement between instructors suggesting scores were rated similarly across raters (ICC = 0.89, n = 42, p = .01).

In summary, I collected two different sources of data during the productive failure task in order to determine the effectiveness of PF instruction compared to LP instruction and transfer of knowledge in solving a novel problem with multiple and conflicting solutions. I evaluated the students’ prior knowledge and moral development using the DIT-2 online assessment. I collected data in written form through the students’ responses to assess students’ knowledge transfer.

The data collection and measurement for this study hinges upon group and individual essays as they worked on each dilemma. These sources include pretests of
moral development and prior knowledge, group assessment, and individual assessment.

**Procedure**

Participants were randomly assigned to either Productive Failure (PF) instruction or Lecture/Practice (LP) instruction one week before the intervention. This was done by using a random assignment computer program:

[http://www.graphpad.com/quickcalcs/randomize1.cfm](http://www.graphpad.com/quickcalcs/randomize1.cfm). The LP group served as the control group and the PF group as the experimental group.

The ethics course is taught in the fall semester in the first 8 weeks of a 16-week semester for two hours. This study took place after the course ended and not during the normally scheduled time period.

As a full-time dental hygiene instructor and having taught this course in the past, I am familiar with its content. Therefore, I decided to be the instructor for this study. Both groups participated in the same number of lessons for a total of two, 50-minute classes in one week. The amount of instructional time was held constant for both conditions. A few days before the unit, students took the Defining Issues Test-2 (DIT-2) online which served as a pretest to measure the student’s level of moral development. Both conditions received the same moral concept to solve however they were presented in different problem designs. The PF problem was designed as ill-structured (see Appendix C), whereas the LP problem was designed as well-structured (see Appendix D). The problem was designed to activate prior knowledge of core values like beneficence, veracity, and maleficence as well as introduce a novel concept of resource allocation. The dental hygiene students in this study are familiar with the core values, but they have not had any formal instruction on the targeted concept of resource allocation.
The other difference between conditions was in the sequencing of the phases. The PF condition consisted of a problem-solving phase first followed by a lecture or consolidation phase. In contrast, the LP condition consisted of a lecture phase first followed by a problem-solving phase. Three days later, students from both conditions individually completed a well-structured moral dilemma (See Appendix E). After the end of the unit, students took the DIT-2 on their own time which served as the post-test. The survey took between 40-45 minutes to complete each time. Table 3.2 describes the study design.

Table 3.2

| Productive Failure Design Compared to Lecture and Practice Instructional Design |
|---------------------------------|---------------------------------|
| **Pretest**                     | **Lecture and Practice**        |
| DIT-2                           | DIT-2                           |
| **Phase 1**                     | **Phase 1**                     |
| Ill-structured moral dilemma with affective support (Dyads and Triads Groups) | Lecture (Dyads and Triads Groups) |
| **Phase 2**                     | **Phase 2**                     |
| Consolidation lecture           | Practice: Well-structured moral dilemma with scaffolding support |
| **Moral Dilemma**               | **Moral Dilemma**               |
| Individual well-structured moral dilemma with support | Individual well-structured moral dilemma with support |
| **Posttest**                    | **Posttest**                    |
| DIT-2                           | DIT-2                           |

**Productive Failure Design**

This study followed Kapur’s (2009) design by implementing the two phases in productive failure: 1. the Invention phase and 2. Consolidation phase. In the PF
condition, the students worked in groups of dyads (2) or triads (2) on their own for 35 minutes to struggle with the concepts in the problem. In the Invention phase, productive failure (PF) students attempted to solve one moral dilemma based on the concept of resource allocation, a concept students had not been taught in their regular course. Specifically, students were asked to make decisions on how resources are allocated where price is a common means of allocating resources based on patient needs and in this problem a dental hygiene student’s right to learn in terms of what is considered equitable and just. The dilemma here is the conflicting objectives of health care systems and training programs. The objective of health care systems is to provide an adequate standard of care and the objective of training programs is to teach a skill that a person may require. Educational programs like dental hygiene are seen as low-cost health care systems and yet are training facilities that teach students how to apply specific techniques or services. Allocation decisions can test moral boundaries on how to effectively and efficiently distribute resources such as low cost dental care to those in need, but who cannot afford dental care and teaching skills to students who need practice, but are required to maintain the standard of care.

This dilemma was designed as an ill-structured problem. Table 3.3 describes the parameters used to create either a well-structured problem or an ill-structured problem. Further, a dental hygiene expert reviewed the problems and used the design elements as a check list. Students were given blank sheets of paper for all their group work. The blank sheets served as documentation and evidence of students’ interaction patterns of moral development by their explanation, critique, and elaboration in their engagement with the ill-structured problem (Kapur, 2008, 2009, Kapur & Bielaczyc, 2012). Following
Kapur’s (2008, 2009, 2011) assertion that offering instructor scaffolds and support can impede knowledge transfer, the instructor refrained from making suggestions or leading the students toward a correct method. Instead, when the groups appeared frustrated or asked questions, the instructor used affective phrases like “Keep working hard on the problem”; “Continue your effort in solving the problem; “Keep stretching yourself”; and “Let’s see what you can come up with on your own” to encourage them to persist and continue to find a solution. This is similar to that used in Kapur’s (2011) design to encourage students to persist in solving the problem.

Table 3.3

<table>
<thead>
<tr>
<th>Ethical Dilemma Problem Design</th>
<th>Ill-structured Problem</th>
<th>Well-structured Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Possess many problem parameters with some unknown, or known only with some degree of confidence</td>
<td>Possess fewer problem parameters than their ill-structured counterparts</td>
<td></td>
</tr>
<tr>
<td>2. Possess problem parameters that interact with each other in interesting ways such that the effect of each could not be examined in isolation</td>
<td>Present problem parameters to the learners with greater degree of confidence,</td>
<td></td>
</tr>
<tr>
<td>3. Possess multiple solutions and solution path</td>
<td>Require the application of a limited number of regular rules and principles that were organized in predictive ways</td>
<td></td>
</tr>
<tr>
<td>4. Possess multiple criteria for evaluating solutions</td>
<td>Have knowable, comprehensible solutions where the relationship between decision choices and all problem states are probabilistically close to being known</td>
<td></td>
</tr>
<tr>
<td>5. Require students to make assumptions, judgments, and express personal opinions or beliefs</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Consolidation phase was led by the instructor for the last 15 minutes of class. According to Kapur (2009), the purpose of the consolidation phase is to assist students on how to organize and assemble key ideas and concepts. For example, the instructor and the PF students discussed and shared their solutions and possibilities to the problem. Thus, the Consolidation phase helps students expand their capacity to learn on what works and what does not work (Kapur, 2008).

Three days later, PF students completed a second moral dilemma. PF students individually worked to solve the problem spending the entire 50 minutes of class time. The second problem intentionally included the concept of resource allocation, but this time through the lens of analyzing cost effectiveness. Specifically, cost effectiveness analyses of a dental preventive program were in conflict with issues of equity and justice between two groups of children needing dental care. Thus, the second problem introduced another novel concept—cost effectiveness. Further, the structure of the second problem was presented in a well-structured manner which possesses different criteria compared to ill-structured problem design. This is further discussed in the Lecture and Practice condition.

**Lecture and Practice Design**

In the Lecture and Practice (LP) condition, the order of phases was reversed in that the LP students were given first a 20-minute lecture on the same novel concept of resource allocation followed by a 15-20 minute practice session on a structured conceptual problem. The last 10 to 15 minutes of class was used to go over the problem solving method. The same instructor presented the lecture and provided scaffolding and assistance as students practiced on their ethical problem. Only this time, the ethical
dilemma was designed as a well-structured problem making the novel concept more obvious to the students. According to Kapur (2008), this type of problem is commonly used in teaching and yet, may not challenge students enough to question why the solution works, thus limiting their decision-making skills and knowledge assembly of key ideas and concepts for future problems. The same dental hygiene expert confirmed content validity of the well-structured problem. Blank sheets were also provided for the LP groups to document their representations and methods for solving the problem. However, instead of leaving students alone to struggle, the instructor prompted them with simple questions and hints to guide attention to features in the problem that were important. Three days later, LP students worked on a second ethical dilemma identical to the well-structured problem given to the PF student to complete.

**Results**

**Effectiveness of Productive Failure**

The DIT-2 pre/posttest was scored by the Center for the Study of Ethical Development at the University of Alabama. Results revealed that 21 dental hygiene students participated in the pretest or first administration of the DIT-2 survey. There were no significant differences between the two groups on the pre-test ($p = .081$). However, for the second administration or posttest of the DIT-2, only ten of the 21 students completed the survey. Moreover, the Center for the Study of Ethical Development purged one of the ten returned scores because the responses failed to meet the Center’s reliability criteria (Rest et al., 1999). Thus, of the 21 students, nine participated in both the pretest and posttest to provide for a matched comparison in the analysis of effectiveness of productive failure.
The severe attrition in this study reduced the power of the statistical analyses considerably. Below, I report the results of all of the analyses, but statistical significance of the results is attenuated because of this loss of power. Since this pilot study was exploratory, I focused on the effect size statistics to reveal more information than the inferential statistics.

I used analysis of covariance (ANCOVA) to assess the effect of two teaching conditions by comparing N2 posttest mean scores, using the N2 pretest score as the covariate. The main purpose of ANCOVA is to reduce error variance (Dimitorv & Rumrill, 2002) that improves the power of analysis because it removes the additional noise to reveal the main effect of treatment (Dugard & Todman, 1995). ANCOVA increases the sensitivity of the F test by removing predictable error. Significance was set at <0.05 for the F statistic in the ANCOVA model.

Before I used ANCOVA, Levene’s test was conducted to determine if the statistical analysis was appropriate. Levene’s test for homogeneity of variance was not significant ($p = .82$) and therefore, verified that variances were equal across the groups. An ANCOVA revealed no statistical differences across teaching conditions, PF and LP on posttest scores ($F(1,7) = 5.16, p = .06, d = .76$). Table 3.4 summarizes the results of the pre and posttest N2 raw scores. As Table 3.4 indicates after controlling for the pretest, PF students had higher N2 scores ($M = 31.83, SD = 9.70$) compared to LP students ($M = 25.85, SD = 5.41$). The use of PF seemed to be effective with a large effect size ($d = .76$) reported. With the standard deviation of 7.85, this means that on average, students in PF had N2 posttest scores that were .76 standard deviation higher than those in
LP. This is considered a substantial difference. The N2 posttest scores for PF students increased from the 20’s to 30’s, which means that students approached the ethical stories at least one educational level higher or in this case, from a vocational/technological level to a junior college or college junior level compared to LP students. The magnitude of the effect sizes in the pilot study suggests that the data are useful and promising in the design of the full-scale study.

**Transfer of Knowledge**

To examine transfer of students’ knowledge on resource allocation, their written responses were scored using Bebeau’s (1995) criterion checklist (Appendix B). Similarly, I used ANCOVA to analyze the data. Posttest scores were designated as the dependent outcome variable, and pretest scores were used as the covariate. Significance was set at <0.05 for the F statistic in the ANCOVA model.

Levene’s test for homogeneity of variance was not significant (p = .93) indicating that variances were equal across the groups. The ANCOVA revealed no statistical differences across teaching conditions, PF and LP, on posttest scores in developing a well-reasoned response after controlling for pretest scores ($F(1, 18) = .33, p = .58$).

### Table 3.4

Comparison of Pretest and Posttest of DIT2 N2 Scores

<table>
<thead>
<tr>
<th></th>
<th>N2 Pretest Score Mean (SD)</th>
<th>N2 Posttest Score Mean (SD)</th>
<th>p</th>
<th>Effect size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecture and Practice</td>
<td>26.06 (12.19)</td>
<td>25.85 (5.41)</td>
<td>.06</td>
<td>.76</td>
</tr>
<tr>
<td>(n = 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Productive Failure</td>
<td>26.45 (10.17)</td>
<td>31.83 (9.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n = 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
However, the effect size was reported as $d = .34$, a moderate effect. With the standard deviation of 2.34, PF students increased their posttest scores by .34 standard deviation compared to LP students, an improvement that is less than one point on the posttest.

I decided to apply a statistical analysis on each criterion to further investigate and examine how each criterion score contributed to the transfer of knowledge score. As presented above, Bebeau’s (1995) checklist evaluated four criteria: 1) issues and points of ethical conflict as they were addressed in the problem, 2) interested party’s expectations which represents skills in perspective-taking, 3) consequences of acting which may be multifaceted in nature because there are positive and negative outcomes on the proposed action, and last 4) duties or obligations of the protagonist which are described and supported by moral considerations. I conducted an independent-samples $t$-test to compare posttest scores on each criterion in productive failure and lecture and practice conditions. Results from the statistical analyses revealed no significant differences in student scores for the Issues criteria in PF ($M = 2.73$, $SD = 1.00$) and LP ($M = 2.20$, $SD = 0.42$) conditions, $p = .14$; Identifying parties criteria in PF ($M = 2.00$, $SD = 0.45$) and LP ($M = 1.80$, $SD = .63$), $p = .41$; Consequences criteria in PF ($M = 2.82$, $SD = 0.75$) and LP ($M = 2.40$, $SD = .966$), $p = .28$; and Obligations and Duties criteria in PF ($M = 3.27$, $SD = 1.10$) and LP ($M = 2.0$, $SD = .48$), $p = .95$. On the other hand, the magnitude of the difference between groups for Issues ($d = .61$) and Consequences of Acting between the two conditions ($d = .45$) had a medium effect.

**Discussion**

This study examined the effectiveness of productive failure, an instructional design, as assessed by two important outcomes in dental hygiene professional education:
moral reasoning regarding ethical dilemmas in professional practice and student learning, specifically in the transfer of knowledge. I followed the productive failure (PF) design approach as described in Kapur and Bielaczyc’s (2011) study and implemented two learning phases which included an initial collaborative problem solving phase followed by a delay in instruction. The lecture and practice (LP) condition was the control group that represented a traditional teaching condition of lecture first followed by practice. In contrast to previous PF studies, the type of problem was an open-ended problem, an ethical or moral dilemma, with many possible pathways and judgments. I compared PF to LP over two class periods in one week, a short time period. Therefore, caution must be used in its interpretation and generalizability. Instead, the findings in this pilot study served a developmental function to increase the chances of success in a larger full-scale study (Leon, Davis, & Kraemer, 2011). As such, findings in this study conflicted with those found in the literature. Although I hypothesized a differential effect, the results were not consistent with previous research on the effectiveness of productive failure (Kapur, 2008, 2009, 2010; Hung, Chen, & Lim, 2009; Kapur & Kinzer, 2009; Kennedy-Clark, 2009; Pathak et al., 2011; Roll et al., 2012; Trueman, 2013; Wiedmann et al., 2012; Westermann & Rummel, 2012) and transfer of knowledge (Kapur, 2008, 2009, 2010; Kennedy-Clark, 2009; Trueman, 2013).

Despite not finding statistical differences, the magnitude of the intervention between students’ N2 scores in the PF condition and LP condition \((d = .76)\) suggests the acquisition of new thinking and the rejection of simplistic thinking (Bebeau & Thoma, 2003) thus approaching complex problems in a more sophisticated moral way of thinking as a result of the educational intervention. Similarly, PF showed moderate effects \((d = \)
.26) on students’ response on the well-structured posttest problem in that PF students provided slightly better arguments using the targeted concept of resource allocation suggesting PF design of problem solving activity enhanced the students’ ability to understand more complex principles of resource allocation. This was certainly seen in posttest scores for addressing issues and points of ethical conflict and describing and incorporating consequences of acting, but not in describing obligations and duties of the protagonist and identifying the interested parties. In other words, students attended to the dilemma with more thought and ideas by describing the moral conflict which often have several possible outcomes. As this was a pilot study, the sample size and study design logistics may be insufficient to reach conclusive effects and caution must be used in interpreting the results of this study. However, the pilot results have identified modifications needed in the pilot design that will help strengthen the feasibility for the full-scale study.

One identified modification was the length of the study. This study encountered curricular time constraints in which the intervention was tested within a short time of a single week, significantly less than previous PF studies (Kapur, 2008, 2009, 2010; Hung, Chen, & Lim, 2009; Kapur & Kinzer, 2009; Kennedy-Clark, 2009; Pathak et al., 2011; Roll et al., 2012; Trueman, 2013; Wiedmann et al., 2012; Westermann & Rummel, 2012). The average length of time for previous experiments ranged between two to sixteen weeks to capture changes in student learning. Originally this study was designed to take place over 3-weeks, but program and student scheduling constraints and course instructor commitment to the length of time precluded me from keeping to the original experiment design. DIT-2 scores have shown to be valid in reporting changes as result of educational
interventions with a typical medium effect size of .40, but in the context of programs that ran three weeks or more. The concern here was whether DIT-2’s sensitivity to educational intervention effects would hold true in this study. Again, there was no statistical difference \( p = .06 \) between the two instructions. The findings in this study were similar to previous PF research (Kapur & Bielaczyc, 2011;) in that PF students \( (n = 5) \) were able to construct deeper conceptual understanding, concepts of moral reasoning and resource allocation, from the PF design to better prepared them to approach another complex problem such as the DIT-2 ethical dilemmas. Notably, differences in posttest scores compared to pretest scores was seen in four N2 scores of PF students resulting in gain scores of 16.64, 6.69, 5.27, and 8.82 with the exception of one individual whose score decreased from pretest by -10.55. Of interest, all LP students \( (n = 4) \) decreased their post N2 scores by -10.75, -2.92, -0.06, and -20.36. However, the sample size and study design logistics may be insufficient to reach conclusive effects of this study.

A second identified modification was the design of the generation task or moral dilemma. Contrary to previous PF studies, in this study the control group did not experience performance success in constructing a well-developed response to the initial moral dilemma. In fact, results show that LP students experienced failure similar to PF students and even produced lower response scores \( (M = 8.1, SD = 1.99) \) compared to LP students \( (M = 9.4, SD = 1.04) \). In other words, both groups in the initial learning phase were unable to fully construct a well-developed reasoning response to the dilemma. In Kapur and Bielaczyc’s (2011) study on designing for productive failure, the generation task or problem should be constructed to include prior knowledge contexts that can activate students’ thinking process about the novel concept even if they have not formally
learned it. The challenge is to find a novel concept that students do not know, but at the same time possess important skills and knowledge from the same domain that could be applied to generation task (Kapur & Rummel, 2012). In this study, students had just completed the semester’s ethics course before entering the study and in the attempt of finding a novel concept meant that I, who was also the instructor, had to research other domains that might not have been familiar. Thus, the design of the well-structured dilemma may not have contained enough problem-solving context for students to successfully process, develop, and articulate a well-supported response even though an expert in dental ethics reviewed the moral dilemmas prior to the time of the study. For the full-scale study, I will follow Kapur’s (2008) problem design more closely and add solutions that are knowable and comprehensible to the students for them to identify the relationship between decision choices and the problem. Again, problem construction is a vital component and concern for the success of PF design (Kapur & Bielaczyc, 2011; Kapur & Rummel, 2012; Loehr et al., 2014, Wiedmann et al., 2012).

This leads to the third modification for the full study. The findings suggested that in descriptive terms, PF and LP students were still unable to organize and assemble key ideas and concepts as seen in their posttest scores, $M = 10.7$, $SD = 2.53$ and $M = 9.9$, $SD = 2.13$, $p = .58$, $d = .26$, respectively. The student responses were evaluated using a 17-point checklist of which total points for PF posttest scores ranged from 7 to 16 points and 7 to 13 points for LP students. The low posttest scores might be due to the fact that the well-structured problem was not designed appropriately for PF students to activate prior knowledge and make the connections to fully conceptualize the solutions as explained in the lecture. Additionally, I may not have introduced the concept well enough to the class.
and thus, compromised the purpose of the consolidation phase in PF condition and lecture phase in LP condition. Specifically, in my attempt to follow instructor facilitation, I might have been unsuccessful in drawing out comparisons and contrasts between resource allocation and student-generated explanations in both groups. It may be that open-ended problems and solutions are more complex to teach and may require more time for students to apply and articulate it in a written assessment. Conceptual knowledge in moral reasoning requires students to develop arguments that appeal to moral ideals for resolving complex moral issues (Bebeua, 2011) and as such may require more time to practice and use knowledge from the instruction as seen in previous studies by Kapur (2008, 2009, 2010) and others (Loehr et al., 2014; Trueman, 2013). Modifications for the subsequent full-scale study will be to research the resource allocation more closely and develop a training session and standard PowerPoint presentation highlighting the main ideas for other instructors to use. This will ensure that students in both groups will receive the same content in the lecture phase of the study design.

**Conclusion**

Productive Failure as an instructional design is based on the premise that learning results from the experience of struggle and failure in solving complex problems with little to no support from teachers which can then lead to students solving subsequent complex problems (Kapur 2008, 2009, Kapur & Bielacyzc, 2012; Kennedy-Clark, 2009; Roll, Holmes, Day & Bonn, 2012). Productive failure offers an alternative method of teaching complex problems by providing the space and opportunities to generate, create, and invent solutions or multiple approaches for problem solving. Consequently, in this pilot
study, statistical differences on conditions between traditional lecture and practice and productive failure could not be established as a result from DIT-2 N2 outcomes and of posttest of written responses. However, this pilot study will help design my dissertation study. Future research with a larger sample size and a higher number of students completing the DIT-2 pre and posttest will help make comparisons more meaningful. Additionally, increasing the duration between administrations and adding more instructional context may also need to be considered to help look for patterns in moral development. Benefits of productive failure were descriptively marginal in this study, but the potential seem to be evident even in the context of a professional program and in an ethics course.
CHAPTER FOUR

PRODUCTIVE FAILURE AS AN EFFECTIVE TEACHING METHOD

IN

DENTAL ETHICS

Karen Tam

University of Arizona

College of Education
Abstract

A quasi-experimental study compared productive failure (PF) instructional design with a traditional lecture and practice (LP) instructional design for three weeks on a moral reasoning unit. Seventy-seven second-year dental hygiene students from four dental hygiene programs participated in a pretest-posttest study design using the Defining Issues Test –N2 Index and moral dilemmas to evaluate the effectiveness of productive failure. Separate multiple regression analyses were conducted. Results found N2 scores of PF students were not significant (PF School A, $p = .79$; PF School B, $p = .81$, and PF School C, $p = .27$) even though PF students demonstrated shifts to Post conventional thinking compared to LP students. In the moral responses, PF students performed similarly when compared to LP School posttest scores. However, PF School C and LP School received scores lower than their pretest problem and the difference between pretest and posttest scores in LP School had a moderate effect ($d = -.64$). Lecture and practice and productive failure have shown benefits in student learning depending on the learning goal at hand. However, this study’s results suggest that productive failure can be effective for deeper conceptual understanding and the transfer of knowledge to novel problems.

*Keywords:* dental hygiene ethics instruction, solving moral dilemmas, productive failure, transfer of knowledge
Productive Failure as an Effective Teaching Method in Dental Ethics

It was fall semester, I started my Dental Prevention class with what I thought was going to be a productive discovery learning activity for my first-year dental hygiene students in which I handed out eight plastic baggies with seven laminated 2” x 2” pictures of teeth in various stages of dental caries. I asked each group to place the pictures in chronological order from stage: no dental caries to stage: HGH or “Huge Gaping Hole” dental caries. With minimal guidance, but supportive comments, I encouraged each group to persist in the activity each time they failed. Once the group successfully placed the pictures in order I asked them to describe to me the enamel and dentinal process of decay for each picture, then explain what preventative measures could be used at each stage and why it was recommended. After the activity, we quickly came together as a class and discussed what was missed, learned, and why. At the end, I thought to myself, “That was fun. Lots of activity and discussion! What a wonderful learning experience for my students!” Feeling pretty happy with myself, I heard a question amidst all the excited talking and discussion, “So, when are you going to lecture to us, Ms. Tam?”

The student’s comment took me somewhat by surprise, but not completely. In fact, the question reflects the change in how teachers teach and the instruction methods they use. Moving away from a lecture format, teachers are focusing on methods that encourage student participation, inquiry, and problem solving. Even curriculum at most dental and dental hygiene schools are moving away from lectures largely in part of a call for reform of existing teaching practices and the adoption of an experiential learning approaches by the ADEA Commission on Change and Innovation in Dental Education in 2006. However, lectures are still used in many subject areas in dental and dental hygiene
education. Ethics is one of those areas. Several studies found that ethics are generally taught using traditional methods of lectures, discussion, quizzes, and written assignments (Duley, 2009; Lanz et al., 2011). These forms of instruction emphasize the memorization of facts with minimal challenge and are inadequate in fostering critical thinking and problem solving skills that are necessary in developing moral and ethical reasoning (Duley, 2009; Lantz et al., 2011; Taichman et al., 2012). The pedagogy for lecture is one that:

- consists of describing an objective, articulating motivating reasons for achieving the objective and connections to previous topics; presenting requisite concepts (if they have not been presented previously; demonstrating how to complete the target problem type; and providing scaffolded phases of guided and independent practice, accompanied by corrective feedback. (Munter, Stein, & Smilty, 2015, p. 7)

Whereas the learning activity, a version of productive failure, represents pedagogy where:

- Students have opportunities to (a) wrestle with big ideas, without teachers interfering prematurely, (b) put forth claims and justify them as well as listening to and critiquing claims of others, and (c) engage in carefully designed, deliberate practice. This requires teachers, first to engage students in two main types of tasks-tasks that introduce students to new ideas and deepen their understanding of concepts, and tasks that help them become more competent with what they already know. (Munter, Stein, & Smilty, 2015, p. 8)
The above definitions of contrasting pedagogies came from a series of discussions among nationally recognized experts who hold opposing points of view on designing instructional environments. While the topic of discussion was mathematics, the debate between the two instructional models can be carried over to other fields (Munter, Stein & Smith, 2015). Understanding the ways the two instructional models function is important and creating an environment in which students engage in discussion and other dialogic activities may not be sufficient evidence that students are learning. In fact, Mayer (2004) argued that, “methods that rely on doing or discussing should be judged not on how much doing or discussing is involved but rather on the degree to which they promote appropriate cognitive processing” (p. 17).

In this study, I investigate and compare two teaching methods, productive failure and lecture and practice. In particular, I explore how the two learning approaches differ from one another and look for evidence of learning in either environment, specifically in teaching ethics in dental hygiene education.

**Dental Hygiene Curriculum Change**

In 2005, the Board of Directors of the American Dental Education Association (ADEA) created the ADEA Commission on Change and Innovation in Dental Education (ADEA CCI) to address challenges of curriculum reform. One area of reform was in teaching environments. The articles posit that faculty members “act as facilitators with active learners”. Active learners are students who no longer memorize information and rely on strategies that are used to “survive” in the curriculum. Instead, students “must have time to reflect and think about their learning” (ADEA CCI, 2005). Since then there have been several studies on instructional methods in dental and dental hygiene
education. However, the empirical evidence to support the use of alternative instruction is weak. Much of the earlier research employs a one group pre and posttest survey design that examined student and/or faculty perceptions after implementing an instructional strategy, thus making it difficult to isolate the effect of the intervention from potentially confounding variables (Johnson & Christensen, 2012). However, in the last few years stronger experimental research designs have been used that include a control group and experimental group.

However, there are mixed results in the effectiveness of alternative teaching methods compared to a lecture format. For example, team based learning (Echeto, Sposetti, Childs, Aguilar, Behar-Horenstein, Rueda, & Nimmo, 2015) and e-learning methods (Ariana, Amin, Pakneshan, Dolan-Evans, & Lam, 2016) were found to be more effective than lecture. While other experimental studies on alternative instruction such as reflective blogs (Wetmore, Boyd, Bowen, Pattillo, 2010), questioning during lectures (Hessheimer, Rogo, Howlett, 2011); online voice-over screen-capture presentations (Schönwetter, Gareau-Wilson, Cunha, Mello, 2016) and small group discussions (Arias, Scott, Peters, McClain, Gluskin, 2016) were found to have no direct influence on student learning. Certainly, many of these studies have yet to be replicated in varied settings and across student cohorts. With this in mind, research needs to continue to examine the effects between alternative instruction and lecture and practice. To date, no studies have examined the effects of productive failure compared to lecture in promoting moral development in dental hygiene students.
Productive Failure

In order to understand the effects that might be found in comparing the two instruction methods, I first describe productive failure (PF) instruction and then compare PF instruction to lecture and practice (LP) instruction. One reason for PF instruction is the notion that there is a hidden efficacy in that students potentially learn through failure and experimentation from their exploration and struggle in solving complex problems (Schwartz & Martin, 2004). Learning is maximized because conceptual understanding goes beyond the superficial, but is deepened by discovering the gaps and inconsistencies in one’s understanding in service of acquiring new knowledge. Kapur (2008, 2009) found that as long as students were able to persist in the problem-solving processes they would find ways of developing representations while producing diverse solution methods.

Productive Failure consists of two phases: 1) the invention phase with students working collaboratively in groups on an ill-structured moral dilemma followed by 2) a subsequent instruction known as the consolidation phase that occurs in the following week. Students are grouped to work together collaboratively. However, the group receives no help or support from the instructor other than encouragement to continue with the problem, thus in the end students fail to find a solution. The consolidation phase is the space with which the instructor facilitates the process of unifying student generated solutions by helping them organize and put together key elements to form the correct solution (Kapur & Bielacyzc, 2011).
Productive Failure Compared to Lecture and Practice

Productive failure (PF) diverges from lecture and practice (LP) in three ways. First, the LP sequence involves *first direct instruction* followed by a problem for practice, in contrast to PF in which sequence involves a *problem first* followed by instruction. For example, in LP, teachers first explain procedures and concepts and then students practice on problem(s). The advantage of LP is the production of more immediate recall of facts with transfer and problem-solving skills while using less cognitive load and working memory on the practice (McMullen & Madelaine, 2014). Using strong guidance, teacher–led activities are used to monitor student progress and mastery which was found to be effective (Kirschner, Sweller & Clark, 2006; McMullen & Madelaine, 2014). However, a growing body of evidence has demonstrated that sequencing the *problem first* prior to instruction can be as effective as, if not moreso than LP (Kapur 2010,2012; Kapur & Bielaczyc, 2011; Loehr et al.,2014; Loibl & Rummel, 2013; 2015; Schwartz, Chase, Oppezzon, & Chin, 2011). Presenting the *problem first* prior to instruction can prepare students for future learning (Schwartz & Bransford, 1998; Schwartz & Martin, 2004). Delaying instruction activates prior knowledge bringing to light important problem features that might not otherwise be noticed if instruction was given first. Initially, students may not know what gaps and inconsistencies exist in their conceptual knowledge and thus, self-assessment of their understanding can be overestimated (Fisher & Keil, 2014; Loibl & Rummel, 2015; Rozenbilt & Keil, 2002). The problem solving process depends on students’ understanding and representation of the problem which can lead them to detect knowledge gaps (Jonassen, 1997). As such, one of the benefits of discovery learning is the process of generating multiple solutions. Generating, creating,
and negotiating in formulating solutions can lead to understanding distinct features of the 
domain that helps students generalize when they are presented with a novel problem.
This process is necessary in the transfer of knowledge (Bransford, Franks, Sherwood, & 

Second, LP is structured, teacher–led instruction using explicit language with 
deliberate implementation and thus, well-structured problem types are used with guided 
practice from the teacher rather than independent practice (McMullen & Madelaine, 
2014). Because LP is goal-oriented and focuses on student achievement, creativity is not 
an important feature in the design of the instruction, which is a complaint made by many 
teachers of the structured rigidity of LP. McMullen and Madelaine (2017) describe this 
point best, “It has to be asked whether it is ultimately in the students’ best interests 
whether the teacher’s need for creativity in lessons outweighs the potential benefits (and 
costs) of choosing or rejecting an instruction methodology on this basis” (p.145). Well-
structured problems fit within the goal of LP because the problem is situated in well-
defined parameters, with a known goal, and the solution is a single, known solution 
(Jonassen, 1997). The argument here is that this type of problem encourages students to 
be passive learners with little engagement in the problem. LP proponents argue that well-
structured problems actually reduce students from “practicing errors or developing 
inaccurate, compensatory strategies that learners otherwise may rely on” (McMullen & 
Madelaine, 2014). If the goal of instruction is student achievement, then this rationale 
fits with LP structure.

However, if the goal of instruction is to construct knowledge, requiring an 
active role of the student, then the problem type may look more like ill-structured
problems, a type of problems used in productive failure. Ill-structured problems have one or more unknown problem element, goals that are vague and unclear, multiple solution paths, and learners expressing their personal opinions or beliefs about the problem (Jonassen, 2017). This type of problem has been known to be not only more authentic, but learners can identify alternative views and gather evidence to support or reject probable causes (Jonassen, 1997; Reed, 2016). The problem space created by presenting an ill-structured problem first allows students to generate as many plausible solutions or solution paths to solve the problem. In this study, moral dilemmas are ill-structured in nature and share characteristics of ill-structured problems. As such, feedback is essential to reduce errors that novice learners experience (McMillian & Madelaine, 2017), but when should feedback be given?

The third way PF diverges from LP is when feedback or scaffolding is given and if it is essential for students to learn. Recent literature has highlighted various approaches in developing problem solving skills. Some researchers and educators feel that students should be guided through the process of problem solving, scaffolded by teachers (Hmelo-Silver, Duncan, & Chinn, 2007; Kirschner, Sweller, & Clark, 2006) or instructed to use structured prompts (McMullin & Madelaine, 2017). For example, mutual interchange and generated dialogue between students and teacher is collaborative learning that helps students explore new possibilities of interacting with the new information (Palincsar, 1998). Notably, most of the strategies and interventions have been used in conditions where performance is maximized in the short term and learning is maximized in the long
Problem-based learning is one example in this category and requires scaffolding from facilitators.

In contrast, other researchers believe that students should approach problems, even novel problems, with minimal to no guidance from teachers or expert (Kapur, 2016, 2013; Loehr et al., 2014) even when students reach an impasse (Kapur, 2016, 2013; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). As such, support structures like scaffolding can impede students from fully learning and understanding the concept being studied (Hung et al., 2009; Kapur, 2008; Schwartz and Martin, 2004; VanLehn et al., 2003). In a study on tutorial behavior that promotes student learning, VanLehn (2011) found that one tutorial behavior that promotes student thinking is to generate opportunities for impasses by giving no content prompts. Creating a feedback loop (other students in the group) keeps students thinking and talking to produce an explanation that satisfies the entire group. The other two effective tutoring strategies is to prompt students to the find the correct step and explain it and provide an explanation if they have tried and failed (VanLehn, 2011). In PF, the last two strategies are combined in the consolidation phase. In the PF condition, performance may not be maximized in the shorter term, because students fail to reach an acceptable conclusion, but learning is maximized in the longer term (Kapur, 2016). Prior literature has examined productive failure condition with problems that have one solution or canonical solutions in science, math, physics, and STEM. However, no studies have examined productive failure in the area of ethics that have not one, but multiple considerations. Using moral dilemmas to examine productive failure offers one way to theorize how students approach and prepare
themselves to solve open-ended problems with multiple solutions. This was the aim in the research described here.

The Current Study

This study aims to better understand productive failure condition in terms of student performance and learning when given a moral dilemma. Specifically, the research questions that guided this study were:

1. What effects does productive failure have on student learning in solving moral dilemmas compared to lecture and practice?
2. To what extent does productive failure help students learn skills in transferable problem solving?

Methods

A quasi-experimental study compared productive failure (PF) instructional design with a traditional lecture and practice (LP) instructional design for three weeks on a moral reasoning unit. Four dental hygiene programs participated in the study. The purpose of using multiple comparison groups was to further explore other possible salient threats to validity to the causal inference and to “triangulate toward a narrower bracket within which the effect is inferred to lie” (Shadish, Cook, & Campbell, 2002, p. 159).

Participants

Second-year associate degree-seeking dental hygiene students participated in this study. These students completed one year of dental hygiene curriculum and passed each course with a C grade or better at four selected community colleges in the southwestern United States. Moreover, students learned early on about the American Dental Hygiene Association Code for Dental Hygienists which include individual autonomy and respect
for human beings, confidentiality, societal trust, non-maleficence, beneficence, justice and fairness, and veracity in the class and clinical labs.

Seventy-seven of the 80 second-year dental hygiene students consented to participate in this study at the start of the semester. Two students did not complete the consent form and one student dropped out of the study. The sample consisted of four intact groups at four dental hygiene schools. The LP school (n = 17) served as the control group and PF School A (n = 20), PF School B (n = 19), and PF School C (n = 21) served as experimental groups. The three experimental treatment groups followed a productive failure design and the fourth group as the control followed a lecture and practice design.

**Program Sites, Context, and Instructors**

Four dental hygiene programs were selected for this study based on multiple criteria. First, the sites were at community colleges that offered an Associate’s in Applied Science degree in Dental Hygiene and were accredited by The Commission on Dental Accreditation of the American Dental Association (CODA), an agency that maintains standards of quality in the education of dental hygiene. With 335 dental hygiene programs in the United States, this study focused primarily on programs offering an associate degree since the majority of the programs (82.7%, 277) were sponsored at this type of institution and yet the majority of dental hygiene research is conducted at universities (14.6%, 49) offering a baccalaureate degree (ADA, 2014-2015). Both types of academic programs allow students to take a licensure examination to work, however baccalaureate programs tend to prepare students for further schooling such as a master’s degree for careers in teaching, research, or clinical practice in school or public health
programs (ADA, 2016). Therefore, the baccalaureate curriculum may have a different focus compared to associate degree dental hygiene programs.

Student admission into the four dental hygiene programs was similar. At the time of the study, student candidates were expected to meet certain criteria and prerequisite courses prior to admissions. For example, students must possess a high school diploma or GED, receive a minimum 3.0 or above grade point average (GPA) for general education courses in English 101 and 102, math, reading and basic science courses in human anatomy and physiology, microbiology, and chemistry. Additional requirements included a level 1 fingerprint clearance card and background check which must be completed prior to admissions. It is important to note that while each program can be found at its own accredited institution, three of the programs and their home institutions are part of a large community college district made up of 10 community colleges in the metropolitan area, in which one admission process is used for all of their students and programs. Furthermore, the three programs also share the same curriculum and sequence. I selected the three programs for this study to control variations that can occur in the admission process and with requirements.

The four dental hygiene programs’ admission requirements were comparable; however the delivery of the programs and teaching experience of instructors varied significantly from one another and are presented in Table 4.1. Despite the differences with respect to the delivery and length of the programs, the contrast between them may help elaborate the productive failure framework and explain differences in student learning across program contexts.
The selected dental hygiene programs were also chosen because of their relative equivalence in population size with one another, ease of proximity to the researcher, and willingness of the district and dental instructors to participate in this study. All instructors hold a master degree and have conducted and/or participated in research studies. The instructors were selected because they had several years of experience teaching dental hygiene courses at their programs and specifically, the ethics course or course with ethics content were used in this study. Indeed, the instructors understood the social dynamics of student groupings and classroom norms to effectively engage students in discussion and problem solving activities. As such, the instructors were also used to providing assistance and support to students. However, one month prior to the first day

<table>
<thead>
<tr>
<th>School</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP School</td>
<td>• Online, lecture and practice format</td>
</tr>
<tr>
<td></td>
<td>• 4 semesters</td>
</tr>
<tr>
<td></td>
<td>• One experienced instructor new to implementation</td>
</tr>
<tr>
<td></td>
<td>• Classroom setting</td>
</tr>
<tr>
<td>PF School A</td>
<td>• Online and hybrid format</td>
</tr>
<tr>
<td></td>
<td>• 15-month accelerated program</td>
</tr>
<tr>
<td></td>
<td>• One experienced instructor new to implementation</td>
</tr>
<tr>
<td></td>
<td>• Classroom setting</td>
</tr>
<tr>
<td>PF School B</td>
<td>• Meta-cognitive focus format</td>
</tr>
<tr>
<td></td>
<td>• 4 semesters</td>
</tr>
<tr>
<td></td>
<td>• One experienced instructor new to implementation</td>
</tr>
<tr>
<td></td>
<td>• Classroom setting</td>
</tr>
<tr>
<td>PF School C</td>
<td>• Lecture and practice format</td>
</tr>
<tr>
<td></td>
<td>• 4 semesters</td>
</tr>
<tr>
<td></td>
<td>• One novice instructor new to implementation</td>
</tr>
<tr>
<td></td>
<td>• Classroom setting</td>
</tr>
</tbody>
</table>
of the semester, one instructor dropped out of the study and was replaced with an instructor who had six months of teaching experience with little to no previous experience teaching an ethics course in dental hygiene education. It was unclear if the new instructor received any mentoring support prior to this study to help apply learning theories to the classroom.

**Sample Size and Power**

Most dental hygiene programs accept approximately 17 to 30 students as one cohort group. In this study, the control condition, LP school consisted of 17 students. The three experimental conditions, PF School A consisted of 20 students, PF School B consisted of 19 students, and PF School C consisted of 21 students. The four intact groups represented the typical number of students one might find in dental hygiene programs. Since randomization was not possible due to restrictions by the institutions and programmatic course sequencing, sampling bias may make generalization of the findings to the targeted population less accurate affecting validity (Johnson & Christensen, 2012). Because sites were predetermined, the groups were intact and could not be expanded rendering an a priori power analysis of no use. Results were interpreted in terms of statistical significance tests and effect sizes measures to look for similarities, differences, strengths and weaknesses.

**Ethical Safeguards**

Each participant signed a consent form (see Appendix A) to participate in the study. No reporting of actual names of the participants occurred. Instead, students’ names were replaced by an assigned number that only the researcher had access to. This research did not affect the grade of any student involved in the study. The students in PF
and LP instructional design received the same content and material. Confidentiality and privacy of students’ actual names, raw data, and codes to the data were stored in a locked file cabinet. At no point in the study were the students placed in danger or exposed to any harm.

Measures

**Pretest and Posttest Assessment Defining Issues Test.** Many empirical studies have used the Defining Issues Test (DIT) to evaluate the effectiveness of an educational intervention or program on the capacity for moral reasoning in student populations (Auger & Gee, 2016; Bebeau & Thoma, 1994; Cannon, 2008; Cummings, Maddux, Cladianos, Richmond, 2010; Dotger, 2010; O’Flaherty, & McGarr, 2014; Waters & Carmichael-Burton, 2008; Schmidt, McAdams, & Foster, 2009; Wilhelm & Czyzowski, 2012). The DIT was derived from Kohlberg’s original cognitive-developmental approach, but James Rest (1999) built upon Kohlberg’s stages and described moral reasoning and development as schemas: 1) Personal interest schema includes stages 2 and 3 which describes decisions made by what is personally at stake, 2) Maintain Norms schema represents stage 4 which describes decisions based on the individual’s orientation of societal cooperation such as following rules and laws, and duty orientation with little negotiation, and 3) Post conventional schema encompasses the more advanced stages 5 and 6 which are based on shared ideals in which moral decisions can be altered and renegotiated using an utilitarian or what is good for all perspective (Rest et al., 1999). Unlike Kohlberg’s “staircase” model, Rest et al. (1999) argued that moral schemas describe how individuals process information by the shift in distributions within the three moral schemas. For example, moral decision making generally includes all three moral
schemas with one moral schema showing a higher distribution overall by how the individual approaches the moral dilemma (Rest et al., 1999). A quantitative instrument, the DIT collects information that is rated and ranked by the individual on how he or she interacts with the stories providing a measure of the level of moral development.

In 1999, the original DIT was replaced with the DIT-2, which is distributed and scored by Center for Ethical Development at the University of Alabama. With updated stories, five rather than six dilemmas, and a bogus check that purges fewer participants, DIT-2 is highly correlated with the original DIT ($r = .79$) and has a high internal consistency (Cronbach’s $\alpha = .90$) (Rest et al., 1999). The DIT-2 has five ethical moral dilemmas followed by issue statements that are rated and ranked (Rest, 1979). The first item asks the individual which action to take: *Do something*, *Can’t decide*, or *Do nothing*. The next 12 items ask the individual to rank the issues as *Great*, *Much*, *Some*, *Little*, or *No*. The last four items rank the importance of 4 of the 12 issues as *Most important issue*, *Second most important*, *Third most important*, and *Fourth most important*. This process continues with the other four dilemmas.

Additionally, the development of a newer index in the DIT-2, the New Index Score index (N2 index) examines the individual’s development in terms of shifts of distribution between the moral schemas. The N2 index uses both rating and ranking data from the DIT and adjusts the Post conventional score up or down by how an individual prioritizes the high stages and how the individual discriminates and rejects the lower stages (Rest et al., 1997). For example, a higher N2 index score reflects an individual's increased capacity for reasoning about moral issues based on a system of fairness that serves the public good; a lower N2 index score tends to reflect reasoning about moral
issues from a self-serving understanding of fairness (Mayhew, Pascarella, Trolian, Selznick, 2015). Furthermore, the N2 index score indicates an individual’s adoption or acquisition toward more sophisticated moral thinking or higher schema such as the Post conventional schema (Rest et al., 1999).

Although there has been recent criticism of DIT-2 for its ability to measure change in an individual’s moral reasoning process (Curzer, Sattler, DuPRee, & Smith-Genthôs, 2014), the DIT-1 and DIT-2 have more than 40 years of empirical evidence starting with James Rest in the 1970’s. The DIT-2 has stood the test of time as a measure within moral psychology (Thoma, Bebeau & Narvaez., 2016) with a broad range of studies demonstrating reliability and validity (Baily, 2011; Rest, Thoma & Edwards, 1997; Thoma & Dong, 2014) across several populations such as students in dental, nursing, medical, and pharmacy (Bebeau, 2002), information technology (Woodward, Davis & Hodis 2007), accounting (Klimek & Wenell, 2011); physical therapists (Edwards, van Kessel, Jones, Beckstead & Swisher, 2012); and educational leadership students (Greer, Searby, & Thoma, 2015).

This study used N2 scores to compared students’ pretest and posttest on the target concept of resource allocation, a system based on fairness in how to distribute resources effectively and efficiently to maximize health benefits for the public good at the lowest cost. Effects of productive failure intervention in the acquisition of new thinking were measured using the N2 index, which is more sensitive to educational interventions that promote moral judgment development (Rest et al., 1997). Specifically, the N2 index measures the degree to which an individual discriminates, prioritizes, and rejects simplistic thinking, (items that appeal to Personal interest which are seen as representing
Kohlberg’s Stage 2 reasoning) and chooses Post conventional reasoning items represented by the higher stages 5 and 6 in Kohlberg’s moral development. The N2 score ranges between 0 and 95 and is determined by adding the Postconventional score to the rating data. Cronbach’s alpha for the N2 index was .83 with a strong correlation between Postconventional and N2 scores at $r = .95$ (Rest et al., 1997).

**Pretest and Posttest Assessment - Checklist for Ethical Decision-Making.** The assessment used to evaluate students’ ethical decision-making was adapted from Bebeau’s (1995) past work. Bebeau (1995) developed a manual with teaching and assessment materials that outlines methodology for teaching ethics to dental students. What started as a grant project 25 years ago sponsored by the American Fund for Dental Health, Bebeau (1995) developed an instructional booklet with ethical dilemma cases and criterion checklists for dental educators to use for an ethics course. The booklet was designed to help dental educators provide instruction to strengthen students’ abilities to develop well-reasoned responses to real world ethical problems and has been extensively tested with several cohorts of dental students in Minnesota (Bebeau, 19995). The teaching material and assessments are available on the Center for the Study of Ethical Development at the University of Alabama (www.centerforthestudyofethicaldevelopment.net). Bebeau’s checklist has four main criteria that identify key points of ethical conflict in ethical analysis: identifying the affected parties, describing the ethical issues, describing the consequences of acting, and describing the relevant obligations or duties. Solutions were scored out of 17 points; 5 points in describing the ethical issues; 3 points in identifying the affected parties; 4 points
in describing the consequences of acting; and 5 points in describing the relevant obligations.

**Scoring the Ethical Dilemma.** For this study, I decided not to use the cases in the booklet since they were unrelated to dental hygiene dilemmas. Instead, I adopted the criterion checklist (Appendix B) and modified it to measure my second research question on problem solving skills and the transfer of knowledge from one problem to another novel problem. I conducted a pilot study on the modified criterion checklist to ensure instrument validation. In the pilot study, a dental hygiene colleague and I independently assessed four pretest and posttest student responses to establish interrater reliability for the four criteria. We reached an interrater reliability of 0.89 using intraclass correlations (ICC = 0.89, n = 42, \(p < .01\), a two-way mixed, absolute agreement measure (Hallgren, 2012). Once we achieved consistency in rating students’ responses, we established agreement on the four criteria for each problem and created a template to score the remaining responses. Each written response was assigned a study ID number and then randomly assigned to an evaluator. In this study, the process was repeated, except I waited to collect all pretest responses before randomly assigning them to evaluator. I did the same for students’ posttest responses. Using the identical modified checklist, we each scored 154 pretest and posttest student responses. In this study, we reached an interrater reliability of 0.91 using intraclass correlations (ICC = 0.91, n = 154, \(p < .01\); Hallgren, 2012). The results showed high agreement between evaluators suggesting scores were rated similarly across raters. Prior to scoring the responses, we reviewed the template for agreement for both problems.
Moral Dilemma Problems (Practice Problem and Transfer Problem)

I developed the moral dilemma problems to be either well-structured or ill-structured. Three problems were developed, an ill-structured practice problem, a well-structured practice problem, and well-structured transfer problem (See Appendices C, D, E). The problems were validated for face validity by two participating instructors. One external evaluator examined the problems and answers for content validity and applicability of the concept being taught. Moreover, the external examiner made sure the concept was beyond the students’ level of knowledge. The moral dilemma problems were validated for content accuracy by a dental hygiene researcher who teaches ethics to dental and dental hygiene students and were found to have content validity. The practice problems addressed identical conceptual information about resource allocation with key features that differentiate the two designs as either ill-structured or well-structured problems (Table 4.2). The architecture of the problem was an important problem-solving feature in productive failure (Kapur, 2008). Kapur and Bielaczyc (2011) explained that problems in productive failure should challenge the student to generate and explore all possible actions unlike most problems in schools and textbooks which are designed as well-structured.

**Well-Structured Problem Design.** Well-structured problems are defined as specific in concepts and principles, the information is straightforward and present, and the problems have knowable and probabilistic solutions (Jonassen, 1997). To solve a well-structured problem, the individual uses normative principles or logical approaches to solve the problem. According to Kapur (2008, 2009, 2012), the well-organized structure of the problem reduces the possibilities of varying solutions giving students a greater...
degree of confidence in their solutions in predictive ways because the solutions were so close to being known and obvious. This type of problem is commonly used in teaching and yet, may not challenge students enough to question why the solution works, thus limiting their decision-making skills and knowledge assembly of key ideas and concepts for future problems (Jonassen, 1997; Kapur, 2008). However, most everyday problems, particularly moral dilemmas, are not well-structured, but are rather messy and muddled with ill-structured parameters.

Table 4.2

_Ethical Dilemma Problem Design_

<table>
<thead>
<tr>
<th>Ill-structured Problem</th>
<th>Well-structured Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Possess many problem parameters with some unknown, or known only with some degree of confidence</td>
<td>Possess fewer problem parameters than their ill-structured counterparts</td>
</tr>
<tr>
<td>2. Possess problem parameters that interact with each other in interesting ways such that the effect of each could not be examined in isolation</td>
<td>Present problem parameters to the learners with greater degree of confidence,</td>
</tr>
<tr>
<td>3. Possess multiple solutions and solution path</td>
<td>Require the application of a limited number of regular rules and principles that were organized in predictive ways</td>
</tr>
<tr>
<td>4. Possess multiple criteria for evaluating solutions</td>
<td>Have knowable, comprehensible solutions where the relationship between decision choices and all problem states are probabilistically close to being known</td>
</tr>
<tr>
<td>5. Require students to make assumptions, judgments, and express personal opinions or beliefs</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The Problem Design is from Kapur (2008)
**Ill-Structured Problem Design.** Ill-structured problems are similar to everyday problems. They are less constrained by specific content domains taught in class, have multiple solutions that are not immediately obvious, and the outcome is uncertain (Jonassen, 1997). Ill-structured problems require students to make judgments about the problem and defend them (Jonassen, 1997). The use of ill-structured problems is well known in dental and dental hygiene education mostly with problem-based learning. However, problem-based learning requires guidance by the instructor while going through the exploration process (Bassir, Sadr-eshkevari, Amirkhoreh, & Karimbux, 2013) and productive failure does not.

**Research Design**

A quasi-experimental study was designed to examine the effects of productive failure compared to traditional lecture and practice for three weeks on a moral reasoning unit. Four dental hygiene programs created four intact groups. In dental educational research, random sampling is not always possible and randomization requirements cannot always be met (Chambers, 2012). Chambers (2012) explained that “the best way to manage this is to use multiple groups, only one of which is exposed to the intervention. A comparison of the outcomes observations in the experimental and the control groups adds confidence to claims regarding interventions. (p. 31)”. Therefore, the design for this study followed Campbell and Stanley’s (1963) non-randomized Pretest-Posttest Group Design with one control group, receiving the lecture and practice (LP) instruction and three experimental groups, each receiving the productive failure (PF) instruction. The intact groups were assigned to either control group or treatment groups. After interviewing each instructor and inquiring about the teaching methods they used, I
learned that two of the schools used LP instruction method. The two schools were matched and assigned to either control group or experimental group. The other two schools used alternative teaching approaches and were assigned to the experimental groups. One school used a hybrid approach, online and face-to-face platforms. The other school used a meta-cognitive approach to teaching. For instance, students are taught self-awareness, personal growth, and well-being in developing them in becoming a more compassionate practitioner. This approach has been recognized as a way to improve attitudes of professionalism (Lovas, Lovas & Lovas, 2008). The control group was identified as Lecture and Practice (LP) School after the assigned instructional design. The experimental groups were identified in the following way, Productive Failure (PF) School A, PF School B, and PF School C.

The four conditions participated in the same number of lessons of three, 50-minute classes in three weeks. The amount of instructional time was held constant for both conditions. Prior to week one, the students took the DIT-2 online which served as a pretest to measure the student’s level of moral development. Students had four days to complete the survey. Students who were unable to complete the survey prior to the first observation day were excluded from the study. For the first two week, the groups were asked to follow either the LP lesson or PF lesson to solve a moral dilemma problem. The problem concept for both groups was the identical, but PF schools practice problem was ill-structured in design and LP school problem was well-structured. Students in PF schools worked on the practice problem with no assistance or guidance in the week one followed by a consolidation or lecture phase in week two. In contrast, students in LP school received the lecture first in week one and worked on the practice problem with
instructor assistance and guidance in week 2. In the last 10 to 15 minutes, the instructor reviewed the answer with the class. All completed practice responses were collected before randomly assigning them to evaluators. Each written response was assigned a study ID number and scored using the established template agreed by evaluators Appendix F). In the week three, students in the four groups completed the transfer problem individually. The grading process was repeated for students’ posttest written responses using the template (see Appendix G) evaluators established agreement. At the end of the three-week intervention, the four groups took the DIT-2 online again which served as the posttest. Quantitative research methods allowed me to analyze the capacity for moral development for variations between groups.

Interventions

Productive Failure Design. Three instructors participated in the PF instructional approach. The PF instructors received a three-hour training session two months prior to the study (See Appendix H). From the discussion in the training, I created a productive failure lesson outline (Appendix I) for the instructors on how to provide instruction on the order of events to follow, materials needed for the day, and procedure on providing feedback to the students. I sent the outline one month prior to the start date. Additionally, support was provided throughout the research study. Support consisted of emails once or twice a week in which questions were explained, intervention implementation was reviewed each week, and student names of those who did not complete the DIT-2 were sent to instructors for follow-up. The instructors and I determined the start date of treatment implementation based on where the problem and
PRODUCTIVE FAILURE

concept fit best in the course and therefore, the three conditions started at different times in the semester.

This study followed Kapur’s (2009) study of implementing the two phases of PF: 1. the invention phase and 2. consolidation phase. In the invention phase, students worked in groups of dyads or triads on their own in week one for 50-minutes. The groups were formed either by the students or instructor. PF School B preassigned students to groups when they first entered into the program unlike PF School A and PF School B where groups were self-selected.

PF students in groups attempted to solve the ill-structured moral dilemma based on the concept of resource allocation, a concept students have not been taught yet in their regular course. The researcher-developed dilemma was designed as an ill-structured problem to stimulate student responses and ethical analysis.

Ill-structured problems are designed for students to engage in the problem and create and generate a moral decision used to explain their answers and position (Kapur & Bielaczyc, 2011). However, the moral dilemma problem was intentionally designed beyond the skills and abilities of the student to which they ultimately failed to approach the moral dilemma using the highest level of approach.

Students were given blank sheets of paper for all their group work. The blank sheets served as documentation and evidence of students’ interaction patterns of moral development by their explanation, critique, and elaboration in their engagement with the ill-structured problem (Kapur, 2008, 2009, Kapur & Bielaczyc, 2012). Following Kapur’s (2008, 2009) study, the instructors refrained from making suggestions or leading the students toward a correct method. Instead, when the groups appeared frustrated or
asked questions, the instructors were instructed to use affective phrases like “Keep working hard on the problem”; “Continue your effort in solving the problem; “Keep stretching yourself”; and “Let’s see what you can come up with on your own” to encourage them to persist and continue to find a solution.

The purpose of the consolidation phase led by the instructor was to assist students on how to organize and assemble key ideas and concepts as well as piece together and close gaps in their knowledge and understanding (Kapur, 2008, 2009).

**Lecture and Practice Design.** Students in the control group, LP, were grouped together in Pre-Clinic, a first semester first year clinical lab using StrengthsQuest, a personality assessment inventory to group individuals to create a positive environment for group work. These groups remained intact in this study. In the LP design, the order of lecture and practice and problem type were manipulated and differed from the PF groups. The LP students were given a lecture first on the same novel concept in week one for the full period. Students in week two then practiced for the first 40 to 45 minutes on a well-structured complex problem on blank sheets. The last 10 to 15 minutes, the instructor reviewed their answer with the class. According to Kapur (2008), the well-structured problem is commonly used in teaching and yet, may not challenge students enough to question why the solution works, thus limiting their decision-making skills and knowledge assembly of key ideas and concepts for future problems. Additionally, unlike the PF instructor who offered little to no help, the LP instructor provided appropriate scaffolding and assistance as students practiced on the complex problem.

Instead of leaving students alone to struggle, the instructor prompted students with simple questions and hints to guide attention to features in the problem that were
important. In week three, LP students worked on a second ethical dilemma identical to the experimental group that was well-structured. Similar to PF, a three-week outline of the LP lesson (see Appendix J) describing the steps to take, the order in which they needed to be taken, how much time for each lesson, and the teacher’s role was emailed one month prior to the first day of the semester.

**Procedure**

I assessed four groups of students who completed the DIT-2 for their N2 scores at the beginning of the PF or LP intervention. N2 scores measured students’ ability to identify, articulate, and reflect critically on ethical issues from their own perspective both personally and professionally, which is specific to the purpose of this study.

The practice problem or pretest problem was designed on the concept of allocating resources based on patient needs and how resources are allocated where price is a common means because the concept is not introduced until the fourth or last semester of the program. The practice problem was confirmed in the pilot study prior to the full-scale study. In the practice problem, the moral dilemma focused on a dental hygiene student’s right to learn in terms of what is considered equitable and just. The dilemma is the conflicting objectives of health care systems and training programs. The objective of health care systems is to provide an adequate standard of care and the objective of training programs is teach a skill that a person may require. Educational programs like dental hygiene are seen as low-cost health care systems and yet are training facilities that teach students how to apply specific techniques or services. Allocation decisions can test moral boundaries on how to effectively and efficiently distribute resources such as low
cost dental care to those in need, but who cannot afford dental care and teaching skills to students who need practice, but are required to maintain the standard of care.

Students received the same conceptual problem to solve, however they were presented in different problem designs. The PF pretest problem was designed as ill-structured, whereas the LP pretest problem was designed as well-structured. The PF problem was designed to activate prior knowledge of core values like beneficence, veracity, and maleficence and introduce a novel concept of resource allocation. Dental hygiene students in this study are familiar with the core values, but they have not had any formal instruction on the targeted concept of resource allocation. The other difference between conditions was the sequencing of the phases. The PF condition consisted of a problem-solving phase first followed by a lecture or consolidation phase, a week later. In contrast, the LP condition consisted of a lecture phase first followed by a problem-solving phase, a week later.

In week three, students from the four conditions completed the same well-structured moral dilemma on their own which served as the post-test for transfer of knowledge. The transfer problem was an extension of the practice problem on the concept of allocating resources based on patient needs and how resources are allocated to poor urban children versus poor rural children according to cost benefit. Furthermore, the transfer problem had a public health approach, thus becoming a novel problem compared to the practice problem which was framed more closely to what students might experience in school. After the end of the unit, students took the DIT-2 again which served as the post-test of the effectiveness of the instruction. The survey took between 40-45 minutes to complete each time.
At the end of the experiment, I held a debriefing session with each instructor and asked about their experience, teaching beliefs and practices in terms of persistence.

**Training for the Instructors**

Findings in the pilot study led me to develop a training session for instructors for which I prepared a slide presentation of my study, purpose, and procedure. I also developed a Resource Allocation slide presentation (see Appendix K) for instructors to use with their students, highlighting the main ideas of the targeted concept. By standardizing the Resource Allocation slide presentation, students received the same content in the lecture phase of the study design. Three of the four instructors met with me three months prior to the initial start date for a 3-hour training session that focused on their role, the targeted concept, and the standardized slide presentation. Feedback from the dental hygiene instructors was incorporated to the moral dilemma problem and slide presentation. The training session also informed the instructors the time allocation for each problem taken as a group or individually. Therefore, time was a constant and not a confounding factor for the differences between the PF and LP intervention. Four weeks later, the instructors received instruction on the procedure of the intervention to which they were assigned to.

As previously described, one instructor dropped out of the study one month prior to the first day of the semester and was replaced with another instructor who had far less teaching experience. The instructor went through a similar training session with me that focused on the role, the targeted concept, and the standardized slide presentation to meet the same 3-hour of contact.
Data Analysis

I conducted separate multiple regression analyses to examine research questions 1 and 2. In this study, the control group was the LP school, which served as the reference group in the analyses, and the regression coefficients reflected the treatment-control mean differences. Also, pre and posttest scores were collected and the pretest scores controlled each student’s prior knowledge about the specific topic on the problem. Therefore, pretest scores were included as independent variables. For research question 1, I compared three PF schools with the LP school to examine the extent to which school and condition predicted N2 posttest scores. The N2 pretest-posttest was scored by staff at the Center for the Study of Ethical Development at the University of Alabama. In addition, descriptive statistics were also performed to examine differences in schools and patterns or shifts of moral schema movement between pretest scores and posttest scores.

For research question 2 on productive failure in the transfer of knowledge in solving a moral dilemma, the PF schools served as the independent variables, with the posttest scores of students’ written responses to a well-structured complex moral dilemma as the dependent variable. As previously stated, pretest scores served to control for variation in each student’s prior knowledge. I evaluated whether the overall regression was significant, what percentage of variance was explained in the outcome measure, and whether each school explained unique variance in the outcome. Furthermore, I also examined descriptive statistics to determine how students performed from pretest to posttest with the learned concept of resource allocation. Results were interpreted using statistical significance tests set at <0.05 for the F statistic and effect sizes measures.
Results

The DIT-2 survey gathered participant demographic data including gender, age, ethnicity, and educational level as seen in Table 4.3. Seventy-seven second-year dental hygiene students participated in the study. To measure the effectiveness of instruction, the DIT-2 was administered at two different times during the study. As a result, 13 students did not complete either the DIT-2 pretest (3 students, 4%) or posttest (10 students, 13%). Additionally, staff at the Center suggested that 14 respondents (18%) be

Table 4.3

Demographic Characteristics of Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General (N = 77)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>74</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>18 – 24</td>
<td>22</td>
</tr>
<tr>
<td>25 – 30</td>
<td>21</td>
</tr>
<tr>
<td>31 – 35</td>
<td>15</td>
</tr>
<tr>
<td>36 – 40</td>
<td>10</td>
</tr>
<tr>
<td>41 – 45</td>
<td>8</td>
</tr>
<tr>
<td>Missing</td>
<td>1</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>2</td>
</tr>
<tr>
<td>Asian/Pacific Islander</td>
<td>6</td>
</tr>
<tr>
<td>Hispanic</td>
<td>19</td>
</tr>
<tr>
<td>White (other than Hispanic)</td>
<td>43</td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
</tr>
<tr>
<td>Education level completed</td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>1</td>
</tr>
<tr>
<td>Vocational/ Technical School</td>
<td>16</td>
</tr>
<tr>
<td>Junior College</td>
<td>45</td>
</tr>
<tr>
<td>Some College in Bachelor degree program</td>
<td>12</td>
</tr>
<tr>
<td>Professional degree</td>
<td>1</td>
</tr>
<tr>
<td>Master's Degree</td>
<td>2</td>
</tr>
</tbody>
</table>
purged from the final results due to unreliable responses, excessive missing data points, or inconsistency in responses (DIT-2 Guide, 2003). Attrition was not equal across groups. For example, School A, School B, and School C lost 10, 3, and 7 students, respectively, compared to LP School that lost 7 students in the DIT-2 survey. The final sample size of 50 students was used resulting in a 65% response rate. The control group, LP School, consisted of 10 dental hygiene students with a mean age of 27. The experimental groups consisted of 10 students in PF School A with a mean age of 32, 16 students in PF School B with a mean age of 28, and 14 students in PF School C with a mean age of 30. The majority of students in LP School, PF School B, and PF School C (70%, 75%, 70%, respectively) received a Junior College education compared to 40% of students in PF School A. To measure transfer of knowledge, all student responses (77) on the ethical dilemmas were evaluated and scored.

DIT-2 results were sent to the Center for the Study of Ethical Development at the University of Alabama to be scored. Because I was unable to apply randomization to the study design, it was essential to control for potential student nonequivalence between treatment and control groups (Theobald & Freeman, 2014). To address the first research question, multiple regression was performed to examine the relationship between LP school N2 posttest scores and PF schools (PF School A, PF School B, and PF School C) N2 posttest scores. As noted previously, dummy variables were used to define group membership. Membership for PF School A was coded as “1” whereas non-membership in the group was coded as “0”. Dummy variables were repeated for PF School B and PF School C with LP School as the reference group. Descriptive statistics of N2 pretest and posttest scores, and effect sizes are seen in Table 4.4. Effect size estimates yielded small
Table 4.4

*N2 Pretest-Posttest Scores Means*

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LP School</td>
<td>26.18 (13.48)</td>
<td>30.52 (10.33)</td>
<td>.36</td>
</tr>
<tr>
<td>PF School A</td>
<td>29.28 (12.97)</td>
<td>34.27 (13.42)</td>
<td>.38</td>
</tr>
<tr>
<td>PF School B</td>
<td>25.58 (12.93)</td>
<td>31.52 (15.59)</td>
<td>.42</td>
</tr>
<tr>
<td>PF School C</td>
<td>24.95 (14.68)</td>
<td>24.73 (16.67)</td>
<td>-.01</td>
</tr>
</tbody>
</table>

*Note:* Using Cohen’s $d$ (1988) $d = 0.20$ small effect; $d = 0.50$ medium; $d = 0.80$ large effect.

to moderate effects among all the groups except with PF School C. This finding highlights that instructional effectiveness was found across instructional design.

The predictors explained 44% of variance $R^2 = .44$, $F (4, 45) = 8.95$, $p < .01$, which was significant. Table 4.5 summarizes results for the regression analysis on the teaching methods of the N2 posttest scores. Only N2 pretest means were found to be significant, $p < .001$ compared to N2 posttest means. As shown, holding all predictors equal, we can expect a student’s posttests score to increase by .63 SD with each 1 SD increase in student scores on the pretest. However, N2 scores of PF schools were not found to be significant compared to LP School posttest scores. With LP school as the baseline and controlling for the students’ scores on the pretest, for each 1 SD increase in the pretest score, scores on the N2 for PF School A and PF School B increased by only .04 SD. Moreover, for each 1 SD increase in the pretest score, scores on the N2 in the PF School C were only -.16 SD lower than LP school.
Table 4.5

*Regression of the N2 Posttest Scores on Pretest Scores and Teaching Methods*

<table>
<thead>
<tr>
<th>Groups (N = 50)</th>
<th>B</th>
<th>Std. Error</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>13.10</td>
<td>4.77</td>
<td></td>
<td>2.75</td>
<td>.009</td>
</tr>
<tr>
<td>Pretest</td>
<td>0.68</td>
<td>0.12</td>
<td>.63</td>
<td>5.58</td>
<td>.001</td>
</tr>
<tr>
<td>PF School A</td>
<td>1.35</td>
<td>5.12</td>
<td>.04</td>
<td>0.27</td>
<td>.79</td>
</tr>
<tr>
<td>PF School B</td>
<td>1.11</td>
<td>4.59</td>
<td>.04</td>
<td>0.24</td>
<td>.81</td>
</tr>
<tr>
<td>PF School C</td>
<td>-5.25</td>
<td>4.71</td>
<td>-.16</td>
<td>-1.11</td>
<td>.27</td>
</tr>
</tbody>
</table>

*Note.* Dependent variable = N2 Posttest scores. The reference group is the Lecture and Practice group.

Although no significant changes were found among the groups in the overall moral judgement or N2 scores, it was worth examining the individual moral schema levels (Personal interest, Maintaining Norms, and Post conventional) to observe how students approached the scenarios and which moral schemas were used primarily in their decision making process. After Rest et al., (1997) explained that the score itself “does not describe information about the extent of whether one schema predominates over the others to a great degree or whether the three schemas are rated more equally” (p. 313). In fact, Bebeau et al. (2002) found that in a study on dental students, Post conventional score had masked acquisition of new thinking and rejection of simplistic thinking.

The shifts of distribution in moral reasoning from pretest scores and posttest scores for each group were analyzed using Paired Sample t-test. Table 4.5 describes shifts of distribution in the moral schemas using change scores to show directional
movement either away (-) or toward (+) one schema to another schema. Effect sizes \(d\) were used to describe the magnitude of the shifts from pretest to posttest.

As indicated by Table 4.6, after receiving PF instruction, PF School A on average

### Table 4.6

*Shift of Distributions Between Moral Schemas Using Mean Differences*

<table>
<thead>
<tr>
<th></th>
<th>LP School</th>
<th>PF School A</th>
<th>PF School B</th>
<th>PF School C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal interest</td>
<td>+ 0.67 (0.05)</td>
<td>- 4.58 (0.28)</td>
<td>- 4.00 (0.33)</td>
<td>+ 0.37 (0.02)</td>
</tr>
<tr>
<td>Maintaining Norms</td>
<td>- 4.11 (0.27)</td>
<td>- 2.88 (0.20)</td>
<td>+ 0.38 (0.04)</td>
<td>+ 1.81 (0.15)</td>
</tr>
<tr>
<td>Post conventional</td>
<td>+ 2.33 (0.17)</td>
<td>+ 10.66 (0.79)</td>
<td>+ 2.50 (0.17)</td>
<td>- 3.53 (0.23)</td>
</tr>
<tr>
<td>N2 Scores</td>
<td>+ 4.77 (0.38)</td>
<td>+ 4.99 (0.38)</td>
<td>+ 4.96 (0.41)</td>
<td>- 0.22 (0.01)</td>
</tr>
</tbody>
</table>

*Note:* Using Cohen’s \(d\) (1988) \(d = 0.20\) small effect; \(d = 0.50\) medium; \(d = 0.80\) large effect.

used less Personal interest and Maintaining norms reasoning and more Post conventional reasoning. With an effect size of \(d = 0.79\), the difference can be interpreted as a large effect (Cohen, 1988). Indeed, the increase in Post conventional scores suggests that students demonstrated judgments and actions that were open to debate and appealed more to furthering the common good (Rest et al., 1997). Similar patterns were found in PF School B which showed a decrease utilization of Personal interest reasoning with more use in Post conventional reasoning with little change in the use of Maintaining Norms reasoning.

Descriptive statistics are presented in Figure 4.1, to illustrate the pattern of change between pretest scores and posttest scores in each schema. This finding suggests that
norms and rules are important to students in PF School B approach to moral dilemmas (Rest et al., 1997). Likewise, LP School saw a shift toward using more Post conventional. However, unlike PF School B, LP School shifted away from Maintaining Norms reasoning while holding onto the use of Personal interest reasoning, a lower level of moral reasoning. The patterns of LP School demonstrated use of the three schemas equally in making decisions on moral dilemmas. This finding suggests that although rules are necessary, they can be challenged if necessary to restructure society, but decisions are guided by a personal stake the individual has in the consequences of an action (Rest et al., 1997). In contrast to the other groups, PF School C was the only group to shift away from using Post conventional moral reasoning from pretest to posttest. Patterns indicated that PF School C used more Maintaining Norms reasoning in resolving ethical dilemmas on DIT-2 while holding onto the Personal interest moral schema. This finding suggests that PF School C students prefer standards and stable norms to make decisions as long as those decisions either positively benefit the self or has little impact on the self or someone close to them (Rest et al., 1997).

**Transfer of Knowledge**

The study also examined the transfer of knowledge in solving a moral dilemma between two instructional designs-Productive Failure and Lecture and Practice. Data from students are used in this section ($n = 77$). The variables being investigated include pretest and posttest scores on students’ written responses on two moral dilemma problems, which were evaluated by Bebeau’s Criterion Checklist.
Figure 4.1. Patterns of Shifts in Moral Schemas from N2 Pretest to Posttest Scores

**Patterns of Personal interest Scores**

- **Pre Personal Interest**
- **Post Personal Interest**

**Patterns of Maintaining Norms Scores**

- **Pre Maintaining Norms**
- **Post Maintaining Norms**

**Patterns of Postconventional Scores**

- **Pre Postconventional**
- **Post Postconventional**
Table 4.7 summarizes the findings from the process analysis. Similar to findings in Kapur and Bielaczyc’s (2011) study, PF students were unable to solve the ill-structured practice problem successfully in groups compared to LP students who worked on the well-structured practice problem. As previously discussed, PF students received no assistance or scaffolding from instructors during this period, unlike LP students who received a lecture first on the learning concept followed by a well-structured practice problem with guidance and heavy scaffolding from the instructor.

I used multiple regression to examine the relationship between two instructional designs, productive failure (PF School A, PF School B, and PF School C) and lecture and practice (LP school) on posttest scores of student responses on the transfer problem using pretest scores as the control variable. It was essential to control the pretest scores for potential student nonequivalence especially when evaluating the impact of a nonrandomized education intervention (Theobald & Freeman, 2013). The LP school

Table 4.7

<table>
<thead>
<tr>
<th>Moral Dilemma Total points (17)</th>
<th>N (77)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP School</td>
<td>17</td>
<td>10.12</td>
<td>2.15</td>
</tr>
<tr>
<td>PF School A</td>
<td>20</td>
<td>7.75</td>
<td>2.27</td>
</tr>
<tr>
<td>PF School B</td>
<td>19</td>
<td>7.25</td>
<td>2.49</td>
</tr>
<tr>
<td>PF School C</td>
<td>21</td>
<td>9.00</td>
<td>2.32</td>
</tr>
</tbody>
</table>
served as the reference group used to compare several PF treatment groups. Table 4.8 summarizes the descriptive statistics and analysis results.

Table 4.8

Posttest Scores of Written Responses ($n = 77$)

<table>
<thead>
<tr>
<th>Response Posttest Score</th>
<th>N (77)</th>
<th>Mean (SD)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP School</td>
<td>17</td>
<td>8.59 (2.60)</td>
<td>.90</td>
</tr>
<tr>
<td>PF School A</td>
<td>20</td>
<td>9.20 (2.73)</td>
<td></td>
</tr>
<tr>
<td>PF School B</td>
<td>19</td>
<td>8.82 (2.74)</td>
<td></td>
</tr>
<tr>
<td>PF School C</td>
<td>21</td>
<td>8.71 (2.59)</td>
<td></td>
</tr>
</tbody>
</table>

The model had an $R^2 = .01$, $F (4, 72) = .17$, $p = .95$, which was not significant. Table 4.9 summarizes results for the regression analysis of the well-structured problem posttest scores on teaching methods to describe the parameter of the regression model and the relationship between LP posttest scores with PF posttest scores while controlling for pretest scores and holding the effects of all other PF schools constant. The practice problem pretest mean was not found to be significant, $p = .75$ compared to posttest mean. As shown, holding all else equal, we can expect a student’s posttest score to increase by .04 SD higher for each SD the student scores on the pretest which is considered negligible. For example, PF School A was about .12 SD higher in posttest scores than LP School scores. The standard deviation of LP School scores was 2.60 and constituted a change of only 1.1 point. Therefore, for every .12 SD increase on PF
Table 4.9

Regression of the Well-structured Problem Posttest Scores on Teaching Methods

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>Std. Error</th>
<th>β</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>8.12</td>
<td>1.57</td>
<td>5.17</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>.05</td>
<td>.14</td>
<td>.04</td>
<td>.32</td>
<td>.75</td>
</tr>
<tr>
<td>PF School A</td>
<td>.72</td>
<td>.93</td>
<td>.12</td>
<td>.77</td>
<td>.44</td>
</tr>
<tr>
<td>PF School B</td>
<td>.37</td>
<td>.95</td>
<td>.06</td>
<td>.39</td>
<td>.70</td>
</tr>
<tr>
<td>PF School C</td>
<td>.18</td>
<td>.88</td>
<td>.03</td>
<td>.21</td>
<td>.83</td>
</tr>
</tbody>
</table>

Note. Dependent variable = Well-structured Problem Posttest scores. Dummy coding was used with the reference group as Lecture and Practice, coded “0”.

A school score, a gain of .31 point was expected compared to LP School. This holds true if the effects of PF School B and PF School C were held constant. On average, for each 1 SD increase in LP School scores, PF School B gained .06 SD PF School C gained .03 SD, which was considered negligible.

Although results did not show significant change across posttest scores of PF schools and LP School, I examined how students performed from pretest to posttest to see if the groups were able to work with the learned concept of resource allocation to carry out their approach and explanation for their next solution method (Kapur & Bielaczyc, 2011). ANOVA was conducted and Table 4.10 shows the mean difference and effect size. A positive score indicates that the posttest score was greater than the pretest score and a negative score indicates that the posttest score was less than the pretest score. On transfer problem, PF School A and PF School B posttest scores showed positive scores. This finding suggests that PF students received scores higher than their pretest problem,
Table 4.10

*Mean Difference of Pretest and Posttest scores on Moral Dilemma Problems*

<table>
<thead>
<tr>
<th>School</th>
<th>Mean Difference (SD)</th>
<th>p</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF School A</td>
<td>+ 1.45 (3.1)</td>
<td>0.05*</td>
<td>.58</td>
</tr>
<tr>
<td>PF School B</td>
<td>+ 1.21(3.5)</td>
<td>0.15</td>
<td>.60</td>
</tr>
<tr>
<td>PF School C</td>
<td>-0.14(3.3)</td>
<td>0.85</td>
<td>-.12</td>
</tr>
<tr>
<td>LP School</td>
<td>-1.53(3.5)</td>
<td>0.09</td>
<td>-.64</td>
</tr>
</tbody>
</table>

*Note: *statistically significant

with PF School A gains as significant \(p = 0.05\) with a moderate effect size \(d = .58\). In contrast, LP School and PF School C showed a negative score which suggests that students received scores lower than their pretest problem and the difference between pretest and posttest scores in LP School had a moderate effect \(d = -.64\). Indeed, when interpreting the results from the analyses, it should be noted there are differences in characteristics of the schools, students, teachers, and teaching methodology that may confound these findings.

**Discussion**

This study was designed to examine the effects of productive failure on student learning in solving moral dilemmas compared to lecture and practice. Moreover, this study was designed to investigate the extent of productive failure in helping students learn skills in transferable problem solving. Specifically, I investigated the effects of productive failure in a community college setting in the area of dental hygiene, and with content in ethics and moral reasoning. This study applied the instructional design created
by Kapur (2008, 2009; Kapur & Bielaczyc, 2012) and extends the previous work of prior investigators to open-ended problems such as moral dilemmas that have multiple perspectives and multiple canonical solutions.

The results showed that productive failure conditions did not produce significant changes in students’ N2 posttest scores on moral judgment development compared to the lecture and practice condition. However, small to moderate effect sizes ($d = .36, .38,$ and $.42$) revealed that despite the instructional design, LP and PF Schools A and B students raised their N2 scores except students in PF School C. These findings are consistent with other ethics courses that used DIT as the criterion measure on instructional effectiveness in science fields ($d = .38$) (Antes, Murphy, Waples, Mumfor, Brown, Connelly, & Devenport, 2009), which may suggest that ethics instruction is beneficial for students in developing moral reasoning.

However, of particular interest, PF School A’s profile largely peaked in post conventional items ($d = 0.79$) to suggest that a greater shift or commitment to the moral perspective (Thoma & Rest, 1997). For example, raw data showed that 5 of the 10 students lowered their personal interest scores, two students saw an increase in personal interest scores, and 3 students did not change their scores from pretest to posttest. Moreover, post conventional scores increased in 8 of the 10 students, but remained the same in one student and decreased in another. On the other hand, the schema profile of students in the LP School appeared flat across all three moral schemas even though there was some movement toward Post conventional items. Four of ten students saw a decrease in personal interest scores, one student did not change scores, and five students showed an increase in Personal interest scores. Likewise, four students showed an
increase in post conventional scores, one student did not see a change, and five students showed a decrease in post conventional scores. This finding indicates that LP students on average appeared to be less committed to a single perspective because no clear distinction could be made on the social situations (Thoma & Rest, 1997). Specifically, LP students were unsure how to approach the stage items causing more conflicting interpretations than a clear definition of which perspective to take (Thoma & Rest, 1997).

In summary, based on the predicted patterns, the stage mixture of LP students suggests that decisions tend to be around self-interest as the underlying force (Thoma & Rest, 1999); whereas PF students in School A and B were more likely to prefer to make judgments beyond their self-interest, but based on norm references with consideration given to the needs and value of the patient (Chambers, 2007). This is an important finding because individuals who make decisions from the post conventional schema are more likely to have access to better conceptual tools to look at moral or ethical issues from a different perspective- a more global perspective (Rest, Narvaez, Bebeau & Thoma, 1999) and make decisions based on the facts of the situation not peer expectations (Klimek & Wenell, 2011). As the dental hygiene profession evolves, more professional responsibilities to the public and dental hygienists will be expected. Challenges of ethical situations that require quick decisions will require dental hygienists to be competent in applying the principles of ethical decision making. This finding suggests that PF could be an instructional design to use for moral development growth, specifically with content in ethics and moral reasoning in dental hygiene programs with different pedagogically approaches. However, since this is the only known productive
failure study in which an open-ended problem was used to test the effectiveness of productive failure, it is difficult to make comparisons with other PF studies.

This study also addressed whether or not productive failure promotes transfer of knowledge on novel complex problems. In phase 1, as predicted, results revealed that all three PF schools scored lower on the ill-structured ethical problem compared to LP School who worked on a well-structured ethical problem. Similar to previous studies (Kapur, 2008, 2009, Kapur & Bielacyzc, 2011), this finding suggests that PF students in groups were unable to converge on causes of the problem or moral dilemma (Kapur, 2008), which may be due in part to the design of the problem. Ill-structured problems are designed to reach beyond surface characteristic recognition unlike well-structured problems (Jonassen, 1997). In contrast, ill-structured problems typically have divergent or multiple solutions that require learners to engage in the problem by using reflective judgment, having good command of domain knowledge, and knowing limitations of their proposed solutions (Jonassen, 1997), thus making it difficult for a group to converge on a solution (Kapur, 2008).

Transfer Problem Posttest results revealed that PF schools were not significantly different than LP School despite having slightly higher scores. This finding suggests that transfer of knowledge in resource allocation was similar across schools irrespective of the type of instructional design. Further analysis revealed that an increase from pretest scores to posttest scores were found in two of the three PF schools, School A and School B. Only PF School A scores reached significant difference ($p = 0.05$) from pretest to posttest and had a moderate effect ($d = 0.58$). Interestingly, PF School C posttest scores declined slightly from pretest, but were still slightly higher than LP scores. A moderate
effect \((d = -0.64)\) was in the direction predicted with LP school showing the largest decline in posttest scores. As evidenced by the findings of gain scores from pretest to posttest, students in PF Schools A and B improved their problem solving skills for deeper understanding of the learned concept to transfer to a new problem. This finding suggests that students were able to carry out key elements of resource allocation concept to a novel complex problem compared to students in LP School and PF School C, thus indicating the problem space in the delay of instruction as well as the design of the problem helpful for students to use again in a new problem.

To sum, the theoretical approaches of lecture and practice and productive failure demonstrated that they helped students to recognize ethical issues and find plausible solutions to moral dilemmas. Although it may appear that PF condition was less effective at supporting learning in a classroom, PF School A and PF School B, after careful examination seemed to engender a slightly deeper conceptual understanding compared to LP students as seen in their schema mixture and specifically, their development of knowledge between ill-structured and well-structured problems. For example, well-structured problems are designed to elicit information processes and are less likely to stimulate the learner to deeply engage and interact with the problem (Jonassen, 1997) which is needed to stimulate working memory capacity for deeper learning and transference of knowledge (Alloway, 2012). Additionally, in well-structured problems characteristics of the problem are more visible and obvious. Therefore, naming principles is not the same as interacting with the principles to guide behavior (Chambers, 2006). Moreover, once the learner finds a successful solution, the problem-solving ends and other possible solutions are explored (Jonassen, 1997). Kapur
(2008) found that students who solved well-structured problems experienced fewer interactions between other students and therefore, did not engage deeper into the problem.

Furthermore, PF students were better prepared to see themselves taking on a professional role as evidenced by students’ response in the transfer problem where the focus was on the social role and responsibility of the profession rather than on the individual interests in the practice problem. Although this was not my initial intent in designing the two problems, this does suggest, however that active learning and presenting problems framed around social issues may help cultivate professional responsibility and raise awareness of social issues (Blue, 2013). This finding suggests that lecture and practice method does little to engage and prepare students to understand key elements of the concept to ask questions about the topic that is necessary in transfer (Bransford & Schwartz, 1999). This finding, though cautious, may also suggest that PF instruction better prepared students to notice critical differences between the two problems with its contrasting-case design of solving an ill-structured problem prior to solving a well-structured problem and thus, helped students to organize and assemble problem-related information to solve the problem (Marton, 2006; Schwartz & Martin, 2004).

Interestingly, in this study the only PF School unable to enhance learning was PF School C. One possibility may be due to the experience of the instructor. As described earlier, there was a change in instructors from the original design. Four weeks prior to first of day of the fall semester, the course was given to an instructor who had less than six months of teaching experience. Although efforts were made to minimize differences
between a novice and experienced instructor by creating a standardized slide presentation and meeting with the instructor to review the role as a facilitator, the experience level of an instructor may have been a significant factor on how resource allocation concepts were represented and presented related to the instructor’s depth of knowledge of the subject. For example, in the open-ended survey after the intervention, I asked the instructor what she noticed about student engagement. The instructor responded positively that students appeared to understand the concept and was amazed they did not ask questions or interact with the moral dilemma during the consolidation phase other than to answer the prescribed questions in the slide presentation. Perhaps that may be a correct assumption; however given the students’ low scores on the post problem, this may not be accurate. In contrast, this response suggests that the novice instructor may have mistaken the students’ silence as a confirmation of their conceptual understanding without the need to probe further with additional questions. Novice instructors may also lack the skills to generate questions for students to elaborate more effectively (Reynolds, 1992) to help students differentiated key knowledge structures to construct or generate multiple scenarios or solutions to the problem and their understanding of resource allocation was more imprecise and superficial (Schwartz & Bransford, 1998). As Reynolds (1992) explains, novice teachers may have difficulty executing tasks like “reading” a class environment and connecting new learning to students’ previous learning and experience, which becomes more apparent if the instructor has a limited understanding of content-specific pedagogy.

Additionally, novice teachers may possess a limited or incomplete problem solving schema in which they struggle to attend to student thinking that is necessary in
providing responsive instruction to help students construct understanding of moral concepts (Levin, Hammer & Coffey, 2009). Preparing for future learning often involves shared learning and interactions with others in the group, questions asked about the topic, and features in the problem become salient to the student resulting in transferable knowledge (Bransford & Schwartz, 1999; National Academy of Sciences, 2012). Moreover, instructors with more experience teaching ethics can identify alternative orientations to moral dilemmas offering alternative approaches to consider (Chambers, 2006a). This could explain why students in PF School C led by a novice instructor did not use more post conventional thinking on the DIT items, make the appropriate connection or identify the important step toward achieving the solution in the moral dilemma compared to other programs led by experienced instructors in teaching ethics. However, it is important to note that in this study, interactional dialogue between instructor and students was not recorded or evaluated.

Finally, in spite of the novice instructor, students in PF School C on average performed slightly better in the well-structured problem than the lecture and practice group, even though both groups did not show gain scores from pretest scores. However, the posttest scores were far from what we consider successful conceptual understanding. To try and understand the low scores across all the groups, I looked closer at the design of the well-structured problem. In reviewing the posttest problem, I noticed the post problem design may have been too complex for students to capture the key elements of resource allocation. In other words, there may not have been enough important characteristics of the well-structured problem space to move beyond analyzing the problem. Vygotsky’s work is perhaps of value here. The zone of proximal development
is the distance between what an individual can accomplish alone and what the individual
can accomplish with the help of a more capable person. In this case, the students may
have been unable to make the necessary connections between prior knowledge and newly
constructed knowledge. In essence, the problem could have been another ill-structured
problem with the same underlying concept of resource allocation. Nevertheless, it should
be noted that this sequencing of problems is not unique to previous PF studies. In fact,
Kapur’s (2008) original PF study included an ill-structured posttest problem. In Kapur’s
(2008) first study on PF, students took two posttest problems-first, a well-structured
problem and then an ill-structured problem. The ill-structured posttest problem required
students not only to use what they learned in the previous phase, but to use more
advanced concepts. Kapur (2008) found that PF students still outperformed LP students
because they were able to discern how to structure the second ill-structured problem and
transfer problem-solving skills to the second posttest. Thus, PF students were able to
notice different things, critical features, and deep structures of the problem elicited by the
first problem that were used to assemble new knowledge for conceptual understanding
and for transfer (Kapur, 2016). This problem solving process could explain why PF
students in the study on average received higher posttest scores producing better
explanations of the targeted content than the LP students. If true, the practice problem
prepared PF students in becoming familiar with the problem structure which enabled
them to learn more effectively in the discovery condition of another ill-structured
problem (Brunstein et al., 2009).

Taken together, these findings on productive failure is promising and should be
considered a viable instructional design for teaching ethics and increasing students’
capacity for moral reasoning. Despite the potential benefits of productive failure approach, the results suggest that it does not always lead to better performance. The findings emphasize the need to explore and identify boundary conditions for the effectiveness of a productive failure approach.

Limitations of the Present Study

While there are several positive contributions of this study, limitations remain. Even though three dental hygiene programs recruit students from the same student pool, there are still multiple confounds between schools, instructors, student ability, and intervention specifications. Additionally, this study is restricted to four community college dental hygiene programs in the southwest; however with 250 dental hygiene programs located in two-year institutions (Survey of Dental Hygiene Programs, 2015-2016), I believe it is possible to glean useful information if the characteristics of the study’s sample are similar to other dental hygiene students in their educational setting. This study also takes place during a period of changes in accreditation guidelines in dental hygiene programs emphasizing ethical proficiency among students (CODA, 2016) and the need for a variety of effective instructional design in teaching ethics (Kacrik et al., 2006). This study can be used as a model for alternative methods for teaching ethics since the research took place in a natural setting, again making for a wider applicability to other settings.

This study suffered from power failure due to small sample sizes. With dental hygiene cohorts consisting of 17 to 33 students, it is difficult to design studies that meet sufficient power. For results to be meaningful, reproducibility of this study and other dental hygiene studies is a key priority along with attention to methodological principles.
With so few dental hygiene studies as experimental in design, it is important to collaborate with other researchers to combine data to increase the total sample size and thus, power.

Additional limitations stem from the ability of the study to relate problem solving skills to critical thinking skills. In this study, I was unable to gather information of students’ cognitive processes in how they critically analyzed the issue, gathered data, evaluated the information, and synthesized the information to make an ethical decision, which are important features of critical thinking theory (Bailin, 2002). For example, I collected and evaluated students’ responses to the problems which provided a glimpse of the cognitive processes, but I was unable to observe students’ processes of negotiating through dialogue and group interaction.

Open-ended questions like in this study require multiple discussions for the development of moral reasoning and critical thinking, especially in teaching for transfer. This study had a short intervention period of three weeks, which may not be enough time to affect and observe any real change. Finding time for the intervention was challenging, especially when the ethics course was eight weeks long instead of an entire 16 week semester. Most PF studies were approximately three weeks to 11 weeks in length, whereas DIT studies usually take the full run of the course in one 16-week semester. Dental hygiene programs have content-heavy curriculum for which students lack the time to fully engage in ill-structured problems without feeling stressed and frustrated. In this study, none of the instructors described the students in that way.
Finally, this study was limited by the problem design available to me. When this research began, I used the one method of designing ill-structured case studies that was familiar to me and what I had researched at the time. Since then, other methods of designing and constructing ethical dilemmas have been developed that can be used to integrate several areas of topics and clinical knowledge across the dental hygiene curriculum (London, Kerber, Beemsterboer, & Garetto, 2017). Nevertheless, this research contributes toward modeling how others can implement productive failure to study open-ended questions and analyze results to both groups and individuals.

Conclusion

Given the appropriate condition for which students can explore, generate, and create solutions to complex novel problems with no guidance, deeper understanding of concepts can occur. This study extends the literature on productive failure by using open-ended complex problems that have multiple solutions. Although the outcomes measure was performance-oriented from pretest and posttest scores, N2 demonstrated that cognitive processes made in productive failure condition were significant compared to lecture and practice condition where minimal cognitive change occurred. Lecture and practice may lead to the problem goal, but productive failure design has the potential to lead to deeper understanding when students analyze their own failure and then use the learned concept to build upon their prior knowledge (Kapur, 2016). Previous studies have shown that students’ approach to solving complex problems improve when prior knowledge and gaps in knowledge are first understood. The first phase of productive failure is designed to activate prior knowledge and the consolidation phase is designed to
help students’ assemble new knowledge (Kapur, 2016). This study reported here reinforces that there is a need to explore various instructional approaches that are effective in teaching moral reasoning. Productive failure can be considered as one instructional design that can produce deeper understanding of learned concepts as well as attend to learning goals of complex open-ended problems.
CONCLUSION

In this dissertation, I set out to examine productive failure design to promote problem solving skills. Specifically, I examined to what extent dental hygiene students learn how to solve moral dilemmas in productive failure compared to lecture and practice and to what extent does instruction in productive failure help students learn skills in transferable problem solving. The findings in the pilot and full-scale studies suggest that, when productive failure was applied, it had the potential to be fairly effective. The goal of these papers was to examine productive failure as an alternative teaching method to traditional lecture and practice on open-ended problems. As such, these studies reveal evidence of instructional effectiveness and transfer using productive failure that is at times easy to miss. In fact, a closer look at students’ approaches revealed small to moderate effects in students’ shifting from self-interest to a post-conventional approach. Researchers that used DIT to measure the effectiveness of ethics instruction in science programs found similar effect sizes of .36 to PF schools A ($d = .38$) and B ($d = .42$) when case-based learning and highly interactive activities were implemented (Antes, Murphy, Waples, Mumford, Brown, Connelly, & Devenport, 2009).

Instructional effectiveness may be explained by how productive failure better prepared students to solve the transfer problem from the experiences generated by their previous experience in solving the practice problem (Bransford & Schwartz, 1999). Bransford and Schwartz (1999) explain that preparing students for future learning largely depends on how they learn from their new experiences in the way they think, perceive, and judge, which influences their learning goal for researching the learned concept. In other words, the practice problem on resource allocation was used as the preparation for
how students thought about the problem, perceived the issue, and judged what solutions might work in the moral dilemma that would later be used in the transfer problem, a different problem of the same concept of resource allocation. Moreover, the experiences gained in solving the practice problem now become a part of or an addition to the prior knowledge schema of what students know and understand for future learning. This was evident in the students’ written responses to the transfer problem. Students in PF Schools A and B had learned key elements of resource allocation that helped shape their approach in solving the transfer problem and therefore, showed positive gains from practice problem to transfer problem. This was compared to the PF School C and LP School that saw a decrease in scores, but more so with LP School. PF School C and LP did little to prepare students for subsequent problems. Here, the students tended to focus on obvious features of moral values like beneficence and autonomy without raising the issue of resource allocation. Indeed, for PF School C, the results may have also been influenced by the instructor’s teaching experience.

As seen in the pilot and full-scale studies, the role of the teacher is a key component in productive failure. An effective teacher helps students examine their failed solutions and detect gaps of knowledge. The teacher is the expert by sharing multiple solutions to the targeted concept with the student and making the necessary connection from the goal of conceptual understanding. Without the teacher’s guidance, the students are discovering the problem and solution on their own. Thus, the instructional design that students experienced unintentionally in PF School C illustrates a learning condition similar to discovery learning. According to Kapur (2016), discovery learning is a design that leads to “unproductive failure” learning in which the condition does not maximize
student performance and learning because students are typically unsuccessful in solving problems and take significantly more time to solve because there was unguided support in student learning. Kapur (2016) explains that students in unproductive failure design may not know what prior knowledge they have and therefore, are less likely to build upon their existing prior knowledge with new learned experiences. This was evident with students in PF School C in the DIT-2 N2 results that showed minimal movement towards a more sophisticated thinking and in the written responses to problems that showed little to no increases in scores.

**Practical Significance**

Productive failure structure can fit within an existing educational structure with minimal changes to school curriculum, teacher training, and technological infrastructure. Implementation can be done within a relatively short timeframe which does speak well of productive failure design’s practical significance (Kapur & Kinzer, 2009).

It was evident that the teacher’s experience of the subject and in educational methodology contributed to results in PF School C. Professional development or even collaboration with other teachers in the same discipline may need to be available for teachers to learn how to design and enact specific structured problems. The absence of such support may impact the effectiveness of the instructional design making the activity disconnected from the goal of conceptual understanding. Productive failure design will require training to learn how to provide affective scaffolding without revealing content. Moreover, teachers may have to teach learners various strategies on how to work collaboratively for productive learning. Although collaborative group work is not new in educational settings there may be scaffolding techniques that work better
than others in productive failure. Teachers using productive failure instructional designs would need to be familiar with the student’s prior knowledge and the learner’s Zone of Proximal Development. Also, if the teacher does not possess conceptual understanding of the subject matter, then the student may not fully understand and learn the key elements that are necessary for the preparation for future learning.

**Future Studies**

For future studies, the field should continue to produce small scale studies like this study to generalize the effectiveness of productive failure. Additionally, future studies should extend this work to dental students with other ethical dilemma designs approaches. Also, exploring and examining instructors’ DIT-2 scores may help us understand how teachers approach and negotiate ethical dilemmas which can affect the depth to which students develop moral reasoning. Lecture and practice and productive failure have shown benefits in student learning depending on the learning goal at hand. However, this study’s results suggest that productive failure can be effective for deeper conceptual understanding and the transfer of knowledge to novel problems.

With changes constantly being made to curriculum, we should be teaching students how to problem solve so that they are better prepared to problem solve the unknown. It is not clear what jobs will look like in the future. Rather than teach to skills that may become obsolete over time, teachers should teach students how to be better problem solvers no matter what problems the future holds.
APPENDIX A

Participant Consent Form

Date
Principal Investigator: Karen Tam
Title: Examining Productive Failure Instruction as an Effective Teaching Method in Dental Ethics
Department: University of Arizona, College of Education, Educational Psychology

Dear Participant:

You are invited to participate in a research study. This consent form will give you information about the study so you can make an informed decision about participation in this research. Participation is voluntary and in no way will participation in this study impact your grades. You can withdraw and stop participating in the study at any time you wish and you will not be penalized in any way. This consent form will give you the information you will need to understand why this study is being done. I encourage you to take some time to think this over and ask questions now and at any other time. If you decide to participate, you will be asked to sign this form and you will be given a copy for your records. Participants must be at least 18 years old and a dental hygiene student enrolled in a community college dental hygiene program.

The purpose of this research study is to examine if productive failure instruction is effective as a teaching method. This study will examine the effects of productive failure instruction. The study will take place at your institution during an ethics class over a three week period of 50 minutes per class time. There are two parts to the research study. In the first part you will be asked to take the Defining Issues Test-2 (DIT-2) online as a pretest before the intervention begins to measure of prior knowledge and your approach to reasoning about ethical issues. During the intervention, you will work in groups to solve an ethical problem. In the second part, you will be asked to take a post-test individually as well as the same DIT-2 online.

You may not receive any benefits from participating, but the results from this study will help us understand how different teaching methods affect learning. I believe there are no known risks associated with this research study; however, a possible inconvenience may be the time it takes to complete the study.

To protect the confidentiality of your study records, I will keep all study records, including any codes to your data, in a locked file cabinet in my work office. Research records will be labeled with a code. A master key that links names and codes will be maintained in a separate and secure location in my work office to ensure security and confidentiality. The documentation will be shredded 6 years after the close of the study. All electronic files containing identifiable information will be password protected. Any computer hosting such files will also have password protection to prevent access by
unauthorized users. At the conclusion of this study, I may publish my findings. Information will be presented in summary format and you will not be identified in any publications or presentations.

I will be happy to answer any question you have about this study. If you have further questions about this project or if you have a research-related problem, you may contact me at 520-206-3102. If you have any questions concerning your rights as a research participant, you may contact the University of Arizona Institutional Review Board Office at (520) 891-0123 or Lori Thorpe, IRB Coordinator, Maricopa County Community College District at (480) 731-8701, lori.thorpe@domail.maricopa.edu or IRB_office@domail.maricopa.edu.

Thank you,

Karen Tam

Consent of Participant
By signing below I indicate that I have read and, to the best of my knowledge, understand the details contained in this document and have been given a copy.

____________________________________________
Print Name:

____________________________________________   ______________________________________
Signature of Person               Date:
APPENDIX B

Problem Evaluation Checklist

<table>
<thead>
<tr>
<th>Problem Evaluation Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date: ______________________</td>
</tr>
<tr>
<td>Student ID: ____________</td>
</tr>
<tr>
<td>Facilitator's Name: ____________________</td>
</tr>
</tbody>
</table>

Instructions: Read each case study and the responses. Use the below criteria to evaluate and score the quality of the answer. Case 1 indicates the pre-study and Case 2 is the post-study.

1. **Describes the ethical issues.** 5 pts
   - **Case # 1**
   - 5 = Describes all pertinent issues; assumptions made are adequately discussed; justified to support the decision
   - 4 = Identifies the ethical issues with some pertinent facts or evidence to justify and support the decision
   - 3 = Identifies the ethical issues to justify and support the decision
   - 2 = suggests only the most obvious issue
   - 1 = Does not identify the ethical issue and does not make a decision

2. **Identifies the consequences of acting** 4 pts
   - 4 = Foresees several possible outcomes; assumptions made are adequately discussed; justified to support the decision
   - 3 = Foresees few possible outcomes; assumptions made are somewhat discussed with little justification to support decision
   - 2 = sees only the obvious outcomes
   - 1 = does not identifies any outcomes

3. **Identifies the interested parties** 3 pts
   - 3 = Names all those directly affected; assumptions made are adequately discussed with viewpoints of key stakeholders; justified to support the decision using their value system
   - 2 = Names all directly affected; assumptions are discussed with little to no mention of viewpoints of key parties.
   - 1 = Suggests only the person most affected

4. **Identifies the relevant obligations** 5 pts
   - 5 = Describes obligation in terms of their basis in moral considerations; assumptions made are adequately discussed; justified to support the decision
   - 4 = Describes obligation with some basis in moral considerations; some assumptions are made and discussed, some justifications to support the decision.
   - 3 = Describes obligation with few basis in moral considerations; few assumptions are made and discussed, few justifications to support the decision.
   - 2 = Describes obligation with little basis in moral considerations; some assumptions are made and discussed, little to no justification to support the decision.
   - 1 = Simply alludes to obligations without rationale

Total / 17
APPENDIX C

Ill-Structured Practice Problem

Bill, a dental hygiene student is in the middle of his 3rd semester of clinic and on rotation at the Ida Kando public health center. In reviewing his schedule, Bill notices he is assigned to Samantha, the same patient he saw 7 months ago. At that appointment, Bill completed the health history and dental and periodontal assessments. He also learned that Samantha is 22 years old, the oldest of four children, a little scared of getting her teeth “cleaned”, but likes her smile. Her dad works 55-60 hours at a manufacturing plant while her mom stays at home caring for their four children as well as an elderly parent. Samantha shared with Bill that they cannot afford to seek dental care for the rest of the children and they will have to wait until next year to receive a dental prophylaxis.

Samantha’s mom is demanding and blunt with her wishes. Bill’s instructor is Ms. Field.

Today, Bill reviewed Samantha’s treatment plan for the dense amounts of subgingival calculus. However, Bill is thinking he needs two more quadrants to meet this semester’s course requirements, but also knows of another student who has similar needs. He checks in with the instructor.

Bill: Ms. Field, I’d like to review Samantha’s treatment plan with you, please.

Ms. Field: Alright. What would you like to do today?

Bill: Well, at the last appointment I completed her oral assessment and found dense

calculus everywhere, moderate gingival bleeding, and infections around #2, #3, and #30. I usually like to start with my hand instruments and maybe I’ll follow-up with my powered instrument to lavage the area.

Ms. Field: How about using your powered instrument before you pick up your hand

instrument?

Bill: Well (with a nervous look on his face), yes, I guess I can do that.

Ms. Field: Good. What do you think you can accomplish at this appointment?

Bill: I think I can finish one quadrant today.

Ms. Field: Then go ahead and proceed. Let me know if you need help.
Bill picks up the powered instrument and attempts to use it as best as he can. He barely passed his lab skills test last semester after repeating the test 2 times. This semester, he has managed to get away with not using it on anyone. Bill doesn’t think Samantha will mind since she has no access to care and whatever Bill does will be better than nothing. Ten minutes into the appointment, the mother demands that Bill hurry up with the treatment and complete as much as he can before picking up the other kids from school. Bill understands the mother’s request, but feels he needs help from Ms. Field with the powered instrument. Instead, he quickly finishes the treatment completing two quadrants before dismissing the patient. Bill believes he benefitted from the experience and thereby more directly met the needs of the patient.

Describe and explain the ethical dilemma and who are involved in this situation?

What, if any, consequences and assumptions can be made in how you approach this dilemma?

What moral obligations are presented in this situation? Explain.
APPENDIX D

Well-Structured Practice Problem

Bill, a dental hygiene student is in the middle of his 3rd semester of clinic and on rotation at the Ida Kando public health center. He has a patient who presents with heavy amounts of subgingival calculus. The patient’s mother is in a hurry and demands that Bill hurry to complete as much as he can before she picks up the other kids from school. He and Ms. Field, the dental hygiene faculty, decide to use a powered instrument. However, Bill does not disclose to Ms. Field that he doesn’t feel ready to use the powered instrument for fear it might slow down the appointment. He decides to use the powered instrument as best as he can without the faculty’s help. He quickly finishes two quadrants before dismissing the patient. Bill believes he benefitted from the experience because he felt an obligation to himself to develop his clinical skills as best he could and thereby directly met the needs of the patient. He believes he didn’t inflict any harm to the patient, but feels he has a duty to the school to be honest and forthcoming in his overall skill level with the powered instrument. Finally, Bill feels an obligation to himself to improve his skills to be competent so he can better serve the Ida Kando center’s patients and the community.

Describe and explain the ethical dilemma and who are involved in this situation?

What, if any, consequences and assumptions can be made in how you approach this dilemma?

What moral obligations are presented in this situation? Explain.
A group of dental hygienists wants to extend their help to children in poor rural areas to receive sufficient preventive care. The group considers sending three dental hygienists each week to various remote areas to place dental sealants and fluoride varnish, but the funds that were donated by Dental Care for Children will not cover all the expenses of the project. Some suggest transferring part of the fund intended for serving poor rural children to be used to serve poor urban children. Their argument is that more children will be able to receive treatment with the savings from travel and time costs.

Describe and explain the ethical dilemma and who are involved in this situation?

What, if any, consequences and assumptions can be made in how you approach this dilemma?

What moral obligations are presented in this situation? Explain.
In the case of Bill, the student dental hygienist, personal and professional conflicts are raised. Bill is confronted with personal conflict when he has to weigh the benefits for himself and patient versus the potential harm of not delivering competent care and to the school and community. The misconduct of this student was his inability to perform the services needed for the patient. His performance may be due to inadequate training, lack of skill and / or experience, or all of the above. Ethical dilemmas materialize with many students because they miscalculate the level of their overall skills and the ability to apply the skills and technique.

1. **Describes the ethical issues.** – (obvious) student needs over the patient needs -substandard of care because patient is low income. There is a conflict of rights AND obligations.
   a) (obvious) Conflicting rights between Bill’s right to learn (Bill has been admitted to DH program and therefore, has a right to learn such as “practice” on patients, but with appropriate safeguards to ensure patient’s right to competent care are not compromised) and rights of the patients to competent care (not substandard care)- patient has the right to know of Bill’s inexperience and substandard preclinical performance.
   b) (Not obvious) Conflicting rights between Bill’s right to learn and Ida Kando’s right to competent help and to know the truth about Mark’s competence- faculty and external clinic has the right to know of Bill’s inexperience and substandard preclinical performance. This is an issue of truth telling.
   c) Conflicting rights between Bill’s right to learn and rights of school to have competent students representing them and their right to have students live up to the regulations the school has established for governing admission to practice in the clinic. School’s right to it reputation. The student represents the DH program and to be accurately represented by students in good standing. School could be harmed by students EVEN THOUGH the school/ faculty failed to check carefully. This does not exempt Bill from his duty to use the ethical principles as guides to action.
   d) Conflicting obligations between Bill’s obligation for service to the less fortunate, and for that reason he ought to do whatever he can do be a volunteer at the offsite.
   e) Bill has an obligation to improve himself to help himself develop the skills needed to pass the course. If he does not develop his skills well enough he may not be permitted to move into the next course. Although he has an obligation to serve society, his duties to develop his competence so he can be of service seem to take precedence.

2. **Identifies the consequences of acting (legal/ moral)**
   a. Positive- Helping the poor family; Before applying the principle of beneficence, Bill must first apply the principle of nonmalefience (to do no harm).
b. Positive-Benefits the patient due to low cost of treatment

c. Negative- potential harm to patient/unintended infection due to inadequate skills;

d. Negative – to faculty and public center. Bill may have forgotten that he is practicing under the license of the DH faulty. If the student makes a mistake resulting in malpractice action, it is the supervisor who would be liable. Some may say that the faculty should have investigated Bill’s competency. It is reminded that faculty may “expect” the student will monitor himself and not present himself for an assignment when he has not met the standards set by the program. When does the student become a “self-regulating” professional?

e. Loss of trust from faculty and societal trust of patients and possibility of the dental hygiene program and school / its reputation, and the profession as well. It is tempting for Bill to think that “the highest quality of service” is reserved for “those who pay the most”.

f. Cementing the patient experience related to dental fear/ anxiety,

g. Cement poor dental experience,

h. Violation of profession’s code of ethics

3. Identifies the interested parties –
   a. Bill, faculty
   b. Patient’s mother
   c. Patient
   d. Ida Kando public health center
   e. DH program/school
   f. Society
   g. Profession

4. Identifies the relevant obligations (we each have)=
   a. Do no harm
   b. Place patient’s rights over self
   c. Practice only when competent
   d. To improve his own clinic skills
   e. Learn in a supervised setting
      Follow implied rules of the beneficence, nonmaleficence, justice.
Due to limited resources, the DH group is faced with limiting their prevention visits to a few rural children. This raises concerns about ethical responsibility among dental hygienists in the group, community, and to the organization, Dental Care for Children.

1. Ethical Dilemma—obvious:
   a. Resource allocation; Fairness to both groups through the lens of feasibility: conflict between treating fewer children (rural) and more children (urban).
   b. Conflict between resources to rural children initially and urban children.
   c. **Less obvious**: Conflict between those who are responsible in keeping with the values and interest of the donors for children dental care and the dental hygienists whose values are all children are equal in needs regardless of where they live (rural vs urban).
   d. Conflict of concern for equality in the worst off (rural children) by making the better-off members worse off. Instead, a priority view looks at which group is the worst off and in this case, the rural children because they have fewer resources (fewer available dental providers, fewer treatments, etc)

2. ID parties:
   a. Dental Hygienists
   b. Profession
   c. poor rural children
   d. poor urban children
   e. donors who funded the project

3. Consequences—
   a. Dental needs for either group could get more serious and / or are not met
   b. Misleading those who donated money
   c. Harm to the profession’s reputation. Damage to the reputation could affect profession’s ability to maintain trust among the community to those they serve
   d. inequity of service/ treatment; inadequate preventive dental care if only dental sealants or fluoride varnish, distrust between providers on which group is more important
   e. loss of DH volunteerism because of conflict between which group needs are important/ or take precedence

4. Moral obligations –
   a. To respect other RDH’s choice of which group is important
   b. To fulfill obligations to the rural kids
c. To educate the public of urban kids’ dental needs

d. To help urban kids too

e. Beneficence, to do as much as you can with what you have.

f. Involve Dental Care for Children organization in the decision-making process. Use ADPIE to help with resource allocation decisions

g. Explain to the community
**APPENDIX H**

Training Slide Presentation

**Tam’s Study**

Karen K. Tam

**Study**

- Dental Ethics courses-Fall Semester
  - 2nd Year Dental Hygiene Students
  - 1st Year Dental Hygiene Students (Preclinical)
- Will have knowledge of ...
  - Core values
  - Autonomy and respect for human beings
  - Confidentiality
  - Sensitive trust
  - Non-malfeasance
  - Beneficence
  - Justice and fairness
  - Veracity
  
  (ACSHA Bylaws and Code of Ethics, 2014, p. 20)

**Purpose**

The purpose of this research study is to examine if productive failure instruction is effective as a teaching method.

**Procedure**

1. The study will take place during an ethics class over a three week period of 50 minutes per class time.
2. Each institution will be assigned to teach the lecture either before or after an activity which will be determined by me.
3. Each institution will receive instruction from the me via email as to which procedure to implement along with specific directions on how to proceed.
4. Pre-assign group in triads or dyads to solve an ethical problem.

**Procedure**

1. Take the Defining Issues Test (DIT-2) online one week before the intervention begins as a pretest to measure prior knowledge and level of moral development.
2. Week 1:
   - Intervention begins
3. Week 2:
   - Intervention continues

The intervention will be one ethical problem and lecture presentation developed by the instructors and researcher. Pre-assigned groups of students in triads or dyads will solve an ethical problem.
Procedure
4. Week 3: Students will work on another ethical problem individually.

5. Students will take the same DIT-2 online the following week. Please assign one day for all students to take the online DIT-2. I have found this needs to be mandatory to ensure 100% completion from those who participated.

Procedure
• Please scan and email me all problems and answers to me as one document. Please make sure the names of the students are on the worksheets/answer sheets.

• DIT-2 will be scored by the Center for the Study of Ethical Development at the University of Alabama.

• All results will be scored by me and another colleague.

• Short survey to all instructors and students for feedback on the teaching method.

Questions?

I believe there are no known risks associated with this research study; however, a possible inconvenience may be the time it takes to complete the study.

If a student chooses not to participate, the individual may continue with the procedure, but his or her work will not be included into the study’s results.

TO DO
1. Preassign groups—send me an email describing how the assigned group members will be decided by the instructor on how well students work together and their social dynamics. Heterogeneous in terms of ability.

If groups are preassigned using an assessment please include it in your description.

2. Which week will the study take place for each institution? What time is the course?
APPENDIX I

Productive Failure Lesson

Day 1: 50 minutes.
The students will work in the preselected groups created by you of dyads (2) or triads (3) on their own for the first period to work to solve the problem. Please provide blank sheets of paper for the students to write their explanation, critique, and elaboration as they engage with the problem. As the students work through the problem, the instructor should refrain from making suggestions or leading the students toward a correct method. Instead, when the groups appear frustrated or ask questions, the instructor will use phrases like “Keep working hard on the problem”; “Continue your effort in solving the problem; “Keep stretching yourself”; and “Let’s see what you can come up with on your own” to encourage them. After 50 minutes, collect the worksheet and any additional sheets and scan them to me. I will provide the answer to the problem for Day 2.

Day 2: 50-minutes.
Have students get into their groups from Day 1. Hand back their worksheet. The instructor will then teach the targeted concept, Resource Allocation, using the presentation provided by me. Encourage the students to discuss and share with the class and instructor their solutions and possibilities to the problem. Be sure to have the students explain their answers first before giving them the answer. Have other students evaluate the other group’s answer. Each group should have time to present their solutions.

Day 3: 50 minutes.
The students are to complete a second problem only this time individually. Please give them 50 minutes of class time to solve the problem.
APPENDIX J

Lecture and Practice Lesson

**Day 1: 50 minutes.**

The instructor will teach the targeted concept, Resource Allocation, using the presentation provided by me. Please allow the full 50 minutes for the presentation.

**Day 2: 50 minutes.**

Ask students to get into their preselected groups that you created. For the first 40 to 45 minutes, the students will work in groups of dyads (2) or triads (3) on their own. Please provide blank sheets of paper for the students to document their representations and methods for solving the problem. The instructor will provide appropriate assistance as students work on the problem. The instructor will prompt students with simple questions and hints to guide attention to features in the problem that are important. For the last 10 to 15 minutes, the instructor will review the answer with the class. Collect the worksheet and any additional sheets and scan them to me.

**Day 3: 50 minutes**

The students are to complete a second problem only this time individually. Please give them students 50 minutes of class time to solve the problem.
APPENDIX K

Resource Allocation Slide Presentation

**Resource Allocation**

**Definition**
A process and strategy involving decisions where scarce resources should be used in the production of goods or services.

To ensure that limited health care resources are used more effectively and efficiently.

It should be cost-effective to maximize the health benefit for the population served.

**Ethical Dilemmas**
- Arises in setting priorities among interventions and among individuals in need of care. Most acute when needs are great and resources are few.
- Dilemmas occur when two rights are in conflict with one another.
  - Patient’s right vs. practitioner’s right

**Ethical Dilemmas**
- Access to health care resources can be limited through unintended barriers, for example...
  - Long wait times for appointments
  - Limited hours of operation

**Conflicts in Decision Making and Rights for Practitioners**
- Obligated to make all reasonable attempts — avoid unnecessary expenditures
- Treatment may not be in the best interest of patients / clients
- Limited third-party coverage
- Manage behavioral and emotional factors in clients, family members, instructors, third-party payers, and governmental regulators
- Obligates ethical dental professional organization to become influential on a policy level and advocate for necessary resource allocation

**Conflicts in Decision Making and Rights for Patients**
- Limited resource on how much can be spent on oral health care
- Limited options available for delivery of dental care
- Patients may have less autonomy over who they see or how often
- Treatment may not be in the best interest of patients / clients
- Fair and just health care
- Public assistance programs have source or limited resources which impact treatment decisions
- Long wait times
- Introduction to new technology is slow

*What other conflicts can you think?*
Focus of Ethics Conflict
- Is usually on competing values of the various stakeholders.
- Internal personal conflict; conflict among professionals; between professionals and the organizations.
- Problem can arise in deciding how to rank the various competing values within the context of priorities.

Decision-Making Process / Questions to Ask
- Are the criteria clear? Are the criteria consistently applied?
- Are there different perspectives among the affected parties?
- What are the impacts of the decisions?
- Was there transparency of all processes and decision-making elements?

Coping with Ethical Challenges and Dilemmas
1. Describe the ethical issues
   - Look at the values at stake
2. Identify the consequences of acting, both short-term and long-term
   - What are the possible outcomes
   - What assumptions can be made by the action
3. Identify the interested parties
   - Who are involved
4. Identify moral obligations
REFERENCES


New York: Cambridge University Press.


Center for the Study of Ethical Development at the University of Alabama (www.centerforthestudyofethicaldevelopment.net.).


Commission on Dental Accreditation (2013). Accreditation Standards for Dental Hygiene Programs, American Dental Educators Association.

Cummings, R., Maddux,C.D., Cladianos, A., Richmond, A. (2010). The effects direct instruction in moral development theory and participation in moral dilemma


Jamison DT, Breman JG, Measham AR, et al., editors. Disease Control Priorities in
Developing Countries. 2nd edition. Washington (DC): World Bank; 2006,
Chapter 14, Brock, D.W. & Wikler, D.

Johnson, B., & Christensen, L. (2012). Educational research: Quantitative, qualitative,
and mixed approaches. Sage.

problem-solving learning outcomes. Educational Technology Research and
Development, 45(1), 65-94.


Kapur, M. (2011). A further study of productive failure in mathematical problem solving:
unpacking the design components. Instructional Science, 39(1), 561–579.


Kapur, M. (2016). Examining productive failure, productive success, unproductive
failure, and unproductive success in learning. Educational Psychologist, 51(2),
289-299.

Learning Sciences, 21(1), 45-83.

Society of Learning Sciences, 4(1), 21 46.


Taichman, R. S., Parkinson, J. W., Nelson, B. A., Nordquist, B., Ferguson-Young, D. C.,


Transforming Dental Hygiene Symposium, September 18-20, 2013, p.7.


