Wayne State University Dissertations

January 2020

Confirmative Evaluation - A New Cipp Model

Tia Lynnece Finney
Wayne State University

Follow this and additional works at: https://digitalcommons.wayne.edu/oa_dissertations

Part of the Educational Assessment, Evaluation, and Research Commons

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

by

TIA L. FINNEY

DISSERTATION

Submitted to the Graduate School of

Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2020

MAJOR: Education Evaluation and Research

Approved By:

_____

Advisor   Date

_____

_____

_____

DEDICATION

This dissertation is dedicated to my family for their undying support and willingness to allow me to miss out on some things while I pursued this degree. To my children for their love and encouragement, my sister for continuously advising me how important obtaining this degree is, to my father for his encouraging words of "when will you be done?" and to my mother (deceased) for making me promise I would finish and not give up on becoming the first Ph.D. in the family; it was her words, push and my promise to her that has even made this possible. There were many days when I wanted to just say "forget it"– but knowing the joy and pride she would have in knowing I accomplished this made me refocus and get back at it.

To ALL (just too many to name) of my friends who have played a part in encouraging me and telling me, "you got this," "WE got this", and "you will not quit"– know your encouragement and push helped me tremendously. But an extended thank you to my "Sisters" Kelly (this really is all for you), Ava, Alisia and Trenia for pushing me and I do mean PUSHING ME and not allowing me to stop! And, then there is Mr. Smith – the GREATEST ENCOURAGER and accountability partner of ALL, I owe you tremendously Delvin!

Finally, I dedicate this to every lifelong learner. There is always something to learn. And to my Ph.D. cohorts in the EER (Education, Evaluation and Research) program who have been on this journey with me, thank you for the venting sessions, the laughter, and yes, even the tears. Somehow, we made it; some faster than others, but we ALL MADE IT!

ACKNOWLEDGMENTS

A huge thanks to everyone for pushing me to finish even when I could not see the worth nor have the dedication and drive to do so.

A special heartfelt thanks to my dissertation committee: Professor Shlomo Sawilowsky, Professor Frank Castronova, Professor Ingrid Guerra-Lopez and Professor Jeffrey Pruchnic. None of this would have ever been accomplished without the direction and numerous conversations, dedication and determination by my Chair, Dr. Sawilowsky. He has given tremendous hours of insight to get me to this point and I could never be more grateful to him for his guidance during this challenging process.

A special thanks to Dr. Donyale Padgett (Dept. of Communications) – for serving as a special mentor and encouraging me even when you did not know you were!

Finally, thank-you to the Finish Line, especially Drs. Zsa Zsa Booker, Darryl Gardner & Leah Robinson. What an inspiration and support you all are for all of us attempting to complete this monstrous task. It is so much easier to say you are going to do it, than it is to accomplish and finish once in it. Thank you!!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1- INTRODUCTION

Background

For many years, the question of what factors are associated with success in medical school has interested medical educators and medical school admissions committees. In fact, some of the earliest publications on the subject appeared as early as the 1940s and 1950s. Although the lines of demarcation among categories are not always clear, predictive factors can be classified into four groups. Demographic factors refer to characteristics of the individual, variables such as age, gender, race/ethnicity, and parental educational attainment. Other factors refer to variables reflecting student experiences or choices, such as a student's college major and the academic reputation/selectivity of the student's college. Performance/aptitude factors include medical college admission test (MCAT) scores and premedical-school grade point average (GPA) (Cleland et al., 2012).

Given the consequences of not detecting problems in students who go on to either fail, drop out or qualify as problem doctors, it is of great importance for medical school teachers and advisers to know what to look out for in their medical students. Struggling students may often pass unnoticed, and continue in their studies with little guidance and feedback (Sayer et al. 2002; Cleland et al. 2005; Denison et al. 2006). When feedback is provided, it often tends to be reactive and aimed at those who have failed a summative assessment (Cleland et al. 2005). In addition, clinical and research commitments and the strain of increasing student numbers further hinder adequate detection and follow-up of students in difficulty, highlighting there is a 'human' gap in the assessment process (Rivis 1996; Challis et al. 1999; Sayer et al. 2002; Cleland et al. 2008) (Tordes, et al. 2012).

Struggling trainees often require a substantial investment of time, effort, and resources from medical educators (Tordes, et al., 2012). A current challenge involves developing effective

ways to identify struggling students and better understand the primary causal factors underlying their poor performance (Artino, et al., 2010). Identifying the potential reasons for poor performance in medical school is a key first step in developing suitable remediation plans (Artino et al., 2010).

Apart from student variables such as gender, nationality, and age, dropout may also be associated with commitment, resilience and motivation to study medicine, as well as medical school factors (e.g., entry requirements, teaching, assessment procedures, curriculum design and delivery) (Maher et al., 2013).

Remediation Efforts

Educational remediation provides a remedy to a problem or a process to correct an academic fault or deficiency (Maize, et al., 2010). Most remediation in medical schools occurs after students have completed didactic courses (Maize et al., 2010). According to the Liaison Committee on Medical Education (LCME), the governing accreditation body for Medical Schools, "each student should be evaluated early enough during a unit of study to allow time for remediation" (LCME, 2010, p. 32). The goals of remediation are to obtain understanding of struggling students, to learn early identification methods, to diagnose learning deficits, to create successful remediation strategies, and to understand remediation outcomes (Winston et al., 2013).

"Progression policies dictate when a student cannot proceed in the curriculum and must either remediate, repeat a section of the curriculum, or be removed from the program" (Maize et al., 2010, p. 2). These policies vary among colleges or schools of medicine; however, they are intended to preserve the high academic standards dictated by the profession. There are a number of remediation approaches used at Medical Schools, including individualized remediation plans, course repetition, and summer restudy programs to assist in the successful progression of students

(Maize et al., 2010). It has been suggested remediation is effective in small groups where dialogue is used for collaborative knowledge construction and social regulation. This requires facilitation by experienced teachers who understand the importance of details for both content and process, and appreciate the use of implementing accurate, timely interventions to encourage overall success for students (Watson et al., 2013).

Remediation

Remediation in medical education is considered the act of facilitating a correction for trainees who started out on the journey toward becoming a physician but have moved off course (Kalet & Chou, 2014). Remediation is important, not only for students who are having academic difficulties, but also for making sure that a competent physician work force is being developed. (Hauer et al. 2009).

Despite a growing scholarly focus on remediation reform, current evidence regarding effective and efficient remediation practice remains limited (Mendel et al., 2013). There is a need to detect and correct deficits earlier in training programs, rather than later when deficits have compounded and the stakes are higher (Winston et al., 2012). It has also been shown remediation usually works: learners who have been remediated are often indistinguishable from their non-remediated peers by the end of their training (Nielson et al., 2015). Research suggests, there can also be problems associated with remediation. For instance, an emphasis on service rather than education can exacerbate the risk of failure for struggling learners (Nielson et al., 2015). Furthermore, "many programs have difficulty placing learners on probation or dismissing them, often because of the fear of legal reprisal or faculty reluctance to judge the learners they have been mentoring. Even when faculty are willing to report struggling learners, institutional barriers and

a lack of common definitions and actions can make it difficult to identify and dismiss failing learners" (Kalet et al., 2017, p. 419).

Wayne State University School of Medicine

Wayne State University School of Medicine's (WSU SOM) remediation programs include small groups, tutoring, organized study groups, and modified curriculum programs. A modified program is designed to provide an alternative to dismissing students facing academic difficulty after they and to give unprepared students opportunities to increase their competence in science courses all while building their self-esteem and learning effective study habits and tools (Maize et. al., 2010). These added components were done in conjunction with a reduced course load. Their schedules were ultimately "maximized to strengthen their science foundation and enhance their study and time management skills" (Maize et al., 2010, p. 4).

The prototype curriculum for Medical Year 1 students (M1) and Medical Year 2 students (M2) at the WSU SOM is rigorous and demanding. Most of the students struggling academically were failing the two major foundational courses in medical school; Gross Anatomy and Histology with a marginal pass (near 70%) or failure grade. It is difficult to successfully complete the SOM curriculum without knowledge of gross anatomy and histology.

The WSU SOM offered modified curriculum programs for Year 1 medical students in an effort to catch struggling students with the purpose of offering a lesser curriculum. Students could request a modified program by contacting their counselor and asking to be modified, or their counselor could recommend a modification to their academic program. The modification is done with a student's schedule in order to reduce the academic load of the first-year curriculum. Some students find greater success if the academic load for M1 and M2 is distributed over three years versus the normal load over two years.

The modified program may be offered following the second Gross Anatomy and Histology exams based upon the students' performance on those exams. A student could accept or decline the offer, and first year students are permitted to self-select the modified program for academic or personal reasons. Students who elected to be in the modified program choose to remain enrolled in either Gross Anatomy or Histology; the curriculum for the remaining coursework is then predetermined. Students who followed the modified curriculum and successfully completed the requirements for Year 1 courses will be promoted to a full time M2 curriculum. Any student participating in the modified program was not reflective in their medical student performance evaluation letter which becomes a part of their medical school academic file and is used for graduation interviews and matching students for their residency programs after graduation.

Each year the Academic Advising Committee (AAC) determined whether the modified program will be offered for the current Academic Year (AY). Discussion surrounding the program and whether it should be continued each year is due to curriculum changes, long-term effectiveness and the overall effectiveness of the program. Ultimately the goal of the modified program is to ensure academic success for at or below marginal pass (MP) students.

CIPP Model of Evaluation

The CIPP Evaluation Model was developed by Stufflebeam (1971) as a decision-oriented approach structured to help administrators make good decisions (Worthen et al., 1997). The CIPP evaluation framework serves managers and administrators facing four different kinds of educational decisions. The first letters of each type of evaluation – context, input, process and product -- form the acronym CIPP, by which Stufflebeam's evaluation model is best known (Worthen et al., 1997):

- Context evaluations

- Input evaluations

- Process evaluations , and

- Product evaluations

"By employing these four interrelated types of evaluation, policymakers, program and project staffs, and individual service providers can conduct or contract for evaluations to help initiate, develop, and install sound programs, projects, or other services; to strengthen existing programs or services" (Stufflebeam et al., 2000, p. 279).

There are several strengths and weaknesses with the CIPP Model. Some strengths include the ease of the model to be applied to multiple evaluation situations and the model's long history of applicability. Some disadvantages include the following: the CIPP model not being widely known and applied in the performance improvement field, and how the model can offer blurred lines between evaluation and other methods such as needs assessments (Guerra-Lopez, 2008).

Confirmative Evaluation

One of the areas the CIPP Model falls short is in identifying if indeed the program evaluation can be confirmed (Powell & Conrad, 2015). Confirmative evaluation goes beyond formative and summative evaluation; it moves traditional evaluation a step closer to full-scope evaluation. It is a "continuation of summative evaluation" (Morrison et al., 2013, p. 337) and "can assist in continual improvement of course materials by determin[ing] the causes of problems and possible remedies" (Morrison, et al., 2013, p. 257). During confirmative evaluation "the evaluation, training, or practitioner collects, analyzes, and interprets data related to behavior, accomplishment, and results" (Hellebrandt & Russell, 1993, p. 22).

Role of Accreditation

National accrediting bodies such as the LCME evaluate Medical Schools on many factors to determine accreditation compliance. One of these factors include types of remediation programs developed and implemented to ensure success in academic programs. Medical schools across the country have remediation programs in place to assist struggling students and many have a version comparable to the modified curriculum program currently offered at WSU SOM.

The Wayne State University SOM implemented a new curriculum for medical year 1 (M1) effective academic year 2018-2019. Within this new curriculum, Gross Anatomy and Histology will be spread out throughout year 1 versus as it has been with the two courses being offered simultaneously at the beginning of the year 1 academic year. By making such a drastic change in the curriculum design, it has been determined the modified program will no longer be an option. Yet, it has not been addressed which other remediation program will replace it when it comes to the identified struggling students. The modified program allows a struggling student to drop either Gross Anatomy or Histology and accept an extra year to be successful in the curriculum. With these two foundational courses being spread out throughout the year 1 curriculum, there is no safety net to catch those students who are academically struggling.

Purpose of This Study

The members of the curriculum committee (course directors, faculty members and administration) recognized this is a major issue needing to be addressed prior to the 18-19 academic year and the implementation of the new curriculum (WSU SOM, Curriculum Committee, 2017). Therefore, the purpose of this study is to incorporate confirmative evaluation into the CIPP evaluation model. . Confirmative evaluation of the current modified program will give stakeholders information needed to determine if a modified program can be incorporated in

the new curriculum. This is important to offer insight in to garnering the best outcomes for modified programs.

This will be carried out by conducting a program evaluation of WSU SOM's modified programs to determine whether they should continue (in some new identified form) or if they should be terminated. For purposes of this study, an evaluation of two groups will be conducted; one using the CIPP model, and the other adding confirmative evaluation component to CIPP. The stakeholders for the program evaluation will include the AAC which is comprised of the Deans of the WSU SOM, course directors, counselors and students. There will be current M1 students who have elected to be placed in the modified program from AY 17/18 and M2 students who were in the modified program as a M1.

The intent of the evaluation is to determine the success of students to ultimately determine if the modified program should continue, or not. Groups will be evaluated on performance; grades/scores in their modified course loads and their STEP 1 scores (an examination given to year 2 students prior to promotion into year 3 which assesses their basic knowledge and foundation from the first 2 years of medical school; it is a precursor of their licensing exam). The added confirmative evaluation phase component will incorporate techniques used to capture results through questionnaires offered to participants and stakeholders.

The notion of evaluation's most important purpose is not to prove but to improve is an idea originally put forward by Egon Guba decades ago (Stufflebeam & Shinkfield, 2007). Kaufman has similarly proposed evaluation data should be used to fix rather than blame (Kaufman & Thomas, 1980). Along these lines, evaluation is simple:

- It compares accomplished results with planned and expected results.
- It can be used to find drivers and barriers to expected performance.

- It should produce actionable recommendations for improving processes, programs, and solutions allowing expected performance to be achieved or maintained (Guerra-Lopez, 2008).

Confirmative evaluation will give insight into if the current modified program has been successful and can still be successful within the new curriculum. Confirmative evaluation will allow the stakeholders to determine if the evaluation plans being put into place are indeed beneficial to the students who have selected to be modified.

Research Questions/Hypothesis

1. How effective is the modified program for student's success in the SOM?

2. Do students benefit from a modified program in medical school?

3. Will the CIPP program evaluation model become more effective by adding confirmative evaluation?

Definitions

Academic Advising Committee: The committee made of leadership at the School of Medicine which looks extensively at the success and failures of medical students. They make recommendations for remediation, modified programs, promotion hearings and dismissals of medical students.

Confirmative Evaluation: Confirmative evaluation "provides continuous quality control over the life cycle of a performance improvement package" (Moseley & Solomon, 1997, p. 12).

Education remediation: The act of providing a remedy to a problem or a process to correct an academic fault or deficiency (Maize, et al., 2010).

Evaluation: A study designed and conducted to assist some audience to assess an object's merit and worth (Stufflebeam, 2001).

Evaluation Standard: "A principle mutually agreed to by people engaged in the professional practice of evaluation, if met, will enhance the quality and fairness of an evaluation" (Joint Committee on Standards for Educational Evaluation, 1994, p. 3).

Modified Program: A remediation plan for struggling Year 1 Medical Students at Wayne State University School of Medicine reduces their curriculum load and allows them to complete the curriculum in 3 years versus the traditional 2 years at in an effort to help them achieve academic success and progression.

Program Evaluation: the systematic application of scientific methods to assess the design, implementation, improvement or outcomes of a program (Rossi & Freeman, 1993; Short, Hennessy, & Campbell, 1996). The term "program" may include any organized action such as media campaigns, service provision, educational services, public policies, research projects, etc. (Center for Disease Control and Prevention [CDC], 1999).

Sustainability: The ability to "maintain programming and its benefits over time" (Hetzel, 2015, p. 1).

Theory-based Program Evaluation: An approach to evaluation examines the theories on which the program is based, activities being conducted, the effects activities will have, and recommendations for the program's next phases (Birckmayer & Weiss, 2000).

CHAPTER 2- LITERATURE REVIEW

Program Evaluation

"Evaluations are, in a broad sense, concerned with the effectiveness of programs" (Spiel, 2001, p.1). Although common sense evaluation has a very long history, evaluation research which relies on scientific methods is a young discipline grown massively in recent years (Spiel, 2001). Patton (1987), stated "evaluation is a systematic process to understand what a program does and how well the program does it. Results can be used to maintain or improve program quality and to ensure future planning can be more evidence-based (p. 2). Guerra-Lopez (2008) argued all components of the evaluation must be aligned with those objectives and expectations that the organization values, and how the decisions will have to be made as a result of the findings. Fundamentally, these decisions are concerned with how to measurably improve performance, at all levels of the organization: internal deliverables; organizational objectives, and external impact on its customers and global environment.

In medical education, program evaluation is an essential responsibility for implementation, oversight, and improvement. Therefore, all programs require a strong evaluation plan (Frye, 2012). This includes programs "as small as an individual class session, a course, or a clerkship rotation. It may be as large as the whole of an educational program. The program can include a medical school setting, post graduate training or via continuing professional development settings" (Frye, 2012, p. 288).

Program Evaluation History and Theory

Program evaluation identifies whether the time, effort, and expense in a program is worthwhile. According to Frye & Hemmer (2012) a program may be "as small as an individual class session, a course, or a clerkship rotation in medical school or it may be as large as the whole

of an educational program. The program might be situated in a medical school, during postgraduate training, or throughout continuing professional development" (p. 289).

Frye and Hemmer (2012) stated educational programs are fundamentally about change and most persons participating in educational programs—including learners, teachers, administrators, other health professionals, and a variety of internal and external stakeholders—do so because they are interested in change. At the fundamental level, evaluation involves making a value judgment about information one has available (Cook, 2010). Thus, educational program evaluation uses information to decide about the value or worth of an educational program.

By asking questions, consulting with partners, making assessments and obtaining feedback, program managers are able to assess the value and impact of their work (Scriven, 1998). The information collected can be used to improve the overall program and these informal assessments fit nicely into a broad definition of evaluation according to Scriven (1998) as the "examination of the worth, merit, or significance of an object" (p. 129).

Program evaluation is "the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future program development" (Patton, 1997, p. 7). It does not occur in a vacuum. "It is influenced by real-world constraints. Evaluation should be practical and feasible and conducted within the confines of resources, time, and political context. Moreover, it should serve a useful purpose, be conducted in an ethical manner, and produce accurate findings" (Patton, 1997, p. 8). Evaluation findings should be used both to make decisions about program implementation and to improve program effectiveness (Centers for Disease Control and Prevention, retrieved from https://www.cdc.gov/eval/guide/introduction).

Educators may have internal and external reasons for evaluating their programs. Internal reasons may include assessment of relevance of program, assessment of changes to be made to a program and the assessment of overall effectiveness of a program. Primary external reasons are typically found in requirements of medical education accreditation organizations, i.e., the Liaison Committee on Medical Education (LCME, 2010). A strong program evaluation process supports accountability while allowing educators to gain useful knowledge about their program and sustain ongoing program development (Goldie, 2006).

According to Patton (2011), evaluation models have not always supported such a range of needs and for many years' evaluation experts focused on simply measuring program outcomes. Stufflebeam and Shinkfield (2007) argued newer evaluation models supported learning about the dynamic processes within the programs, allowing an additional focus on program improvement.

There are several theories that inform educational program evaluation models; reductionism, system theory and complexity theory. Medical education programs are affected by many factors both internal and external to the program: program participants' characteristics, influence of stakeholders or regulators, the ever-changing nature of the knowledge of which a discipline is based, professional practice patterns, and the environment in which the educational program functions to name only a few (Geyer et al., 2005). Frye and Hemmer (2010) suggested medical education programs are best characterized as complex systems, given they are made up of diverse components with interactions among those components.

Mennin (2010) stated complexity theory and complexity science attempted to embrace the richness and diversity of systems in which ambiguity and uncertainty are expected. Complexity theory can inform the choice of program evaluation models. For example, Stufflebeam and Shinkfield (2007) suggested the concept of "program elements" (p. 20). It was suggested that this

relationship is prominent in the CIPP evaluation model in which "the context studies play a critical role in shaping the approach to evaluating program effectiveness and in which program process studies are separate but of equal importance" (Frye & Hemmer, 2012, p.291).

Roles of Evaluation Theory

Knowledge of evaluation theory can help evaluators become better ambassadors for the profession of evaluation and educators of potential clients. Because professional evaluation now offers a range of acceptable approaches and perspectives, it is critical sponsors and users understand there are variations and how they differ. Evaluation approaches and services may differ rather dramatically across evaluation teams. Finding an optimal fit between an evaluation team and the needs and interests of evaluation sponsors and stakeholders could arguably be one the most important factors in determining whether an evaluation will ultimately be useful (Donaldson, 2004).

Effective evaluation practice has the potential to help prospective clients and other stakeholders dramatically improve their work. For example, professional evaluation can help stakeholders make better decisions about service, policy, and organizational direction; build knowledge, skills, and develop a capacity for evaluative thinking; facilitate continuous quality improvement and organizational learning; and provide accountability or justify a program, policy, or organization's value to investors, volunteers, staff, and prospective funders (Donaldson, 2004).

Beyond the general benefits of evaluation, however, is the question of how appropriate a particular evaluation is for a particular program at a particular time. It is important to consider who could be negatively affected by an evaluation of a given sort, how much time and resources may be taken away from program services while the evaluation is being conducted, and the ways in which the evaluation process might be uncomfortable and disruptive for some project team

members and other stakeholders (Donaldson, 2001b; Donaldson, Gooler, & Scriven, 2002). It must also be recognized the questions a particular evaluation asks and the way in which it goes about answering those questions will have repercussions and will not always be constructive. When evaluators and stakeholders fully explore the potential benefits and costs of doing a specific evaluation and consider other options and approaches (based on other theories of practice), their expectations and plans become more realistic and the evaluation is much more likely to reach its potential (see Donaldson, 2001b).

Distinguished evaluators as Scriven (1998, 2004) and Stufflebeam (2001, 2004) have asserted there is little need for theory or, at least, some forms of theory, in evaluation. Scriven (2004) claimed "it's possible to do very good program evaluation without getting into evaluation theory or program theory" (p. 29), and stated "the most popular misconception amongst currently politically correct program evaluators is the evaluation of a program (a) requires you have, or (b) is much benefited by having, a logic model or program theory" (p. 29).

Stufflebeam (2001), in a review of evaluation models and theories of evaluation practice, remarked "there really is not much to recommend theory-based evaluation, since doing it right is usually not feasible and since failed or misrepresented attempts can be counterproductive" (p. 31). More recently, Stufflebeam (2004) described the "now fashionable advocacy of 'theory-based evaluation'" (p. 31) as a situation herein one "assumes the complexity of variables and interactions involved in running a project in the complicated, sometimes chaotic conditions of the real world can be worked out and used a priori to determine the pertinent evaluation questions and variables" (Stufflebeam, 2004, p.32).

In contrast, other evaluators argued program, evaluation, and social science theory play important roles in modern program evaluation (e.g., Alkin, 2004a; Chen, 1990; Donaldson, 2003;

Fetterman, 2003; Lipsey, 1990; Mark, 2003; Rossi, Freeman, & Lipsey, 2004; Shadish, Cook, & Campbell, 2004; Weiss, 2004a, 2004b. For example, in the case of evaluation theory, Shadish (1998) asserted:

> All evaluators should know evaluation theory because it is central to our professional identity. "It is what we talk about more than anything else, it seems to give rise to our most trenchant debates, it gives us the language we use for talking to ourselves and others, and perhaps most important, it is what makes us different from other professions. Especially in the latter regards, it is in our own self-interest to be explicit about this message, and to make evaluation theory the very core of our identity" (p. 1).

Scriven (1967) introduced formative and summative roles of evaluation. These two terms were accepted amongst practitioners in the evaluation field. Formative evaluation is "conducted to provide program staff evaluative information useful in improving the program" (Worthen et al., 1997, p. 14) and summative evaluation is "conducted and made public to provide program decision makers and potential consumers with judgments about a program's worth or merit in relation to important criteria." (Worthen, et al., 1997, p. 14).

The phrase formative assessment was not used consistently in the literature (Black & William, 1998; Sebatane, 1998). Sebatane (1998) stated some saw all classroom assessment as formative and discuss summative assessments primarily in terms of external assessments, others agreed all classroom assessment can be formative, but only if students use the information for formative purposes, while others recognized some classroom assessment can serve summative purposes, too (p. 132).

Informational feedback, information students can use to improve their performances, is intrinsically motivating (Ryan et al., 1985; Ames & Archer, 1988; Covington, 1992; Pintrich & Schrauben, 1992). This is important, given the nature of the assessment process. Black and Williams (1998) defined the core of formative assessment as two actions: 1) the student

recognizing there is a gap between current understanding or skill level and 2) the desired understanding or skill level; the student taking effective action to close the gap. Sadler (1983, p. 63) articulated three steps in the formative feedback loop: (1) attending to goals, (2) devising strategies to reach them, and (3) monitoring the discrepancy between actual and desired performance.

Accurate self-assessment and appropriate use of feedback are necessary for the process. Feedback, however detailed, will not lead to improvement until a student understands both the feedback itself and how it applies to his or her work. This appraisal is a part of the learning process (Sadler, 1989; Black, 1998). Self-assessment is essential for progress as a learner: for understanding of selves as learners, for an increasingly complex understanding of tasks and learning goals, and for strategic knowledge of how to go about improving (Sadler, 1983). Learners are motivated both by intrinsic interest and by the desire to succeed at school (Ames & Archer, 1988).

Summative assessment is an "overview of previous learning" (Black, 1998, p. 28). Gipps (1994), building on the work of others (e.g., Black & William, 2003; Broadfoot, 2008; Gipps & Stobart, 1997; Stiggins, 2002), considered two different summative processes; summing up and checking up. Summing up meaning creating a picture of achievement based on accumulating assessments intended to be originally formative. Checking up means tests or tasks at the end of learning, assigned specifically to collect information for summative judgements.

The relationships between formative and summative assessments were examined in an attempt to determine if it is possible to have a summing-up process using information originally intended as formative assessment for a summative purpose (Sebatane, 1998). This might obviate the formative purpose, especially if students pay less attention to feedback and more to the grade

or score counting in the final grade, and thus learn less from the feedback than they might otherwise (Crooks, 1988; Sadler, 1989).

There is a counter-argument to this point of view. "Sensible educational models make effective use of both FA [formative assessment] and SA [summative assessment]" (Biggs, 1998, p. 105). Formative and summative assessment need not be mutually exclusive if one's model of assessment is inclusive:

> Instead of seeing FA and SA up close as two different trees, zoom to a wider angle conceptually. Then, in the broad picture of the whole teaching context— incorporating curriculum, teaching itself … and summative assessment—instead of two tree-trunks, the backside of an elephant appears (Biggs, 1998, p. 108).

Summative assessment is often assumed to have entirely negative consequences, but if it "is aligned to instruction and deeply criterion-referenced, incorporating the intended curriculum, which should be clearly salient in the perceived assessment demands" (Biggs, 1998, p. 107), then "classroom summative assessment, such as a test at the end of a teaching episode or unit, can have positive effects" (Biggs, 1998, p. 107). Black (1998) argued teachers have to be involved in both formative and summative assessment, and must keep the two in tension. Formative assessment focuses on the needs of the learner, while summative assessment focuses on the need for accountability (Black, 1998).

The basic concept of formative assessments seems simple, but can be complex, as formative assessments can be both formal and informal in nature. The underlying purpose of formative assessment is to "contribute to student learning through the provision of information about performance" (Rowntree, 1987, p. 4-5). Formal formative assessments can be defined as those to a specific curricular assessment framework. According to Rowntree (1987), they can include activities required of the student and of the assessor. When students have been surveyed

about the value they place on organized formative assessment sessions, such evidence points to an overwhelmingly positive response (Carroll, 1995; Rolfe & McPherson, 1995; Vaz et al., 1996).

Following (Bloom et al., 1971), the distinction is typically made between formative and summative assessment, the latter being concerned with determining the extent to which a student has achieved curricular objectives. As some have observed, the distinction between formative and summative assessment is however far from sharp (Yorke, 2003). Some assessments, according to Yorke (2003), (e.g., in course assignments) can be "deliberately designed to be simultaneously formative and summative – formative because the student is expected to learn from whatever feedback is provided, and summative because the grade awarded contributes to the overall grade at the end of the study unit" (p. 480). Summative assessments in relation to a curricular component (the student passes or fails a module, for example) can act formatively if the student learns from them (Yorke, 2003).

Medical School Remediation

Academic remediation is a "near universal problem faced by medical schools and residency training programs" (Bennion, et al., 2018, p.1). Studies suggest the need for remediation is significant given 10% of medical students encounter an academic failure at some point during their training (Bennion, et al., 2018). Szumacher, et al. (2007) argued academic difficulty can often be a significant problem for students in health professional programs and the magnitude of learning problems lead to remediation, and efficacy of the remedial programs must be identified. Students in difficulty are often identified late in their training and run the risk of dismissal if remediation is not successful. Such students have been described by Yates & James (2006) as "strugglers," "marginal students," and "problem learners" (p. 338). They also include students presented to an academic progress committee who have not met expectations in a course or clerkship, students

who meet some but not all requirements of a clinical evaluation, as well as students with characteristics such as poor interpersonal skills, excessive shyness, poor integration skills, and a lack of personal responsibility (Frellson et al., 2008).

Frellson et al. (2008) found as much as 15% of medical students were identified as "struggling" (p. 15), based on written examinations, clinical performance evaluations, clinical evaluations formal evaluations of professionalism, peer assessments, group reviews, or grading sessions or using a combination method such as clerkship pretest combined with marginal clinical indicators. One of the major issues with struggling students is they are most times not identified. Cox et al. (2008), suggested accurately identifying struggling medical students is important if we are to have the opportunity to remediate deficiencies and to ensure inadequately performing students do not advance to the next stage of training. According to a study conducted by Maher et al., (2013) academic struggling, decelerated curriculum (US), "failing at least one basic science course in Year 1 (US) and low Year 1 grade point averages were strongly associated with dropout. Very few dropout studies have undertaken a detailed analysis of repeat examinations and repeat years" (Maher, et al. 2013, p. 2).

Reported graduation rates for medical students historically have been very high with an ultimate graduation rate of 96 percent throughout the 1990s. However, only about 81% of medical students graduated from medical school in four years (Maher, 2013). The graduation rate for M.D. remains very high, there is a clear trend over the past 30 years showing a drop in the four-year graduation rate for single degree medical students (Maher, 2013). The complex patterns of assessment in medicine mean academically deficient students may continue with little guidance or specific educational interventions (Sayer et al., 2002). They are at risk of becoming incompetent doctors (Challis et al., 1999) whose colleagues are left with the responsibility of spotting dangers,

reporting clinical errors and bad practice to the authorities. Early support may stop students experiencing a cycle of failure. For staff, the short-term investment of time required to explore and develop systems is likely to be of benefit in the long-term (Cleland et al., 2005).

According to Hauer et al., (2009) medical educators and accrediting organizations have "shifted emphasis from what is taught in the curriculum to what a medical student, resident, or practicing physician can perform. Remediation begins with the identification of trainees or physicians in practice who fail to demonstrate competence during assessments of their skills" (p. 1822). Identification of trainees needing remediation may be easiest at the undergraduate level because the performance expectations of students are relatively homogeneous, and students are frequently tested within their schools (Hauer et. al., 2009). If "deficits go undetected or unaddressed, physician performance and patient safety are jeopardized" (Hauer, et al., 2009, p. 1823). For example, performance problems in the domains of knowledge and professionalism have been linked to subsequent disciplinary action by state medical boards (Hauer, et al., 2009).

Frellson et al. (2008) identified various remediation plans for students who failed the National Board of Medical Examiners (NBME) internal medicine subject exam. These remediation actions included "self-directed study, faculty-led tutorials with directed reading, attendance at teaching conferences, and problem-based discussion sessions" (p. 878).

The remediation process allows students to be more successful in their academic curriculum. The institution benefits as well from the student's success because the student is still enrolled, paying tuition and can contribute to the overall culture of the campus (Maher, 2013).

It was contended in the 1998 Higher Education Policy report remedial education is more cost-effective and far less expensive for society than such alternatives as unemployment, low-wage jobs, welfare participation, and incarceration (Section 131). Saxon & Boylan (2001) reviewed the

literature on the cost of remedial education. Although researchers warned against "sweeping strategic decisions about delivery, modification, or elimination of remedial education based on this data, they recognized common findings. (1) Remediation costs are in the range of 1%-2% of the overall cost of education. (2) Revenues generated by remedial course work either fully cover or exceed the cost of delivering the service. (3) There is no ongoing research tracking growth and/or cost of remedial education" (Saxon & Boylan, 2001, p. 4).

Evaluation of Remediation Programs

Remediation efforts "must be evaluated to determine whether goals are being met, and assessed to make effective decisions to optimize and improve programs. This is important because there are no validated, turnkey models for remediation" (Maize, et al., 2010, p. 22). The best assessment and evaluation plans should include an array of data drawn from both formative and summative assessments, which incorporate standardized as well as locally developed methods (Maize, et. al., 2010). Maize et. al., (2010) argued the "effectiveness of a remediation plan for colleges can be evaluated by preventative strategies to minimize the need for remediation, and remediation approaches to correct deficiencies" (Maize, et. al., 2010, p. 23).

The CIPP Model

The CIPP approach includes "four complementary sets of evaluation studies allowing evaluators to consider important but easily overlooked program dimensions" (Frye & Hemmer, 2012, p. 296). CIPP components accommodate the ever-changing nature of most educational programs as well as educators' appetite for program-improvement data. By alternately focusing on program Context, Inputs, Process, and Products (CIPP), the CIPP model addresses all phases of an education program: planning, implementation, and a summative or final retrospective assessment if desired (Frye & Hemmer, 2012). The first three elements of the CIPP model are

useful for improvement-focused (formative) evaluation studies, while the Product approach, the fourth element, is very appropriate for summative (final) studies (Alqahtani, 2016).

Educational evaluation, "a family of approaches to evaluating educational programs. The following discussion of selected evaluation models places them in relationship to the theoretical constructs that informed their development" (Frye & Hemmer, 2012, p. 292). Thoughtful selection of a specific evaluation model allows educators to structure their planning and to assure important information is not overlooked (Frye & Hemmer, 2012).

Development of the revised CIPP Model

The CIPP model is comprised of four major components:

1. Context

2. Input

3. Process

4. Product

Context intends to answer the question such as what needs to be done versus were important needs addressed? The Input addresses how it should be done versus if a defensible design was employed. The Process considers if it is being done versus if the design was well executed. The Product addresses if the program is succeeding versus if the effort succeeded.

The CIPP Model is a "comprehensive framework for guiding formative and summative evaluations of projects, programs, personnel, products, institutions, and systems. The model is configured for use in internal evaluations conducted by an organization's evaluators, self-evaluations conducted by project teams or individual service providers, and contracted or mandated external evaluations. The model has been employed throughout the U.S. and around the

world in short-term and long-term investigations—both small and large" (Stufflebeam, 2003, p. 2). Applications have spanned various disciplines and service areas, including education, housing and community development, transportation safety, and military personnel review systems (Stufflebeam, 2003).

Stufflebeam (2003) explained "the model's core concepts are denoted by the acronym CIPP, which stands for evaluations of an entity's context, inputs, processes, and products. Context evaluations assess needs, problems, assets, and opportunities to help decision makers define goals and priorities and help the broader three groups of users judge goals, priorities, and outcomes" (p. 2-3).

A CIPP context evaluation study is typically conducted when a new program is being planned but context studies can also be conducted when decisions about cutting existing programs are necessary (Frye & Hemmer, 2012). A CIPP model input evaluation study is "useful when resource allocation (e.g., staff, budget, time) is part of planning an educational program or writing an educational proposal. When applied to a program already in place, an input evaluation study can help the educator to assess current educational practices against other potential practices" (Frye & Hemmer, 2012, p. 296-297). Its focus on feasibility and effectiveness allows a developing program to remain sensitive to the practices most likely to work well (Frye & Hemmer, 2012).

A CIPP process evaluation study is typically used "to assess a program's implementation. For programs operating in the complex environment typical of medical education programs, this attention to process issues allows an ongoing data flow useful for program management and ongoing effective change. This kind of evaluation study can also be conducted after a program concludes to help the educator understand how the program actually worked. A CIPP process study explicitly recognizes an educational model or program adopted from one site can rarely be

implemented with fidelity in a new site" (Aziz, et al., 2018, p. 193-194). This is because contextual differences usually dictate minor to major adaptations to assure effectiveness (Prasetyono, 2016).

Educators may seem familiar with the CIPP Model because of how it focuses on the outcomes of a program. Zhang, et. al., (2011), stated this type of evaluation study aims to identify and assess the program outcomes, including both positive and negative outcomes, intended and unintended outcomes, short-term and long-term outcomes. It also "assesses, where relevant, the impact, the effectiveness, the sustainability of the program and/or its outcomes, and the transportability of the program. Finally, a CIPP model product evaluation study also examines the degree to which the targeted educational needs were met" (Zhang, et al., 2011, p. 58-59).

The bases for judging CIPP evaluations are pertinent professional standards, including the Joint Committee (1988, 1994, 2003) standards for evaluations of personnel, programs, and students. These standards require evaluations to "meet conditions of utility (serving the information needs of intended users), feasibility (keeping evaluation operations realistic, prudent, viable, and frugal), propriety (conducting evaluations legally, ethically, and with due regard for the welfare of participants and those affected by results), and accuracy (revealing and conveying technically sound information about the features that determine merit, worth, probity, and/or significance)" (Joint Committee on Standards for Educational Evaluation, 1994, p. 282).

The CIPP approach "consists of four complementary sets of evaluation studies which allow evaluators to consider important but easily overlooked program dimensions" (Frye & Hemmer, 2012, p. 296). According to Frye & Hemmer (2012), the CIPP components "accommodated the ever-changing nature of most educational programs as well as educators' appetite for program-improvement data. They stated the CIPP model addresses all phases of an education program: planning, implementation, and a summative or final retrospective assessment, if desired" (p. 296).

The first three elements of the CIPP model are useful for improvement-focused (formative) evaluation studies, while the Product approach, the fourth element, is very appropriate for summative (final) studies (Frye & Hemmer, 2010).

According to Aziz, et al., (2018), the CIPP model studies can "be used both formatively (during program's processes) and summatively (retrospectively)" (p. 192). Stufflebeam's CIPP model is consistent with system theory and, to some degree, with complexity theory it is flexible enough to incorporate the studies and support ongoing program improvement. (Zhang, et al, 2011).

Summarized in Table 1 (Stufflebeam, 2003) are uses of the CIPP Model for both formative and summative evaluations. "The matrix's eight cells encompass much of the evaluative information required to guide enterprises and produce summative evaluation reports" (Stufflebeam, 2003, p. 6). Beyond context, input, process, and product evaluations—set in both formative and summative contexts (Stufflebeam, 2003).

Table 1. The Relevance of Four Evaluation Types to Formative and Summative Evaluation Roles

EVALUATION ROLES	*Context*	*Input*	*Process*	*Product*
Formative Evaluation: Prospective application of CIPP information to assist decision making and quality assurance	Guidance for identifying needed interventions and choosing and ranking goals (based on assessing needs, problems, assets, and opportunities).	Guidance for choosing a program or other strategy (based on assessing alternative strategies and resource allocation plans) followed by examination of the work plan.	Guidance for implementing the work plan (based on monitoring and judging activities and periodic evaluative feedback).	Guidance for continuing, modifying, adopting, or terminating the effort (based on assessing outcomes and side effects).
Summative Evaluation: Retrospective use of CIPP information to sum up the program's merit, worth, probity, and significance	Comparison of goals and priorities to assessed needs, problems, assets, and opportunities.	Comparison of the program's strategy, design, and budget to those of critical competitors and to the targeted needs of beneficiaries.	Full description of the actual process and record of costs. Comparison of the designed and actual processes and costs.	Comparison of outcomes and side effects to targeted needs and, as feasible, to results of competitive programs. Interpretation of results against the effort's assessed context, inputs, and processes.

Note: Excerpted from Stufflebeam, 2003, p. 6.

Stufflebeam (2003) stated, all evaluations have the potential for the emergence of misunderstandings and disputes concerning a wide range of matters. Some of these concerns can include who is allowed to edit and release report findings, the agreed upon deliverables, how much money will be spent on tasks and deliverable due dates. These are things that should be contractually agreed upon by all stakeholders prior to beginning an evaluation (Stufflebeam, 2003). Given in Table 2 (Stufflebeam, 2003) are illustrations of methods of potential use in CIPP Evaluations.

Table 2. Illustration of Methods of Potential Use in CIPP Evaluations

METHODS	Context	Input	Process	Impact	Effectiveness	Sustainability	Transportability
Survey	X		X	X	X	X	
Literature Review	X	X					
Document Review	X	X	X	X	X		
Visits to Other Programs		X		X	X		X
Advocate Teams (to create & assess competing action plans)		X					
Delphi Technique	X	X					
Program Profile/Database		X	X	X	X	X	
On-Site Observer			X	X	X	X	
Case Studies			X	X	X	X	
Comparative/ Experimental Design Studies		X			X	X	
Stakeholder Interviews	X		X	X	X	X	X
Focus Groups	X	X	X	X	X	X	X
Hearings	X	X			X		
Cost Analysis		X	X		X	X	
Secondary Data Analysis	X				X		
Goal-Free Evaluation			X	X	X	X	X
Photographic Record	X		X	X	X	X	X
Task Reports/Feedback Meetings	X	X	X	X	X	X	X
Synthesis/Final Report	X	X	X	X	X	X	X

Note: Excerpted from Stufflebeam, 2003, p. 16.

CIPP Model Strengths and Limitations

Guerra-Lopez (2008) identified some of the strengths and limitations of the CIPP Model. Some of the strengths identified included the model being well established and having a long history of applicability, its comprehensive approach of evaluation can be applied from program planning to outcomes and fulfillment of core values and because the model was not designed with any specific program or solution it can be easily applied to multiple evaluation situations. Worthen and Sanders (1987) indicated this approach is designed to serve the needs of management in the decision-making process and this approach takes advantage of opportunities as they arise and allows management to make informed decisions need to be made (Barrett, 1998). The CIPP, is widely used for educational programs or projects in many fields for not only accountability but

also improvements. The whole CIPP model is suitable for universities which are under accreditation because the model provides chances for evaluators to assess not only programs' implementation but also universities as the system (Vol, 2018).

According to Guerra-Lopez (2008), the limitations included the model not being as widely known and applied in the performance improvement field as other models and it could be said to blur the line between evaluation and other investigative processes such as needs assessment. This type of management-oriented evaluation model is limited, because it serves the needs of the decision maker and may restrict or impede the evaluator's exploration of other issues which arise through the course of the evaluation. Although of these potential issues may be important, they may be overlooked in favor of complying with the objectives and directions of the decision makers (Barrett, 1998).

This type of evaluation may also be subject to political or personal agendas which could shape the outcome of an evaluation. Another limitation is the cost factor related to conducting an evaluation of this type in its entirety. It was stated, "if followed in its entirety, the management-oriented approach can result in costly and complex evaluations" (Worthen and Sanders, 1987, p. 85.

Improving the CIPP Model

The first step of improving the CIPP Model is adding a confirmative evaluation component. This can be done by including evaluation instruments e.g. knowledge tests, interviews, questionnaires, attitude scales, self-reports, observations and checklists (Dessinger & Moseley, 2003). For purposes of this study, checklists, knowledge tests and interviews will be conducted. Confirmative evaluation is the "marriage of evaluation and continuous improvement" (Dessinger & Moseley, 2003, p. 5) and "it tests the endurance of outcomes, the return on investment, and

establishes the effectiveness, efficiency, impact, and value of the program over time" (Dessinger & Moseley, 2003, p. 6).

Rationale for Evaluating Remediation Programs

There are several benefits for engaging in program outside of the budgetary benefits. Program evaluation is "part of a larger culture of evidence approach to decision making and quality assurance which regional accrediting agencies find attractive. Thus, engaging in careful program evaluation enhances the chances campus accreditation or reaccreditation proceedings will be successful. Program evaluation is the student support services counterpart to learning outcomes assessment in the curricular realm" (Fairris, 2012, p. 4-5). Both require clear goals are enunciated and there is careful assessment of whether those goals are achieved (Fairris, 2012). Ultimately programs should be assessed more frequently and become an intricate part of evaluation in higher education, which will guide the way stakeholders, educators and decision makers think about how well they are actually doing (Fairis, 2012).

Fairris (2012) suggested understanding the methodological features of program evaluation—such as the need, when making causal inferences, to compare groups are "similar in every relevant respect"—affects the way staff come to understand and interpret data. "Evaluating the effectiveness of academic support services is the next frontier in the effort to ensure educational quality and student success in higher education. Institutions ahead of the curve in this regard can benefit enormously" (Fairis, 2012, www.aacu.org).

CHAPTER 3- METHODOLOGY

The aim of this study is to determine if adding confirmative evaluation to the CIPP model will make it a more useful and powerful tool. The revisions to the CIPP Model will be based on the Joint Committee on Standards for Educational Evaluation (1994) and Stufflebeam's (2001) research. After confirmative evaluation has been designed, it will be used to evaluate the modified remediation program within a local School of Medicine in order to compare the current CIPP evaluation model to the revised CIPP evaluation model. The purpose of this research is to determine how effective the modified remediation program will be by adding confirmative evaluation to the CIPP evaluation model. This is useful in determining if an institution should keep, revise or eliminate the modified remediation program for identified at risk students, with a practical example being the implementation in the WSU School of Medicine. By adding confirmative evaluation as steps inclusive to the product phase of the original CIPP Model has the ability to make the original CIPP Model of Evaluation a stronger and more robust tool.

Testing the Original and Newly Revised CIPP Models

<u>Procedures</u>

The original CIPP model will be completed by using the CIPP Evaluation Model Checklist a "comprehensive framework for guiding evaluations of programs, projects, personnel, products, institutions, and systems and "is focused on program evaluations, particularly those aimed at affecting long-term, sustainable improvements" (Stufflebeam, 2003, p. 2). Confirmative evaluation checklist criteria will be added to the revised CIPP evaluation model checklist. Due to the length and steps involved with each checklist criteria, the evaluator will not complete each step in its entirety as time resource will serve as an issue.

During the evaluation process, information will be collected and decisions made based on this information. There are three types of judgements that are made: decisions related to the program, decisions that are related to the strategies of the program, and those decisions that are related to the outcomes of the program (Cranton & Legge, 1978).

<u>Participants</u>

The participants of interest for this study will be students who participated in the modified remediation program from academic years 2015 through 2019 at a local SOM (School of Medicine). These students participated in the modified program meaning they elected to take five years to complete a four-year medical education program n = 165 out of a population of n = 1495.

<u>Research Design</u>

The objective of this study is to determine which CIPP model is more effective; stronger and robust by incorporating a confirmative evaluation method according to the standards created by the Joint Committee on Standards for Educational Evaluation (1994). The evaluation of the modified remediation program will be conducted as a non-experimental design. Participants are not randomly assigned to groups in non-experimental design, which is used with intact groups. Because internal and external threats to validity exist according to Campbell and Stanley (1963), attempts to control extraneous data are necessary (Salkind, 2012).

One approach to non-experimental designs is to limit them to simply answering questions about intact groups. There will be no attempt to generalize the results or claim causal relationships.

It is common to perform a meta-evaluation to determine the effectiveness of the original CIPP model and the newly revised CIPP model in order to determine if the evaluation met acceptable levels of quality and established standards (Patton, 2013). A summative meta-evaluation is used to determine if there is a statistically significant difference between the original

CIPP model and the revised CIPP model, and will ultimately help to determine which evaluation model is more effective based on program evaluation standards and guidelines (Patton, 2013). A meta-evaluation checklist will be used to determine which evaluation standards were met for each of the models (Scriven, 1969).

<u>Meta Evaluation</u>

Scriven (1969) introduced the term meta-evaluation in the Educational Products Report, and applied the underlying concept to the assessment of a plan for evaluating educational products (Stufflebeam, 1978). Meta Evaluation assesses the extent that an evaluation is as follows:

1. Technically Adequate in revealing the merit of some object;

2. Useful in guiding decisions;

3. Ethical in dealing with people and organizations; and

4. Practical in using resources.

Meta-evaluations bare three main characteristics (Woodside & Sakay, 2001): 1. They are syntheses of findings and inferences of evaluative research about the program performance. They report the effectiveness of managing the goals achieved by the programs and provide information about two characteristics: Well managed programs and poorly managed programs. 2. They inform about the validity and utility of evaluation methods, offering guidance regarding useful evaluation methods. 3. They provide strong evidence regarding the program impact, subsiding the decision-making process regarding it. Hence, the results of the meta evaluation assist and justify the increase of trust by the interested parts and managers of programs in the evaluation results.

The meta-evaluation will serve as a method used to determine the effectiveness of the original CIPP model and the newly revised CIPP model. The meta-evaluations of both the original and newly revised CIPP models will be conducted by using the standards from the Joint Committee on Standards for Education Evaluation (1994). The evaluation standards being used in the meta-evaluation from the Joint Committee on Standards for Educational Evaluation (1994) include five different standards categories:

1. Utility Standards help to assure that stakeholders find program evaluation processes valuable to their needs (Joint Committee on Standards for Educational Evaluation, 1994, p. 1).

2. Feasibility Standards are intended to increase evaluation effectiveness and efficiency (Joint Committee on Standards for Educational Evaluation, 1994, p. 1).

3. Proprietary Standards support legal, fair and just evaluations (Joint Committee on Standards for Educational Evaluation, 1994, p. 1-2).

4. Accuracy Standards intended to increase the dependability and truthfulness of evaluation representations, propositions, and findings, especially those that support interpretations and judgments about quality (Joint Committee on Standards for Educational Evaluation, 1994, p. 2-3).

5. Evaluation Accountability Standards encourage adequate documentation of evaluations and a meta-evaluative perspective focused on improvement and accountability for evaluation processes and products. (Joint Committee on Standards for Educational Evaluation, 1994, p. 3)

The New CIPP Model

The new CIPP Model of Evaluation will include additional checklist and interviews as well as a survey to for the participants to determine the usefulness of confirmative evaluation steps to the CIPP Model. Those participating in the study will be given a survey about their experience in the modified program. The survey will consist of scaled and open-ended questions.

The end of year (EOY) scores will be assessed, along with STEP 1 and 2 scores to determine the overall rank in the medical school program. These scores will be compared to a random number of students n=165 in the general population who did not participate in the modified program to ultimately determine if the modified program was an overall successful program.

The students who participated in the modified program will be given an additional assessment to determine the effectiveness and impact of confirmative evaluation. Confirmative evaluation occurs months after the program, and those participating in this study are students who would have already graduated from the medical school. This will give students the opportunity to feel confident and safe in their responses and participation not thinking that it will impact their status within the medical school. Students will be reached via phone and email to discuss participating in the study to determine the validity of value of the new revised CIPP Model.

Dependent Variables

The dependent variables for this study are students' MCAT score, Year 1 and 2 medical exam scores, and USMLE STEP 1and STEP 2 scores.

Independent Variable

The independent variables include participation in the original vs modified remediation program at both SOM's.

Data Collection

Data needed to perform the evaluation of the modified remediation program will be collected using STARS (an application used at both SOM's), a web application tool for faculty and advisors that interacts with a collection of WSU databases, designed to enable convenient access to university data at both an individual and aggregate level for advising, retention efforts, curriculum tracking, and program evaluation, and documented in an Excel database.

Data Analyses

The Statistical Package for the Social Sciences for Windows (SPSS v. 26) will be used. A nominal alpha level of 0.05 will be used as the significance level. A Hoteling's Two Sample T^2 analysis will be conducted. It is the multivariate extension of the two-group Student's t-test. In a t-test, differences in the mean response between two populations are studied. T^2 is used when the number of response variables are two or more. Hotelling's T-squared has several advantages over the t-test (Fang, 2017):

- The Type I error rate is well controlled,

- The relationship between multiple variables is considered,

- It can generate an overall conclusion even if multiple (single) t-tests are inconsistent. While a t-test will tell you, which variable differ between groups, Hotelling's summarizes the between-group differences (Hotelling, 1931).

The null hypothesis is that the group means for all response variables are equal, with the alternative hypothesis the centroids differ:

$$H_o: \boldsymbol{\mu}1 = \boldsymbol{\mu}1$$

$$H_a: \boldsymbol{\mu}1 \neq \boldsymbol{\mu}1$$

where bold type indicates multi-dependent variables x=traditional and y=modified.

<u>Underlying Assumptions</u>

When conducting the Hotelling T2 statistic, four data assumptions are important to consider when computing centroid mean differences:

1. The data are sampled from a population with mean vector μ. This assumption implies that there are no subpopulations with different population means.

2. The data from both populations have a common variance-covariance matrix – Σ.

3. The data subjects are independent. This means that the subjects from both populations are independently sampled.

4. Both populations of data are multivariate normally distributed.

<u>Limitations</u>

A retrospective cohort study, also called a historical study, considers events that have already taken place (Mann, 2003). Cohort studies are designated by the timing of data collection, either prospectively or retrospectively, in the investigator's time. Studies collecting data on events that have already occurred have been labeled as historical, retrospective, and non-concurrent (Samet & Munoz, 1998). Some of the advantages of a retrospective cohort study include the following: they are cheaper and tend to take less time to complete; there is a lack of bias because the data was collected in the past and a single study can test various outcome variables (Mann, 2003).

Threats to validity that are relevant to a retrospective study include history, maturation, selection bias, and single group threat (Tofthagen, 2012). The validity threat known as history occurs when an event is unrelated to intervention during a study (Campbell & Stanley, 1963). Maturation is the process of systematic changes occurring naturally during a study (Campbell & Stanley, 1963). Selection bias occurs when a comparison group is selected non-randomly, which is a concern because this group could differ from the intervention group and ultimately affect the

study outcome (Campbell & Stanley, 1963). The single group threat occurs when there is a lack of comparison or control group in the study (Tofthagen, 2012). The participants are all from Wayne State University School of Medicine and have accepted a modified academic program for their medical education. Therefore, this homogeneity serves as a threat to validity.

In order to analyze the results of the meta-evaluation, an independent Samples t-Test will be conducted. This will help to determine which CIPP model is the best as it relates to evaluation standards (Salkind, 2010). If population normality is violated, he Wilcoxon Rank Sum Test, the non-parametric alternative will be used, because it is robust and more powerful: "When normality is met or nearly met (which occurs rarely), the t test maintains a very small power advantage over the Wilcoxon Rank Sum / MannWhitney U-Test. When normality is violated, the Wilcoxon Rank Sum Test can be three or four times more powerful than the independent samples T-Test" (Sawilowsky, 2005, p. 598).

CHAPTER 4- RESULTS

The aim of this study is to determine if adding confirmative evaluation to the CIPP model will make it a more useful and powerful tool. The proposed revisions to the CIPP Model are based on the Joint Committee on Standards for Educational Evaluation (1994) and Stufflebeam's (2001). A confirmative evaluation was used to assess the modified remediation program within a local School of Medicine in order to compare the current CIPP evaluation model to the revised CIPP evaluation model. The purpose of this research is to determine how effective the modified remediation program will be by adding confirmative evaluation to the CIPP evaluation model. Data were collected from STARS, which included MCAT scores, exam scores, STEP 1 scores as well as questionnaires conducted with administration, students and staff. The results will be laid out first, from data collected from STARS then by the conducted questionnaires which offer more insight and detail on to the confirmative evaluation method.

The general linear model approach to the primary data analysis was pursued. In this design, the multivariate Hotelling's T^2 on group (original vs. remediation) by the three dependent variables MES2, USMLE1, and USMLE2 was conducted with MCAT and MES1 serving as covariates. There were n = 81 participants in the remediation group and n = 84 participants in the non-remediation group, for a total of N = 165 participants (Sawilowsky, 2020).

Descriptive statistics are compiled in Table 1.

Table 3. Descriptive Statistics

	Grp	M	SD	N
	0	84.19	3.392	84
MES2	1	74.44	3.950	81
	Total	79.41	6.109	165
	0	237.55	8.545	84
USMLE1	1	220.86	4.764	81
	Total	229.36	10.863	165
	0	256.40	10.197	84
USMLE2	1	247.67	8.025	81
	Total	252.12	10.161	165

Box's test of equality of covariance matrices was statistically significant ($F = 7.79$, $df_1 = 6$, $df_2 = 191803.2$, $p = 0.000$), indicating a violation of the underlying assumption of multivariate normality. However, this test is hyper-sensitive to small departures of homogeneity for large sample sizes, and is not particularly robust for population nonmorality. Nevertheless, Levene's test of equality of error variances for all three dependent variables were also statistically significant, as noted in Table 2. Therefore, the results of the Hoteling's T^2 test must be interpreted with caution (Sawilowsky, 2020).

Table 4. Levene's Test of Equality of Error Variances

	F	df_1	df_2	Sig.
MES2	22.960	1	163	.000
USMLE1	21.849	1	163	.000
USMLE2	7.744	1	163	.006

The primary results are compiled in Table 3. Pillai's trace (as well as the comparable Wilk's Lamda, Hotelling's Trace, and Roy's Largest Root) were not statistically significant ($F = 2.48$, $df = 3, 159$, $p = 0.063$).

Table 5. Multivariate Tests[a]

Effect		Value	F	Hypothesis *df*	Error *df*	Sig.
Intercept	Pillai's Trace	.444	42.367[b]	3.000	159.000	.000
	Wilks' Lambda	.556	42.367[b]	3.000	159.000	.000
	Hotelling's Trace	.799	42.367[b]	3.000	159.000	.000
	Roy's Largest Root	.799	42.367[b]	3.000	159.000	.000
MCAT	Pillai's Trace	.014	.766[b]	3.000	159.000	.515
	Wilks' Lambda	.986	.766[b]	3.000	159.000	.515
	Hotelling's Trace	.014	.766[b]	3.000	159.000	.515
	Roy's Largest Root	.014	.766[b]	3.000	159.000	.515
MES1	Pillai's Trace	.278	20.360[b]	3.000	159.000	.000
	Wilks' Lambda	.722	20.360[b]	3.000	159.000	.000
	Hotelling's Trace	.384	20.360[b]	3.000	159.000	.000
	Roy's Largest Root	.384	20.360[b]	3.000	159.000	.000
Grp	Pillai's Trace	.045	2.479[b]	3.000	159.000	.063
	Wilks' Lambda	.955	2.479[b]	3.000	159.000	.063
	Hotelling's Trace	.047	2.479[b]	3.000	159.000	.063
	Roy's Largest Root	.047	2.479[b]	3.000	159.000	.063

a. Design: Intercept + MCAT + MES1 + Grp

b. Exact statistic

Although the multivariate approach was taken (in order to avoid experiment-wise Type I error inflation), the univariate breakdown analyses indicated there was no statistically significant difference between the remediation and un-remediation group for MES2 ($F = 1.57$, $df = 1, 165$, $p = 0.212$) and USMLE1 ($F = 1.69$, $df = 1, 165$, $p = 0.196$), but was statistically significantly different for USMLE2 ($F = 5,951$, $df = 1, 165$, $p = 0.016$) (Sawilowsky, 2020).

The upshot of this portion of the study was a demonstration of the effectiveness of remediation programs. Although it was never expected to leapfrogging performance of lesser prepared students over better prepared students, the remediation program examined did lead to raising the most typical medical standardized scores of these students to a level statistically comparable to those who did not require remediation (Sawilowsky, 2020).

Questionnaire Results:

There were 140 email questionnaires were sent out between faculty, administration and students of which 65 (46.4%) were returned. In each section (administration, faculty, student) there were questions of which there were a combination of open ended, yes or no and scaled questions.

The open-ended questions were grouped into themes. The questionnaires were sent out anonymously through a developed email system in which it randomly selects participants, no demographic data was collected from participants and there were no advantages or disadvantages for participation. The students selected were graduates of the medical school and some faculty and administration that were part of the random selection were previous employees of the medical school. These questionnaires were presented as confirmative evaluation methods because it was reflective and evaluative of processes that were already in place.
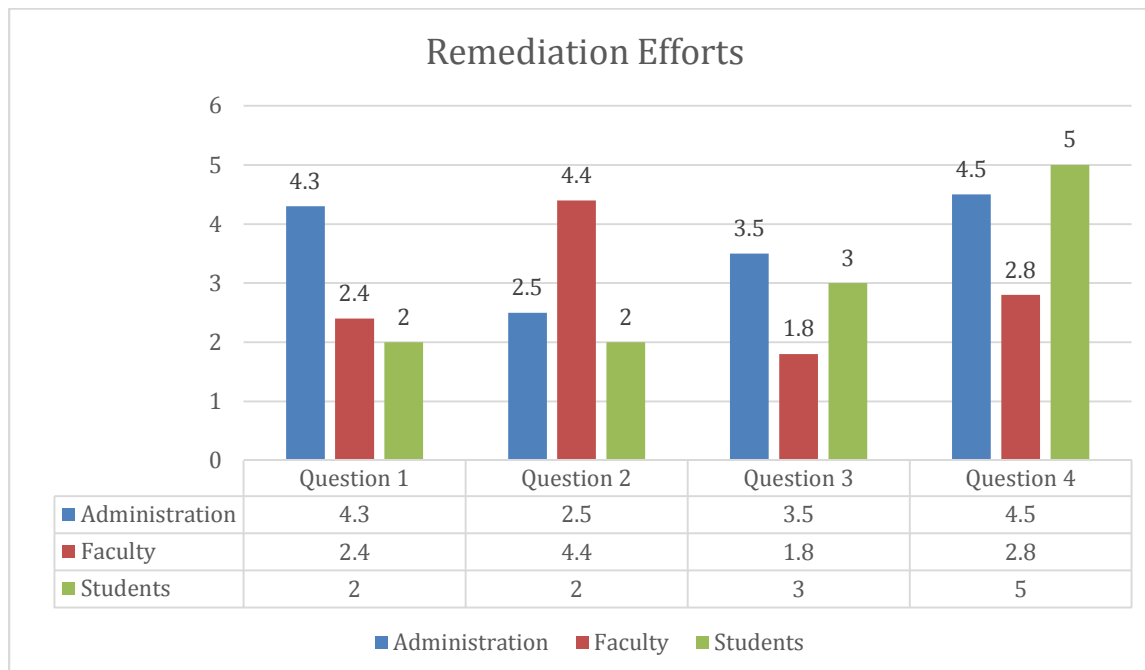
Questionnaire Data

Figure 1. Remediation Efforts

Question 1. On a scale of 1 – 5 with 1 being none to 5 being high, how would you rate the resources and support made available for struggling students?

Question 2. On a scale of 1 – 5 with 1 being non-existent to 5 being highly effective, how would you rate the SOM remediation efforts for struggling students?

Question 3. On a scale of 1 – 5 with 1 being completely unimportant and 5 being very important, how would you rate the importance of having remediation plans for struggling students?

Question 4. As administration, faculty or a student on a scale of 1 – 5 with 1 being low to 5 being high, how would you rate the willingness of students to participate in remediation programs?
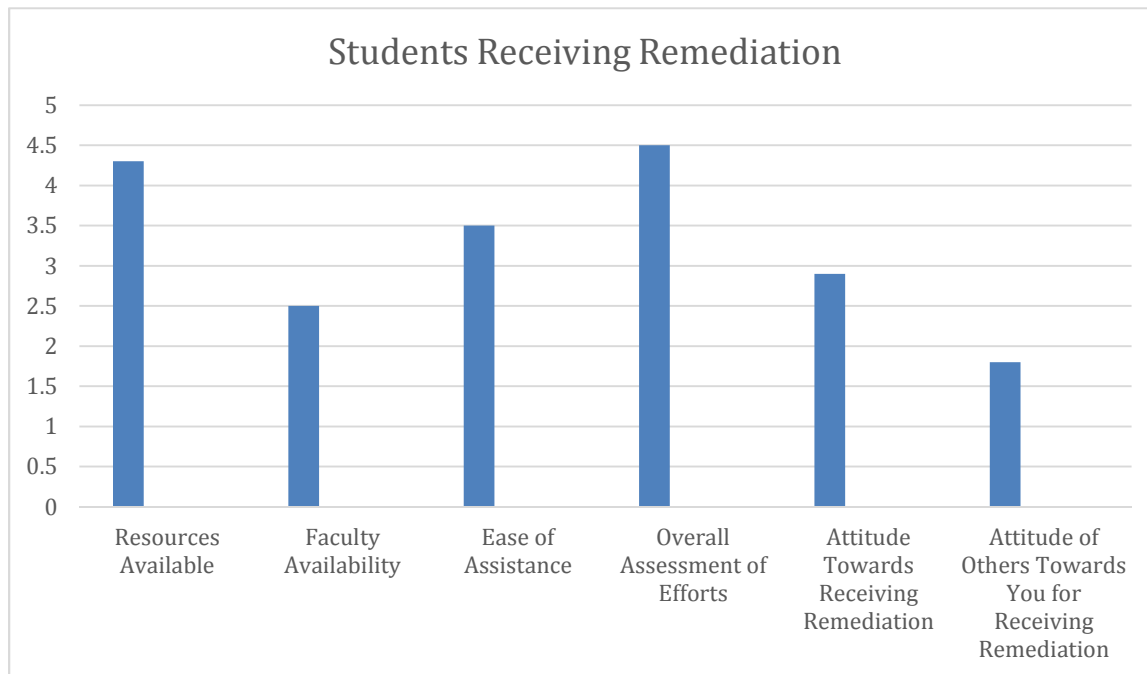

Figure 2. **Students Receiving Remediation**

Question 1. On a scale of 1 – 5, with 1 being low and 5 being high, how would you rate the resources available for struggling students?

Question 2. On a scale of 1 – 5, with 1 being low and 5 being high, how would you rate faculty availability to struggling students?

Question 3. On a scale of 1 – 5, with 1 being low and 5 being high, how would you rate the ease of assistance for struggling students?

Question 4. On a scale of 1 – 5, with 1 being low and 5 being high, how would you rate the overall assessment of efforts given to struggling students?

Question 5. On a scale of 1 – 5, with 1 being low and 5 being high, how would you rate your success as a medical doctor due to the remediation program?

Question 6. On a scale of 1 – 5, with 1 being low and 5 being high, how would you rate your attitude towards receiving remediation efforts?
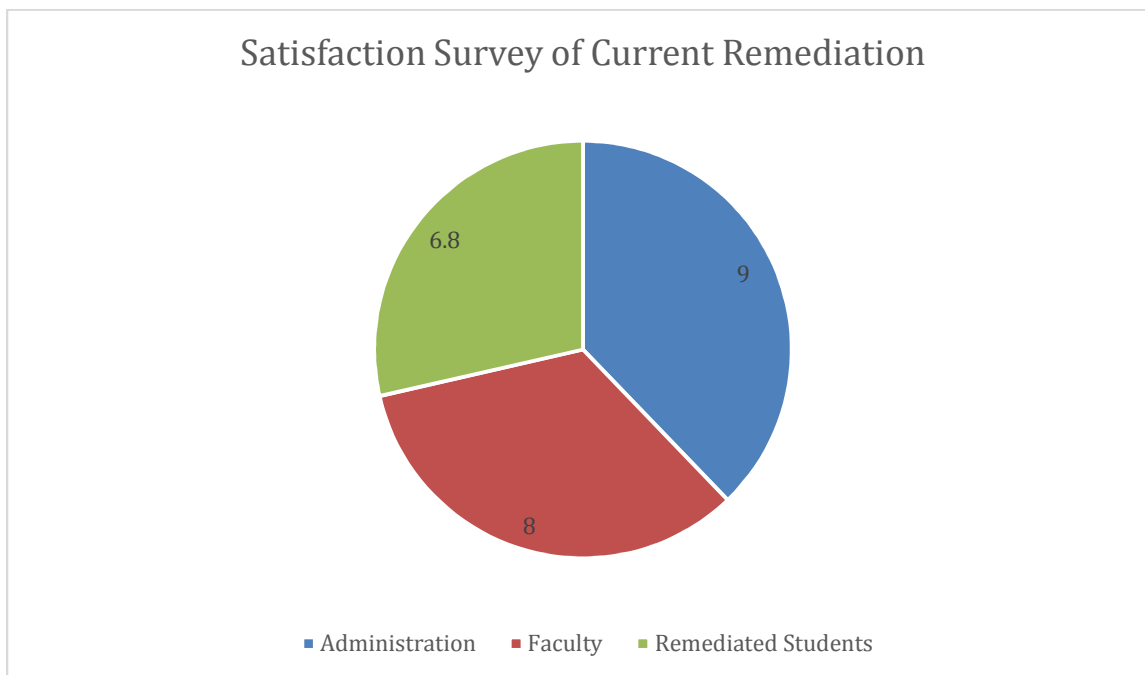

Figure 3. **Satisfaction Survey**

Question 1. How satisfied are you with the current remediation efforts?

Question 2. How satisfied are you with the success rate of the graduating medical students?

Question 3. How satisfied are you with the willingness of students to participate in remediation efforts?

Question 4. How satisfied are you compared to other SOM's in regard to the offered remediation plans for student?
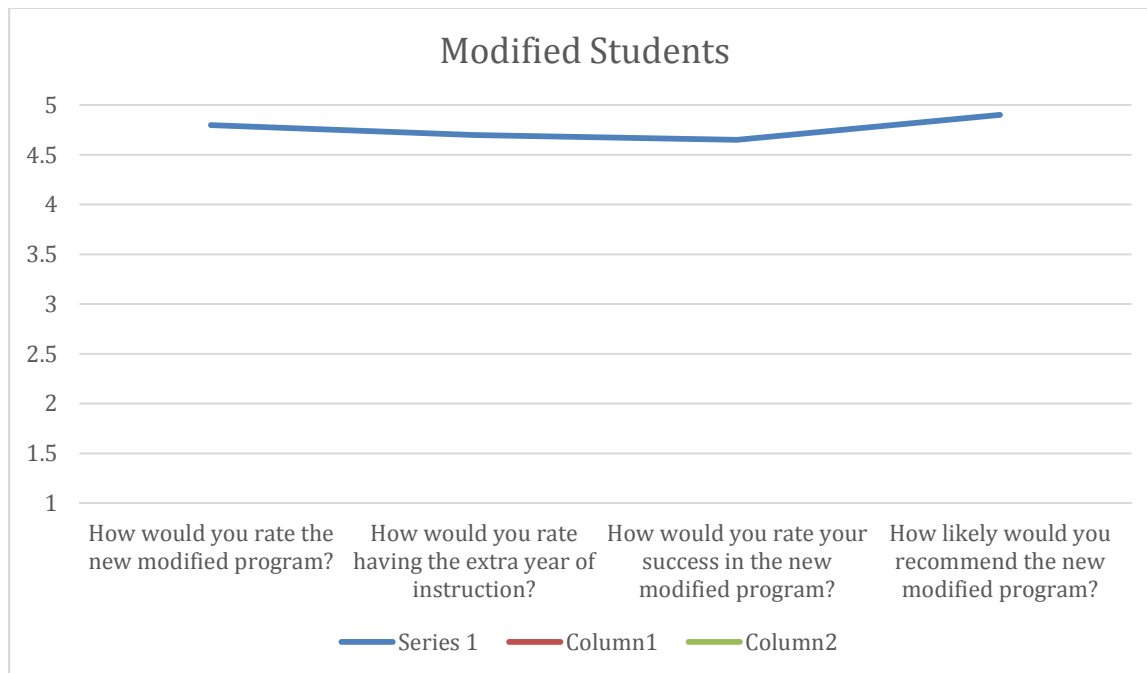

Figure 4. **Modified Student Satisfaction**

Question 1. On a scale of 1 -5, with 1 being poor and 5 being excellent how would you rate the new modified program?

Question 2. On a scale of 1 -5, with 1 being poor and 5 being excellent, how would you rate having the extra year in medical doctor?

Question 3. On a scale of 1 -5, with 1 being unlikely and 5 being very likely, would you recommend the new modified program to at risk medical students?

Revised Model
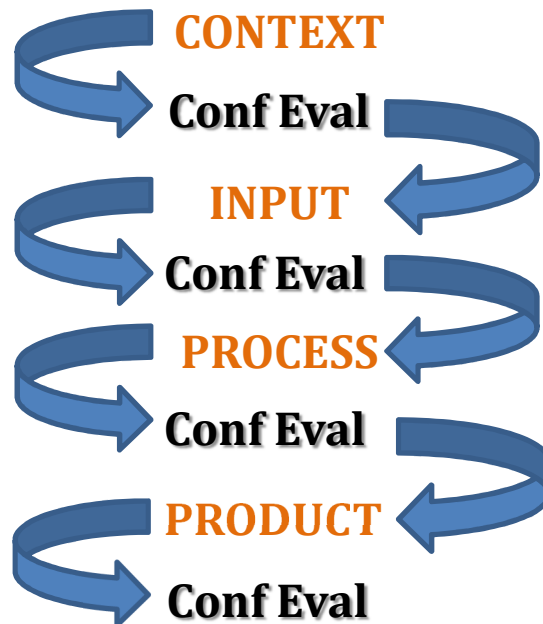
The revised model consists of adding continuous confirmative evaluation throughout the entire CIPP process. After each stage of the evaluation, confirmative evaluation is conducted. It is a circular process represented as a continuous flow, meaning in order for each stage to be complete, some form of confirmative evaluation must be conducted before proceeding to the next stage of the evaluation process. Confirmative evaluation is a continuous process. "Confirmative evaluation identifies, explains, and confirms the value of the performance improvement intervention over time" (Misanchuk, 1978, p. 15). To incorporate confirmative evaluation after each phase of the CIPP Model of Evaluation, 1) data should be reviewed and incorporated into actual activities ongoing activities, 2) continuous interviews should be conducted from participants in the current phase of the cipp model, 3) information should be continually reviewed to verify the content of the phase is still valid, timely and aligned with the overall intent of the evaluation.

CHAPTER 5 - DISCUSSION

The purpose of this research was to determine if an additional step – Confirmative Evaluation to the CIPP Model would make it a more robust and powerful evaluation mode. The backdrop of the research was based on the effectiveness of remediation efforts at the School of Medicine. Looking at close to 1500 students, it was determined that remediation plans are critical to the success of the academic program. Struggling students who were having difficulties in course work, passing course exams and passing major medical STEP exams needed remediation efforts to be successful. Students struggling medical schools is not anything new; most schools see a number of students who require some additional assistance in order to be successful. The key is to ensure that as school begin developing and implementing new academic medical curriculums that they plan for and implement strategies for remediation. It is a critical component for the success of the program.

Remediated students often feel like they are being labeled and that people will not consider them to be able to be successful physicians. This study has shown that though they do score lower than their peers on exams, and in overall course work, with an effective remediation program they are capable of matching well as their peers in terms of standardized test scores. In responses from the questionnaire's, graduated medical students were very appreciative of the remediation efforts of the medical school and believe that overall, these programs were beneficial to their overall success as a medical student and as a physician.

This research objective was to find if adding confirmative evaluation to the CIPP Model would make it a more robust and powerful evaluation model. The confirmative evaluation step was conducted via use of a questionnaire to administration, faculty and students. The collected information shows that going back to review programs and processes after some time has lapsed

gives credence to determining if a program is still performing in the manner it was originally designed. Students and faculty alike, expressed that the continuation of looking at data to ensure the effectiveness of the remediation programs at medical school was beneficial and had great value to the overall evaluation of the long-term success of the medical school. Most important, it was determined as indicated in the revised model that confirmative evaluation needs to occur after each step in the CIPP Model allowing for the presence of continuous quality improvement. It is important to not wait until the program has been completely evaluated to conduct confirmative evaluation. It is more effective as an ongoing step in the complete process of program evaluation. Faculty believed that the continuous collection of data requires them to consciously think of best practices to ensure that they are creating curriculum and course work that will benefit every student within the medical school. Students expressed that the continuous collection would alleviate any bias felt by students who needed to participate in remediation programs. They also expressed that as faculty was working more diligent to add remediation efforts in coursework more students were willing to acknowledge their need for help and seek it out without the need to be first identified. Students suggested that more students are openly asking for help and they believed the measures being put in place were to ensure the success of all medical students regardless of background, ethnicity, etc.

Confirmative evaluation for many institutions of higher learning can be time consuming and costly. Like with other organizations, once the process has been implemented they want to leave it there and not continuously review to ensure it still works. Confirmative evaluation is a step that requires much effort and commitment and can be quite costly to an organization.

More research is needed to accurately determine if adding confirmative evaluation to the CIPP Model makes it a more robust and powerful tool. Based on the information from this

research, it has been determined that it a "satisfactory" and pragmatic step in the overall process of program evaluation, it cannot be determined however, if it makes it a more robust and powerful tool. Limitations to the CIPP Model is that it is tedious and outdated, however as it currently stands as a solid program evaluation model, adding confirmative evaluation as an additional step to the process does not prove to make it a more sustainable evaluation model. More research and testing needs to be done to find the best addition to the model to make it a more robust and powerful tool.

Conclusion

The research question posed was does adding confirmative evaluation to the CIPP Model of evaluation make it a more powerful and robust tool. According to the questionnaire data it makes it a more useful tool. Does it make it a more powerful and robust tool, the answer according to the Wilcoxon Rank Test is the old CIPP Model and the newly revised model are equal. This study shows that more research is needed in the area of confirmative evaluation and its use to make evaluation models more comprehensive, effective and efficient. There remains a critical need to expand the CIPP Model of Evaluation. As the data showed, there was not significant difference between the remediation and the un-remediated groups. While it is showing that there is room for improvement in the remediation group, the positive take away is that though there is room for improvement, the remediation program and remediation efforts are important and relevant to the students in the medical school.

APPENDIX

Stakeholder (Administration, Faculty, Student) Interview

Question	Answer Summary
How do you assess academic performance?	Reviewing quiz scores. Reviewing exam scores (instructor written exams and National Board exams). Reviewing STEP1 national exam scores.
How important is academic remediation?	Academic remediation is very important. We have a task force in place as part of the new curriculum efforts to address the needs of students who identify as at risk. Brainstorming initiatives to catch these students prior to them falling behind and possibly forced to withdraw or be dismissed from the SOM. Addressing the needs and concerns of students to have the necessary resources available for them to be successful in the curriculum.
How do you encourage students/faculty members to solicit the necessary help to ensure remediation?	Faculty are held to the highest integrity standards to ensure that they are instructing our students in a manner that is fair across the board. We encourage our faculty to identify students early in the course who are having difficulty with the material and exams. We encourage our students to reach out for assistance and to not feel that they are less than because they need assistance. In order for us to help them, we have to recognize they require help. Some students feel ashamed and don't want to reach out for help, they believe they can figure it out for themselves. Unfortunately, medical school is very difficult and it very difficult to catch up once you have fallen behind in a subject area as each subject area builds from the previous. Repeating exams makes it very difficult to stay on track so we continuously encourage students, especially those we identify as struggling in the curriculum to accept the necessary resources. In fact in the new curriculum, students who are identified as struggling will be mandated to participate in additional learning activities such as small groups, tutoring and remedial assistance to get them

	back on track. If students refuse to accept the assistance they will be putting themselves in a position harmful t their success and to their defense when they have to present in front od the academic boards to discuss if they should be allowed to continue as a student.
What are your expected outcomes of the new curriculum in regards to ensuring students don't fall through the cracks?	There will be more resources in place to catch struggling students. There will be more checkpoints built within the curriculum to ensure students are identified earlier in the process. There will be more hands-on deck – meaning more eyes on the progress of students. These efforts will ensure students do not fall through the cracks, are quickly identified and placed in remediation programs to get them back up to speed and on track for success.
Do you feel comfortable recommending students for remediation?	We believe it is important that all students have the ability to participate in activities that will allow them to be successful in medical school. Based on formative assessments as well as summative assessments is how students are recommended for remediation efforts.
What remediation plans do you think would be most beneficial to medical students?	More resources are always needed. More tutors, the creation of the Office of Teaching and Learning and more eyes on the data will allow us to be able to identify students who are in need. Additional assignments, more faculty willing to serve as tutors and open up more hours of assistance to students will be a great help.
Do you think the SOM offers the necessary resources for struggling students?	There has been an increase in the number of resources available to students, but more resources would be helpful. We are a very large SOM and not all of the resources are readily available to students. The faculty student ratio is high so there is not enough resources if we all needed the assistance. However, the OLT office offers great resources and tutoring opportunities for struggling students as well as other initiatives to assist.
What resources are needed that have not been implemented by the SOM?	More staff in the OLT office and more faculty willing to be more accessible to students.

REFERENCES

Accreditation Standards and Guidelines (2007). In *Accreditation Council for Pharmacy Education*. Retrieved February 4, 2018, from http://www.acpeaccredit.org/pdf/ACPE_Revised_PharmD_Standards_Adopted_Jan152006.pdf.

ACGME. (2010a). Accreditation council for graduate medical education: Glossary of terms. *Accreditation Council for Graduate Medical Education* Retrieved from http://www.acgme.org/acWebsite/about/ ab_ACGMEglossary.pdf

ACGME. (2010b). Program director guide to the common program requirements. *Accreditation Council for Graduate Medical Education*. Retrieved from http://www.acgme.org/ acWebsite/navPages/nav_commonpr.asp

Alqahtani, K. (2016). Decision-oriented evaluation: A review of various models of evaluation. *International Journal of Scientific & Engineering Research*, 7, (5), May-2016 929 ISSN 2229-5518.

Association of American Medical Colleges. (2004, March 19). The status of the new AAMC definition of "underrepresented in medicine" following the Supreme Court's decision in Grutter. In *AAMC*. Retrieved August 25, 2018, from https://www.aamc.org/download/54278/data/urm.pdf

American Evaluation Association. (2003). *American Evaluation Association guiding principles for evaluators*. Retrieved from http://www.eva;.org/d/do/594

Ames, C. & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80 (3), 260-267.

Astin, A. W. (1999). Student involvement: A Developmental theory for higher education.

Journal of College Student Development, 40 (5), 518-529.

Audetat M-C, Laurin S, Dory V. Remediation for struggling learners: putting an end to 'more of

the same' Med Educ. 2013;47:230–231. doi: 10.1111/medu.12131. [PubMed] [CrossRef]

[Google Scholar]

Avgerou, C. (1995). Evaluating information systems by consultation and negotiation.

International Journal of Information Management, 15 (6), 427- 436.

Bennion et al. *BMC Medical Education* (2018) 18:120

Bierer SB, Dannefer EF, Tetzlaff JE. Time to loosen the apron strings: cohort-based evaluation

of a learner-driven remediation model at one medical school. J Gen Intern Med.

2015;30:1339–1343. doi: 10.1007/s11606-015-3343-1. [PMC free article] [PubMed]

[CrossRef] [Google Scholar]

Birckmayer, J. D. & Weiss, C. H. (2000). Theory-based evaluation in practice what do we learn?

Evaluation Review, 24 (4), 407-431.

Booker, Zsa-Zsa Lashawn-Marie, "*A New Logic Model for Change*" (2016). Wayne State

University Dissertations. https://digitalcommons.wayne.edu/oa_dissertations/1627.

Boyce, E. G. (2008). Finding and Using Readily Available Sources of Assessment

Data. *American Journal of Pharmaceutical Education,*72 (5), Article 102.

Brandenburg, D. C. (1989). Evaluation and business issues: Tools for management decision

making. *Special Issue: Evaluating Training Programs in Business and Industry*, 83-100.

Retrieved September 4, 2018, from

https://onlinelibrary.wiley.com/doi/abs/10.1002/ev.1536

Campbell, D. & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*.

　　　Chicago, IL: Rand-McNally.

Chen, H. T. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.

Chen, H. T. (2015). *Practical Program Evaluation: Theory-Driven Evaluation and the*

　　　Integrated.

Cleland, J. (2005). *Medical Teacher*, 27 (6), 504-508.

Cleland J, Leggett H, Sandars J, Costa MJ, Patel R, Moffat M. The remediation challenge:

　　　theoretical and methodological insights from a systematic review. Med Educ.

　　　2013;47:242–251. doi: 10.1111/medu.12052. [PubMed] [CrossRef] [Google Scholar]

Cook, D. A. (2010). Twelve tips for evaluating educational programs. *Med Teach*, 32, 296-301.

Cranton, P. A. & Legge, L.H. (1978). Program evaluation in higher education. *Journal of Higher*

　　　Education, 49 (5), 464-71.

Dessinger and Moseley (2003).

Donaldson, S.I., & Lipsey, M.W. (2006). Roles for theory in evaluation practice. In I. Shaw, J.

　　　Greene, & M. Mark (Eds.), *Handbook of Evaluation*. Thousand Oaks, CA: Sage.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2012). *Program evaluation: Alternative*

　　　approaches and practical guidelines (4th ed.). Boston, MA: Pearson.

Frellsen, S., Baker, E. A.,Papp, K. & Durning, S. [abbreviate first name] (2009). *Academic*

　　　Medicine, 83 (9) 876-88.

Goldie, J. (2006). AMEE Education Guide no. 29: Evaluating educational programmes, *Medical*

　　　Teacher, 28:3, 210-224. doi [use lowercase "doi"]: 10.1080/01421590500271282

Guerrasio J, Garrity MJ, Aagaard EM. Learner deficits and academic outcomes of medical

　　　students, residents, fellows, and attending physicians referred to a remediation program,

2006–2012. Acad Med. 2014;89:352–358. doi: 10.1097/ACM.0000000000000122.
[PubMed] [CrossRef] [Google Scholar]

Guerrasio J, Furfari KA, Rosenthal LD, Nogar CL, Wray KW, Aagaard EM. Failure to fail: the institutional perspective. Med Teach. 2014;36:799–803. doi: 10.3109/0142159X.2014.910295. [PubMed] [CrossRef] [Google Scholar

Hanssen, C. E., Lawrenz, F., & Dunet, D. O. (2008). Concurrent meta-evaluation: A critique. *American Journal of Evaluation*, 29 (4), 572-582.

Hauer, K., Ciccone, A.,Henzel, T. R., Katsufrakis, P.,Miller, S. H. & Norcross, W., Papadakis, M. A. & Irby, D (2009). Remediation of the deficiencies of physicians across the continuum from medical school to practice: A thematic review of the literature. *Academic Medicine*: 84 (12), 1822-1832. doi:10.1097/ACM.0b013e3181bf3170.

Hawthorne M, Chretien K, Torre D, Chheda S. Re-demonstration without remediation— a missed opportunity? A national survey of internal medicine clerkship directors. Med Educ Online. 2014;19:25991. doi: 10.3402/meo.v19.25991. [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Hirschheim, R., & Smithson, S. (1998). Evaluation of information systems: A critical assessment. In L. P. Willcocks & S. Lester (Eds.), *Beyond the IT productivity paradox* (pp. 381-410). West Sussex, U.K.: Wiley and Sons.

Hotelling, H. (1931). The generalization of student's ratio. *Annals of Mathematical Statistics*, 2(3), 360-378.

Huberty, C. J. (1984). Issues in the use and interpretation of discriminant analysis. Psychological Bulletin 95, 156-171.

Johnson, B. (2001). Toward a new classification of non-experimental quantitative research. *Educational Researcher*, 30(2), 3-13.

Josepha, Esdras, Galeanob, Pedro & Lilloc, Rosa E (2015). Two-sample Hotelling's T2 statistics based on the functional Mahalanobis semi-distance. *UC3M Working Papers Statistics and Econometrics* 15-03 ISSN 2387-0303.

Joint Committee on Standards for Educational Evaluation (1981). *Standards for evaluations of educational programs, projects, and materials*. New York, NY: McGraw-Hill.

Joint Committee on Standards for Educational Evaluation (1994). *The Program evaluation standards* (2nd ed.). Thousand Oaks, CA: Sage.

Kalet, A. Chou, C. & Ellaway R. (2017). To fail is human: remediation in medical education. *Perspect Medical Educ.* 2017 Dec; 6(6): 418–424. Published online 2017 Oct 25. doi: 10.1007/s40037-017-0385-6

Kalet A, Chou CL. Remediation in medical education. New York: Springer; 2014. p. xvii. [Google Scholar]

Keppel, G. & Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook* (4th ed.). Upper Saddle River, NJ.

LCME. (2010). Functions and structure of a medical school. *Standards for accreditation of medical education programs leading to the M.D. degree.* Washington, DC: Liaison Committee on Medical Education. [Accessed 20 April 2017] Retrieved from http://www.lcme.org/ standard.htm.x

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance f test. *Review of Educational Research*, 66 (4), 579-619.

Madaus, G. F., Scriven, M., & Stufflebeam, D. L. (1983). Evaluation models. Norwell, MA:
    Kluwer Nijhoff.

Maher, B., Hynes, H., Sweeney, C., Khashan, A., O'Rourke, M., Doran, K., Harris, A. & Flynn,
    S. (2013). Medical school attrition -- Beyond the statistics a ten year retrospective study.
    BMC Medical Education, 13 (13), 10.1186/1472-6920-13-13

Maize, D., Fuller, S., Hritcko, P., Matsumoto, R. R.,Soltis, D., Taheri, R. R. & Duncan, W.
    (2010). A review of remediation programs in pharmacy and other health professions.
    American Journal of Pharmaceutical Education, 74, (25) 10.5688/aj740225

Mann, C. J. (2003). Observational research methods. Research design II: Cohort, cross sectional,
    and case-control studies. *Emergency Medicine Journal*, 20, 54-60.

Matt, G. E., & Cook, T. D. (1994). Threats to the validity of research syntheses. In H. Cooper &
    L. V. Hedges (Eds.). *The Handbook of Research Synthesis* (pp. 503-520). New York,
    NY: Russell Sage Foundation.

Mendel D, Jamieson A, Whiteman J. Remediation. In: Walsh K, editor. Oxford textbook of
    medical education. Oxford: Oxford University Press; 2013. pp. 362–371. [Google
    Scholar]

Moseley, James L.; Solomon, David L. *Performance Improvement*, v36 n5 p12-16 May-Jun
    1997.

Nielsen DG, Jensen SL, Neill LO. Clinical assessment of transthoracic echocardiography skills :
    a generalizability study. BMC Med Educ. 2015;15:9. doi: 10.1186/s12909-015-0294-5.
    [PMC free article] [PubMed] [CrossRef] [Google Scholar]

Patten, M. L. (2007). *Understanding Research Methods: An Overview of the Essentials* (6th ed.).
    Glendale, CA: Pyrczak Publishing.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand

    Oaks, CA: Sage.

Patton, M. Q. (2013). Meta-evaluation: Evaluating the evaluation of the Paris Declaration. *The*

    Canadian Journal of Program Evaluation, 27 (3), 147-171.

Powell, B. & Conrad, E. (2015). Utilizing the CIPP Model as a means to develop an integrated

    service-learning component in a university health course. *Journal of Health Education*

    Teaching, 6(1): 21-32.

Prasetyono, H. (2016). Graduate program evaluation in the area leading educational, outlying

    and backward. *Journal of Education and Practice*. ISSN 2222-1735 (Paper) ISSN 2222-

    288X (Online) Vol.7, No.36, 2016.

Rencher, A. C. and Scott, D. T. (1990). Assessing the contribution of individual variables

    following rejection of a multivariate hypothesis. *Communications in Statistics-*

    Simulation and Computation 19, 535-553.

Samet, J. M. [add space after period] & Munoz, A. (1998). Evolution of the Cohort Study.

    Epidemiol Rev, 20 (1), 1-14.

Sawilowsky, S. (2005). Misconceptions leading to choosing the t-test over the Wilcoxon Man-

    Whitney test for shift in location parameter. *Theoretical and Behavioral Foundations of*

    Education Faculty Publications, 4 (2), 598-600.

Saxon, D. P. & Boylan, H. R. (2001). The cost of remedial education in higher

    education. *Journal of Developmental Education*, 25 (2), 2.

Schroeder, C. C., Minor, F. D., & Tarkow, T. A. (1999). Freshmen interest groups: Partnership

    for promoting success. *New Directions for Student Success*, 87, 37-49.

Schumacker, R. (2014). *Learning statistics using R*. Thousand Oaks, CA: Sage.

Scriven M. (1998). Minimalist theory of evaluation: The least theory that practice requires. *American Journal of Evaluation*, 19, 57-70.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). New York: Sage.

Sebatane, E. M. (1998) Assessment and classroom learning: A [capitalize "a"] response to Black & Wiliam. *Assessment in Education: Principles, Policy and Practice*, 5:1, 123-130. doi: 10.1080/0969595980050108

Sidle, M. W. & McReynolds, J. (1999). The freshmen year experience: Student retention and student success. *NASPA Journal*, 36 (4), 288-300.

Stufflebeam, D. L. (1978, April 1). Meta evaluation: An overview [Electronic version]. *Evaluation and the Health Professions*, *1*(1), 17-43.

Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 89, 7-98.

Stufflebeam, D. L. (2011). Meta-evaluation. *Journal of Multi-Disciplinary Evaluation*, 7 (15), 99-158.

Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass.

Stufflebeam, D. L. (2017, March). CIPP evaluation model checklist [Second Edition]: A tool for applying the CIPP model to assess long-term enterprises. In *Western Michigan University*. Retrieved August 25, 2018, from https://www.wmich.edu/sites/default/files/attachments/u350/2014/cippchecklist_mar07.pdf

Suchman, E. A. (1967). *Evaluative Research*. Russell Sage Foundation.

Stufflebeam D.L. (2000) The CIPP Model for Evaluation. In: Stufflebeam D.L., Madaus G.F., Kellaghan T. (eds) Evaluation Models. *Evaluation in Education and Human Services, vol 49.* Springer, Dordrecht

Symons, V. (1991). A review of information systems evaluation: Content, context and process. *European Journal of Information Systems,* 1 (3), 205 - 212.

Symons, V. & Walsham, G. (1988). Overcoming theory-practice inconsistencies: Critical realism and information systems research. *Information and Organization*, 16 (3), 191 - 211.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn & Bacon.

Tofthagen, C. (2012). Threats to validity in retrospective studies. *Journal of Advanced Practitioners in Oncology*, 3 (3), 181-183.

Vo, Thi. (2018). Evaluation Models in Educational Program: Strenghts and Weaknesses. VNU Journal of Foreign Studies. 34. 10.25073/2525-2445/vnufs.4252.

Walsham, G. (2006). *European journal information systems*, 15, 320. https://doi.org/10.1057/palgrave.ejis.3000589

Weiss, C. H. (1993). Where politics and evaluation research meet. *Evaluation Practice*, 14 (1), 93- 106.

Wholey, J. S. (1981). Using evaluation to improve program performance. *Evaluation Studies Review Annual*, 6, 55-69.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010). *Handbook of practical program evaluation (3rd ed.).* San Francisco, CA: Jossey-Bass.

Winston KA, Van Der Vleuten CPM, Scherpbier AJJA. The role of the teacher in remediating at-risk medical students. Med Teach. 2012;34:e732–e742. doi: 10.3109/0142159X.2012.689447. [PubMed] [CrossRef] [Google Scholar]

Worthen, B. (1990). Program evaluation. H. Walberg & G. Haertel (Eds.). *The International Encyclopedia of Educational Evaluation* (pp. 42-47). Toronto, ON: Pergamon.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A (2011). The program evaluation standards: A guide for evaluators and evaluation users (3rd ed.) Thousand Oaks, CA: Sage.

Yates, J. & James, D. (2006). Predicting the "strugglers": A case-control study of students at Nottingham University Medical School. *BMJ*, 332, 1009. doi:10.1136/bmj.38730.678310.63

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, *45*(4), 477-501. Retrieved from https://nrtevaluator.wordpress.com/2016/05/11/what-is-the-role-of-formative-assessment/

Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. Using the Context, Input, Process, and Product evaluation model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. *Journal of Higher Education Outreach and Engagement*, *15*(4), 57-84. Retrieved from https://eric.ed.gov/?id=EJ957107

Zientek, R. M. (2008). *The impact of themed learning communities on academic performance and retention* (Doctoral Dissertation). Retrieved from UMI. (UMI No.92 3320408).

ABSTRACT

CONFIRMATIVE EVALUATION: A NEW LEVEL TO THE CIPP MODEL

by

TIA L. FINNEY

December 2020

Advisor: Dr. Shlomo Sawilowsky

Major: Education, Evaluation and Research

Degree: Doctor of Philosophy

Struggling trainees often require a substantial investment of time, effort, and resources from medical educators. An emergent challenge involves developing effective ways to accurately identify struggling students and better understand the primary causal factors underlying their poor performance. Identifying the potential reasons for poor performance in medical school is a key first step in developing suitable remediation plans (Artino, et al., 2010). The SOM Modified Program is a remediation program that aims to ensure academic success for medical students. The purpose of this study is to determine the impact of modifying the CIPP evaluation model by adding a confirmative evaluation step to the model. This will be carried out by conducting a program evaluation of Wayne State University's School of Medicine Modified Program to determine its effectiveness for student success. The key research questions for this study are 1) How effective is the Modified Program for student's success in the SOM? 2) Do students benefit from a modified program in medical school? 3) Will the CIPP program evaluation model become more effective by adding confirmative evaluation component?

AUTOBIOGRAPHICAL STATEMENT

Tia Finney is a native of Detroit and a proud graduate of Detroit's Cass Technical High School. She holds a Ph.D. in Education, Evaluation and Research from Wayne State University. Her education also includes a B.A. in Communications, a M.Ed. in Instructional Technology and Design and a M.A. in Communications (Organization) – all from WSU.

Ms. Finney worked for WSU from 2002 – 2010 on a grant funded program in the WSU School of Medicine (SOM), where she went from a training specialist to the Director in less than five years. Due to a significant decrease in funding for the grant, she elected to release herself from the project. She believed this would be the best time to redirect her focus and pursue her lifelong passion of ministry where she held numerous positions for Triumph Church in Detroit, MI under the great leadership of Rev. Solomon W. Kinloch, Jr. (Senior Pastor). In 2015, following her calling, Ms. Finney became a licensed minister and in February 2020 became an Ordained Minister under the leadership of Rev. Dr. John Marks (Bethany Baptist Church). Rev. T (as she likes to be called) will pursue her passion for ministry as she intends to enroll to earn a Master of Divinity from Ashland Theological Seminary in Detroit, MI. She also desires to earn an MBA (she says her final degree) from Wayne State University. It is ultimately her goal to become a Pastor.

Currently, Ms. Finney is the Supervisor of Medical Testing at WSU SOM where she has been working since 2017. Ms. Finney operates daily in accordance to her motto: "live in such a way that if someone spoke badly of you, no one would believe it!" She is a world traveler, a proud member of Alpha Kappa Alpha Sorority, Inc for over 28 years, a member of the NAACP, the Progressive Baptist Convention and she is most proud of her involvement with the community where she serves as a positive role model, motivational speaker and mentor to many students and young adults throughout metro Detroit.