

12-7-2020

Confirmative Evaluation: New CIPP Evaluation Model

Tia L. Finney

Wayne State University, tiafinney@wayne.edu



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Finney, T. L. (2019). Confirmative evaluation: New CIPP evaluation model. *Journal of Modern Applied Statistical Methods*, 18(2), eP3568. doi:

SPECIAL SECTION ON QUANTITATIVE PROGRAM EVALUATION

Confirmative Evaluation: New CIPP Evaluation Model

Tia L. Finney

Wayne State University
Detroit, MI

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. 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?

Keywords: CIPP model of evaluation, modified programs, confirmative evaluation

Introduction

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, 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-López & Hutchinson, 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).

Application to Evaluation Medical School Remediation

The factors associated with success in medical school has interested medical educators and medical school admissions committees. 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

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

often pass unnoticed and continue in their studies with little guidance and feedback (Sayer et al., 2002; Cleland et al., 2005). 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 (Sayer et al., 2002; Cleland et al., 2005).

Struggling trainees often require a substantial investment of time, effort, and resources from medical educators (Sayer et al., 2002). 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).

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, the governing accreditation body for Medical Schools, “each student should be evaluated early enough during a unit of study to allow time for remediation” (Liaison Committee on Medical Education, 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., 2012).

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 (Mendel et al., 2013).

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.

Purpose of This Study

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.

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

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 an idea originally put forward by Egon Guba decades ago ([Stufflebeam & Shinkfield, 2007](#)). Kaufman 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-López & Hutchinson, 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. Will the CIPP program evaluation model become more effective by adding confirmative evaluation?
2. Do students benefit from a modified program in medical school?

Literature Review

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 of the most important factors in determining whether an evaluation will ultimately be useful (Conrad & 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 (Conrad & 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, 2001; Donaldson et al., 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).

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 (2003) described the "now fashionable advocacy of 'theory-based evaluation'" (p. 31) as a situation wherein 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, 2003, p. 32).

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

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).

Informational feedback, information students can use to improve their performances, is intrinsically motivating (Ames & Archer, 1988; Covington, 1992; Pintrich & Schrauben, 1992). This is important, given the nature of the assessment process. Black and Wiliam (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 (1989, 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). 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, 1989). 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 & Wiliam, 1998, p. 28). Building on the work of others (e.g., Black et al., 2003), two different summative processes are considered; 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 (Sadler, 1989).

There is a counterargument 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 and Wiliam (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 & Wiliam, 1998).

The 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, pp. 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).

Following Bloom (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

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

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).

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).

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” (pp. 2-3).

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. A CIPP model product evaluation study also examines the degree to which the targeted educational needs were met” (Zhang et al., 2011, pp. 58-59).

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 and 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, 2012).

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).

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

CIPP Model Strengths and Limitations

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 & 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).

Methodology

Procedures

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).

Participants

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 $n = 1495$.

Research Design

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.

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.

Meta Evaluation

The term meta-evaluation was introduced in 1969 by Scriven 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;

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

3. Ethical in dealing with people and organizations; and
4. Practical in using resources.

Meta-evaluations have three main characteristics: 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, subsidizing 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 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 1 and STEP 2 scores.

Independent Variable

The independent variables include participation in the original vs modified remediation program at both SOMs.

Data Collection

Data needed to perform the evaluation of the modified remediation program will be collected using STARS (an application used at both SOMs), 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 version 26) will be used. A nominal alpha level of 0.05 will be used as the significance level. A Hotelling's Two Sample T^2 analysis will be conducted.

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

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

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

Limitations

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

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

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 (Toftthagen, 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 (Toftthagen, 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.

Results

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.

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 nonnormality. Nevertheless, Levene's test of equality of error variances for all three dependent variables were also statistically significant. Therefore, the results of the Hotelling's T^2 test must be interpreted with caution.

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

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 ($F = 1.57$, $df = 1165$, $p = 0.212$) and USMLE1 ($F = 1.69$, $df = 1165$, $p = 0.196$), but was statistically significantly different for USMLE2 ($F = 5951$, $df = 1165$, $p = 0.016$).

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.

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.

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. 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

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

content of the phase is still valid, timely and aligned with the overall intent of the evaluation.

Conclusion

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.

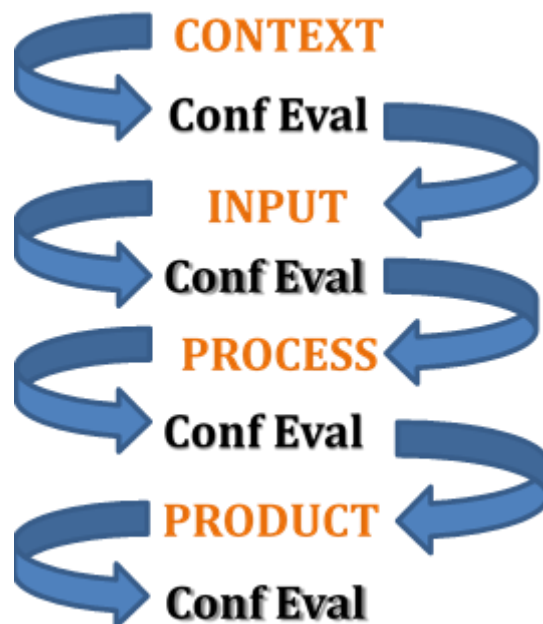


Figure 1. Revised CIPP model

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.

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

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.

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.

References

- Alqahtani, K. (2016). Decision-oriented evaluation: A review of various models of evaluation. *International Journal of Scientific & Engineering Research*, 7(5), 929-391. Retrieved from <https://www.ijser.org/onlineResearchPaperViewer.aspx?Decision-oriented-Evaluation--A-Review-of-Various-Models-of-Evaluation.pdf>
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, 80(3), 260-267.
- Artino, A. R., & La Rochelle, J. S., & Durning, S. J. (2010). Second-year medical students' motivational beliefs, emotions, and achievement. *Medical Education*, 44(12), 1203-1212. doi: 10.1111/j.1365-2923.2010.03712.x

- Aziz, S., Mahmood, M., & Rehman, Z. (2018). Implementation of CIPP model for quality evaluation at school level: A case study. *Journal of Education and Educational Development*, 5(1), 189-206. doi: [10.22555/joeed.v5i1.1553](https://doi.org/10.22555/joeed.v5i1.1553)
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education Principles, Policy & Practice*, 5(1), 103-110. doi: [10.1080/0969595980050106](https://doi.org/10.1080/0969595980050106)
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. doi: [10.1080/0969595980050102](https://doi.org/10.1080/0969595980050102)
- Black, P. J., Harrison, C., Lee, C. S., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.
- Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47-63). New York: Holt, Rinehart & Winston.
- Campbell, D., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.
- Carroll, M. (1994). Library-based projects for first-year medical and dental students: A critical assessment. *Biochemical Education*, 22(2), 75-78.
- Cleland, J., Arnold, R., & Chesser, A. (2005). Failing finals is often a surprise for the student but not the teacher: Identifying difficulties and supporting students with academic difficulties. *Medical Teacher*, 27(6), 504-508. doi: [10.1080/01421590500156269](https://doi.org/10.1080/01421590500156269)
- Conrad, R. M., & Donaldson, J. A. (2004). *Engaging the online learner: Activities and resources for creative instruction*. San Francisco: Jossey-Bass
- Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. Cambridge: Cambridge University Press. doi: [10.1017/CBO9781139173582](https://doi.org/10.1017/CBO9781139173582)
- Cranton, P. A., & Legge, L. H. (1978). Program evaluation in higher education. *Journal of Higher Education*, 49(5), 464-471. doi: [10.1080/00221546.1978.11780396](https://doi.org/10.1080/00221546.1978.11780396)
- Dessinger, J. C., & Moseley, J. L. (2003). *Confirmative evaluation: Practical strategies for valuing continuous improvement*. New York: John Wiley & Sons.
- Donaldson, L. (2001) *The contingency theory of organizations*. New York: Sage Publications, Inc. doi: [10.4135/9781452229249](https://doi.org/10.4135/9781452229249)
- Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation.

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

American Journal of Evaluation, 23(3), 261-273. doi:

10.1177/109821400202300303

Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE guide no. 67. *Medical Teacher*, 34(5), e288-e299. doi: 10.3109/0142159X.2012.668637

Guerra-López, I., & Hutchinson, A. (2008). Measurable and continuous performance improvement: The development of a performance measurement, management, and improvement system. *Performance Improvement Quarterly*, 26(2), 159-173. doi: 10.1002/piq.21151

Hellebrandt, J., & Russell, J. D. (1993). Confirmative evaluation of instructional materials and learners. *Performance and Instruction*, 32(6), 22-27.

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd edition). Thousand Oaks, CA: Sage.

Kaufman, R. A., & Thomas, S. (1980). *Evaluation without fear*. New York: New Viewpoints.

Liaison Committee on Medical Education. (2010). *Functions and structure of a medical school: Standards for accreditation of medical education programs leading to the MD degree*. Washington, DC: Association of American Medical Colleges.

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(2), 25. doi: 10.5688/aj740225

Mann, C. J. (2003). Observational research methods. Research design II: Cohort, cross sectional, and case-control studies. *Emergency Medicine Journal*, 20(1), 54-60. Retrieved from <https://emj.bmj.com/content/20/1/54>

Mendel, D., Jamieson, A., & Whiteman, J. (2013). Remediation. In K. Walsh (Ed.), *Oxford textbook of medical education* (pp. 362-371). Oxford, UK: Oxford University Press. doi: 10.1093/med/9780199652679.003.0031

Morrison, G. R., Ross, S. J., Kalman, H. K., & Kemp, J. (2013). *Designing effective instruction* (7th edition). Hoboken, NJ: Wiley & Sons, Inc.

Patton, M. Q. (2013). Meta-evaluation: Evaluating the evaluation of the Paris declaration. *The Canadian Journal of Program Evaluation*, 27(3), 147-171.

Pintrich, P. R., & Schrauben, B. (1992). Students' motivational beliefs and their cognitive engagement in academic tasks. In D. H. Schunk & J. L. Meece

(Eds.), *Student perceptions in the classroom* (pp. 149-183). Hillsdale, NJ: Erlbaum.

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. Retrieved from http://www.jhetonline.com/2015_issue_of_the_jhet

Rowntree, D. (1987). *Assessing students: How shall we know them?* (2nd edition). London: Harper and Row.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144. doi: 10.1007/BF00117714

Samet, J. M., & Munoz, A. (1998). Evolution of the cohort study. *Epidemiologic Reviews*, 20(1), 1-14. doi: 10.1093/oxfordjournals.epirev.a017964

Sayer, M., Bowman, D., Evans, D., Wessier, A., & Wood, D. (2002). Use of patients in professional medical examinations: Current UK practice and the ethicolegal implications for medical education. *BMJ*, 324, 404-407. doi: 10.1136/bmj.324.7334.404

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (pp. 39-83). Chicago: Rand McNally & Co.

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

Stufflebeam, D. L. (1971, February 24). *The relevance of the CIPP evaluation model for educational accountability*. Paper presented at the annual meeting of the American Association of School Administrators, Atlantic City, NJ.

Stufflebeam, D. L. (1978). Meta evaluation: An overview. *Evaluation and the Health Professions*, 1(1), 17-43. doi: 10.1177/016327877800100102

Stufflebeam, D. L. (2000). The CIPP model for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models* (2nd edition, pp. 279-317). Dordrecht, Netherlands: Springer. doi: 10.1007/0-306-47559-6_16

Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 2001(89), 7-98. doi: 10.1002/ev.3

Stufflebeam, D. L. (2003). The CIPP model for evaluation: An update, a review of the model's development, a checklist to guide implementation. 2003

CONFIRMATIVE EVALUATION: NEW CIPP EVALUATION MODEL

Annual Conference of the Oregon Program Evaluators Network (OPEN), Portland, OR.

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

Toftthagen, C. (2012). Threats to validity in retrospective studies. *Journal of the Advanced Practitioner in Oncology*, 3(3), 181-183. doi: [10.6004/jadpro.2012.3.3.7](https://doi.org/10.6004/jadpro.2012.3.3.7)

Winston, K. A., Van Der Vleuten, C. P. M., & Scherpbier, A. J. J. A. (2012). The role of the teacher in remediating at-risk medical students. *Medical Teacher*, 34(11), e732-e742. doi: [10.3109/0142159X.2012.689447](https://doi.org/10.3109/0142159X.2012.689447)

Worthen, B. R., & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman Press.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program evaluation: Alternative approaches and practical guidelines* (2nd edition). New York: Longman.

Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501. doi: [10.1023/A:1023967026413](https://doi.org/10.1023/A:1023967026413)

Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misulis, K. (2011). 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://openjournals.libs.uga.edu/jheoe/article/view/901>