

1-1-2014

A Case Study Of The Impact Of A Sytematic Evaluation Process In A Graduate Medical Education Residency Program

Heidi T. Kromrei
Wayne State University,

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



Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Other Education Commons](#)

Recommended Citation

Kromrei, Heidi T., "A Case Study Of The Impact Of A Sytematic Evaluation Process In A Graduate Medical Education Residency Program" (2014). *Wayne State University Dissertations*. Paper 896.

This Open Access Dissertation is brought to you for free and open access by DigitalCommons@WayneState. It has been accepted for inclusion in Wayne State University Dissertations by an authorized administrator of DigitalCommons@WayneState.

**A CASE STUDY OF THE IMPACT OF A SYSTEMATIC EVALUATION
PROCESS IN A GRADUATE MEDICAL EDUCATION RESIDENCY
PROGRAM**

by

HEIDI T. KROMREI

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

2014

MAJOR: INSTRUCTIONAL TECHNOLOGY

Approved By:

Advisor

Date

**© COPYRIGHT BY
HEIDI T. KROMREI
2014
All Rights Reserved**

DEDICATION

This dissertation is dedicated to my mother, Diana Kromrei, and to my daughters, Ingrid and Elise who are my inspiration for all things.

ACKNOWLEDGMENTS

The author would like to express gratitude for the exceptional patience and helpful guidance of the dissertation committee. The committee Chair, Ingrid Guerra-López PhD, provided the intellectual foundation for much of this study through her “Impact Evaluation Process” model and is a greatly valued and appreciated mentor. Willhelmine Wiese-Rometsch, MD is an irreplaceable colleague and guide for all things GME as well as major contributor to the concepts utilized in this study. Timothy Spannaus PhD has been a serene and reassuring resource throughout this process. James Moseley EdD is a role model for student mentorship and an invaluable resource of exceptional knowledge of medical education, evaluation and performance improvement.

Life’s journey is infinitely richer when accompanied by friends. Special thanks to my dissertation support system, my oldest friend and mentor Kip and my dearest twibling brother, Jack. Thank you to my colleagues, Carol Bartley and Vanessa Sykes, who encouraged and supported me through the data collection and analysis years.

This research would not be possible without the support and extraordinary efforts of those associated with the Kresge Eye Institute Ophthalmology residency program. The diligent efforts and contributions of Deborah Chesney and Ingrid Ciungan were crucial to this project, as they collected mountains of data and provided support during every year of the annual program evaluation process and beyond. The faculty and residents of the KEI program contributed extraordinary efforts toward program evaluation and I am most grateful for their outstanding work. Lastly, without the dedication, vision, wisdom, support and enthusiasm of Dr. Mark Juzych, this research would not exist.

TABLE OF CONTENTS

Dedication.....	ii
Acknowledgments.....	iii
List of Tables.....	vii
Chapter 1 Introduction.....	1
<i>Antecedents</i>	1
<i>Description of the Problem</i>	10
<i>Research Questions</i>	12
<i>Justification of the Problem</i>	12
<i>Practical Support for Program Evaluation in GME</i>	13
<i>Theoretical Support for Program Evaluation in GME</i>	16
<i>Theoretical Framework</i>	20
<i>Operational Definitions</i>	23
<i>Limitations and Delimitations</i>	27
<i>Summary</i>	27
Chapter 2 Literature Review.....	29
<i>Self-Assessment</i>	30
<i>HPT, Performance Measurement, and Evaluation</i>	32
<i>Opportunities for Methodological Improvement in Evaluation</i>	34
<i>Participant Evaluation</i>	35
<i>Building Evaluation Capacity</i>	37
<i>Program Evaluation in GME</i>	38

<i>Summary</i>	39
Chapter 3 Method.....	41
<i>Overview</i>	41
<i>Case Study Design</i>	41
<i>Setting: Graduate Medical Education</i>	42
<i>Case Selection for the Study</i>	44
<i>Data Collection and Analysis Plan</i>	45
<i>Description of Data Sources and Research Instruments</i>	51
<i>Data Collection Instruments Developed for the Study</i>	53
<i>Data Analyses</i>	54
<i>Reliability and Validity</i>	57
<i>Additional Procedures</i>	60
<i>Study Limitations</i>	60
<i>Summary</i>	61
Chapter 4 Results	
<i>Annual Program Evaluation Results</i>	62
<i>Program Outcomes</i>	70
<i>Stakeholder Perspectives</i>	74
<i>Summary</i>	84
Chapter 4 Discussion	
<i>Conclusions</i>	86
<i>Annual Program Evaluation Results</i>	87
<i>Program Outcomes</i>	93

<i>Stakeholder Perspectives</i>	98
<i>Study Limitations</i>	101
<i>Suggestions for Improvements to the Evaluation Process</i>	102
<i>Summary</i>	103
References.....	106
Abstract.....	120
Autobiographical Statement.....	122

LIST OF TABLES

Table 1: Adaptation of Impact Evaluation Process.....	23
Table 2: Data Collection and Analysis Plan.....	48
Table 3: Qualitative Analysis Purpose.....	55
Table 4: Performance Improvement Recommendations by Year, Performance Domain, Recommendation Type, Actionable Status, and Alignment to Performance Deficit.....	66
Table 5: Performance Improvement Recommendation Types and Completion Rates 2009-2013.....	69
Table 6: ACGME Accreditation Outcomes.....	70
Table 7: Program OKAP (In-Service) Examination Results 2007 to 2013.....	71
Table 8: Program Graduate Board Examination Results Evaluation 2007 to 2011.....	72
Table 9: ACGME Ophthalmology Resident Case Log Report Results.....	72
Table 10: ACGME Annual Resident Survey Data 2010 to 2013.....	73
Table 11: 2011 Stakeholder Perceptions of Anticipated Benefits of Systematic Evaluation Process.....	75
Table 12: 2011 Stakeholder Perceptions of Actual Benefits of Systematic Evaluation Process.....	76
Table 13: 2011 Stakeholder Perceptions Anticipated Challenges of the Systematic Evaluation Process.....	77
Table 14: 2011 Stakeholder Perceptions Actual Challenges of Systematic Evaluation Process.....	78
Table 15: 2011 Participant Lessons Learned.....	79
Table 16: 2011 Stakeholder Suggestions for Improving the Systematic Evaluation Process.....	80
Table 17: 2013 Program Administrator of 3 years of Systematic Evaluation Process: Program Improvements.....	81

Table 18: 2013 Program Administrator of 3 years of Systematic Evaluation Process:
Benefits of the Process.....81

Table 19: 2013 Program Administrator Perceptions of 3 years of Systematic
Evaluation Process: Challenges of the Process.....82

Table 20: 2013 Program Administrator of 3 years of Systematic Evaluation Process:
Personal Impact.....83

Chapter I. Introduction

Educating physicians is a complex and crucial practice that is highly regulated and closely monitored in the United States. Institutions engaged in undergraduate and graduate medical education must consistently measure and monitor performance at multiple levels, institutional, program, and individual. Annual program evaluation is one of the numerous regulatory requirements for graduate medical education programs (residency and fellowships). Many residency and fellowship programs do not comply with this requirement and published outcomes from this mandatory process are lacking.

The current study sought to examine the utility and efficacy of the application of a systematic self-evaluation process in a single graduate medical education residency program and to compare the results of this process to previous program self-evaluation efforts.

Antecedents

The path to becoming a board certified physician in the United States is long and arduous. Medical education encompasses up to 15 years study, including undergraduate (4 years of medical school); graduate training (3 to 5 years of initial specialty training); and post-graduate subspecialty training (up to 6 years of additional fellowship training), all of which is typically completed after obtaining a four-year baccalaureate degree.

The *business* of educating physicians is equally laborious; regulations and requirements for educational institutions are comprehensive and exacting and the stakes are high for the institutions and stakeholders invested in the medical educational process. Medical schools are an important source of revenue for Universities and the communities that surround them. The American Association of Medical Colleges report that in 2012

the median revenue of the 126 medical schools in the United States was \$574 million dollars, with private medical school revenue median of \$648 million (“LCME I-A, Annual Financial Questionnaire, FY2012”). According to the American Medical Association (“Critical condition,” n.d.) current funding for graduate medical education includes dollars from Medicare (\$9.5 billion); Medicaid (\$2 billion); and the Department of Veterans Affairs (\$1 billion) generating approximately 12.5 billion dollars in resource dollars largely tied to hospital settings. Educating physicians generates significant income for Universities, hospitals, and the communities that surround them while simultaneously providing essential medical services. The current economic climate is necessitating consideration of drastic cuts to universities and Medicare payments to teaching hospitals; President Obama’s 2014 federal budget proposes a reduction in graduate medical education payments in the amount of \$780 million in 2014 and close to \$11 billion over ten years (Lubell, 2013; Miesen, 2013). During these stringent economic times for graduate medical education, hospital and University administrators must find ways to maximize resources and react to budget cuts while simultaneously continuing to produce excellent educational and patient care outcomes.

Undergraduate Medical Education in the United States

Undergraduate medical education comprises the four years of education students receive during medical school. There are two types of medical schools in the United States, allopathic (MD) and osteopathic (DO). Both types of medical schools are overseen by national organizations that hold them to a rigorous set of accreditation standards. The majority of medical schools in the United States are allopathic; the Liaison Committee on Medical Education (LCME) accredits these programs in the United States

and Canada. LCME accreditation standards include comprehensive requirements for the institution; the educational program structure; curriculum design and management; and evaluation of program effectiveness, among others (Functions and Structure of a Medical School, 2013). The American Osteopathic Association's Commission on Osteopathic College Accreditation (COCA) accredits Osteopathic medical school programs in the United States and Canada. COCA mandates eight accreditation standards for medical schools; similar to the LCME, COCA requirements include institutional and curricular in addition to a self-study component.

Graduate Medical Education Accreditation in the United States

Graduate medical education (GME), encompassing the years of specialty training after medical school is completed, is also a highly regulated educational system in the United States. Similar to the undergraduate process of accreditation, allopathic and osteopathic residency and fellowship programs (post-doctoral training programs) must adhere to rigorous accreditation standards imposed by non-governmental agencies composed of peers. Osteopathic residency programs are accredited and evaluated by the Council on Osteopathic Postdoctoral Training Institutions (COPTI); allopathic residency and fellowship programs are accredited and evaluated by the Accreditation Council for Graduate Medical Education (ACGME). Both councils include infrastructure to evaluate and accredit GME sponsoring institutions (e.g., hospitals, universities, medical schools, public health agencies, etc.) as well as individual residency training programs (e.g. internal medicine, ophthalmology, general surgery, radiology, etc.). One of the functions of accreditation is to provide medical school and post-doctoral programs an opportunity for critical self-analysis, which is expected to lead to improvements in quality.

Allopathic graduate medical education programs must adhere to the ACGME “common program requirements” (applicable to all specialty and subspecialty training programs) as well as “specialty-specific requirements” (additional ACGME requirements particular to the medical specialty or subspecialty). A Residency Review Committee (RRC) is established for each of the major specialty areas with responsibility to accredit programs in the general specialty and related subspecialties. Residency programs must demonstrate substantial compliance with both common and specialty program requirements to maintain ACGME accreditation.

Osteopathic post-doctoral training institutions (OPTI) must adhere to the AOA basic standards. The AOA Program and Trainee Review Council (PTRC) is the body that monitors and oversees DO residency training programs and determines program accreditation status. According to the AOA, “The accreditation process involves systematic examination and peer examination and evaluation of all aspects of the educational impact and effectiveness of an OPTI as measured against AOA-approved standards” (“The Basic Documents,” 2013, p.4).

Accreditation of residency programs governed by the ACGME includes evaluation by the ACGME RRC, resulting in a determination of program accreditation status (e.g., initial accreditation, probationary accreditation, and maintenance of accreditation) with commendations for exceptional compliance, and citations for substantial lack of compliance. Programs with a significant number of citations for non-compliance are required to submit additional progress reports and may be subject to additional documentation requirements and/or a “focused site visit” (assessment of selected program aspects conducted by ACGME field representatives). Repeated citations

may lead to a reduction in resident compliment, program probationary accreditation status, or withdrawal of program accreditation. Given the potential for significant loss of funding to support residency programs and the lower-cost health care services residents provide, the loss of residency program accreditation can significantly impact a hospital's bottom line as well as the ability to provide safe and effective patient care.

Graduate Medical Education Accreditation

Abraham Flexner's (1910) compelling report critiquing medical education in the United States led to what eventually became the regulatory process of accreditation of medical education in the United States and Canada. Regulation and oversight has evolved to include multiple accreditation committees and commissions that oversee undergraduate and graduate medical education programs. The American Association of Medical Colleges notes that there are currently 141 U.S. and 17 Canadian accredited medical schools (AAMCb, 2013). As of July 2013 the ACGME oversees more than 9,040 accredited Graduate Medical Education Specialty Programs in the United States (ACGME Report, 2013) and the AOA oversees 718 programs in the United States (AOA, Summary of Positions Offered and Filled by Program Type, 2013).

ACGME aims to improve health care through the accreditation process, citing the need for a structured approach to competency evaluation and the provision of customized formative feedback (ACMGE Mission, Vision and Values, accessed online 09/01/13). The AOA Council on Postdoctoral Training also aims to ensure optimal health outcomes while enhancing educational quality and improving compliance (AOA Basic Documents for Postdoctoral Training, p. 6). The progression of this accreditation process led to a focus on educational and patient outcomes, while ensuring that physicians in training are

not only properly and thoroughly educated, but that the institutions overseeing this process are consistently monitoring the training process and results.

Residency Program Directors are charged with responsibility for the oversight and administration of GME specialty and/or subspecialty GME programs and must ensure that their educational programs comply with ACGME and/or AOA requirements. One such residency program accreditation requirement relates to program evaluation and improvement. Both accrediting bodies require that residency program directors ensure that the educational program is evaluated at least annually, and both provide further evaluation specifications.

As of 2011, both the ACGME and the AOA accreditation requirements mandated that residency programs must conduct program evaluations. Since 2007, the ACGME required a “formal systematic evaluation of the curriculum at least annually,” to include monitoring and tracking of resident performance, graduate performance, faculty development and program quality (“Common Program Requirements, V.C.” pp. 11-12). Further, the ACGME required that a performance improvement action plan be developed if program deficiencies were found. The 2011 AOA Basic Documents for Postdoctoral Training also included a mandate for evaluation of training programs and faculty (AOA Basic Documents for Postdoctoral Training, BOT 7/2011, pp. 50-51) requiring that “provisions should be made for various levels of program evaluation...the results of these evaluations should be used to continually improve the program...the Medical Education Committee shall evaluate the intern training program quarterly.”

ACGME recently initiated a revised accreditation system entitled, the “Next Accreditation System” (NAS) that began implementation in seven medical specialty

residency programs July 1, 2013 and continues with all medical specialty residencies and fellowships as of July 1, 2014. Annual program evaluation will take on increased significance as the process of accreditation moves to a continuous accreditation model where performance indicators are analyzed annually and in-person external audits (site visits) are performed every ten years (unless data reported to the ACGME warrant a more frequent schedule). Updated requirements include the formation of a “Program Evaluation Committee” (PEC) and more explicit guidelines for annual program evaluation. The PEC must now not only prepare a written action plan for program improvement each year, but must also note how these actionable items will be measured and monitored (ACGME Common Program Requirements NAS, effective 07/01/13, section V.C.3, p. 13).

According to the ACGME Accreditation Policies and Procedures Manual (Section: 17:30 a, p. 75), effective July 1, 2013, the ACGME will initiate “Self-Study Visits” which will include expectation of residency program documentation of continuous program self-evaluation:

The 10 year Self-Study site visit is based on a comprehensive self-study, which includes a description of how the program or sponsoring institution creates an effective learning and working environment, and how this leads to desired educational outcomes, and an analysis of strengths, weaknesses, and plans for improvement.

As per the accreditation requirements of both the ACGME and AOA regulatory bodies, GME residency program directors are required to develop a method for analyzing, designing, developing, implementing, and self-evaluating their educational training program performance typically without the benefit of evaluation or performance management expertise. Program evaluations are expected to facilitate continuous

improvement of residency program performance, yet published evidence of program evaluation outcomes in GME are lacking.

A Human Performance Technology Approach to Evaluation and Improvement

The Human Performance Technology (HPT) field offers numerous performance improvement models that take evaluative (Kirkpatrick, 1994; Kaufman, 1996) and strategic (Rummler & Brache, 1995; Kaufman & Watkins, 2000; Kaufman, 2005; Kaufman, 2006) approaches. According to Guerra López (2012, p.43), professionals in the field of performance improvement (PI) are in a unique position to “help the field grow further and achieve sustainability” through the utilization of methods proven to add value to any industry, that is, through effective needs assessment and evaluation techniques. The performance improvement field provides ample models and interpretations of needs assessment (Leigh, Watkins, Platt & Kaufman, 1998; Watkins & Guerra, 2003) and evaluation methods (Kirkpatrick, 1994; Dessinger & Moseley, 2004; Guerra-López, 2007a,b, c) with an emphasis on continuous evaluation and performance measurement and management.

Guerra-López’s Impact Evaluation Process (2007, 2011) exemplifies a systematic PI evaluation process that provides organizations effective steps for planning and implementing evaluation that leads to high impact performance outcomes. The Impact Evaluation process allows stakeholders to answer important evaluation questions concerning the efficacy and impact of projects, interventions, and solutions while simultaneously determining which, if any, internal targets were reached (Guerra López, 2007).

Dessinger & Moseley's (2004) emphasis on confirmative evaluation processes demonstrates HPT practitioners' rationale for taking a long-term view of the effectiveness, efficiency, impact and value of a training program. Doing so helps decision makers manage the instructional performance system and focuses on the program's continuing impact and value (Van Tiem, Moseley, & Dessinger, 2000). Medsker (2006) illustrates that PI-focused evaluation emphasizes strategy and by doing so addresses how the results will be used and encourages continuous improvement. These evaluation processes provide a road map for the development of tailored interventions for meeting the ACGME and AOA requirements for annual residency program evaluation and ensuring that such evaluation processes can lead to meaningful performance improvement.

Further, the effective application of performance improvement-focused evaluation models in a variety of fields and disciplines provides opportunity for continued expansion and sustainability in the field of PI, as Guerra López (2012, p.44) posits:

It would behoove practitioners to explore opportunities beyond their familiar boundaries and challenge themselves to solve important problems across all sectors of society. Likewise, it is important for researchers to explore cross-disciplinary research where improvement methods can be applied, tested, improved, and showcased.

Kaufman (2012, p.7) also professes that expanding the scope of HPT is essential to the "...future validity, ethics, and usefulness of our field..."

Description of the Problem

Medical education programs at both the undergraduate and graduate levels offer scant, if any, physician training in business practices such as performance measurement, management, or performance evaluation methodology. Yet, as noted previously, the ACGME and AOA require that programs must adhere to requirements that mandate “Systematic collection and analysis of information related to the design, implementation, and outcomes of a resident education program, for the purpose of monitoring and improving the quality and effectiveness of the program ” (“ACGME Glossary”, 2013 p.78).

A variety of tools have been developed for residency program evaluation including program report cards (Phitayakorn, Levitan, and Shuck, 2007) and surveys of faculty and residents, (Bellini, Shea and Asch, 1997; Liebelt, Daniels, Farrell and Myers, 1993), but there is a lack of published evidence indicating the utility and efficacy of residency program evaluation methods. Musick (2006) proposed a five step conceptual model for GME program evaluation, 1.) determining the evaluation need, 2.) determining the evaluation focus, 3.) determining the evaluation method, 4.) presenting the evaluation findings, and 5.) documenting the evaluation results. Durning, Hemmer, and Pangaro (2007) suggest a “Before, During, and After” model for undergraduate and graduate medical education program evaluation using baseline, process, and product measurements. Other models and structures have been proposed (Vassar, Wheller, Davison, and Franklin, 2010) and some have begun to collect system-wide surveys regarding program performance (McOwen, Bellini, Morrison, and Shea, 2009). While these tools and models offer suggestions for evaluation processes, they lack published

outcomes of the relationship between evaluation results and improved performance outcomes.

Peer-reviewed publications of applications and resulting outcomes of a systematic program evaluation in graduate medical education are rare; to date only one such published study could be found. System-wide use of a Duke University institutional template for program evaluation resulted in increased compliance with ACGME requirements for annual program evaluation, easier documentation for site visits, and fewer ACGME citations (Andolesek, Nagler, and Weinerth, 2010); however, specific evaluation results and program performance improvement outcomes were not reported.

Purpose

The purpose of the present evaluation research was to examine the utility, efficacy, and challenges of applying a systematic evaluation process to the required annual program evaluation of a residency program. Unlike previous work, which proposes theoretical evaluation models or the utility of an evaluation plan that ensures only compliance with regulatory requirements, this study analyzed the utility of a self-evaluation process in a case study as it relates to the performance improvement plans generated, the performance outcomes resulting from this self-evaluation process, and the perspectives of the participants about the self-evaluation process in a ACGME accredited residency program.

Research Questions

The study aims to answer the following research questions:

1. Do the evaluation results generated using a systematic evaluation process (e.g. evaluation findings and recommendations for improvement) differ from previous years' annual program evaluation results?
2. Does the utilization of a systematic evaluation process lead to action-based performance improvement plans tied to specific performance gaps?
3. Does using a systematic evaluation process result in improved program outcomes (e.g., adherence to requirements, management of program performance, educational outcomes, implementation of solutions)?
4. What are the stakeholders' perceptions of barriers to and benefits of implementation of a systematic evaluation process?

Justification of the Problem

As noted previously, regulatory agencies that oversee graduate medical education require residency programs to conduct an annual program evaluation. Proposed residency program evaluation models (Musick, 2006; Durning, Hemmer, and Pangaro, 2007; Vassar, Wheller, Davision, and Franklin, 2010) fail to report either action-based recommendations for measurably improved performance or improved program outcomes, essential findings for the performance improvement and evaluation fields. The field of performance improvement offers practical and theoretical support for designing, developing, implementing and evaluating a systematic process for residency program evaluation in graduate medical education and has called for an increase in research and

publications demonstrating added value to our clients (Stolovitch and Keeps, 1999; Kaufman and Clark, 1999; Stolovitch, 2000; Sugrue and Stolovitch, 2000; Guerra-Lopéz and Leigh 2009).

Practical Support for Program Evaluation in Graduate Medical Education

Evaluation researchers continue to explore the concept of utility of formal evaluation, the ways in which stakeholders intend and ultimately use evaluation results (Patton, 2002; Guerra-Lopez, 2007). Evaluation processes, if implemented systematically and with proactively derived objectives, can result in "...action-based recommendations for measurably improving performance" (Guerra-Lopez, 2007, p.33). It is essential that both researchers and practitioners of evaluation methods consider the practical advantages when designing and implementing evaluations. The potential for practical support offered by this evaluation study includes 1.) the efficient use of program resources, 2.) effective medical professional training, 3.) improved graduate medical education program performance, and 4.) increased compliance with regulatory requirements.

Efficient use of resources: The institutions that sponsor GME programs receive the majority of their public funding for physician training from Medicare. Residency training program administrators subsequently receive program-level funding from the hospitals and universities that sponsor physician training. These funds must cover a variety of educational expenses including resident salary and benefits, faculty teaching and administration salaries, educational expenses, and some portion of the clinical costs associated with training. Universities and hospitals note that Medicare funding alone does not fully support the cost of physician training. For example, a recent report from the

University of California cites that it receives approximately \$100,000 of Medicaid funding per resident per year but incurs direct and indirect costs of approximately \$200,000 per year to train each resident (UCHealth, 2012).

Restricted budgets demand that residency program administrators make the most of the funding they receive through the efficient use of the resources available to them. Methods that engage programs in systematic evaluation to improve performance are ideally suited to ensure that limited funding is used in ways that will provide the best training and educational opportunities while simultaneously using publicly funded dollars wisely. Further, the highest performing programs are likely to experience fewer accreditation site visits, leading to less time, money, and resources expended on the substantial preparation and documentation required.

Effective medical professional training: The charge of ensuring that physicians are adequately trained is an enormous responsibility; patients' lives literally depend upon it. The provision of effective physician training requires that programs understand and identify measurable performance objectives and evaluate their success at meeting these objectives. Many residency program performance objectives are explicitly stated in the regulatory requirements (e.g., ACGME common program requirements, AOA basic standards) but the path to compliance is largely left to the residency program administrators to forge. The utilization of systematic evaluation processes can ensure that residency program performance outcomes are analyzed and compared to objectives (Guerra-Lopéz, 2007) and that changes to educational programs are made as a result of data driven decisions. Data-driven decisions should result in the implementation of

improvements that will lead to more effective graduate medical education and professional training for the physicians enrolled.

An additional anticipated benefit of implementing a systematic evaluation process to examine residency program performance is participation of physicians and residency program staff in the evaluation process itself.

Improve graduate medical education program performance: Residency programs are educational systems designed to ensure that physicians are trained to practice medicine competently and independently. The utilization of systematic evaluation processes to analyze residency program performance provides an opportunity to compare the current program performance to the desired program performance and identify the needed program improvements. The proper use of evaluation methods can ensure that the solutions chosen for program improvement are a reflection of analysis of the right data, increasing the likelihood of improved educational outcomes and, thus, better-educated, more competent program graduates. Improving program performance may also lead to fewer accreditation site visits, less program citations, and more potential commendations (external indicators of educational quality).

Increase organizational and stakeholder competence with evaluation and improving performance: Involving program stakeholders in the process of evaluation has been shown to accrue multiple benefits. Johnson, Greenseid, Toal, King, Lawrenz, and Volkov's (2010) review of the empirical literature on evaluation use from 1986 to 2005 noted that stakeholder involvement in the evaluation process strengthened commitment in the evaluation process and lead to greater use of evaluation results. It is hoped that stakeholders' exposure to and participation in a systematic evaluation process

in a residency program will increase administrators' ability to perform evaluations that lead to improved performance.

Although physicians receive a great deal of training in algorithms to diagnose patient pathology, they receive very little, if any, training in conducting evaluations. Yet, they are charged with developing an entire evaluation system for their residency program. Engaging stakeholders at all program levels in the process of program evaluation provides an opportunity to introduce both physicians and administrators to the systematic processes involved in evaluation and the potential impact such methods hold for improving performance. In this research study, stakeholders at all levels are involved in the evaluation process from identifying the questions to be answered to analyzing the data and making recommendations for program improvement.

Building evaluation capacity within residency programs is not only an essential business practice, but also a regulatory requirement (ACGME Common Program Requirements, 2011, 2013). Clearly defined roles for evaluation participants and support for the necessary components of a systematic evaluation process are fundamental. The process of building evaluation capacity offers opportunity to decrease the likelihood of participants feeling threatened by evaluation, increase program staff knowledge, improve understanding of evaluation issues and improve data tracking systems (McDonald, B, Rogers, P., & Kefford, B., 2003). Indeed, the potential rewards of building evaluation capacity are many.

Theoretical Support for Program Evaluation in Graduate Medical Education:

Improving performance in the field of medicine is a topic of utmost importance in the United States. Estimates of economic loss due to waste in the U.S. health care system

reached \$750 billion dollars in 2009 (Smith, Saunders, Stuckhardt, McGinnis 2012). A recent report from the Institute for Medicine, “Best Care at Lower Cost: The Path to Continuously Learning Health Care in America” asserts that “The foundation for a learning health care system is continuous knowledge development, improvement, and application.” (Smith, et al, 2012, p. Ab-2). The American Medical Association formed a Physician Consortium for Performance Improvement[®] with the aim of “identifying and developing evidence-based performance measures and measurement resources that enhance the quality of patient care and foster accountability... promoting the implementation of effective and relevant clinical performance improvement activities... and advancing the science of clinical performance measurement and improvement” (AMA, 2010). The medical field is highly receptive to performance improvement initiatives and the field of human performance technology is uniquely poised to assist through the provision of performance improvement models, processes, research, and theory.

According to the Association of American Medical Colleges, enrollment in medical schools in the United States has consistently risen in the last decades and will likely increase enrollment by the targeted 30 percent by 2017 (AAMC, 2013). The seemingly ever-growing industry of health care is in need of performance improvement expertise. Introducing performance improvement approaches, such as systematic evaluation processes, during physician training years provides opportunity to educate a new generation of physicians in performance improvement and evaluation methodology.

The call for validation of HPT’s assertion of adding value and achieving desired results for clients has been heard for more than a decade with less than ideal results

(Stolovitch and Keeps, 1999; Kaufman and Clark, 1999; Stolovitch, 2000; Sugrue and Stolovitch, 2000). The types of articles published in *Performance Improvement Quarterly* from 1997 to 2000 were analyzed by Klein (2002) who found only 36% were publications of empirical research, suggesting, “more research on the effects of non-instructional performance interventions should be conducted and published in the literature.” (p.105). Replication of Klein’s study reveals a rise in the percentage of data-based articles published in *Performance Improvement Quarterly* (to 54% from 2001 to 2005), but stress a continued need to encourage empirical work in the field of HPT (Conn and Gitonga, 2004).

Performance improvement researchers and practitioners must also consider broadening the contexts and settings in which they choose to work and publish. Huglin (2009) reviewed the citation patterns of references cited in the *International Society for Performance Improvement* journal articles (1962-2007) and noted primary subject categories for performance improvement citations. The most frequently cited primary subject categories included psychology (161 cites); business and economics: management (133 cites); education: (105 cites); business and economics (74 cites); education: teaching methods and curriculum (62 cites); education: higher education (56 cites); business and economics: personnel management (41 cites); business and economics: marketing and purchasing (38 cites); medical sciences (35 cites); and sociology (35 cites). Hughlin’s study also revealed that human performance practitioner publications tend to cite their own literature more than that of other cognate fields (2009). If the academic prowess of the field will be judged through the depth of publication in peer reviewed journals, the

field of performance improvement must seek to broaden the horizons in which we work, consult, and publish.

Performance improvement professionals are poised to apply the theories, models and methods of their field in a variety of contexts. However, reviews of the empirical literature reveal that the field will benefit from expanding the application of performance improvement applications and research to a broader variety of settings utilizing more rigorous research methods. Guerra-López and Leigh's (2009) analysis of the performance improvement literature and the current use of evaluation and measurement in the field of performance improvement note that "The data, in particular those related to our practitioner journal, *PIJ* [Performance Improvement Journal], reveal that our attention to evaluation and measurement is not at a level that supports our claims to add measurable value to our clients." (p. 107). It is essential therefore, that empirical performance improvement studies are conducted in a variety of fields and subsequently published in peer reviewed journals to demonstrate the utility of the application of performance improvement theories, models, processes and methods.

The current study attempted to establish the degree to which self-evaluation of a residency program using a systematic, performance improvement-focused evaluation process improved the quality of recommendations generated, leads to action-based performance improvement plans tied to specific performance gaps, reduces non-compliance with regulatory requirements, and improves educational outcomes. The current study explored the use of a modified version of Ingrid Guerra-López's Impact Evaluation Process (Guerra-López, 2007b, 2007c, 2011) as a method for evaluating a residency program's performance under the guidance of an experienced evaluator.

Theoretical Framework

Human performance technology combines a process orientation with focus on results and outcomes; people; efficiency, effectiveness, impact and value; improvement and accomplishments; and measurement/quantifiable results. Guerra-Lopéz posits that building an “authentic, collaborative, and productive partnership with stakeholders” is essential to every performance improvement project (2007c, p.36). The process of evaluation is inherent in HPT models (Van Tiem, Mosely & Dessinger, 2000; Pershing, 2006); indeed, evaluation is central to the field of performance improvement. Meaningful evaluation requires that an organization and its individuals understand, value and place priority on the evaluation process and use of the results. Developing internal evaluation capacity is a means to both increasing the utility of evaluation and culture change. Partnering with and engaging stakeholders in the process of systematic performance evaluation should therefore lead to impactful and valuable improvements.

HPT theory related to study

HPT theoretical and practice models provide guidance for evaluating training and performance (Kirkpatrick, 1987; Dick & Carey, 1996; Moseley & Solomon, 1997; Sleezer, Zhang, Gradous & Maile, 1999; Dunlap, 2008; Guerra-Lopez, 2007a, 2007 b, 2007c and 2008). Sleezer, Zhang, Gradous & Maile, (1999) note that a systems view conceptualizes performance improvement as a “problem-defining process, a problem-solving process, or a product of various evaluative activities that focus on current and future performance...”(p.129).

The ISPI HPT model (Van Tiem, Mosely & Dessinger, 2000, p.3) depicts the evaluation element in a multidirectional relationship with all other model elements (performance analysis, cause analysis, intervention selection/design/development, and interventional implementation/change) and notes the need for formative and summative evaluation. Dessinger & Moseley posit the need for “confirmative evaluation” to support continuous improvement while “meta” evaluations focuses on the worth of evaluation processes and whether or not we are obtaining reliable and valid evaluation results (2006, p.320). Indeed, if performance is to be improved it must also be measured and evaluated; evaluation is essential to performance improvement.

Evaluation capacity building

Evaluation capacity can be thought of as the ability to conduct an effective evaluation. Building evaluation capacity in an organization can result in increased understanding of evaluation practices and utilization of evaluation results (Peters, Bagget, Gonzales, DeCotis, & Bronfman, 2007). Building evaluation capacity seeks to expand stakeholders understanding of evaluation concepts and practices. Aims for sustainable improvements in an organization’s evaluation capacity include 1.) increased stakeholder participation in the evaluation process, 2.) improved understanding of evaluation methods and practices, and 3.) increased use of evaluation results.

Participatory Evaluation

Participatory Evaluation is thought to foster evaluation capacity building (McDonald, Roger, Kefford, 2003; Peters, Baggett, Gonzales, DeCotis, & Bronfman, 2007; Overcast, Schmidt, Lei, Rodgers & Chung, 2009) and evaluation utility (Brown-McGowan, 1992; Cousins & Earl, 1992; Cousins & Earl 1995; Papineau & Kiely, 1996;

Plottu & Plottu, 2009; Greene, 1987;). The process of participatory evaluation aims to engage stakeholder participants in decision-making, increase abilities to plan and conduct evaluations, and increase evaluation utility.

Adaptation of Impact Evaluation Process

Systematic evaluation processes are engineered to allow for the conduction of useful evaluations that lead to recommendations that will measurably improve performance. The systematic evaluation process selected for the current study is an adaptation of Guerra-López's Impact Evaluation Process (2007), a seven-step evaluation process derived from foundations of the performance improvement field including Roger Kaufman's Organizational Elements Model (OEM) (1999, 2006); the ADDIE model: Analysis, Design, Development, Implementation and Evaluation (Branson 1975; Dick and Carey, 1996); and the modified A2DDIE model, which adds the essential element of a needs "assessment" to the original ADDIE model (Guerra, 2003), as well as more contemporary approaches to design thinking (Cross, 2011). The adaptation of the Impact Evaluation Process includes the addition of enlisting professional evaluator assistance at the onset of the project and engaging the stakeholders in the process of practical participatory evaluation. The adaptations are designed to support the process of working with physicians who have little or no formal training in evaluation methodology; by enlisting a professional evaluator we provide additional aid to the stakeholders in the evaluation design and implementation process.

Table 1 summarizes the adapted Impact Evaluation Process (Adapted from Guerra-López, 2007b, 2007c; Guerra-López, 2008).

Table 1 *Adaptation of Impact Evaluation Process*

<u>Impact Evaluation Process Step</u>	<u>Adaptation</u>
1. Enlist professional evaluator assistance	Step added to allow for the provision of evaluation expertise to aid in systematic evaluation process implementation.
2. Identify stakeholders and expectations	No adaptation.
3. Determine and develop evaluation question	Participatory evaluation methods employed.
4. Design evaluation plan <ul style="list-style-type: none"> a. Define key decisions and performance objectives b. Define measurable performance indicators 	Combined two steps in original model. Participatory evaluation methods employed.
5. Select data collection instruments and identify data sources	Participatory evaluation methods employed.
6. Select data analysis methods	Participatory evaluation methods employed.
7. Communicate results and recommendations	Participatory evaluation methods employed.

Adapted from Guerra-López, 2007 b, c; Guerra-López, 2008.

Operational Definitions

Graduate medical education refers to the educational period (residency and fellowship) post-medical school completion. In the United States this period can range from three years (e.g. Internal Medical Residency Program) to up to seven years (e.g. Neurological Surgery Program).

Residency program refers to the regulated process of training physicians upon graduation from medical school. Further training in highly specialized fields may be required for some medical sub-specialties upon completion of a residency program; this training period is referred to as fellowship program. The program represented in this case study is the Ophthalmology Residency Program sponsored by the Detroit Medical Center with clinical activities and administration housed at the Kresge Eye Institute (KEI). The program requires completion of a “Transitional Year” of broad clinical education (completed prior to enrollment in the ophthalmology residency program) followed by three years of ophthalmology residency training. The KEI Ophthalmology Residency program enrolls seven residents each year with 21 trainees practicing concurrently (7 first year residents, 7 second year residents, and 7 third year residents).

Residency program director refers to the administrative head of the residency program; he or she is charged with the oversight of the operation of the residency program and is held accountable for the quality of the graduate medical education of the physician trainees in addition to bearing responsibility for compliance with regulatory requirements (e.g. ACGME, AOA, Joint Commission, etc.)

Accreditation in graduate medical education is a voluntary process that involves an initial application process and subsequent audits to ensure that residency programs maintain compliance with regulatory requirements. The ACGME is a private, nonprofit council that evaluates and accredits residency programs in the United States. The AOA accredits institutions and programs and approves osteopathic postdoctoral training programs. ACGME or AOA accreditation is required for residency programs to receive graduate medical education funds from the federal Center for Medicare and Medicaid

Services. Graduation from ACGME/AOA-accredited programs allows residents and fellows to be eligible to sit for board certification examinations in their chosen medical specialty. The KEI Ophthalmology Residency Program is accredited by the ACGME.

Annual program evaluation is a regulatory requirement noted by both the ACGME common program requirements (ACGME, 2011; 2013) and AOA Basic Standards for Residency Programs (AOA, 2011) for residency program accreditation. While the elements noted by each accreditation body differ, each requires that residency programs conduct an evaluation of program quality and that data be collected and evaluated as part of the evaluation process. The ACGME defines program evaluation as the “Systematic collection and analysis of information related to the design, implementation, and outcomes of a resident education program for the purpose of monitoring and improving the quality and effectiveness of the program” (ACGME Glossary, 2013, p.8).

Chief Resident is a residency position typically held in the final year of residency for surgical programs. It is often, as in the case of the program participating in this research study, a peer-elected position that beholds the resident to additional administrative and leadership roles in the program.

ACGME Case Log System is an electronic web-based system that allows residents to record their procedural experiences and affords the ACGME ability to track compliance with volume and variety requirements as specified by specialty Residency Review Committees.

ACGME Citation is “a finding of a Review Committee that a program or institution is failing to comply substantially with a particular accreditation standard or ACGME policy or procedure” (ACGME Glossary of Terms, 2013).

Evaluation capacity refers to the ability to conduct an effective evaluation (Millstein, Chapel, Wetterhall, & Cotton, 2000). In the context of this study, the aims for building evaluation capacity include introducing the concept of systematic evaluation practices and encouraging the development of the skills required to conduct rigorous self-evaluations.

Participatory evaluation involves the active participation of major stakeholders in evaluation planning and process and assumes that stakeholder participation will contribute to decision-making (Plottu & Plottu, 2011) as well as enhance participant ability to think evaluatively (MacLellan-Wright, Patten, Cruz, & Flaherty 2007; Patton, 1998). In the proposed case study major stakeholders conduct all aspects of the evaluation process with the resident trainees responsible for evaluating the data collected and recommending solutions for improvement.

The “Impact Evaluation Process” (Guerra-Lopez, 2007b) is a seven-step model based on systems theory concepts. It is a process “based on a systematic approach to evaluation and performance improvement” (Guerra-Lopez, 2008, p. 83). A modified version of the “Impact Evaluation Process” is proposed as the evaluation plan for the proposed case study. Two modifications are proposed; first, the enlistment of an experienced evaluator to facilitate the evaluation process and act as both coach and researcher and, second, the utilization of this process as a means for participatory evaluation.

Action-based performance improvement plan is a plan that includes recommendations for improvement that 1.) can be acted upon, and 2.) the success of which can be measured or determined.

Performance gap is a discrepancy between the actual and desire performance.

Self-assessment is defined as “assessment of evaluation of oneself or one’s actions and attitudes, in particular, of one’s performance at a job or learning task considered in relation to an objective standard” (Oxford Dictionaries, n.d.).

Limitations and Delimitations

The current study employed a single case study design. According to Yin (1990), case studies differ from experimental designs that deliberately impose a treatment on a group of randomized subjects. Limitations to case studies include potential biases due to the lack of ability to control for outside variables, lack randomization, lack of generalizability and challenges of establishing reliability (Yin, 1990). Because conclusions about cause and effect relationships cannot be inferred when using case studies, results must be limited to descriptions.

Delimitations include the selection of a single residency program to serve as the focus of the study. Doing so is a practical, rather than empirical, decision. The evaluation capacity of the researcher and the residency program are constrained by resource limitations (e.g., time, burden, and monetary).

Summary

This chapter presented key background information for the current case study evaluation research to examine the results of a participatory evaluation approach utilizing

a modified version of the “Impact Evaluation Process” (Guerra-Lopez, 2007b, 2007c). This study addresses the dearth of empirical evidence supporting HPT evaluation practices while simultaneously addressing the accreditation requirement (need) of residency programs to design, develop, and evaluate sound program evaluation practices that lead to program improvement.

Chapter II. Literature Review

The general purpose of the present evaluation research study was to examine the difference in outcomes when utilizing a systematic evaluation process, an adapted version of the “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008), to analyze performance compared to less rigorous evaluation methods. The specific purpose was to examine the impact of a systematic evaluation process on evaluation findings; determine if the process led to action-based performance improvement plans tied to specific performance gaps and improved outcomes; and identify the benefits of and the barriers to implementation of this process in a graduate medical education residency program.

The purpose of the literature review is to examine research on implementing a systematic evaluation process and to extrapolate those findings to potential implications relevant for human performance technologists and graduate medical education residency program administrators. Empirical findings related to the use of systematic evaluation processes are explored and relevant HPT theories, models, and recommendations examined. Leaders in the field of human performance technology have called for empirical research to examine the contributions of HPT theories, models, and processes in applied settings (Kaufman & Clark, 1999; Guerra-Lopez & Leigh, 2009; Stolovitch, 2000; Sugrue & Stolovitch, 2000; Brethower, 2000; Kaufman & Bernardez, 2012). Graduate medical education residency program administrators are challenged with the task of performing annual program evaluations and improving residency program performance without the benefit of evaluation training or, in the vast majority of cases, the assistance of a professional evaluator.

This section begins with an overview of research examining physicians' ability to self-assess and examines general challenges faced in the self-assessment process. The role of evaluation in HPT theories and models is explored and published evidence that HPT evaluation models result in improved performance examined. Opportunities for improvement in evaluation methodology are presented, highlighting the importance of stakeholder participation and building organizational evaluation capacity. Finally, we examine the current empirical findings related to annual program evaluation of graduate medical education programs. The literature review then ends with a summary of the relevance of the literature review findings as they relate to the ACGME requirements for annual program evaluation and the aims of the proposed study.

Self-Assessment

The annual program evaluation mandate for GME residency programs requires program administrators to conduct an evaluation and effectively “self-assess” their program’s performance. The Oxford Dictionaries (n.d.) defines self-assessment as “assessment or evaluation of oneself or one’s actions and attitudes, in particular, of one’s performance at a job or learning task considered in relation to an objective standard.”

Self-assessment is often thought to be an effective means to performance assessment; however, studies indicate there is evidence to the contrary. Dunning, Heath, & Suls (2004) reviewed the empirical findings of self-assessment in health, education, and the workplace and found the accuracy of self-assessment lacking. People tend to overrate themselves, students tend to exhibit overconfidence, employees overestimate their skills, and CEOs display overconfidence in judgment (Dunning et al., 2004). The authors specifically note the unrealistic expectation that medical students should be able to

develop life-long learning abilities that are dependent upon accurate self-assessment to identify knowledge deficits. They suggest the need for interventions that will provide objective evidence of performance to increase the accuracy of improvement plans (Dunning, et al., 2004).

Similarly, practicing physicians exhibit limited ability to accurately self-assess. Davis, Mazmanian, Fordis, Van Harrison, Thorpe, & Perrier, (2006) compared physician self-assessment to external observations and found them lacking; thirteen of the 20 studies examined found little, none, or an inverse relationship between physician self-assessment and external observations. Further, the least skilled, most confident physicians demonstrated the worst accuracy in self-assessment (Davis, et al., 2006).

Researchers have noted significant barriers to effective self-assessment. Self-assessment is not a stable skill, but one that varies depending on context, content, and perspective (Eva & Regehr, 2005). In the context of self-assessment of residency program performance, we may presume that most physicians lack training and expertise in program evaluation methodology due to lack of education and training in their chosen field of medicine. Second, self-assessment is a difficult task and we often lack crucial information (Dunning, et al., 2004), frequently due to less than optimal evaluation strategies. Finally, Ward, Gruppen and Regehr (2002) note that there are significant methodological issues that challenge the findings of current self-assessment literature due to problems with measurement.

Annual evaluation of residency programs, although mandated by the ACGME since 2007, has yielded little published data regarding program outcomes. Although models have been proposed for this purpose, (Musick, 2006; Durning, Hemmer, Pangaro,

2007; Vassar, Wheller, Davision, and Franklin, 2010) very few have published data reflecting their utility in application.

As noted, there are significant challenges to residency programs conducting a successful self-assessment to meet the annual program evaluation requirement, the accuracy of self-assessment methods is often lacking, physicians have been found to demonstrate poor self-assessment skills and self-assessment has been shown to be a challenging endeavor. However, there are a plethora of evaluation models and processes that may be utilized to simultaneously inform the design of a self-administered systematic evaluation of a residency program and support the aim to improve the utility of residency program self-evaluation. The Human Performance Technology Field offers theoretical and practical support for such an endeavor, as do research studies in evaluation methodology.

HPT, Performance Measurement and Evaluation

Evaluation is a central premise in the Human Performance Technology field. From training needs assessment (Rossett, 1987), to performance measurement and management (Guerra-Lopez, 2007a), to evaluation methods and types (Kirkpatrick, 1997; Dessinger & Moseley, 2004), evaluation is an expected part of methods, processes, and models in HPT.

The Human Performance Technology field emphasizes the importance of evaluation in every step of the educational process from design to educational outcomes. The value of evaluation is noted at all stages of performance; in needs assessment to identify if a performance gap exists (Gilbert, 1978; Gordon, 1994; Kaufman, 1992, Watkins, Leigh, Platt, & Kaufman, 1998; Leigh, et al, 2000; Mager & Pipe, 1997;

Rummler & Brache, 1990); in formative evaluation to provide “real time” feedback (Scriven, 1967); in summative evaluation to determine merit or worth (Scriven, 1967); in confirmative evaluation to determine lasting utility and added value (Dessinger & Moseley, 2004); and in meta-evaluation to verify the validity of the evaluation process itself (Moseley & Dessinger, 1998).

Evaluation is an essential element in many HPT models such as Branson’s “ADDIE” model (1975); Rummler, Brethower & Geis’ (1974) “Human Performance System” model; Rummler’s “Performance Planned and Performance Managed” model (2004); and Van Tiem, Moseley & Dessinger’s “Human Performance Technology” model (2004). Guerra-Lopez & Leigh (2009, p.97) note that “measurement and evaluation are at the core of reliably improving performance,” and emphasize that evaluation and measurement are a means to providing evidence of the value of our contributions to our clients and our field.

While the field of HPT strongly supports performance measurement and evaluation, its journal publications indicate there is much room for improvement in providing evidence that our evaluation and performance improvement models, processes and interventions actually improve performance and add value. Guerra-Lopez and Leigh (2009) examined the publications in *Performance Improvement Journal (PIJ)* and *Performance Improvement Quarterly (PIQ)* journal, the premiere journals in the HPT field, to determine the subject emphasis on evaluation and performance improvement and types of formats presented (analysis of a 10 year period from 1997-2006). Their findings indicate that only one-tenth of *PIJ* publications contain an evaluation component while nearly half of *PIQ* articles contain some evaluation component. The preponderance of

articles offered guidance on performance measurement (66%), while others focused on models and/or advocating evaluation in general (Guerra-Lopez & Leigh, 2009). These findings demonstrate that there is great opportunity, and in fact a great need, for HPT practitioners and researchers to publish data-driven work that highlights the utility of HPT performance measurement and evaluation process outcomes.

Opportunities for Methodological Improvements in Evaluation

Traditional evaluation practices routinely depend upon the design of external evaluator experts who offer outside interpretation of the worth or merit of a program. High-ranking authorities determine the purpose of the evaluation, the evaluation expert collects data, and a report is generated by an outside observer. Traditional evaluation strategies include scientific-experimental models (e.g., quasi-experimental design, objectives-based research), management oriented systems models (e.g., Program Evaluation Review Technique, Critical Path Method, CIPP Model), qualitative/anthropological models (e.g., “Fourth Generation”), and participant-oriented models (e.g., client-centered and stakeholder approaches). The implementation of evaluation strategies is varied among industries and sectors, with a great deal of published literature focused on governmental funded evaluations of programs and interventions (perhaps because both funding and other resources are earmarked for evaluation). The focus of such evaluation strategies is often on the evaluation process and methods themselves with less concern for the concepts of continuous measurement/management and performance improvement.

An assessment of evaluation designs examining case studies of 12 large federal evaluations (Howell & Yemane, 2006) scrutinized characteristics deemed essential to

evaluation success, 1.) design, 2.) evaluation expertise, and 3.) dissemination. The authors cite a need for multiple components to broad based evaluation (process analysis, program monitoring, rigorous impact assessment, and cost-effectiveness analysis), a need for the utilization of experienced evaluators, and a requirement for timely, wide distribution of evaluation results. Among other findings, Howell & Yemane (2006) noted a lack of available quality data (resulting in decreased utility of evaluation results), lack of impact assessment component, insufficient attention to the design phase, insufficient evaluation expertise, and limited broad dissemination of evaluation findings. Such evaluation characteristic are key to ensuring that the evaluation process adds value to organizations and maximizes the utilization of the resources expended toward the process of evaluation and their resulting impact on performance.

Participant Evaluation

There is a movement toward evaluation methods aimed to be more inclusive of multiple stakeholders' perceptions to reflect a more transparent culture where information is shared. House and Howe (2003) define one such "deliberative democratic evaluation" method as an attempt to make evaluation practices more democratic by 1.) representing a wide array of views and interests in evaluation studies, 2.) encouraging stakeholder participation in the evaluation process, and 3.) providing opportunities for extended deliberation. The "deliberative democratic evaluation" process proposes significant engagement of stakeholders in the evaluation process while the professional evaluator(s) retains responsibility for adherence to appropriate data collection and analysis techniques. House and Howe (2003, p.80) equate stakeholder involvement with genuine democracy

whereby issues are discussed and deliberated at length. Aims of this process include collective decision-making, inclusion, and stakeholder transformation.

Participatory evaluation research methods espouse somewhat similar aims. A utilitarian approach to evaluation that relies upon stakeholder participation in the evaluation process as a means to increase the use of the evaluation findings; participatory evaluation aims to engage stakeholder participants in decision-making and increase stakeholder abilities to plan and conduct evaluations.

Participatory evaluation methods have benefits beyond transparency. Plottu & Plottu (2009) note that the principles espoused by House (2005) intend to result in increased external validity, greater use of the evaluation results, engagement, empowerment, and increased analytical democracy. Greene (1987) reports the benefits of stakeholder participation in evaluation design include learning about the program and evaluation; providing opportunity for reflection and analysis; generating credibility for the program; positive feelings about the process of being heard/hearing others, while costs included time and negative feelings about participation.

Stakeholder participation in the evaluation process has been found to increase evaluation utilization. Cousins and Leithwood (1986) examined sixty five evaluation studies and found that when evaluation users were involved in the evaluation process, when findings were consistent with user beliefs and expectations, and when the data reported was relevant to users' problems evaluation use was greater. Thirteen years later, Cousins and Leithwood's work was replicated by Johnson, Greenesid, Toal, King, Lawrenz, and Volkov (2009) and stakeholder participation in the evaluation process was noted to be a new evaluation category of evaluation use with particular implications for

utilization. Specifically, the authors posit that “findings point to the importance of stakeholder involvement in facilitating evaluation use and suggest that engagement, interaction, and communication between evaluation clients and evaluators is critical to the meaningful use of evaluations” (p. 377).

Building Evaluation Capacity

Evaluation capacity building provides a means for organizations to increase stakeholder capabilities and understanding of evaluation methods. Preskill & Boyle (2008) report that participatory, collaborative and stakeholder forms of evaluation are more common than ever before and that outcomes of evaluation capacity building activities include increased knowledge and understanding of evaluation concepts, increased commitment to evaluation practices, and improved program quality.

Lennie (2005) examined outcomes of an Australian effort to build evaluation capacity in rural communities’ ability to evaluate local communication and information technology initiatives. Reported strengths of this process included improved knowledge and skills; participant empowerment and increased evaluation capacity; involvement of a broad diversity of community members; flexibility, transparency and flexibility of the process; improved objectives and decisions on priorities for action; maintained interests and motivation; effective utilization of technology, and; mutual learning and understanding. Limitations of the process included time/resource burden; disempowering impact on some participants (capacity building limited to small group of participants); challenges in obtaining participant involvement; domination of agenda by subgroup; lack of process fit to all participants values; technological challenges due to lack of community resources, and; learning impact limited to actively involved participants.

The process of building evaluation capacity depends upon stakeholder participation in the evaluation process. The potential for building organizational and participant evaluation capacity lies in the motivation and ability of the organization to provide time and resources toward the evaluation agenda, design, development and implementation.

Program Evaluation in GME

A variety of tools have been developed for residency program evaluation including program report cards (Phitayakorn, Levitan, and Shuck, 2007) and surveys of faculty and residents, (Bellini, Shea and Asch, 1997; Liebelt, Daniels, Farrell and Myers, 1993). Musick (2006) reported that a unified approach to program evaluation in Graduate Medical Education is lacking and offered a conceptual model for GME program evaluation that requires five steps, 1.) determining the evaluation need, 2.) determining the evaluation focus, 3.) determining the evaluation method, 4.) presenting the evaluation findings, and 5.) documenting the evaluation results. Other models and structures have been proposed (Durning, Hemmer, Pangaro, 2007; Vassar, Wheller, Davision, and Franklin, 2010) and some have begun to collect system-wide surveys regarding program performance (McOwen, Bellini, Morrison, and Shea, 2009).

Duke University (Andolsek, Nagler, & Weinerth, 2010) improved adherence to the ACGME requirement for annual program evaluation through the utilization of a program evaluation report template, resulting in a significant reduction in the number of ACGME citations for residency program lack of compliance in annual program evaluation. The template and subsequent monitoring of its utility did not, however, analyze the quality of the program evaluation or the performance outcomes associated

with it, but simply noted that the utilization of the evaluation template report ensured that the programs met the minimal accreditation requirements (Andolsek, Nagler, & Weinerth, 2010).

Summary

The challenges facing graduate medical education residency programs as they attempt to evaluate their educational program's outcomes are many. There is a need, mandated by accrediting bodies, for graduate medical education programs to evaluate the effectiveness of their training programs. Medical school training increases students' medical knowledge and prepares them for additional training in specialty and subspecialty clinical practice, but does little to prepare physicians to assess and evaluate performance. We have noted research highlighting the significant challenges to residency programs aiming to meet the mandated annual program evaluation requirement, 1.) the accuracy of self-assessment methods is often lacking, 2.) physicians have been found to demonstrate poor self-assessment skills and, 3.) self-assessment has been shown to be a challenging endeavor. As mandated by the ACGME, annual program evaluation requires high-level evaluation and assessment skills and the empirical evidence reveals the challenges for physicians to do so.

Human Performance Technology provides a plethora of tools for practitioners aiming to assist in evaluative efforts and, as in all scholarly pursuits, there is a need to validate these tools and processes and demonstrate their efficacy in the field. Indeed, the literature review reveals that the HPT scholars have called for research examining the application of HPT and PI models and processes in a variety of fields. Graduate medical education, with its regulatory requirements for evaluation and focus on outcomes, is an

excellent context in which to conduct such research. Further, there is opportunity to explore evaluation methodologies aligned with the ACGME requirement for annual program evaluation including participatory evaluation (as per ACGME requirements, programs must include representative residents in the APE process) and building evaluation capacity (the ACGME Next Accreditation System places strong emphasis on self-evaluation practices).

Chapter III. Method

The general purpose of the present evaluation research study was to examine the difference in outcomes when utilizing a systematic evaluation process, an adapted version of the “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008), to analyze performance compared to less rigorous evaluation methods. The specific purpose was to examine the impact of a systematic evaluation process on evaluation findings; determine if the process led to action-based performance improvement plans tied to specific performance gaps and improved outcomes; and identify the benefits of and the barriers to implementation of this process in a graduate medical education residency program.

Overview

The purpose of this chapter is to present the proposed study methods. First, the research design, study setting and the program utilized for the case study are described. The study sample and data collection plans are detailed, including the procedures to be followed. The Institutional Review Board application is discussed and, finally, the data analysis plan described.

Case Study Design

The proposed design is a case study. According to Yin (2009), case study design provides an application to explain causal links in real-life interventions that may present challenges for survey or experimental designs. Yin (2009, p.18) defines case studies as follows:

A case study is an empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident...The case study inquiry copes with the technically distinctive situation in which there will

be many more variables of interest than data points, and as one result relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result, benefits from the prior development of theoretical propositions to guide data collection and analysis.

Further, case studies may be employed to “describe an intervention and the real-life context in which it occurred” (Yin, 2009, p.20). In the current study, the application of a case study design provides a means for asking multiple research questions such as, “How do each year’s annual program evaluation results differ (prior to the initial implementation of the systematic evaluation process and each subsequent year of implementation)?” The case study design also affords the opportunity to use a multiplicity of data sources and, as a result, employs both quantitative and qualitative data analysis.

Setting: Graduate Medical Education at the Detroit Medical Center

The Detroit Medical Center (DMC) is the largest healthcare provider in southeast Michigan, including 8 hospitals, more than 2,000 licensed beds, and 3,000 affiliated physicians (“Organization History and Profile,” n.d.). The Detroit Medical Center sponsors 90 Graduate Medical Education Programs (“Training Programs” n.d.) training over 1,000 Residents and Fellows each year. In the 2011-2012 academic year, the DMC sponsored 53 ACGME accredited residency programs training 740 residents (“ACGME Data Resource Book”, 2012, p.98). Through its Graduate Medical Education Committee (GMEC) and the Graduate Medical Education Office (GMEO), the DMC has ultimate responsibility for the sponsored residency and fellowship programs. This responsibility includes demonstrating an overall commitment to GME, maintaining affiliation agreements with other institutions participating in GME, monitoring the Joint

Commission (JC) status of participating institutions, ensuring that formal quality assurance programs are conducted at participating institutions, monitoring eligibility and selection of residents, monitoring all aspects of resident appointment, monitoring resident participation in educational and professional activities, and monitoring the residents' work environment.

The GMEO tracks program outcome measures, conducts extensive internal reviews of each GME program, and ensures that each program teaches and assesses the ACGME general competencies of patient care, medical knowledge, practice-based learning, interpersonal and communication skills, professionalism, and systems-based practice or the AOA competencies of osteopathic philosophy and osteopathic manipulative medicine, patient care, medical knowledge, interpersonal and communication skills, practice-based learning and improvement, professionalism, and systems-based practice.

ACGME accredited institutions' GMEC must monitor their individual residency program compliance with ACGME requirements, including oversight of an annual evaluation of program and improvement activities (ACGME Institutional Requirements, 2013, p. 3). The Detroit Medical Center began tracking ACGME citations for program evaluation in September 2010 when the first ACGME citation tracking reports became available. Since that time, 12 (23%) of the DMC sponsored ACGME accredited programs received citations for non-compliance with the evaluation of program requirement. Citations reflect poor program performance and may lead to reduced accreditation cycle lengths, more frequent site visits, and poor institutional level accreditation performance.

Although the ACGME requirements for annual program evaluation have been in effect since 2007, internal reviews (mandatory program audits) of residency and fellowship programs at the DMC during 2010 and 2011 reveal that 26 of the 28 (93%) programs reviewed did not meet all of the requirements for annual program evaluation (unpublished report, 2012). Without a means of systematically self-evaluating program performance, it is unlikely that residency programs are engaged in the continuous measurement and management processes required to ensure high-level performance outcomes. The results of the ACGME citations and DMC GME internal review findings provide evidence supporting the need for the development of a systematic process for residency program evaluation.

Case Selection for the Study

The Detroit Medical Center sponsors 53 ACGME accredited residency and fellowship programs. The proposed case for study is a single residency program, the Kresge Eye Institute (KEI) ophthalmology residency training program located in Detroit, Michigan. The KEI residency program is a long-established training program that was founded in 1951 and is sponsored by the Detroit Medical Center. The Accreditation Council for Graduate Medical Education (ACGME) accredits the ophthalmology residency training program. The program sponsors 21 enrolled residents (7 enrolled per year for a 3 year long program).

The KEI ophthalmology residency program's mission is, "to provide an optimal clinical education to physicians in the science and art of the specialty of ophthalmology" (KEI, n.d.). The program provides a 36-month curriculum with structured clinical and basic science courses, clinical conferences and independent study. The program employs

32 physician faculty members and two non-physician faculty members (PhD and OD) encompassing a broad range of ophthalmic subspecialties. A Program Coordinator and Assistant Program Coordinator also provide administrative support for the program. The program's educational and clinical activities are centered at the KEI while the residents also see patients in the metropolitan Detroit area at the John D. Dingell Veterans Administration Hospital, Children's Hospital of Michigan, Sinai-Grace Hospital, Detroit Receiving Hospital, and Harper University Hospital.

The selection of the case was not randomized; being granted full access to the researcher as well as invested program administration interest in conducting a systematic program evaluation were the major determining factors in program selection. The current KEI ophthalmology program director served in multiple GME leadership roles both locally and nationally (Designated Institutional Official for Graduate Medical Education at the DMC, Associate Dean for Graduate Medical Education at Wayne State University's School of Medicine, Chair of the ACGME Ophthalmology Residency Review Committee) and, as such, possesses a high level understanding of ACGME requirements related to annual program evaluation.

The case selection process limits the generalizability of the study. However, the richness of the data collected and the mixed methods used to examine the same performance dimensions related to the research questions provides opportunity to fully explore the context of annual program evaluation in a natural setting over a three year time period.

Data Collection and Analysis Plan

One of the hallmarks of case studies is the collection and analysis of data from multiple

sources, often referred to as data triangulation, an effort that aims to increase the validity and reliability of the study. External and internal data sources were used to examine program outcomes, adding depth and validity to the research findings. Utilization of both qualitative inquiry and quantitative methods to study the program allows for cross validation of results, a potentially deeper understanding of the results, and opportunity to increase confidence in the study findings.

Patton (2001) posits that qualitative research must be concerned about validity and reliability during the study design and analysis phases and that these concerns extend to judging the quality of the study. Jick (1979) asserts that accuracy of judgments can be improved by collection of different kinds of data relating to the same phenomenon, or “triangulation.” The proposed study utilizes a variety of methods to examine both the process and results of annual program evaluation in a residency program. These methods allow for the collection of multiple stakeholder perspectives using different data collection modalities (e.g., internal/external surveys and face-to-face semi-structured interviews) and analysis of a variety of archival data relating to process and outcomes (e.g., meeting minutes, reports, program performance tracking, and accreditation results). This “triangulation” of data will provide a more complete contextual depiction of the stakeholder perspectives and the residency program’s performance over time, thus increasing confidence in the study results.

The KEI Ophthalmology residency program utilized the adapted “Impact Evaluation Process” for three consecutive years (2011, 2012, and 2013) as a means to fulfill the ACGME requirement for annual program evaluation. All of the program’s residents participated in the process each year and in 2011 and 2012 residents conducted all of the

analyses and were fully responsible for preparing recommendations based on their findings. The APE reports and program's reported improvement progress provide essential data sources for the research study. Table 2 includes descriptions of the proposed data to be collected, data collection methods, data source/documentation, and analysis plan.

Table 2 *Data Collection and Analysis Plan*

Research Question	Data to be collected	Data Collection Method	Data Source/Documentation	Analysis Plan
<p>1. Do the evaluation results generated using a systematic evaluation process (e.g. evaluation findings and recommendations for improvement) differ from previous years' annual program evaluation results?</p>	<ul style="list-style-type: none"> • Annual program evaluation (APE) results (2007 – 2013) • Stakeholder perceptions (Program Director) 	<ul style="list-style-type: none"> • Archival data review • Semi-structured interview 	<ul style="list-style-type: none"> • Educational Committee Meeting minutes • APE Reports • Semi-structured interview transcript 	<p>Qualitative analysis will be utilized to:</p> <ul style="list-style-type: none"> • Describe evaluation findings and recommendations using general indicative approach described by Thomas (2006) • Compare evaluation results from previous years to those generated in the years using the systematic evaluation process
<p>2. Does the utilization of a systematic evaluation process lead to action-based performance improvement plans tied to specific performance gaps?</p>	<ul style="list-style-type: none"> • APE results (2011, 2012, 2013) 	<ul style="list-style-type: none"> • Archival data review 	<ul style="list-style-type: none"> • APE reports (2011, 2013, 2013) 	<p>Qualitative analysis will be utilized to:</p> <ul style="list-style-type: none"> • Examine proposed evaluation results to ascertain if performance improvement plans are 1.) action-based, and 2.) tied to specific performance gaps

Research Question	Data to be collected	Data Collection Method	Data Source/Documentation	Analysis Plan
<p>3. Does using a systematic evaluation process result in improved program outcomes (e.g., adherence to requirements, management of program performance, educational outcomes, implementation of solutions)?</p>	<p>Program Performance Data</p> <ul style="list-style-type: none"> • In-service examination scores • Board Examination pass rates • ACGME site visit results • ACGME Resident Survey results • Program performance tracking results • Resident Surgical Volume • Stakeholder perceptions (Program Director/Department Chair, Faculty member, Chief Residents, Program Coordinator) 	<ul style="list-style-type: none"> • Archival data review • Semi-structured interviews 	<ul style="list-style-type: none"> • ACGME Resident Survey reports (2010, 2011, 2013, 2013) • ACGME Site Visit report • Program performance tracking documentation • In-service examination and Board Examination reports • ACGME/Program Surgical Case Log Reports • Semi-structured interview transcripts 	<p>Qualitative and Quantitative analysis will be utilized to:</p> <ul style="list-style-type: none"> • Compare in-service and Board Examination scores from 2009-present (quantitative) • Compare program surgical volume from 2009-present (quantitative) • Compare recommendations for program improvement to program outcomes (year to year - qualitative)

Research Question	Data to be collected	Data Collection Method	Data Source/Documentation	Analysis Plan
<p>4. What are the benefits of and the barriers to the implementation of a systematic evaluation process?</p>	<ul style="list-style-type: none"> • Stakeholder perspectives (Program Director/Chair, Faculty, Residents, Program Coordinator) 	<ul style="list-style-type: none"> • Archival data review • Pre-Post implementation confidential survey of 2011 APE participants • Semi-structured interviews 	<ul style="list-style-type: none"> • Pre-post intervention survey results 2011 • Semi-structured interview transcripts 	<p>Qualitative analysis will be utilized to:</p> <ul style="list-style-type: none"> • Categorize benefits of and barriers to implementing a systematic evaluation process

Description of Data Sources and Research Instruments

The majority of data evaluated in this study was archival. Research instruments included data analysis tools designed to provide documentation of findings relative to research questions. A list of archival data is provided below.

Archival Data

1. Annual Program Evaluation (APE) Reports and Program Performance

Tracking Documentation: The Program Coordinators maintain electronic copies of annual program evaluation agendas, attendance, and reports. Since 2011, the first year the adapted Impact Evaluation Process was utilized, evaluation methods and questions have also been documented as part of the evaluation planning process. Since 2007 the ACGME requires that each annual program evaluation report must include an update of program progress since the last evaluation period. APE reports were analyzed to determine if findings/recommendations were actionable and if they were acted upon. Further, APE reports from years 2007-2010 (prior to implement of the adapted Impact Evaluation Process) were compared to reports generated in years 2011, 2012, and 2013 (years when the adapted Impact Evaluation Process was utilized).

a. Research Instrument: Annual Program Evaluation Report and Performance Tracking Form

2. ACGME Resident Survey Reports: The ACGME provides the Program Director annual reports of resident survey results via their website (requires a program ID and password to gain access). ACGME provides mean scores for constructs (e.g. educational content, faculty engagement, etc.).

- a. Research Instrument: ACGME Resident Survey Report Evaluation Form
3. **ACGME Site Visit Report:** The ACGME website maintains records of the results of their external audit of residency programs available to Program Director and Coordinator (requires a program ID and password to gain access).
 - a. Research Instrument: ACGMME Site Visit Report Evaluation Form
4. **In-service examination results:** The program residents participate in an annual medical knowledge examination proctored by the American Academy of Ophthalmology, entitled the Ophthalmic Knowledge Assessment Program (OKAP). Test results are available in individual (available to the resident) and program summary format (available to the Program Director). Resident scores will be stratified into pass (at or above the 33rd percentile) or fail (below the 33rd percentile) in line with the benchmarks identified in Johnson, Bloom, Szczotka-Flynn, Zauner, & Tomsak (2010) and Chen & Bhandari (2010).
 - a. Research Instrument: In-service Examination Results Evaluation Form
5. **Board Examination Results:** Program graduates participate in a Board Certification Examination proctored by the American Board of Ophthalmology. There are written and oral components to the two-part examination. The Program Director is provided a summary report of the examination results indicating if graduates have passed or failed each component. The ACGME utilizes resident Board examination performance as an indicator of program quality and mandates that at least 80 percent of eligible program graduates in the preceding five years must take the exam, and of those taking it for the first time, 60 percent must pass

(ACGME Program Requirements for Graduate Medical Education in Ophthalmology, 2013, p. 18).

a. Research Instrument: Board Examination Results Evaluation Form

6. **Surgical Case Log Reports:** As per ACGME requirements, Ophthalmology residents must record their surgical volume to enable the ACGME Residency Review Committee to ensure compliance with specifications for surgical volume and variety.

a. Research Instrument: Changes in surgical volume were analyzed in the time period from 2009 to 2013 utilizing the ACGME annual case log reports and the KEI program “Surgical Report Card” tracking tool.

7. **Pre-and Post Annual Program Evaluation Survey:** In 2011 (the first year a systematic evaluation process was implemented in the Ophthalmology residency program that is the focus of this study), a pre-and post survey of participants was conducted.

a. 2011 Pre and Post Intervention Survey Responses Evaluation Form

Data Collection Instruments Developed for the Study:

1. An online survey was conducted with program administrators who had participated in all three years of the systematic evaluation process implementation (Program Director/Chair, Former Chief Residents, Faculty Member, and Program Coordinator) to address the research question pertaining to perceived benefits and challenges of the systematic evaluation process.

a. Data Analysis Research Instrument: Program Administrator Survey

Data Analysis

Qualitative Analysis was utilized for each of the research questions using a “general inductive approach” identified by Thomas (2006). Thomas (2006, p.237) defines a “general inductive approach for analyzing qualitative evaluation data,” as a process to determine the core meanings evident in the text as they relate to the research questions. Using Thomas’ method as a guide the qualitative evaluation process was employed as described in Table 3.

Table 3. *Qualitative Analysis Purpose (as prescribed in Thomas, 2006, p.237)*

<u>Purpose</u>	<u>Analytic Strategies</u>	<u>Analytic Tasks</u>
1. <u>Condense text data into summary format</u>	1. <u>Multiple readings and interpretation of the raw data as it relates to research questions</u>	a) <u>Rigorous reading and coding of documents/transcripts to allow major themes to emerge</u>
2. <u>Establish links between research questions and summary findings derived from the raw data</u>	2. <u>Categories are identified from the raw data into a framework or mode with key themes and processes identified</u>	a) <u>Identify text segments related to research questions</u> b) <u>Label text segments (categories)</u> c) <u>Create a model incorporating most categories</u>
3. <u>Develop theory about the underlying structure of experiences evident in the data</u>	3. <u>Multiple interpretations are made from the raw data resulting in findings</u>	a) <u>Similarities across groups explored as applicable</u> b) <u>Summary of findings resulting from following analytic strategies described</u>

As per the methods proposed by Thomas (2006, p.237 listed in Table 3), a content analysis of archival data documents and survey results (e.g. APE reports, Pre-and Post Annual Program Evaluation Survey) were completed and categories were identified into a framework with key categories recorded. Text was analyzed assigned to categories. Continual comparison of text segment categorical assignments ensured continuity of this process. As posited by Thomas (2006) this process allowed a categorical structure to develop “naturally and intuitively” and provided reasonable opportunity for classification. The APE reports were analyzed to identify themes and categories and to determine whether the planned improvements were “action-based” and aligned with performance gaps. Program performance tracking information (updates on planned improvement progress) was analyzed to determine if planned improvements were implemented. Finally, the ACGME Site Visit Reports (2007 pre-implementation and 2012 one year post-implementation) were analyzed and compared to identify changes in program performance as indicated by an external evaluation process. Accreditation results were evaluated using the awarded accreditation cycle length, commendations, and citations as evaluation criteria.

Quantitative analysis was utilized to identify changes in program performance as it relates to resident perceptions (ACGME annual resident survey) and educational outcomes (OKAP in-service and Board Examination results). Results of the 2011, 2012, and 2013 ACGME annual resident surveys were compared to the 2010 pre-intervention survey. The ACGME provides a mean score for resident responses to survey questions regarding program performance across a variety of domains (duty hours, faculty, evaluation, educational content, resources, patient safety and

teamwork). Because no raw response data is provided, upward and downward mean trends for survey constructs was noted. Resident scores on the OKAP in-service examination were compared from years 2010 to the present and percentage of residents in “pass” and “fail” categories were noted using criteria developed by Chen & Bhandari (2010) and Johnson, Bloom, Szczotka-Flynn, Zauner, & Tomsak (2010). Board examination results were analyzed to determine the number of residents who pass both written and oral examination the first time it is taken (an indicator tracked by the ACGME with explicit quantitative requirements).

Reliability and Validity

Case studies pose unique challenges related to reliability and validity. Yin (2009) posits that there are four tests that are commonly used to establish the empirical quality of social research, 1.) construct validity, 2.) internal validity, 3.) external validity, and 4.) reliability. These four tests are presented below following by a description of a triangulation approach (Patton, 1990) that is utilized as a framework to enhance the reliability and validity of this case study approach.

Reliability is limited due to the historical nature of the project; the KEI Ophthalmology Residency Program is evolving naturally over time, as are the ACGME accreditation requirements. During the course of the three years during which the adapted “Impact Evaluation Process” was implemented, there were administrative changes within the Ophthalmology Department and the ACGME changed their accreditation system. Historical events cannot be controlled and it is not possible to return in time. It would be impossible for another researcher to *exactly* replicate this study for these reasons.

Construct validity is limited due to the incomplete set of standardized measures for the study. When available, standard data sets are utilized (e.g. OKAP in-service examination result report, Board Examination Report, ACGME Surgical Case Log Report). Subjective interpretation of the much of the data is required.

Internal validity is limited due to the constraints imposed by the inability to directly observe, record, and analyze all possible factors contributing to program performance and improvement.

External validity is limited due to the case study design and the sample selection process. The case study nature of the design requires that conclusions must be limited strictly to the Kresge Eye Institute Ophthalmology Residency Program.

Triangulation Approach to Increase Validity and Reliability (Patton, 1990)

Patton (1990, p.245) posits that, “A multi-method, triangulation approach to field work increases both the validity and the reliability of evaluation data.” The research study employed a variety of means to increase study validity and reliability. Using Michael Quinn Patton’s guidelines for reducing biases and increasing study validity and reliability, the following methods are represented in the data collection and analysis plan. References to Yin’s (2009) case study tactics to address construct, internal/external validity, and reliability are also presented as appropriate.

Methods Triangulation: Methods triangulation involves utilization of mixed methods, including qualitative inquiry and quantitative analysis. Patton (1990) asserts that such “comparative analysis” can strengthen the reliability by using different measures of the same concept. Data collection methods include analysis of archival data as well as semi-structured interviews with stakeholder participants. Qualitative inquiry

was utilized for all of the research questions. Interview transcripts, contents of APE reports, program performance improvement tracking reports, and open-ended survey results will be analyzed using the general inductive approach (Thomas, 2006). Qualitative analysis will be used to compare in-service examination and board examination results from 2007 through 2013 as one of the means to determine if the use of a systematic evaluation process resulted in improved program outcomes.

Triangulation of Sources: According to Patton (1990) the triangulation of sources method requires the researcher to crosscheck the consistency of information gathered *within* qualitative methods. This can be achieved through comparing observational and survey data and comparing the perspectives of stakeholders with different views. A mix of external and internal reports provides data for this study. ACGME site visit reports and annual resident surveys provide external evaluation of program performance. KEI Ophthalmology residents' standardized test results, in the form of in-service and Board Examination reports, provide markers of individual resident and program performance as compared to national averages. APE and program performance tracking reports provide internal evaluation of program performance gathered by program stakeholders. Finally, the 2011 pre and post-implementation (of the adapted Impact Evaluation Process) confidential participant survey combined with the longitudinal survey of program administrators provide perspectives about the process and impact from a participant point of view. Yin (1990, p.41) also notes that the use of multiple sources is an appropriate case study tactic to increase construct validity.

Analyst Triangulation: Analyst triangulation can be achieved by a review of the findings by the study participants (also referred to as member or stakeholder checks). According to Patton (1990, p.468), “Evaluators can learn a great deal about the accuracy, fairness, and validity of their data analysis by having the people described in that data analysis react to what is described.” The research study findings were reviewed with stakeholders representing multiple points of view including the Program Director/Chair of the Department, Faculty member, former Chief Residents, and the Program Coordinator. Each of these stakeholders was present for the APEs conducted in 2011, 2012, and 2013 and each participated in member check meetings. Yin (1990, p. 41) asserts that having key case informants review drafts of case study reports as a tactic to increase construct validity.

Additional Procedures

The proposed study is part of the “Graduate Medical Education Leadership Academy Curriculum Evaluation” approved by the Wayne State University Human Investigation Committee on November 12, 2010 (HIC#092510B3X, Protocol # 1009008756).

Study Limitations

The current study employed a single case study design. According to Yin (1990), case studies differ from experimental designs that deliberately impose a treatment on a group of randomized subjects. Limitations to case studies include potential biases due to the lack of ability to control for outside variables, lack randomization, lack of generalizability and challenges of establishing reliability (Yin, 1990). Because conclusions about cause and effect relationships cannot be inferred when using case studies, results are limited to descriptions.

Delimitations include the selection of a single residency program to serve as the focus of the study. Doing so is a practical, rather than empirical, decision. The evaluation capacity of the researcher and the residency program are constrained by resource limitations (e.g., time, burden, and monetary).

Summary

This chapter presented the methods to be used in the proposed study, which examined the utility, efficacy, and challenges of applying a systematic evaluation process to the required annual program evaluation of a residency program. A description of the study design and the instruments were described, as were the procedures. The rationale for utilizing qualitative and quantitative statistical analyses was discussed. The study limitations were noted and methods to overcome said limitations proposed.

CHAPTER IV. Results

The general purpose of the present evaluation research study was to examine the difference in outcomes when utilizing a systematic evaluation process, an adapted version of the “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008), to analyze performance compared to less rigorous evaluation methods. The specific purpose was to examine the impact of a systematic evaluation process on evaluation findings; determine if the process led to action-based performance improvement plans tied to specific performance gaps and improved outcomes; and identify the benefits of and the barriers to implementation of this process in a graduate medical education residency program. The purpose of this chapter is to present the study results. The evaluation research study attempted to answer the four following questions:

1. Do the evaluation results generated using a systematic evaluation process (e.g. evaluation findings and recommendations for improvement) differ from previous years’ annual program evaluation results?
2. Does the utilization of a systematic evaluation process lead to action-based performance improvement plans tied to specific performance gaps?
3. Does using a systematic evaluation process result in improved program outcomes (e.g., adherence to requirements, management of program performance, educational outcomes, implementation of solutions)?
4. What are the stakeholders’ perceptions of barriers to and benefits of implementation of a systematic evaluation process?

Annual Program Evaluation Results

The data presented in Table 4 describes the categorized performance improvement recommendations delineated by year, performance domain evaluated, action items completed, number of actionable items generated, and indicates if recommendations were aligned to program deficits. The program evaluations conducted during years 2009 and 2010 represent baseline data, that is, evaluations conducted prior to instituting a systematic evaluation process for the Kresge Eye Institute Ophthalmology Residency Program. The number of recommended actions generated in these baseline years (10 in 2009 and 10 in 2010) were less than in the years where a systematic evaluation process was utilized (69 in 2011, 49 in 2012, and 32 in 2013). The number of performance domains evaluated each year varied slightly (6 in 2009, 5 in 2010, 7 in 2011, 8 in 2012, and 5 in 2013), while the types of recommended actions proposed increased with the implementation of the systematic evaluation process (5 types in 2009, 5 types in 2010, 14 types in 2011, 11 types in 2012, and 9 types in 2013).

The recommendation types in years 2009 and 2010 included called for improvements in communications, meetings, schedule changes, policy development, curriculum development and remediation protocol development. The recommendations types in years 2011, 2012 and 2013 called for improvements in communication, curriculum development, didactic schedule, provision of food for the residents during didactic sessions, education technology, evaluation protocol, faculty responsibilities, leadership succession, leadership culture, OKAP curriculum development, online curriculum development, online resident scheduling, remediation protocol development, resident

clinic operations/teaching/patient scheduling, resource allocation, surgical curriculum development, surgical evaluation protocol development, and surgical protocol adherence.

The number of recommended actions completed differed by year. Nine of ten recommended actions were completed from the 2009 report; seven of ten recommended actions were completed from the 2010 report; thirty-seven of 69 recommended actions were completed from the 2011 report; thirty-six of the 49 recommended actions were completed from the 2012 report; and twenty-three of the recommended actions were completed from the 2013 report. Some of the recommended actions took more than one year to complete.

The percentage of performance improvement recommendations completed decreased as the number of recommendations increased, except in 2010. In 2009 and 2010 when 10 recommendations were made each year, 90% (9/10) and 70% (7/10) of recommended actions were completed. In 2011, the first year the systematic evaluation process was implemented, 54% of the recommended action items (37/69) were completed, in 2012 78% of the recommended action items (38/49) were completed, and in 2013 72% of the recommended action items (23/32) were completed. Time to completion of recommended actions varied from one to two years, inferring that 2013 completed recommended actions may increase by the time the next annual program evaluation is conducted in 2014.

The number of actionable items (recommendations that were specific and measurable) increased from 30% in 2009 to 90% in 2010, 93% in 2011, 100% in 2012, and 94% in 2013. Nearly all of the performance improvement recommendations were

aligned to a performance deficit, in 2009 and 2010 alignment to deficits was noted in 100% of recommendations, in 2011 96% were aligned, in 2012 96% were aligned and in 2013 94% were aligned to performance deficits.

One hundred and fourteen (67%) of the one hundred and seventy recommended action items were completed since 2009; 9 were completed in 2009; 7 in 2010; 37 in 2011; 36 in 2012; 23 in 2013.

Table 4 *Performance Improvement Recommendations by Year, Performance Domain, Recommendation Type, Actionable Status, and Alignment to Performance Deficit*

Academic Year	Performance Domain Evaluated and Number of Recommendations Generated for Each	Recommendation Types and Number Generated for Each	Action Items Completed	Actionable Items	Aligned to Deficit
2009 (n*=10)	Communication (1); Education/Didactics (1); Education/Electronic Medical Records (1); Education/Rotation Schedule (1); Resident Clinic (1); Surgical Curriculum (5)	Communication (3); Meeting (5); Report (1); Rotation Schedule Revision (1)	9 (90%)	3 (30%)	10 (100%)
2010 (n=10)	Education/Didactics (5); Education/Policy (1); Education/Rotation (1); Remediation (1); Surgical Curriculum (2)	Curriculum Development (2); Didactic Schedule Change (4); Policy Development (1); Remediation Protocol Development (1); Schedule Revision (2)	7 (70%)	9 (90%)	10 (100%)
2011 (n=69)	Education/Didactics (21); Evaluation (6); Leadership (7); Remediation (6); Resident Clinic Education (6); Resident Clinic Operations (8); Surgical Curriculum (15)	Communication (1); Curriculum Development (23); Didactic Schedule (2); Evaluation Protocol Development (6); Faculty Responsibilities (1); Leadership Succession (5); Leadership Culture (2); Remediation Protocol Development (6); Resident Clinic/Operations (8); Resident Clinic/Patient Scheduling (1); Resident Clinic/Teaching (6); Resource Allocation (1); Surgical Curriculum Development (1); Surgical Evaluation Protocol	37 (54%)	64 (93%)	66 (96%)

Academic Year	Performance Domain Evaluated and Number of Recommendations Generated for Each	Recommendation Types and Number Generated for Each	Action Items Completed	Actionable Items	Aligned to Deficit
		Development (6)			
2012 (n=49)	Education/Didactics (12); Education/Curriculum (9); Education/OKAPs (4); Education/Technology (8); Remediation (5); Surgical Case Volume (1); Surgical Curriculum (6); Surgical Volume Tracking (4)	Curriculum Development (8); Didactic Schedule (4); Didactics/Food (1); Education Technology (3); Faculty Responsibilities (2); OKAP Curriculum Development (6); Online Curriculum Development (8); Online Resident Scheduling (1); Remediation Protocol Development (4); Surgical Curriculum Development (7); Surgical Evaluation Protocol Development (5)	38 (78%)	49 (100%)	47 (96%)
2013 (n=32)	Education/Didactics (8); Education/OKAPs (9); Education/Technology (5); Surgical Curriculum (8); Surgical Volume Tracking (2)	Curriculum Development (2); Didactic Schedule (2); Evaluation Protocol Development (1); OKAP Curriculum Development (9); Online Curriculum Development (6); Remediation Protocol Development (1); Surgical Curriculum Development (4); Surgical Evaluation Protocol Development (6); Surgical Curriculum Protocol Adherence (1)	23 (72%)	30 (94%)	30 (94%)

* n refers to the number of recommendations

Table 5 demonstrates the type of recommendations generated, number of recommendation actions completed, and the percentage completion rate for each type of recommended action. Recommendations for curriculum development improvement dominated the types of recommendations generated across all years with 76 recommendations for improvement (35 general, 15 OKAP-specific, 14 online, 12 surgical) representing 44

% of all recommendations. Other types of recommended actions proposed most frequently include surgical evaluation protocol development (17), didactic schedule changes (12), and remediation protocol development (12).

The recommendation types with the highest number of actions completed include curriculum development (32 general curriculum-related actions completed; 14 OKAP-specific curriculum development actions completed), surgical evaluation protocol development (14 actions completed), and didactic schedule changes (10 actions completed). The recommended action type completion rate varied widely from 0% complete (resident clinic teaching) to 100% (didactics/food; education technology; faculty responsibilities; leadership culture/succession; meetings; online resident scheduling; policy development; report generation; resident clinic/patient scheduling; resident schedule revision; resource allocation; and surgical curriculum protocol adherence). Recommendation types with only one recommended action were completed in all cases, in all years. Recommendations types with two recommended actions were completed 67% of the time, (2/3 categories); recommended types with three recommended actions were completed 100% of the time (1/1 category); recommended types with four recommended actions were completed 75% of the time (3/4 categories);

and recommended types with five recommended actions were completed 100% of the time (1/1 category). The recommendation action types with completion rates of less than 50% include improvements to evaluation protocol (43% complete), resident clinic teaching (0% complete), and surgical curriculum development (42% complete).

Table 5 *Performance Improvement Recommendation Types and Completion Rates 2009-2013*

Recommendation Type	Number of Recommendations	Completed Recommendations	Completion Rate
Communication	4	3	75%
Curriculum Development	35	18	51%
Didactic Schedule	12	10	83%
Didactics/Food	1	1	100%
Education Technology	1	1	100%
Evaluation Protocol	7	3	43%
Evaluation Technology	2	1	50%
Faculty Responsibilities	3	3	100%
Leadership Succession	5	5	100%
Leadership: Culture	2	2	100%
Meeting	5	5	100%
OKAP Curriculum Development	15	14	93%
Online Curriculum Development	14	7	50%
Online Resident Scheduling	1	1	100%
Policy Development	1	1	100%
Remediation Protocol Development	12	9	75%
Report	1	1	100%
Resident Clinic Operations	8	4	50%
Resident Clinic: Patient Scheduling	1	1	100%
Resident Clinic: Teaching	6	0	0%
Resident schedule revision	2	2	100%
Resource Allocation	1	1	100%
Rotation Schedule Revision	1	1	100%
Surgical Curriculum	12	5	42%

Development			
Surgical Curriculum Protocol Adherence	1	1	100%
Surgical Evaluation Protocol Development	17	14	82%

Program Performance Outcomes

Table 6 denotes the program outcomes related to the ACGME accreditation site visit (external regulatory audit process). The ACGME conducted site visits with the KEI Ophthalmology Residency program in December 2006 (prior to the institution of the annual program evaluation requirement) and in again February 2012 (ten months after the first systematic evaluation process report was distributed). The results of the 2006 site visit include a 5-year accreditation cycle, 2 program citations and 1 program commendation. The results of the 2012 site visit include a shorter 4-year accreditation cycle, 3 program citations (1 repeat of a 2007 citation), and 1 program commendation.

Table 6 *ACGME Accreditation Outcomes*

Outcome	2007	2012
ACGME Cycle Length	5 years	4 years
Number of Citations	2	3
ACGME Citation Type	<ol style="list-style-type: none"> 1. Patient Care/Minimum Operative #s, equitable distribution of cases 2. Didactics/Insufficient instruction in ethics 	<ol style="list-style-type: none"> 1. Patient Population Volume & Variety 2. Patient Care/Minimum Operative #s 3. Resident Scholarly Activity/Participation
Number of Commendation(s)	1	1
Commendation Type	1. Substantial Compliance with ACGME	1. Substantial Compliance with ACGME

	requirements	requirements
--	--------------	--------------

Table 7 illustrates the OKAP (In-service examination) results for years 2007 through 2013. The threshold for passing was determined by a 33rd percentile or higher rank, failing by a 32nd percentile or lower rank. As noted, the percentage of residents with a passing OKAP score is as follows, 62% pass rate in 2007; 67% pass rate in 2008; 52% pass rate in 2009; 52% pass rate in 2010; 62% pass rate in 2011; 38% pass rate in 2012; and a 71% pass rate in 2013. The OKAP scores in years post-implementation of the systematic evaluation process were somewhat similar to previous years in 2011, declined in 2012, then were higher than previous years in 2013.

Table 7. *Program OKAP (In-Service) Examination Results 2007 to 2013*

Year	Pass/Total	Percent Passed
2007	13/21	62%
2008	14/21	67%
2009	11/21	52%
2010	11/21	52%
2011	13/21	62%
2012	8/21	38%
2013	15/21	71%

*Percentile scores provided by the test examination board

**Pass = 33rd percentile or higher; Fail = 32nd percentile or lower

Table 8 notes the ophthalmology board examination rates for the seven graduates of the program in 2007, 2008, 2009, 2010, and 2011. Pass rates vary each year with a range of 71 to 100% passing the written and oral board examinations. The ACGME requires that 80% of each program's graduating Ophthalmology residents take the ophthalmology board examination each year and that 60% pass their examination (written and oral). Data from 2012 and 2013 (years post-implementation of the systematic evaluation process) are not yet available for review.

Table 8. *Program Graduate Board Examination Results Evaluation 2007 to 2011*

Year	Written Examination Passed/Taken	Pass Rate %	Oral Examination Passed/Taken	Pass Rate %
2007	5/7	(71%)	5/7	(71%)
2008	7/7	(100%)	7/7	(100)
2009	6/7	(86%)	6/7	(86%)
2010	5/7	(71%)	5/7	(71%)
2011	6/7	(86%)	5/7	(71%)
2012	Data not available			
2013	Data not available			

Table 9 presents the percentage of graduating residents meeting ACGME minimum surgical case requirements. Seven residents graduated from the program in each year reported. Only one (14%) of the seven graduates in 2009 met the minimum requirements for surgical volume in all categories. In 2010 and 2011, two (29%) of the seven graduates (per year) met the minimum requirements for surgical volume in all categories. In 2012 and 2013 seven (100%) of the seven graduates (per year) met the minimum requirements for surgical volume in all categories.

Table 9. *ACGME Ophthalmology Resident Case Log Report Results*

Year	Graduating Residents Who Met ACGME Minimum Surgical Requirements (All Categories)	% of Graduating Residents Who Met ACGME Minimum Surgical Requirements
2009	1/7	14%
2010	2/7	29%
2011	2/7	29%
2012	7/7	100%
2013	7/7	100%

Table 10 examines the results of the annual ACGME Resident Survey, administered electronically to the KEI Ophthalmology program by the ACGME each

year. The table compares survey results in multiple constructs and notes if the means score for the construct has increased or decreased as compared to 2010 (the year prior to the systematic program evaluation process implementation). The ACGME significantly revised the survey in 2010 such that comparison to previous years' surveys is not feasible. Compared to the 2010 data, results were favorable in 2011, 2012, and 2013 in multiple areas. In 2011 improved mean scores were noted in duty hours, resources and clinical education performance as compared to 2010 results. In 2012 improved mean scores were noted in duty hours, resources, didactics, and clinical education as compared to 2010 results. In 2013 improved mean scores were noted in duty hours, educational content, resources, didactics, and clinical education as compared to 2010 results. Decreased mean scores (as compared to 2010 data) were reported for the following faculty (2011, 2012, 2013), evaluation (2011, 2012, 2013), and educational content (2011). In sum, since 2010 the KEI residents reported increased compliance with ACGME requirements (as demonstrated in increased survey mean scores) 12 times, and reported decreased compliance 7 times.

Table 10. *ACGME Annual Resident Survey Data 2010 to 2013*

Survey Construct	2010	2011	*	2012	*	2013	*
Duty Hours	4.8	5	+	5	+	5	+
Faculty	4.5	3.8	-	4	-	4	-
Evaluation	4.7	4.1	-	4.6	-	4.5	-
Educational Content	4.2	3.8	-	4.2		4.5	+
Resources	3.4	4	+	4.5	+	4.4	+
Patient Safety	No data	No data		4.5		4.3	
Teamwork	No data	3.8		4.5		4.1	
Didactics	3.3	3.3		3.7	+	3.8	+
Clinical	3.32	3.5	+	3.8	+	3.9	+

Total increased since 2010	NA		3		4		5
Total decreased since 2010	NA		3		2		2

*Indicates increase or decrease from 2010 baseline mean score for the section.

Stakeholder Perspectives: Benefits and Challenges of the Systematic Evaluation Process

Three surveys were conducted for the current study, 1.) 2011 pre-intervention survey, 2.) 2011 post-intervention survey, and 3.) 2013 program administrator survey. The 2011 pre-intervention survey was presented to residents, faculty and program administrators in attendance at the February 2011 meeting scheduled to introduce the “Impact Evaluation Process” to the program. The 2011 post-intervention survey was presented to the residents, faculty, program administrator and program support staff in attendance at the KEI annual retreat in April 2011 where the residents presented the findings of their evaluation efforts as prescribed in the systematic evaluation process instructions.

Thirty-eight program stakeholders (100% of the residents, faculty, and program administrators in attendance) completed the pre-intervention survey and twenty-three stakeholders (60% of the residents, faculty, and program administrators in attendance) completed the post-implementation survey. The pre-intervention survey asked the following questions:

1. What are the benefits you expect from using the Program Performance Portfolio* to evaluate your program?
2. What are the challenges you expect from using the Program Performance Portfolio* to evaluate your program?

The post-intervention survey asked the following questions:

1. What were the benefits of the method used for program evaluation (Program Performance Portfolio*) for the 2011 Ophthalmology Resident Program Annual Retreat?
2. What were the challenges of the method used for program evaluation (Program Performance Portfolio*) for the 2011 Ophthalmology Resident Program Annual Retreat?
3. What, if anything, did you learn from your participation?
4. How can we improve this method of program evaluation?

*The “Program Performance Portfolio” was the name utilized for the application of the implementation of the “Impact Evaluation Process” (Guerra-Lopez, 2007) in the 2011 annual program evaluation instructions and in the 2011 surveys.

Table 11 presents the anticipated benefits of implementation of the systematic evaluation process expressed by residents, faculty, and program administrators who took part in the 2011 systematic evaluation process. The most frequently reported anticipated benefits included program improvement (15 responses), improved communication (5 responses), and improved education (4 responses).

Table 11 2011 Stakeholder Perceptions of Anticipated Benefits of Systematic Evaluation Process

Pre-intervention Survey Question 1:	
What are the benefits you expect from using the Program Performance Portfolio* to evaluate your program?	
Response Category	Number of Responses in this Category
Constructive feedback/criticism	2
Direction	1
Implemented changes	2
Improved academic performance	1
Improved clinic function	1
Improved communication	5
Improved education	4
Improved evaluation process	1
Improved program culture	1
Improved resident performance	2
Increased faculty involvement	1
Program improvement	15
Rigorous evaluation	1
Strategic Development	1

Stakeholder perceptions of the actual benefits of implementation of the systematic evaluation process are presented in Table 12. The most frequently reported actual benefits reported included improvements to the evaluation process itself (11 responses), communication between stakeholders (7 responses), multiple stakeholder involvement (7 responses), resident input/perspectives (5responses), and program improvement (4 responses).

Table 12 2011 Stakeholder Perceptions of Actual Benefits of Systematic Evaluation Process

Post-Intervention Survey Question 1:	
What were the benefits of the method used for program evaluation (Program Performance Portfolio*) for the 2011 Ophthalmology Resident Program Annual Retreat?	
Response Category	Number of Responses in this Category
Collaboration between residents, faculty, administrators	2
Communication: dialogue between residents, faculty, administrators	7
Critical thinking	1
Data used for improvement	3
Improved evaluation process	11
Everyone together	1
Improved clinical rotations	1
Informative	3
Interactive meeting	1
Multiple stakeholder involvement	7
Program improvement	4
Increased awareness of issues	1
Resident input/perspectives	5
Strategic planning	1

Residents, faculty, and program administrators shared their perceptions of anticipated challenges of implementing the systematic evaluation process as demonstrated in the survey results presented in Table 13. The most frequently reported anticipated challenges included burden of the evaluation process (7 responses), lack of anticipated changes (7 responses), faculty investment in the process (4 responses), and challenges to the organization of the evaluation process (4 responses).

Table 13 2011 Stakeholder Perceptions Anticipated Challenges of the Systematic Evaluation Process

Pre-Intervention Question 2:	
What are the challenges you expect from using the Program Performance Portfolio* to evaluate your program?	
Response Category	Number of Responses in this Category
Burden (i.e., too much work)	7
Communication	3
Faculty Investment	4
Implementation Challenges	2
Lack of Anticipated Changes	7
Organization of Evaluation	4
Time	3
Timely implementation of recommendations	1
Timing of evaluation (OKAP in-service exam)	1

Stakeholder perceptions of actual challenges to the 2011 systematic evaluation process are reported in Table 14. The most frequently reported actual challenges include not having enough time to prepare the evaluation results (9 responses) burden of the evaluation process (8 responses), limitations of the data provided (5 responses), and challenges to implementing the recommended actions (4 responses).

Table 14 *2011 Stakeholder Perceptions Actual Challenges of Systematic Evaluation Process*

Post-Intervention Question 2:	
What were the challenges of the method used for program evaluation (Program Performance Portfolio*) for the 2011 Ophthalmology Resident Program Annual Retreat?	
Response Category	Number of Responses in this Category
Being candid during process/confidentiality	2
Burden	8
Challenges to implementing recommendations	4
Commitment to program changes	3
Data limitations	5
Faculty investment	2
Organization of evaluation process	2
Not enough time to prepare	9
Timing of Evaluation (Before OKAP exam)	2
Tracking Progress	2

Table 15 denotes the stakeholder perceptions of the lessons learned from their participation in the 2011 systematic evaluation process. The most frequently reported lessons learned included increased awareness of program issues (10 responses) and faculty learning about resident perspectives of the program (5 responses).

Table 15 2011 Participant Lessons Learned

Post-Intervention Question 3:	
What, if anything, did you learn from your participation?	
Response Category	Number of Responses in this Category
Barriers to successful improvement	1
Clarification of misconceptions	1
Difference in stakeholder perspectives	1
Evaluation process is good way to organize ideas	1
Faculty learned about resident perspectives	5
Faculty were invested	1
Good opportunity for dialogue	1
Improved evaluation	1
Increased awareness of program issues	10
Multiple changes to improve program	1
Multiple stakeholders' commitment to process	3
Need to change my teaching	1
Organization is important part of evaluation process	3
Positive perceptions of the program	1
Program improvement opportunities	3
Program problems can be resolved	1
Resident evaluations of faculty are anonymous	3

The most frequently reported stakeholder suggestions for improvement to the systematic evaluation process implemented in 2011 are illustrated in Table 16 and include communication about progress of recommended actions (10 responses), changing the timing of the evaluation process to after the OKAP in-service examination (5 responses), allowing more time for the evaluation process itself (4 responses), and ensuring that the recommended changes are implemented (4 responses).

Table 16 2011 Stakeholder Suggestions for Improving the Systematic Evaluation Process

Post-Intervention Question 4:	
How can we improve this method of program evaluation?	
Response Category	Number of Responses in this Category
Communication about progress	10
Change meeting time/day	1
Continue this process annually	2
Implement suggested changes	4
Limit evaluation scope	2
Increase faculty involvement	1
Allow more time for evaluation process	4
Satisfied with this method	4
Streamline process	3
Change timing of evaluation process (schedule it after OKAP in-service exam)	5

2013 Program Administrator Survey Results

In November 2103 a survey was sent via email using “Survey Monkey” to the Program Chair/Director, a KEI Residency Program Faculty member, the two Chief Residents from academic year 2012-2013, and the Program Coordinator. Each of these program stakeholders were participants in all three years of the implemented systematic evaluation process (years 2011, 2012, and 2013). This survey with member check follow up to discuss the findings was used in lieu of the planned semi-structured face-to-face interviews due to the fact that two of the participants no longer live in the state of Michigan. The results of this survey are presented in Tables 17 through 20. The number of responses in each category may exceed the number of respondents due to the fact that questions 2-4 allowed for open-ended commentary and participants could make multiple statements that fell into the same category.

As illustrated in Table 17, when asked the question, “In your opinion, has using a systematic evaluation process in the required annual program evaluation of the Ophthalmology residency program resulted in program improvement?” all 5 of the respondents responded “Yes.”

Table 17 2013 Program Administrator of 3 years of Systematic Evaluation Process: Program Improvements

Question 1:	
In your opinion, has using a systematic evaluation process in the required annual program evaluation of the Ophthalmology residency program resulted in program improvement? (n=5)	
Yes = 5	No = 0

Table 18 denotes reported benefits of using the systematic evaluation process as reported by the program administrators previously listed. The most frequently reported benefits included improved evaluation process (8 responses), and increased stakeholder engagement (3 responses).

Table 18 2013 Program Administrator of 3 years of Systematic Evaluation Process: Benefits of the Process

Question 2:	
What are the benefits of utilizing the systematic evaluation process in the required annual program evaluation of the Ophthalmology residency program? (n=5)	
Response Category	Number of Responses in this Category
Evaluating the evaluation process	1
Improved program documentation	1
Improved program focus	1
Improved performance tracking	1
Improved evaluation process	8
Increased status of evaluation process	1

Signal of program investment in education	1
Increased stakeholder engagement	3

Table 19 presents program administrator reported challenges to implementing the systematic evaluation process. The most frequently reported challenges included stakeholder investment in the process (6 responses), time burden (5 responses), and organization of the process (3 responses).

Table 19 2013 Program Administrator Perceptions of 3 years of Systematic Evaluation Process: Challenges of the Process

Question 3:	
What are/were the challenges of this process? (n=5)	
Response Category	Number of Responses in this Category
Lack of departmental evaluation skills	1
Organization of the process	3
Participant burden	1
Process required professional evaluator	2
Resource burden	2
Stakeholder investment	6
Stakeholder participation	1
Time burden	5
Time constraints	1
Time management	1
Timely reporting	1

Program administrators were asked to express the personal impact of three years of implementation of the systematic evaluation process. Table 20 demonstrates that the most frequently reported impact included using the evaluation process in other areas of work (6 responses), recognizing challenges in obtaining data (5 responses), need for increased departmental evaluation capacity (2 responses), and need for more frequent communication of findings/progress (2 responses).

Table 20 2013 Program Administrator of 3 years of Systematic Evaluation Process: Personal Impact

Question 4:	
How has engagement in the process impacted you personally, that is, in what ways (if any) did the process change your behaviors and approach to your work as a physician, faculty member, or program administrator? (n=5)	
Response Category	Number of Responses in this Category
Challenges obtaining data	5
Data management challenges	1
Illustrated program investment	1
Improved departmental vision	1
Improved performance tracking	1
Improved stakeholder knowledge of other perspectives	1
Increased awareness of importance of measurement	1
Increased evaluation capacity	1
Increased knowledge of program strengths and weaknesses	1
Informed decision making	1
Need for increased department evaluation capacity	2
Need for more frequent communication of findings/progress	2
No impact	1
Process positively impacted resident learning	1
Use the evaluation process in other areas of work	6

Summary

This chapter presented the results of the study in three main areas, annual program evaluation results, program outcomes, and stakeholder perceptions about the systematic evaluation process. Annual program evaluation results in the years when the systematic evaluation process were implemented were different from the previous years, indicating that the answer to the first research question is yes, the evaluation results generated using

a systematic evaluation process differed in multiple aspects. The second research question was addressed by examining the recommended actions for each year and determining if they led to action-based performance improvement plans tied to specific performance gaps. Again, the answer was yes, and results indicated that previous years' plans were also action-based and tied to performance gaps, although in smaller number and with simpler to implement recommendations. The third research question was addressed by examining a variety of program outcome indicators and showed improvement in all but one area, accreditation results (complete board examination results were not available). The fourth research question was addressed by the utilization of three separate surveys of program stakeholders. Reactions to the systematic evaluation process were described prior to the initial implementation, immediately after the implementation, and again in 2013 after three years of implementation of the systematic evaluation process. The next chapter discusses these results.

CHAPTER IV. Discussion

The general purpose of the present evaluation research study was to examine the difference in outcomes when utilizing a systematic evaluation process, an adapted version of the “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008), to analyze performance compared to less rigorous evaluation methods. The specific purpose was to examine the impact of a systematic evaluation process on evaluation findings; determine if the process led to action-based performance improvement plans tied to specific performance gaps and improved outcomes; and identify the benefits of and the barriers to implementation of this process in a graduate medical education residency program. The purpose of this chapter is to discuss the results.

Conclusions

The evaluation results generated using the “Impact Evaluation Process” as a means for systematic annual program evaluation (Guerra-López, 2007b, 2007c; Guerra-López, 2008) differed from previous years’ annual program evaluation results in multiple ways. The volume and types of recommended actions generated dramatically increased in the years when the systematic evaluation process was employed when compared to previous years’ results and the number of completed recommended actions also increased.

The utilization of a systematic evaluation process did lead to action-based performance improvement plans tied to specific gaps. This was true in the years prior to the implementation of the systematic evaluation process as well as in those years post-implementation.

The majority of the program outcomes examined in the current study improved and one worsened compared to the years prior to implementation of the systematic evaluation process. Ophthalmology certification board examination results for 2012 and 2013 were not available at the time of this report. This program outcome was revealed to be inadequate for analysis in the context of this study timeline (summary board reports are released to the programs every four years).

Stakeholders reported a variety of anticipated benefits of and barriers to implementation of a systematic evaluation process. The 2011 pre-and post-intervention results reveal that stakeholder preliminary expectations of the systematic evaluation process implementation did not frequently match stakeholder final perceptions post-implementation. Stakeholders expected program improvements, improved communication and improved education as a result of the intervention. Reported benefits included improvements to the evaluation process, improved communication, stakeholder engagement and multiple stakeholder involvement. Stakeholders expected to be challenged by the burden of the process, organization of the evaluation, and level of faculty investment. Reported challenges included not having enough time to prepare, burden of the evaluation process, data limitations, stakeholder investment, time burden, and challenges to implementing the recommendations.

Annual Program Evaluation Results

The utilization of a systematic evaluation process resulted in the generation of increased volume and type of recommended actions for improving program performance. In the years prior to implementation of the systematic evaluation process the annual report did not include an action plan for performance improvement as required by the

ACGME nor did the meeting include resident representation. Instead, meeting minutes documented faculty member discussions about program performance with actionable items embedded in the recorded meeting dialogue.

The results of the systematic evaluation process approach to the KEI Ophthalmology residency program's annual program evaluations in 2011, 2012, and 2013 differed from the previous years in multiple ways. The evaluation process itself was different, the number and type of participants larger, and the responsibilities of the analysis of data fell largely to the residents to complete with faculty facilitators assigned to assist. It is important to note in each year that the systematic evaluation process was employed all 21 of the Ophthalmology residents participated in the evaluation in addition to many of the faculty members, the program coordinators, program director, and other program staff. An evaluation expert was engaged to facilitate the evaluation process.

Differences in results include increased volume and type of generated recommended actions as well as increased number of completed recommended action items. The number of generated recommended actions increased from 10 each in 2009 and 2010 to 69 in 2011, 49 in 2012, and 32 in 2013. Completion rates decreased as the number of recommended actions increased (except in 2010) with 90% of recommended actions completed in 2009, 70% in 2010, 54% in 2011, 78% in 2012, and 72% in 2013. The recommended actions in the 2010 meeting minutes were more complex than those recorded in the 2009 meeting minutes (e.g. email a reminder vs. develop a curriculum). Although the percentage of completed recommended actions were lower in years where the systematic evaluation process was employed, the number of completed recommended actions was greater in all of those years. Thus, the employment of a systematic evaluation

process improved the results of the annual program evaluation meeting by increasing the number of improvements to the program when compared to previous years.

The depth and range of recommended actions was improved during the years that a systematic evaluation process was employed. In 2009 five of the ten recommended actions were to schedule a meeting for further dialogue, two were to “keep in contact” or “keep the committee posted” on plans for improving an identified program deficit, one was to send out an email reminding residents to complete a required course, one was to provide a surgical case report, and another to revise a rotation schedule. In 2010 the recommended actions were slightly more robust, with two recommendations for curricular development in areas with identified performance deficits (i.e., contact lens, cornea, anterior segment lectures and development of a structured simulator curriculum); four recommendations for changes to the didactic schedule, one to change a pertinent policy, one to develop a remediation protocol to assist residents with sub-par OKAP in-service examination scores, and two to change residents schedules to improve surgical case volume. Although the majority of the recommended items were deemed actionable (specific and measurable) in 2009 and 2010, none of these actionable items were formally tracked over time and reported to the program stakeholders.

In 2011 recommended actions for improvement included 14 types, the majority of which were related to curriculum development, recommended changes to the resident clinic, specific and measurable suggestions for the development of a formal remediation protocol, and calls for evaluation protocol development to assess resident competencies. In 2012 recommended actions for improvement included 12 types with emphasis on general curriculum development (including specifics for improving the poor OKAP in-

service exam scores in 2012), online curriculum development, surgical curriculum development, and continued call for remediation and evaluation protocol development. In 2013 recommended actions for improvement included 9 types with focus on surgical activities (curriculum development, evaluation protocol development, and protocol adherence), and continued call for OKAP and online curriculum developments.

More robust recommendations led to tangible, documented improvements that continue to enhance the educational program. For example, recommended actions from the 2012 annual evaluation report for improving the surgical curriculum, development of improved surgical evaluation protocols, and resident adherence to surgical evaluation protocols led to an innovative performance-tracking tool referred to as the “Surgical Report Card.” This report card provides regular monitoring of surgical activities at the individual resident, resident cohort, and program levels. The program administrators and residents are provided monthly reports of surgical volume, compliance with surgical simulator training requirements, compliance with newly developed online real time surgical evaluation procedures, and compliance with other surgical performance data tracking requirements (e.g., surgical outcomes entered into an online database). This tool directly addresses three of the four ACGME citations received in the past two audits, and has resulted in increased reported surgical volume in all cohorts in addition to increased resident compliance with other surgical requirements. In the two years since this tool has been implemented 100% of graduating residents have met ACGME minimal requirements for surgical cases in all required categories, a performance indicator that directly addresses the ACGME citations for surgical volume.

Analysis of performance deficits in the activities associated in the resident-operated Ophthalmology clinic led to a significant number of recommended actions in the 2011 report. Recommendations for improvement to the resident clinic resulted in the formation of a “Resident Operations Committee” that meets monthly to review the operational activities in the very busy residents’ Ophthalmology clinic. Results of these meetings include implementation of protocols that increased compliance with supervisory requirements, significantly improved financial performance of the clinic, and a resident award from the Detroit Medical Center’s QuESST (Quality Improvement and Safe Systems Training) 2012 Resident Research Day Competition, recognizing excellence in quality improvement projects at the DMC.

Repeated calls for curriculum development led to the implementation of multiple changes to the didactic schedule and lecture processes as well as increased emphasis on OKAP and board examination educational activities. Implemented improvements designed to improve examination scores include the incorporation of resident study halls into the lecture schedule; three pre-OKAP examination study days off; Friday morning “Breakfast Club” presentations by the residents, for the residents where each provide presentations and self-generated questions for the audience on OKAP topics on which they performed poorly (requiring higher level learning to produce) with the guidance of a faculty facilitator. More than 37 recommended actions related to curriculum development alone have been instituted since 2011 compared to 4 in 2009 and 3 in 2010. Ten recommended actions related to the didactic schedule have been implemented, as have 14 recommended actions related to surgical evaluation protocol. These and many other data

driven implemented performance improvement projects were guided by recommended actions derived from the analysis of data related to documented performance deficits.

The high volume of recommendations generated in 2011 raises concern about the scope of the systematic evaluation process as implemented. In 2011, the year with the greatest number of recommended actions, only 54% of the recommended actions were completed. A significant proportion of these items were deemed actionable (94%) and most were aligned to an identified program deficit (96%). The percentage of recommended actions completed in 2012 was greater (78%) with 49 actionable items generated and 96% of those aligned to a program deficit. The number of recommended actions decreased further (to 32) in 2013, with 72% of those completed in the six months since the evaluation took place.

Reflecting on Table 5, which depicts the full volume of recommended actions completed over all years analyzed, it becomes evident that categories with fewer recommended actions are more likely to be completed. For example, 100% of types with a single recommended action were completed (9/9 categories); 67% of types with two recommended actions were completed (2/3 categories); 100% with three recommended actions were completed (1/1 category); 75% with four recommended actions were completed (2/2 categories); and 100% with five recommended actions were completed (1/1 category). Further, these recommended actions with lower numbers of suggested improvements fall into categories that are non-curricular, that is, requiring less intellectual capital to complete (e.g., ordering food, buying mini iPads, changing schedules, writing a policy, spending dollars, etc.). Some of the recommended actions took more than a year to complete. Determining a reasonable scope for the systematic

evaluation process may lead to improved perceptions of the process; reduce burden on both the program and participants; and provide increased opportunity to focus program analyses and resources more challenging program deficits.

Program Outcomes

Program outcome data demonstrated measurable improvements and one report of decreased performance. Program accreditation performance declined in 2012 compared to 2007, with a shorter 4-year accreditation cycle length (vs. 5-year in 2007) and three citations (vs. 2 in 2007). Resident performance on the OKAP in-service examination improved to 62% pass rate in 2011 (compared to 52% in 2009 and 2010), decreased in 2012 to 38% pass rate, then increased again in 2013 to a 71% pass rate, the highest in the past seven years. Significant improvement in graduating resident surgical case volume was reported in years 2012 and 2013, with 100% of graduating residents meeting minimal surgical requirements in those years compared to 14% on 2009, and 29% in 2011 and 2012. Significant progress was also made in completing recommended actions for program improvements, with the number of improvements implemented increasing dramatically in the years post implementation of the systematic evaluation process. One hundred and fourteen (67%) of the one hundred and seventy recommended action items were completed since 2009; 9 were completed in 2009; 7 in 2010; 37 in 2011; 36 in 2012; 23 in 2013.

In their report of the February 2012 ACGME site visit results, the ACGME Residency Review Committee cited the program for three performance deficits, two related to surgical volume and one related to resident research. The program director's

response to ACGME citation for surgical volume referenced the departure of an essential faculty member, a cornea surgeon, and described the program's efforts to replace her with expected improvement in surgical volume in cornea. The citations related to patient population and surgical variety can be directly traced to insufficient data entry on the part of the residents. Program surgical records indicate that the residents were not logging all of their surgical case volume and this deficit was demonstrated in the citation by the ACGME. The citation for lack of resident research was perplexing; each resident is required to participate in a research project every year in the program and all present their projects at an annual KEI Clinical Conference. Program director's query of the residents regarding this citation revealed that they reported the program noncompliance in the ACGME resident survey and during the site visit because they understood the question in terms of ranking KEI program performance as compared to others with renowned and more robust Ophthalmology research departments.

Improvement to the residents' OKAP in-service examination scores was reported in 2013. The OKAP examination is proctored in late March of each year. The program improved to a 62% pass rate in March 2011 (compared to 52% in 2009 and 2010), decreased in March 2012 to a 38% pass rate, and then increased again in March 2013 to a 71% pass rate, the highest in the past seven years. Recommended actions for OKAP curriculum development were all completed in either 2012 or 2013, none were completed in 2011. Completed annual program evaluation recommended actions related to this topic include recommendations for "more OKAP-centric lectures," "provide residents with subject specific results on the practice OKAP exam," and "enroll residents with OKAP scores below the 30th percentile in a remediation program," "one lecture per month

focused on OKAP-style questions,” “integration of Wayne State University basic science course on the biology of the eye into the program,” “develop and implement OKAP review course,” “purchase OPHTHO questions (online program) for residents,” “develop OKAP summary report,” and “develop individual resident education plans for the next academic year,” “engage faculty to incorporate 5-10 OKAP-style questions at the end of Grand Rounds,” “implement survey monkey questionnaire to evaluate previous OKAP interventions,” “maintain current OKAP review session curricula and schedule,” “align 2013-2014 Breakfast Club presentations to revised didactic schedule using missed key words, both individual and institutional,” and “maintain online OPHTHO questions as resident resource.” Since 2011, evaluation of resident OKAP scores has occurred during each of the annual program evaluations and 13 recommended actions have been completed. As the number of OKAP-centric completed recommended actions for program improvement has increased, so have the resident OKAP examination scores.

Board examination results do not yield any significant data for the purposes of this study. Board passage rates for years 2012 and 2013 are not yet available, and given the years of study required to master the materials presented in the three-year residency program, impact is not yet discernable for this indicator. Residents take the written Ophthalmology board examination nine months post-graduation, with the oral examination taken up to fourteen months post-graduation. The program does not receive the results of the exam until two years post-graduating year.

Significant improvement in graduating resident surgical case volume was reported in years 2012 and 2013, with 100% of graduating residents meeting minimal surgical requirements in those years compared to 14% on 2009, and 29% in 2011 and 2012. The

“Surgery Report Card” tool, initiated in 2011 and improved upon each year since, has resulted in consistent and frequent monitoring as well as routine reporting of resident surgical-related activities. The report card includes multiple performance metrics reported by individual resident, resident cohort, and program level. Resident logging of surgical cases, adherence to surgical simulator requirements, completion of surgical courses, adherence to surgical self-evaluation requirements, and ratings of professionalism (compliance with surgical tracking) are all measured, monitored and reported each month. Surgical boarding privileges are withheld for non-compliance. The report is distributed to individual residents and discussed at the monthly resident meeting. Significant improvements in case logging and adherence to protocols have resulted in much higher surgical volumes reported.

ACGME resident survey results have improved in some areas and declined in others. Since 2010 the KEI residents reported increased compliance with ACGME requirements (as demonstrated in increased survey mean scores) 12 times, and reported decreased compliance 7 times. The foci of the systematic evaluation process, that is the evaluation questions, performance indicators selected, and data analyzed in the annual program evaluation reflect the deficits noted in the survey each year. Completed recommended actions correlate to improvements in resident survey results. As noted earlier, the recommendation types with fewer recommended actions per category and high completion rates (e.g. resources, clinical) correlated to domains with the largest increase in ACGME resident survey scores. Other survey performance domains also reflect program improvements, such as didactics and educational content (areas with lower rates of completion, but a significant number of recommended actions completed).

Survey performance domains that remain lower than 2010 results are categories that the program has not expended significant efforts and resources to improve. For example, there are two recommendation types that require faculty involvement, faculty responsibilities and resident clinic: teaching. The first, “faculty responsibilities” recommended “get more faculty to attend grand rounds,” “increase faculty conference attendance,” and “faculty to present grand rounds once per month.” These actions were relatively simple to complete and all were done. The more complex recommended actions related to improving faculty engagement in the resident clinic remain incomplete (0 of 6 recommended actions complete).

As previously discussed, the implementation of a systematic evaluation process in 2011, 2012, and 2013 generated one hundred and fifty recommended actions for program improvement. The volume of recommendations from these years is large and ninety-seven were completed; 37 from 2011; 36 from 2012; and 23 from 2013. The recommendation types with the lowest completion rates include resident clinic teaching, surgical curriculum development, and general curriculum development. It is worth noting that all of these recommended actions require significant human resources, “brain capacity” as noted by the program director.

Stakeholder Perspectives: Benefits and Challenges of the Systematic Evaluation Process

Three surveys were conducted in this study. The pre-intervention survey was completed in February the 2011 by residents, faculty, program administrators and program staff prior to the intervention and asked about anticipated benefits and challenges to implementing the process. Residents, faculty, program administrators and program staff completed the post-intervention survey in April 2011. The final longitudinal survey was completed in November 2013 by program administrators who participated in all three years of the implementation of the systematic evaluation process (program director, faculty member, two chief residents, and the program coordinator).

2011 Survey Results

The 2011 pre-and post-intervention results reveal that stakeholder preliminary expectations of the systematic evaluation process implementation did not frequently match stakeholder perceptions post-implementation. Frequently anticipated benefits included expected improvements in communication, improved education, and program improvement. Frequently reported realized benefits post-implementation included improved communication, improved evaluation process and engagement of multiple stakeholders in the process. Most frequently anticipated challenges of the proposed process included burden (i.e., too much work), faculty investment, lack of anticipated program changes resulting from the evaluation, and challenges to the organization of the systematic evaluation process. Frequently reported challenges post-implementation included burden, data limitation, not having enough time to prepare (two weeks were allotted to evaluation teams), and challenges to implementing recommendations. In 2011, stakeholders' expectations about the systematic evaluation process matched reported

results in three areas, improved communication; the burden of the evaluation process; and concerns about changes actually occurring.

Reported benefits indicated that some of the anticipated challenges were overcome during the implementation process. The expectation that faculty investment would be lacking was not realized, in fact, multiple stakeholder involvement was a reported benefit of the process. Preliminary concerns regarding challenges associated with the organization of the evaluation were somewhat ameliorated by the reported benefits of an improved evaluation process, the most frequently reported benefit of all. The anticipated and actual challenge concerning lack of anticipated changes to the program and challenges of doing so reported post-implementation are countered by the nearly 100 completed recommended actions since the systematic evaluation process was employed in 2011.

Enlistment of stakeholders in the evaluation process, “participatory evaluation” was expected to engage stakeholders in decision making, increase abilities to plan and conduct evaluations, and increase evaluation utility. When asked, “What, if anything, did you learn from your participation?” stakeholders resoundingly reported that the process increased awareness of program issues while also allowing faculty to learn about resident perspectives. Participants also noted the multiple stakeholders’ commitment to the evaluation process, recognized that organization is an important part of the evaluation process, and that the process generated program improvement opportunities. Finally, residents reported a realization that their evaluations of the faculty are anonymous, an issue of concern reported in previous program surveys.

The final question of the 2011 post-intervention survey asked for recommendations to improve the systematic evaluation process. Stakeholders requested more frequent communication about progress on recommended actions, asked to change the timing of the evaluation to after the OKAP in-service examination (held at the end of March each year), wished that the program would actually implement the recommended changes, and allow more time for the evaluation process. All but one of these recommendations were completed, the suggestion for more frequent communication about progress on recommended actions was not accomplished. Results are shared with residents every six months and are not routinely shared with faculty members.

2013 Survey Results

Five program administrators were surveyed in 2013, all had participated in the 2011, 2012, and 2013 annual program evaluations using the “Impact Evaluation Process” (Guerra-López, 2007). Each of the participants were involved in every step of systematic evaluation process (although the chief residents were engaged in steps 3-5 only in 2013, their final year of education as part of their administrative role of chief resident). All five (100%) administrators reported that using the systematic evaluation process resulted in program improvement. Most frequently reported realized benefits of the process include improvements to the evaluation process and increased stakeholder engagement. Most frequently reported challenges include stakeholder investment, time burden, organization of the process, the need to engage a professional evaluation, and resource burden. The perceived benefits expressed by administrators matched two of those reported in the 2011 survey, improvement to the evaluation process and stakeholder involvement. The burdens of the process, time and otherwise, were echoed in both the 2011 and 2013 survey.

Administrators were challenged by engagement of the stakeholders, a responsibility that the residents and faculty did not face as they were not responsible for the evaluation process design and task assignments.

Study Limitations

Limits to the current study include both internal and external validity. The single case sample was one of convenience and, lacking randomization, results may not be generalized beyond the KEI Ophthalmology residency program. Without the controlled conditions indicative of experimental designs, conclusions about cause and effect relationships cannot be drawn. Although the Ophthalmology residency program at KEI shares attributes with other residency programs, attempts to generalize the results beyond this program risks drawing conclusions that cannot be supported by the data collected in this case study.

The participatory nature of this study (researcher participating in the evaluation process and analyzing the data) presents challenges to the internal validity of this study. Case studies are reported to be susceptible to the introduction of biases due to the inability to control for outside variables.

Suggestions for Further Research

A multiple case study design might be employed to compare results of utilizing a systematic evaluation process across residency programs either within an institution or across multiple institutions. Enlisting an outside evaluator to facilitate the systematic evaluation process would increase objectivity. A single institution with a large number of residency programs could randomize programs into a case/control experimental design

that would allow for increased methodological rigor and afford opportunity for greater confidence in the study results.

Data from the current study could be evaluated using different methods. The introduction of additional researchers to concurrently analyze study data would provide opportunity for inter-rater reliability in the qualitative analyses.

Improving the Systematic Evaluation Process

Three years have passed since the adapted “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008) was first implemented in the KEI Ophthalmology Residency Program in 2011. Since then, the program has modified the annual program evaluation process in alignment with stakeholder requests and implemented nearly 100 recommended actions for program improvement.

The current study employed participatory evaluation practices with the hope of realizing some of House and Howe’s (2003, p.80) stated aims of the “deliberate democratic process” inclusion, collective decision-making, and stakeholder transformation. Each of these aims has been met on some level, a multiplicity of stakeholders were involved in the process, collective decisions were made based on input from stakeholders at multiple levels, and some stakeholders were transformed in that they learned more about their program, learned more about the evaluation process, and recognized its benefits and limitations.

Efforts to build evaluation capacity in the KEI Ophthalmology residency program through utilization of the adapted “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008) also seem to be realized. Results of the 2013 program administrator survey indicate that the impact of the evaluation goes beyond the confines

of the residency program, “The preliminary systematic evaluation process dialogue extends beyond the annual program evaluation to what are we doing overall at KEI (within the residency program and also as a department). It extends to not just thinking about the retreat event, I think it is actually positive because it extends to departmental vision.”

As performance improvement professionals posit, evaluation results are not the end of the story, but the beginning. Efforts to improve this systematic evaluation process are ongoing within the KEI Ophthalmology residency program. A “Program Evaluation Committee” has been formed according to ACGME 2014 requirements and this committee will be charged with evaluating the results of the current study to determine the best means to improve upon it and planning the 2014 annual program evaluation. Major considerations will include determining a reasonable scope for the evaluation process, increasing faculty engagement in the preliminary process, and utilizing the recommendations of the stakeholders for improving the evaluation process itself. It is essential that the KEI Ophthalmology residency program be more informed about the progress made as a result of their evaluation efforts and it is highly recommended that a structure be in place for regular reports on the progress of program improvement efforts.

Summary

The general purpose of the present evaluation research study was to examine the difference in outcomes when utilizing a systematic evaluation process, an adapted version of the “Impact Evaluation Process” (Guerra-López, 2007b, 2007c; Guerra-López, 2008), to analyze performance compared to less rigorous evaluation methods. The specific purpose was to examine the impact of a systematic evaluation process on evaluation

findings; determine if the process led to action-based performance improvement plans tied to specific performance gaps and improved outcomes; and identify the benefits of and the barriers to implementation of this process in a graduate medical education residency program. This chapter discussed the results of the case study and the following conclusions were formulated:

1. The evaluation results generated using a systematic evaluation process differed from previous years' annual program evaluation results. The results differed in multiple ways. More recommendations were generated using a systematic evaluation process and more types of recommended actions were proposed. The types of proposed actions were more robust than previous years' and more improvements were made to the program. Program outcomes improved over the course of the three years of implementation.
2. Utilization of a systematic evaluation process led to action-based performance improvement plans tied to specific gaps. The "Impact Evaluation Process" (Guerra-López, 2007b, 2007c; Guerra-López, 2008) adapted for use in the current study, ensured that the evaluation questions, performance indicators, and data sources were aligned with identified performance deficits. Engaging a professional evaluator ensured that the program adhered to the intended design. Although previous years evaluations yielded action-based performance improvement plans tied to specific gaps, these plans were more simplistic in content and lacked documented follow up.
3. Program outcomes improved in multiple performance domains during the three years that the current study was conducted. Significant improvements in resident

performance on the OKAP in-service exam were noted in 2013, post-implementation of multiple actions recommended in the evaluation process. In 2012 and 2013 all graduating residents met the surgical minimum volumes in all surgical categories as required by the ACGME, effectively addressing three ACGME program deficits that resulted in ACGME citations. Fifteen instances of increased scores on the ACGME Resident Survey were noted since the implementation of the systematic evaluation process.

4. The benefits of and barriers to implementation of the systematic evaluation process are represented in stakeholder perceptions and outcomes of the evaluation process itself. Stakeholders appreciated the improved evaluation process, engagement and commitment of multiple stakeholders, and improved communication between residents, faculty and administrators. Challenges perceived by the stakeholders included the burdens of implementing a rigorous evaluation process, time constraints (too much or too little), data limitations, and challenges to implementing the recommended changes. The results of using a systematic evaluation process included a greater volume of program improvements that were aligned to program deficits.

REFERENCES

AAMC^a Association of American Medical Colleges: Medical Schools. Accessed online on 07/14/13 at <https://www.aamc.org/about/medicalschoools/>

AAMC^b Association of American Medical Colleges. Medical School Enrollment on Pace to Reach 30 Percent Increase by 2017. Posted on May 2, 2013, Washington, D.C. Accessed online July 15, 2013 at:

<https://www.aamc.org/newsroom/newsreleases/335244/050213.html>

ACGME Report. Number of Accredited Programs for the Current Academic Year (2013-2014) United States: Report Date 09/14/2013 at 5:47:17 PM. Accessed online at: <https://www.acgme.org/ads/Public/Reports/ReportRun?ReportId=3&CurrentYear=2013&AcademicYearId=2013>

ACGME Data Resource Book: Academic Year 2011-2012. Accessed online on July 16, 2013 at:

http://www.acgme.org/acgmeweb/Portals/0/PFAssets/PublicationsBooks/2011-2012_ACGME_DATABOOK_DOCUMENT_Final.pdf

Accreditation Council for Graduate Medical Education Glossary of Terms. Effective July 1, 2013. Accessed online on 07/14/13 at:

http://acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/ab_ACGMEglossary.pdf

Accreditation Council for Graduate Medical Education Common Program Requirements Next Accreditation System. Accessed online 07/14/13 at http://www.acgme-nas.org/assets/pdf/CPR_Categorization_07012013.pdf

Accreditation Council for Graduate Medical Education Common Program Requirements

Effective 07/01/2011. Accessed online on 07/14/14 at: <http://www.acgme->

[2010standards.org/pdf/Common_Program_Requirements_07012011.pdf](http://www.acgme-2010standards.org/pdf/Common_Program_Requirements_07012011.pdf)

Accreditation Council for Graduate Medical Education Common Program Requirements

Effective 07/01/2013. Accessed online on 09/04/14 at:

<http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/CPRs2013.pdf>

ACGME Fact Sheet. Accessed online on 07/14/13 at

<http://acgme.org/acgmeweb/tabid/121/About/Misson,VisionandValues.aspx>

<http://acgme.org/acgmeweb/About/Newsroom/FactSheet.aspx>

ACGME Mission, Vision and Values. Accessed online 09/01/13 at American Medical

Association. Physician Consortium for Performance Improvement: About the

PCPI. Accessed online July 15, 2013 at: <http://www.ama->

[assn.org/ama/pub/physician-resources/physician-consortium-performance-improvement/about-pcpi.page?](http://www.ama-assn.org/ama/pub/physician-resources/physician-consortium-performance-improvement/about-pcpi.page?)

ACGME Program Requirements for Graduate Medical Education in Ophthalmology,

2013, p. 18. Accessed on 09/05/13 at:

http://www.acgme.org/acgmeweb/Portals/0/PFAssets/ProgramRequirements/240_ophthalmology_07012014.pdf

American Osteopathic Association. The Basic Documents for Postdoctoral Training.

Effective July 1, 2013. Accessed online 07/14/13 at

<http://www.osteopathic.org/inside-aoa/accreditation/postdoctoral-training->

- approval/postdoctoral-training-standards/Documents/aoa-basic-document-for-postdoctoral-training.pdf
- Andolsek, K. M., Nagler, A., & Weinerth, J. L. (2010). Use of an Institutional Template for Annual Program Evaluation and Improvement: Benefits for Program Participation and Performance. *Journal of graduate medical education*, 2(2), 160-164.
- AOA Basic Documents for Postdoctoral Training, Effective July 1, 2013. Accessed online 7/14/13 at: <http://www.osteopathic.org/inside-aoa/accreditation/postdoctoral-training-approval/postdoctoral-training-standards/Documents/aoa-basic-document-for-postdoctoral-training.pdf>
- AOA Summary of Positions Offered and Filled by Program Type: Results of the 2013 Match. Accessed online on September 4, 2013 at: <https://natmatch.com/aoairp/stats/2013prgstats.html>
- Bellini, L., Shea, J. A., & Asch, D. A. (1997). A new instrument for residency program evaluation. *Journal of general internal medicine*, 12(11), 707-710.
- Branson, R.K. (1975). *Interservice procedures for instructional systems development*: Tallahassee, FL: Center for Educational Technology, Florida State University.
- Brethower, D. (2000). Integrating theory, research and practice in human performance technology. *Performance Improvement*, 39(4), 33-43.
- Brown-McGowan, S. (1992). Effects of decision maker and context variables on evaluation utilization (Doctoral dissertation, University of Pittsburgh).
- Conn, C. A., & Gitonga, J. (2004). The status of training and performance research in the AECT journals. *TechTrends*, 48(2), 16-21.

- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational evaluation and policy analysis*, 14(4), 397-418.
- Cousins, J. B., & Earl, L. M. (1995). Participatory evaluation in education: What do we know? Where do we go. Participatory evaluation in education: *Studies in evaluation use and organizational learning*, 159-180.
- Cousins, J. B., & Leithwood, K. A. (1986). Current empirical research on evaluation utilization. *Review of educational research*, 56(3), 331-364.
- Critical Condition: The call to increase graduate medical education funding. AMA Center for Transforming Medical Education and AMA Advocacy Resource Center accessed online 07/14/13 at <https://www.ama-assn.org/ama1/pub/upload/mm/40/graduate-medical-education-funding.pdf>
- Cross, N. (2011). *Design thinking: Understanding how designers think and work*. New York: NY, Berg.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence. *JAMA: the Journal of the American Medical Association*, 296(9), 1094-1102.
- Dessinger, J.C. & Moseley, J.L. (2006). The full scoop on full-scope evaluation. In J. A. Pershing (ed.), *Handbook of Human Performance Technology* (3rd ed., pp. 312-376). San Francisco: CA. Pfeiffer.
- Dessinger, J. C., & Moseley, J. L. (2004). *Confirmative evaluation: Practical strategies for valuing continuous improvement*. San Francisco: Jossey-Bass/Pfeiffer.

Dick, W., & Carey, L. (1996). *The systematic design of instruction* (4th ed.) New York: Addison Wesley Longman.

Detroit Medical Center. *Organization History and Profile*. Accessed online on July 16, 2013 at: <http://www.dmc.org/organization-history-and-profile.html>

Detroit Medical Center. *Training Programs*. Accessed online on July 16, 2013 at: <http://www.dmc.org/training-programs>

Dunlap, C. A. (2008). Effective evaluation through appreciative inquiry. *Performance Improvement*, 47(2), 23-29.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3), 69-106.

Durning, S. J., Hemmer, P., & Pangaro, L. N. (2007). The structure of program evaluation: an approach for evaluating a course, clerkship, or components of a residency or fellowship training program. *Teaching and Learning in Medicine*, 19(3), 308-318.

Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: a reformulation and research agenda. *Academic Medicine*, 80(10), S46-S54.

Functions and Structure of a Medical School: Standards for the Accreditation of medical education programs leading to the M.D. Degree. Liaison Committee on Medical Education publication, June, 2013. Accessed online 07/14/13 at <http://centinel600.cardinalcommerce.com>

Gilbert, T. F. (1978). Human competence—Engineering worthy performance. *NSPI Journal*, 17(9), 19-27.

- Greene, J. C., (1987). Stakeholder participation in the evaluation design: Is it worth the effort? *Evaluation and Program Planning*, 10, pp. 379-394.
- Guerra, I. (2003) Key competencies required of performance improvement professionals. *Performance Improvement Quarterly*, 16(1), 55-72.
- Guerra-López, I. (2007a). *Evaluating Impact*. Amherst, MA: HRD Press, Inc.
- Guerra-López, I. (2007b). The impact evaluation process, part 1: Building a case for demonstrating the worth of performance improvement interventions. *Performance Improvement*, 46(7), 33-38.
- Guerra-López, I. (2007c). Planning a responsive evaluation: Establishing solid partnerships by clarifying expectations and purpose, part 2. *Performance Improvement*, 46(8), 32-36.
- Guerra-López, I. J. (2008). *Performance evaluation: proven approaches for improving program and organizational performance* (Vol. 21). San Francisco, CA: Jossey-Bass.
- Guerra-López, I., & Leigh, H. N. (2009). Are performance improvement professionals measurably improving performance? What PIJ and PIQ have to say about the current use of evaluation and measurement in the field of performance improvement. *Performance Improvement Quarterly*, 22(2), 97-110.
- Guerra-López, I. (2012). The prevalence of performance improvement as a central topic in the professional literature. *Performance Improvement Quarterly*, 25(1), 35-45.
- Howell, E. M., & Yemane, A. (2006). An Assessment of Evaluation Designs Case Studies of 12 Large Federal Evaluations. *American Journal of Evaluation*, 27(2), 219-236.

- House, E.R. (2005). Promising practices: The many forms of democratic evaluation. *The Evaluation Exchange* (The Harvard Family Research Project's evaluation periodical, Cambridge, MA) 11(3): 7.
- House, E.R. and Howe, K. R. (2005). Deliberate Democratic Evaluation. In T. Kellghan & D.L. Stufflebeam (eds), *International Handbook of Educational Evaluation*, volume 9 (pp. 79-100). Kluwer Academic Publishers: Norwell, MA.
- Huglin, L. M. (2009). HPT roots and branches: Analyzing over 45 years of the field's own citations. Part 1: Journal citations. *Performance Improvement Quarterly*, 21(4), 95-115.
- Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, 24(4), 602-611.
- Johnson, K., Greenesid, L.O., Toal, S.A., King, J.A., Lawrenz, F., Bolkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30, 377-410.
- Kaufman, R., Keller, J., & Watkins, R. (1996). What works and what doesn't: Evaluation beyond Kirkpatrick. *Performance + Instruction*, 35(2), 8-12.
- Kaufman, R. (1992). *Mapping Educational Success: Strategic Thinking and Planning for School Administrators*. Successful Schools: Guidebooks to Effective Educational Leadership. Volume 1.
- Kaufman, R., & Clark, R. (1999). Reestablishing performance improvement as a legitimate area of inquiry, activity, and contribution: Rules of the road. *Performance Improvement*, 38(9), 13-18.

- Kaufman, R., & Watkins, R. (2000). Getting serious about results and payoffs: We are what we say, do, and deliver. *Performance improvement*, 39(4), 23-32.
- Kaufman, R. (2005). Defining and delivering measurable value: A mega thinking and planning primer. *Performance Improvement Quarterly*, 18(3), 6-16.
- Kaufman, R. (1999). *Mega planning: Practical tools for organizational success*. Sage.
- Kaufman, R. (2006). *Change, choices, and consequences: A guide to mega thinking and planning* (Vol. 1). HRD Products.
- Kaufman, R., & Clark, R. (1999). Re-establishing performance improvement as a legitimate area of inquiry, activity, and contribution: Rules of the road. *Performance Improvement*, 38(9), 13-18.
- Kaufman, R. and Bernardez, M. L. (2012), Human performance technology and its future. *Performance Improvement Quarterly*, 25: 5–11.
- Kirkpatrick, D. (1997). Evaluation of training. In R.L Craig (Ed.), *Training and development handbook: A guide to human resource development* (3rd ed., pp. 301-319). New York: McGraw-Hill.
- Kirkpatrick, D. (1994). *Evaluating Training Programs; the 4 levels*. San Francisco: Berrett-Koehler.
- Klein, J. D. (2002). Empirical research on performance improvement. *Performance Improvement Quarterly*, 15(1), 99-110.
- Kresge Eye Institute “Residency Program” Retrieved September 3, 2013 from:
<http://www.kresgeeye.org/Residency-Program>
- “LCME I-A, Annual Financial Questionnaire, FY2012” accessed online 07/14/13 at
https://www.aamc.org/download/344944/data/fy2012_revenue_by_source.pdf

- Liebelt, E. L., Daniels, S. R., Farrell, M. K., & Myers, M. G. (1993). Evaluation of pediatric training by the alumni of a residency program. *Pediatrics*, *91*(2), 360-364.
- Leigh, D., Watkins, R., Platt, W. A., & Kaufman, R. (2000). Alternate models of needs assessment: Selecting the right one for your organization. *Human Resource Development Quarterly*, *11*(1), 87-93.
- Lincoln, Y. S., & Guba, E. G. (1985). Establishing trustworthiness. *Naturalistic Inquiry*, 289-331.
- Lubell, J. (2013). Obama budget signals quicker trigger on Medicare pay cuts. Posted April 22, 2013. Accessed online 07/14/13 at:
<http://www.amednews.com/article/20130422/government/130429972/1/>
- Mager, R. F., & Pipe, P. (1997). *Analyzing performance problems*. Atlanta, GA: The Center for Effective Performance.
- Medsker, K. (2006). Strategy Streamlines Evaluation. *Performance Improvement Quarterly*, *19*(1), 3-5.
- McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching people to fish? Building the evaluation capability of public sector organizations. *Evaluation*, *9* (1), 9-29.
- MacLellan-Wright, M. F., Patten, S., Cruz, A. D., & Flaherty, A. (2007). *A participatory approach to the development of an evaluation framework: process, pitfalls, and payoffs*. *Canadian Journal of Program Evaluation*, *22*(1), 99.
- McOwen, K. S., Bellini, L. M., Morrison, G., & Shea, J. A. (2009). The development and implementation of a health-system-wide evaluation system for education activities: Build it and they will come. *Academic Medicine*, *84*(10), 1352-1359.

- Miesen, M. (2013). Obama Budget 2014: The most significant health care cut you know nothing about. Accessed online on September 4, 2014 at:
<http://www.policymic.com/articles/38907/obama-budget-2014-the-most-significant-health-care-cut-you-know-nothing-about>
- Moseley, J. L. (1998). The Dessinger-Moseley evaluation model: A comprehensive approach to training evaluation. *Performance improvement interventions: Methods for organizational learning*. Washington, DC: International Society for Performance Improvement.
- Moseley, J. L., & Solomon, D. L. (1997). Confirmative evaluation: A new paradigm for continuous improvement. *Performance Improvement*, 36(5), 12-16.
- Milstein, B., Chapel, T. J., Wetterhall, S. F., & Cotton, D. A. (2002). Building capacity for program evaluation at the centers for disease control and prevention. *New Directions for Evaluation*, 93, 27-46.
- Musick, D. W. (2006). A conceptual model for program evaluation in graduate medical education. *Academic Medicine*, 81(8), 759-765.
- Overcast, S., Schmidt, T., Lei, K., Rodgers, C., & Chung, N. A. (2009). A case example of assessment and evaluation: Building capability in a corporate university. *Performance Improvement*, 48(6), 5-15.
- Oxford Dictionaries. "self-assessment". Oxford University Press, n.d. Web.
 <http://oxforddictionaries.com/us/definition/american_english/self--assessment>.
 Accessed on 09/06/13.

- Papineau, D., & Kiely, M. C. (1996). Participatory evaluation in a community organization: Fostering stakeholder empowerment and utilization. *Evaluation and program planning*, 19(1), 79-93.
- Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4(2), 225-233.
- Patton, M. Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *The American Journal of Evaluation*, 22(3), 329-336.
- Patton, M. Q. (2002). Utilization-focused evaluation. In *Evaluation models* (pp. 425-438). Springer Netherlands.
- Pershing, J.A. (2006). Human Performance Technology Fundamentals. In J. A. Pershing (ed.), *Handbook of Human Performance Technology* (3rd ed., p.15). San Francisco: CA. Pfeiffer.
- Peters, J., Baggett, S., Gonzales, P., DeCotis, P., & Bronfman, B. (2007). How organizations implement evaluation results. In Proceedings of the 2007 Energy Program Evaluation Conference (pp. 35-47).
- Phitayakorn, R., Levitan, N., & Shuck, J. M. (2007). Program report cards: evaluation across multiple residency programs at one institution. *Academic Medicine*, 82(6), 608-615.
- Plottu B., & Plottu, E. (2009). Approaches to participation in evaluation: Some conditions for implementation. *Evaluation*, 15 (3), 343-359.
- Plottu, B., & Plottu, E. (2011). Participatory Evaluation: The Virtues for Public Governance, the Constraints on Implementation. *Group Decision and Negotiation*, 20(6), 805-824.

- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.
- Rossett, A. (1987). *Training needs assessment*. Englewood Cliffs: NJ, Educational Technology
- Rummler, Brethower, & Geis (1974)
- Rummler, G. A., & Brache, P. A. (1990). *Improving performance: Managing the white space on the organization chart* (2nd ed.). San Francisco: Jossey-Bass.
- Rummler, G. A., & Brache, A.P. (1995). *Improving performance: How to manage the white spaces on the organizational chart*. San Francisco: Jossey-Bass.
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M. Scriven (Eds.), *Perspectives on curriculum evaluation*, New York: McGraw-Hill.
- Shulha, L. M., & Cousins, J. B. (1997). Evaluation use: theory, research, and practice since 1986. *American Journal of Evaluation*, 18(3), 195-208.
- Smith, M., Saunders, R., Stuckhardt, L., & McGinnis, J.M. (Eds.) (2012). *Best care at lower cost: The path to continuously learning health care in America*. Washington DC, National Academies Press.
- Sleezer, C. M., Zhang, J., Gradous, D. B., & Maile, C. (1999). A Step Beyond Univision Evaluations Using a Systems Model of Performance Improvement. *Performance Improvement Quarterly*, 12(3), 119-131.
- Stolovitch, H. D., & Keeps, E. J. (1999). What is human performance technology? In H. Stolovitch & E. Keeps (Eds.), *Handbook of human performance technology* (2nd ed. Pp.3-23) San Francisco: Jossey-Bass.

- Stolovitch, H. D. (2000). Human performance technology: Research and theory to practice. *Performance Improvement*, 39(4), 7-16.
- Sugrue, B., & Stolovitch, H. (2000). Report of 1999 ISPI symposium appropriate inquiry in human performance technology. *Performance Improvement*, 39(1), 33-36.
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), 237-246.
- UC Health, G. M. E. (2012). Training physicians: Medicare support provides vital funding to train our doctors and maintain patient access.
- Van Tiem, D.M., Moseley, J.L., & Dessinger, J.C. (2004).
- Van Tiem, D.M., Moseley, J.L., & Dessinger, J.C. (2000). Fundamentals of performance technology: A guide to improving people, process, and performance. Washington, DC: International Society for Performance Improvement.
- Vassar, M., Wheeler, D. L., Davison, M., & Franklin, J. (2010). Program evaluation in medical education: An overview of the utilization-focused approach. *Journal of educational evaluation for health professions*, 7.
- Ward, M., Gruppen, L. & Regehr, G. (2002). Measuring self-assessment: Current state of the art. *Advances in Health Sciences Education*, 7 63-80.
- Watkins, R. Leigh, D., Platt, W., & Kaufman, R. (1998) Needs assessment: A digest, review, and comparison of needs assessment literature. *Performance Improvement*, 37(7), 40-53.
- Watkins, R., & Guerra, I. (2003). Assessing or evaluating: Determining which approach is required. In M. Silberman & P. Phillips (Eds.), *The 2003 team and organizational development sourcebook*. New York: McGraw-Hill.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). CA: Sage.

ABSTRACT**A CASE STUDY OF THE IMPACT OF A SYSTEMATIC EVALUATION PROCESS IN A
GRADUATE MEDICAL EDUCATION RESIDENCY**

by

HEIDI T. KROMREI

May 2014

Advisor: Ingrid Guerra-López, PhD**Major:** Instructional Technology**Degree:** Doctor of Philosophy

The Accreditation Council for Graduate Medical Education has charged institutions that sponsor accredited Graduate Medical Education programs (residency and fellowship specialty programs) with overseeing implementation of mandatory annual program evaluation efforts to ensure compliance with regulatory requirements. Physicians receive scant, if any, training in program evaluation methodology. Human Performance Technology (HPT) offers models suitable for residency program evaluation as well as trained evaluators who are experts in evaluation. Leaders in the field of HPT have called for empirical studies to examine the impact of HPT models in a variety of contexts.

This single case study examined the impact of using a systematic evaluation process, the “Impact Evaluation Process,” (Guerra-López, 2007b, 2007c; Guerra-López, 2008), as a means for annual program evaluation in an ophthalmology residency program sponsored by large healthcare institution in the Midwest. Outcome data from 2011, 2012, and 2013 (the years in which the “Impact Evaluation Process,” was utilized), was

analyzed and compared to prior years evaluation efforts. Surveys with residency program stakeholders were conducted in 2011 and 2013. Results indicate that the number of recommendations for program improvement, types of recommendations and completed recommended actions increased in years that the systematic evaluation process was implemented. Recommendations generated using the systematic evaluation process were actionable (specific and measurable) and aligned to program deficits. Some program outcomes improved during the three years of systematic evaluation process implementation, while one performance outcome declined during this time. Stakeholder perceptions about the process indicated that anticipated and realized benefits of the process differed.

AUTOBIOGRAPHICAL STATEMENT

Heidi T. Kromrei has been the Academic Director for Graduate Medical Education at the Detroit Medical Center since 2010 and held the same position at Wayne State University School of Medicine from 2007 until 2010. Her prior work includes positions related to Health and Behavioral Research held at the University of Illinois at Chicago, University of Michigan, and Henry Ford Health Systems. She currently teaches program evaluation courses at Oakland University in Rochester, Michigan. Heidi has co-authored a book chapter on strategies for training survey research interviewers and published articles on educating residents and fellows in quality improvement methods and the utilization of systematic evaluation processes in preparing portfolios for learning and performance assessment. She has spoken widely on a variety of topics related to Graduate Medical Education at international, national, region and local professional meetings and conferences. She has served on national evaluation committees for the International Society for Performance Improvement, American Association of Medical Colleges Group of Resident Education and Generalists in Medical Education as well as regional and local committees pertaining to Graduate Medical Education.