THE EMERGENCE OF AN AUTONOMY-ORIENTED ASSESSMENT CULTURE

IN PEDIATRIC RESIDENCY EDUCATION: A CASE STUDY

by

Laura April McEwen

A dissertation submitted to the Faculty of Education
in conformity with the requirements for
the degree of Doctor of Philosophy

Queen's University
Kingston, Ontario, Canada
December, 2012

Copyright© Laura April McEwen, 2012
Abstract

This case study examines the emergence of an autonomy-oriented assessment culture in Pediatric residency education in the School of Medicine at Queen’s University. Through a case study approach this research explores how an assessment system to capture residents’ performance in the clinical environment was initiated, how that process supported a shift in assessment cultural, and how assessment innovation is eclipsing departmental boundaries.

The case has instrumental value in illuminating how an autonomy-oriented assessment system and culture can be cultivated in residency education. The analytic frame for the case was constructed based on research literature that identified weaknesses in assessment practices in residency education more generally. The approach was theoretical, with the intent to explain how a shift in assessment culture is emerging in Pediatrics. A longitudinal approach was adopted to expose shifts in the culture. The narrative structure distills the journey into a manageable story. Three watershed events that exemplified change were systematically selected from data collected over a three-year period and constitute the findings of this research.

The central contribution this research makes is that it is possible to shift the culture of assessment within a Pediatric residency program. That shift can be understood to unfold over a prolonged period through a process of mediating both social and regulatory requirements. Beginning to shift the assessment culture in Pediatrics was achieved by: recognizing the need for change in assessment practice, re-conceptualizing and realizing that change, and engaging and empowering the community to support a shift in assessment culture. Strong leadership, widening community engagement and the Rubric Descriptor Bank supported this process.

Five theoretically informed principles guided the emergence of the autonomy-oriented assessment culture in Pediatrics including: (a) conceptualizing learning as a social, active
process: (b) focusing attention on residents’ multidimensional growth; (c) moving away from a high-stakes orientation to assessment based on the false dichotomy between formative and summative assessment and embracing it as a tool for supporting and monitoring growth over time and across contexts; (d) actively supporting residents’ learning strategy and assessment skill development; and (e) fostering a growth orientation to learning, embracing the concept of graduated autonomy.
ACKNOWLEDGMENTS

I came to Queen’s University to pursue doctorial studies in assessment and evaluation under the guidance of leaders in the field. I am grateful for the opportunities and relationships that decision afforded both in the Faculty of Education and the School of Medicine.

This dissertation is the culmination of three years of work. It would not have been possible without the generosity of educational leaders, clinical teachers, and residents in the Department of Pediatric. I thank them for welcoming me into their world so we could explore it together. In particular, I wish to acknowledge my very good friend and respected colleague Dr. Robert Connelly who drew me into the residency education fold and kept me there.

I wish to thank my supervisor, Dr. Lyn Shulha for her unwavering support, inquisitive mind, and assessment expertise. Doctorial supervision is a challenging task of balancing guidance with freedom: I dare say we negotiated that balance well. I also wish to acknowledge Dr. Don Klinger’s significant contribution to the evolution of my thinking. His on-going critical appraisal and guidance were and continue to be invaluable. I thank Dr. Brian Hodges for agreeing to be my external examiner and turning his Foucauldian gaze to my work.

I thank Dr. Michelle Gibson who befriended me during my first graduate course at Queen’s in 2006, and continues to help me straddle the education/medicine divide. I also acknowledge Dr. Elaine Van Melle for opening the door to medical education and Drs. Leslie Flynn and Ross Walker for their foresight and willingness to support my research process.

I am grateful to Sheila Pinchin and Dr. Sue Fostaty Young for their educational perspectives couched in the context of medical education. Their insights were invaluable.

Finally, I thank my family for their confidence in my ability, patience, and sacrifice. I dedicate this work to my dear daughter Ehis, the person who has competed with my educational aspirations her entire life – I am done Sweet Pea!
CHAPTER THREE: CASE STUDY METHODOLOGY

INTRODUCTION ........................................................................................................................................ 47

CASE STUDY METHODOLOGY .................................................................................................................. 47

CASE STUDY DESIGN .................................................................................................................................. 48

PARTICIPANTS ........................................................................................................................................ 49

Research Partners .................................................................................................................................. 49

Research Participants .............................................................................................................................. 50

Critical Friends ...................................................................................................................................... 50

DATA COLLECTION ................................................................................................................................. 52

Document Archiving ............................................................................................................................... 52

Active Recording .................................................................................................................................. 53

Journaling .............................................................................................................................................. 53

Interviewing Residents ............................................................................................................................ 53

DATA ANALYSIS AND SELECTION .......................................................................................................... 54

PRESENTATION OF FINDINGS .................................................................................................................. 54

ETHICAL ISSUES .................................................................................................................................... 55

QUALITY .................................................................................................................................................. 56

CHAPTER FOUR: FINDINGS

INTRODUCTION ........................................................................................................................................ 63

CREATING AN ASSESSMENT SYSTEM

Watershed Event 1: Emergence of the First EPA Rubric ........................................................................ 64
RESEARCH BOUNDARIES................................................................................................................................. 119
OPPORTUNITIES FOR FURTHER RESEARCH.................................................................................................. 120
CONCLUSION.......................................................................................................................................................... 121
REFERENCES....................................................................................................................................................... 123

APPENDICES

APPENDIX A: LIST OF ACRONYMS.................................................................................................................... 135
APPENDIX B: PEDIATRIC DATA CATALOG........................................................................................................ 136
APPENDIX C: ETHICS CLEARANCE LETTER..................................................................................................... 143
APPENDIX D: RESIDENT CONSENT FORM....................................................................................................... 144
APPENDIX E: RESEARCH PARTNER CONSENT FORM.................................................................................... 146
APPENDIX F: RUBRIC DESCRIPTOR BANK (LIMITED VERSION)................................................................. 148
LIST OF FIGURES

Figure 1: Four phases of the Self-regulated Learning (SRL) cognitive model..............................21
Figure 2: The Physician’s Learning Cycle..................................................................................37
Figure 3: The Resident’s Learning Cycle.................................................................................41
Figure 4: Common Problems Rubric......................................................................................64
Figure 5: Pediatric Assessment Committee EPA Concept Map..............................................69
Figure 6: Neonatal/Perinatal Medicine PGY1 Rotation EPAs..................................................78
Figure 7: Management of Common Respiratory Presentations.............................................80
Figure 8: Visual Representation of Development Phases.........................................................84
Figure 9: July 2011 Assessment Table Handout....................................................................91
Figure 10: Encounter Card Rubric for Clinics and Minor Emerge...........................................94
Figure 11: Senior Consult Clinic Rubric..................................................................................95
Figure 12: July 2012 Assessment Table Handout....................................................................97
Figure 13: First Excel Workbook Format..............................................................................101
Figure 14: Second Iteration of the Excel Workbook..............................................................102
Figure 15: Current Rubric Descriptor Bank...........................................................................103
CHAPTER ONE: INTRODUCTION

Prologue

In October of 2009, the Vice-Dean of Education for the Faculty of Health Sciences at Queen’s University gathered a group of postgraduate educational leaders to discuss emergent technology needs related to assessment. As the Assessment Specialist for the School of Medicine, I sat across the table from Dr. Robert Connelly, a Neonatologist and chair of the newly formed Pediatric Assessment Committee (PAC). During the meeting he mentioned Entrustable Professional Activities (EPAs), something he had recently encountered in our journal club. EPAs were an idea proposed by a group of researchers in the Netherlands as a means of organizing competency-based medical education (Scheele, et al., 2008). Specifically, EPAs are defined as an agreed upon list of professional activities that a specific medical specialist can be asked to do (ten Cate & Scheele, 2007). Dr. Connelly’s reference to the concept was reassuring as it was an idea I planned to introduce in an upcoming assessment workshop for postgraduate faculty members.

I was pleased to see Dr. Connelly at the workshop a few days later. Although we had little time to engage one-on-one, the interest he expressed in the concept of EPAs made a positive contribution to group discussion. He approached me after the workshop suggesting we meet to discuss his department’s assessment needs. Up to that point my involvement at the postgraduate level had been almost exclusively limited to policy development. I welcomed the prospect of working more closely with faculty to address the day-to-day challenges of resident assessment.

When Dr. Connelly and I met on November 10, 2009, he described current assessment practices in Pediatrics as principally organized around the In-Training Evaluation Report (ITER)
system and shared his perceptions of the weaknesses of the system: namely, time constraints in the clinical context, inconsistent standards among teachers, and a lack of continuity in learner supervision. These were all issues that were emerging in the literature around professional learning in medicine (Dudek, Marks, J., & Curtis, 2008; Govaerts, van der Vleuten, Schuwirth, & Muijtjens, 2006; Watling, et al., 2010). He then invited me to attend his committee’s next meeting in early December and present some ideas about designing an assessment system to address the weaknesses he outlined. As an enthusiastic novice, I prepared a handout for the committee outlining what I considered to be seven key components of an assessment system: (a) Defined assessment objectives; (b) Data collection tools; (c) A data management system; (d) Data synthesis and review processes; (e) Progress and promotion protocols; (f) Remediation mechanisms; and (g) A structure to monitor overall system function. From a theoretical perspective (e.g., formative assessment and self-regulated learning theory) this system had all the elements of a comprehensive approach to assessment. From a practical perspective, however, it was daunting.

Although only Drs. Connelly and Acker were able to attend the December committee meeting, we discussed the current state of assessment in Pediatrics and reviewed my handout. They expressed a sense of urgency in addressing departmental assessment concerns, as the end of our accreditation cycle was less than two years away. Dr. Connelly prepared a summary of our meeting for his sub-committee, which he included with his report to the Pediatric Residency Program Committee (RPC) the following week.

In the subsequent 3 years, Dr. Connelly and I established a solid collaborative learning relationship characterized by mutual professional respect, trust and growth. At a fundamental level, this working relationship was my gateway into the world of residency education. It
afforded me the opportunity to see the clinical learning environment through the eyes of a clinician and fostered a deep understanding of the challenges faced in assessment within that context. Furthermore, my participation in the sub-committee was instrumental in establishing my credibility among clinicians, and invaluable in refining in me the crucial communication skills that facilitate the sharing of expertise. In effect, the events and developments that underpin the learning and findings captured in this dissertation work would not have been possible without the collaborative efforts of Dr. Connelly, the clinicians, and resident members of the committee. I wish to acknowledge that although I am the author of this work, all those referenced here, share ownership of the processes and outcomes.

**Purpose of the Research**

This case study documents the emergence of autonomy-oriented assessment in Pediatric residency education at Queen’s University. Although the idea of assessment systems in medical education is gaining ground from a theoretical perspective (Cooke, Irby, & O'Brien, 2010; Govaerts, Van der Vleuten, Schuwirth, & Muijtjens, 2007; Hauer, Holmboe, & Kogan, 2011; B. D. Hodges & Kuper, 2011; Holmboe, Sherbino, Long, Swing, & Frank, 2010; van der Vleuten, Schuwirth, Scheele, Driessen, & Hodges, 2010), there is little published research examining how this goal can be achieved in practice and less about how such systems are negotiated amongst stakeholders. This research addresses this gap by examining how one group of medical professionals organized themselves to address the multiple purposes of assessment and documents and analyzes the outcomes of that work. More specifically, this dissertation describes the evolution of a collaborative partnership between Pediatric faculty and residents, and an assessment specialist as we worked to develop an assessment system to capture residents’ performances in the clinical learning environment, shift the culture of assessment in the
department, and ultimately came to impact the wider postgraduate medical education (PGME) community at Queen’s. This case study documents the struggles, setbacks, and successes we encountered as we navigated this process and critically examines the artifacts generated so far. This work is significant because the effective assessment of learners in medicine is widely recognized as an important educational challenge that must be addressed (FMEC, 2011).

Furthermore, assessment researchers support the value of embedded and ongoing assessments as pedagogical devices to support and promote learning (Anderson, et al., 2010; Boud & Falchikov, 2006; Colliver, et al., 2010; Gibbs, 2006; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Price, Carroll, O'Donovan, & Rust, 2010; Regehr, Eva, Ginsburg, Halwani, & Sidhu, 2011; Shute, 2008).

The research question that guided this case study research was: How does a residency program cultivate an autonomy-oriented assessment culture? The following 3 enabling questions served to focus this case study report:

1. How do Pediatric educational leaders and an assessment specialist co-create an assessment system to capture residents’ performance in the clinical learning environment?

2. How does the culture of assessment within the Department of Pediatrics shift as a result of evolving understandings about assessment?

3. Can assessment innovation in Pediatrics eclipse Departmental boundaries?

This case study documents the emergence of an autonomy-oriented assessment culture, examining the context, theories, practices, decisions, and professional relationships that informed and enabled that emergence and the artifacts that are a legacy of the process so far. Re-conceptualizing assessment in postgraduate medical education and fostering a culture to support
this new conceptualization is a formidable task. It required us to confront and shift long established traditions and find innovative solutions to complex assessment challenges. This work explores the pressures, events, and processes that led the Pediatric Department to adopt a new orientation to assessment in residency education. In explicitly documenting our experience, decisions, and their consequences in the early stages of this change we hope to support those engaged in similar efforts and encourage those who have yet to embark.

A case study approach was adopted to document and recount the story of our lived experience. This approach is appropriate given that the insights that have emerged about the relationship between assessment and residency education are (and continue to be) best represented as the products of ongoing collaborative inquiry in response to questions and dilemmas arising out of purposeful local action. This dissertation makes a substantial contribution to the field of educational assessment because it offers a model to support thinking about how professional learning can unfold in the context of PGME and what the role of assessment can and should play in supporting that learning. The work is unique in that it bridges educational theory and professional learning in medicine in ways that are accessible to both educational scholars and physicians.

**Context: The Challenge of Assessment in Residency Education**

Assessment in PGME serves three primary stakeholder groups: (a) educational leaders, responsible for ensuring program quality, (b) clinical teachers, supporting and documenting learners’ progress, and (c) residents, engaged in professional learning. Each stakeholder group requires different but interrelated types of information from assessment activities. Satisfying the needs of all three groups defines the challenge of assessment in residency education. The
following sections explore the standards of program quality set out by accrediting bodies and the assessment needs of individual stakeholder groups in more detail.

**Accrediting Bodies**

Responsibility for setting standards and monitoring the quality of PGME is shared among three accrediting bodies in Canada: The Royal College of Physicians and Surgeons Canada (RCPSC), The College of Family Physicians of Canada (CFPC), and the Collège des médecins du Québec (CMQ). Canadian schools of medicine are required to meet nationally established standards to maintain program accreditation. The credentialing of graduates hinges on their ability to do so. Regular surveys of all residency training programs are conducted on a six-year cycle (RCPSC, 2006). Pre-survey questionnaires explaining how standards are met within the context of individual programs and the medical school as a whole are submitted to the responsible accrediting body in advance of a site visit. The Accreditation and Specialty committees (e.g., Pediatric, Cardiology, Critical Care, etc.) review these documents and identify program weaknesses prior to the survey team’s arrival at the school. During the site visit, a team of surveyors examines documented evidence of compliance provided by the school’s administration and individual specialty programs. Team members also tour facilities and interview staff, faculty, and learners. Following the visit they submit a report to the accrediting bodies outlining recommendations for accreditation status.

The assessment of learners is a particular focus during the accrediting process. This is reflected in Accreditation Standard B6: Evaluation of Resident Performance. According to this standard “There must be mechanisms in place to ensure the systematic collection and interpretation of evaluation data on each resident enrolled in the program” (RCPSC, 2011). The interpretation of this standard provided by the RCPSC states that the system “must be based on
the goals and objectives of the program”, “clearly identify the methods by which residents are to be evaluated and the level of performance expected” and “be compatible with the characteristics being assessed” (p. 20). It also stipulates a requirement to provide residents with “honest, helpful and timely feedback” that is both explicitly discussed and formally documented, and ensure residents are promptly informed when “serious concerns exist and given the opportunity to correct their performance” (p. 21). From an assessment perspective these are reasonable demands: However, the extent to which they are met in residency program across Canada has been questioned.

The In-Training Evaluation Report (ITER) system “serves as the assessment backbone for all Royal College programs.” (Bullock, et al., 2011). ITERs are completed for every resident at the end of a designated learning experience (e.g., a month long rotation). They are developed locally and list objectives grouped by each of the seven CanMEDS Roles (e.g., Medical Expert, Communicator, Collaborator). Accompanying each of these objectives is a Likert scale that allows clinical teachers to indicate the degree to which residents have met stated objectives. The number of objectives is determined departmentally (e.g., Pediatrics, Cardiology, Internal Medicine) and varies both across departments and universities. However, departments across all universities must conform to the Objectives of Training set out by their specialty.

In cases where an individual resident’s ratings on an objective either falls below the mid-point of the scale or is pegged at the very top of the scale, an explanatory narrative rationale is to be added to that resident’s ITER form. Many forms also require a global performance judgment to indicate whether or not expectations for the rotation have been met. The global performance judgment is the summative outcome of a resident’s performance for a given period of time that may be a month or more in duration.
In theory, the completion of ITERs is based on documented performance data collected over the course of an entire learning experience, although there is significant evidence now that this is rarely the case in practice (Chou, Lockyer, Cole, & McLaughlin, 2009; Dudek, et al., 2008; Norcini & Burch, 2007; Williams, Dunnington, & Klamen, 2005). Furthermore, while the intent of ITERs organized in CanMEDS format is to ensure all roles are assessed, there is evidence that both physicians and residents struggled to link compartmentalized objectives to their daily practice (Verma, Flynn, & Seguin, 2005; Zibrowski, et al., 2009). Overall, the ITER system has been described as “seriously flawed” because it “is rarely populated with reliable or objective data, allows faculty to focus on restricted performance domains, and is often completed long after the training experience has ended” (Bullock, et al., 2011).

The Assessment Needs of Educational Leaders

During my initial meeting with Dr. Connelly on November 10th 2009, we discussed his perceptions of the weaknesses of the ITER system. The weaknesses he identified then would later emerge in the research literature around practice. Specifically, he mentioned time constraints in the clinical context, inconsistent standards among teachers, and a lack of continuity in learner supervision (Dudek, et al., 2008; Govaerts, et al., 2007; Watling, et al., 2010).

In relation to time constraints, Dr. Connelly talked about the tension clinical teachers experience between the demands of patient care and teaching. While he acknowledged the challenge of balancing service and teaching, he strongly advocated for the need to allocate time to resident assessment processes arguing that it was an important professional responsibility (FIELDNOTE: 10/10/09). His position aligned with emergent thinking about the need to formally recognize clinicians’ investment of time in training (Collins, 2010; Hayes, 2011). In effect, there are tangible costs associated with clinical teaching. As our Associate Dean of
Undergraduate Medicine poignantly put it when talking about the normal pace of clinical practice, “having learners slows you down” (FIELDNOTE: Dr. S, FMEC town hall meeting, 02/09/2011). Researchers now advocate for adjustments to reward structures that would explicitly acknowledge the multiple responsibilities associated with teaching in the clinical environment and provide protected time for clinicians to fulfill them (Bindal, Wall, & Goodyear, 2011; Hayes, 2011; Setna, Jha, Boursicot, & Roberts, 2010).

When asked about standards that individual teachers bring to bear when completing ITERs, Dr. Connelly referred to the struggle between “hawks and doves” and the impact this had on assessment. He was describing a situation referred to in the medical education literature as stringency and leniency bias respectively, whereby “hawks” consistently apply higher standards for performance than “doves” (McManus, Thompsom, & Mollon, 2006). He talked about the challenge this presented when faced with reconciling these differing performance standards on ITERs.

In Dr. Connelly’s opinion, the challenge of inconsistent standards among clinical teachers was exacerbated by the sequential nature of relationships between teachers and learners in the clinical learning context that limit opportunity for clinicians to observe learners over time. When probed further, he described the dilemma in the Neonatal Intensive Care Unit (NICU) as an example of the lack of continuity in supervision. Here residents are assigned to a four-week rotation and their clinical teachers changed on a weekly basis. Standard protocol at the time required the clinical teacher on service during the fourth week of a rotation to complete a resident’s ITER.

This situation becomes particularly problematic given the lack of documented performance data. As a consequence, clinical teachers were known to rely on informal
discussions with colleagues about a resident’s performance to inform ITER completion (FIELDNOTE: 10/10/09). This practice is not unique to this program and has been reported as a common occurrence in PGME (Dudek, Marks, & Regehr, 2005).

Dr. Connelly recognized all of these weaknesses and wished to address them, but our conversation revealed additional weaknesses in the ITER system. Specifically, we agreed that the system of completing ITERs did not support learning or adequately value the role of feedback in the learning process, two elements we saw as central to the kind of assessment system we wanted to create.

Dr. Connelly characterized the process of completing ITERs as having been reduced to “a meaningless tick box exercise” (FIELDNOTE: 10/10/09). When asked why this was the case, he talked about the “laundry list of objectives” set out by the RCPSC and how these informed the content of ITERs currently in use (FIELDNOTE: 10/10/09). Aside from the sheer number of items on an ITER, he explained that organizing them in CanMEDs format made it difficult for clinical teachers to relate them to what they saw residents do on a daily basis (FIELDNOTE: 10/10/09). The CanMEDs framework defines seven “meta-competencies” or physician roles that reflect the essential abilities physicians need to ensure optimal patient outcomes (J.R. Frank, 2005). The seven thematic groups of competencies include: (a) Medical Expert, (b) Communicator, (c) Collaborator, (d) Manager, (e) Health Advocate, (f) Scholar, and (g) Professional (See The CanMEDs Physician Competency Framework for a full description of Roles available at http://rcpsc.medical.org/canmeds/). The intention of the framework was to explicitly acknowledge the multidimensionality of physician competence beyond medical expertise and emphasize the interconnectedness among these multiple domains. However, there
is evidence that compartmentalizing clinical performance in this way poses a challenge for clinical teachers (Verma, et al., 2005; Zibrowski, et al., 2009).

Beyond the challenge of linking compartmentalized CanMEDs objectives to clinical practice, Dr. Connelly conceptualized the ITER system as primarily serving an administrative function (FIELDNOTE: 10/10/09). At a fundamental level he questioned the educational value of ITERs, and especially their worth as a vehicle for feedback. Part of his skepticism stemmed from the delay between clinical performance and feedback available on ITERs. Dr. Connelly explained that the organization of learning in segmented blocks meant that residents often received ITER results at the very end of a block or after beginning a new one. In his opinion, this delay made it difficult for residents to act on feedback. He hastened to add that feedback was an on-going occurrence in the clinical context, even though it was not formally documented at that time, but reminded me that this feedback varied in quality depending upon the clinical teacher. Interestingly, while clinical teachers have reported they provide quality feedback using ITERs, Dr. Connelly’s thinking about the value and quality of the feedback available on ITERs aligned more with how residents generally perceived it (Sender-Liberman, Liberman, Steinert, McLeod, & Meterissian, 2005). There is also evidence that clinical teachers and residents differ in their opinions about the impact of feedback provided on ITERs. Where clinical teachers conceptualize ITERs as a mechanism for promoting professional development, residents report this to be contingent upon their perceptions of clinical teachers’ level of engagement in the assessment process (Watling, et al., 2010; Watling, et al., 2008).

The Assessment Needs of Clinical Teachers

Clinical teachers function in a demanding environment where they must balance patient care and teaching responsibilities. Depending on the context (e.g., ward, clinic, office), it is not
unusual for a single clinical teacher to simultaneously manage multiple learners of varying levels of experience. Faced with the responsibility of assessing and supporting a diverse range of learners, they are continually adjusting their expectations and standards for performance to align with the learners they are interacting with in any given moment. This complexity paired with the demands of patient care represents a significant cognitive load. In fact, researchers have only begun to theorize what the impact of managing this heavy cognitive load has on assessment processes and consider ways this burden might be better managed in the clinical learning environment (Tavares & Eva, 2012). In spite of the cognitive challenges, there is evidence that clinical teachers value assessment as a lever for learning, are motivated to do it well, but frustrated by the fragmented ITER system (Watling, et al., 2010). They report inadequate opportunity to directly observe learners, little access to learners’ performance history, and struggling to provide useful feedback while preserving learners’ self-efficacy (Watling, et al., 2010).

**The Assessment Needs of Pediatric Residents**

Residents enter PGME programs having completed an undergraduate (UG) medical degree. They are matched to a specialty program (e.g., Pediatrics, Surgery, Family Medicine) through the Canadian Resident Matching Service (CaRMS). On July 1st of each year a new cohort of residents enter into residency training programs across Canada. They will spend up to six years in the program learning their chosen medical specialties. During that time residents’ clinical performance will be a predominant focus of assessment, and feedback will play a significant role in shaping their learning.

The catalytic potential of feedback to enhance learning is widely recognized (Anderson, et al., 2010; Ericsson, 2004; Hattie & Timperley, 2007; Krackov, 2011; Norcini & Burch, 2007).
Although evidence for the direct impact of feedback on residents’ clinical performance is limited at present, residents’ self-report data about its educational value is positive (Miller & Archor, 2010). However, there is evidence that this potential depends, in large part, on learners’ perceptions of themselves (self-efficacy beliefs), the source of feedback (clinical teachers’ credibility), and the context in which it is provided (characteristics of the social learning environment). These factors are discussed with reference to the wider research literature in the following sections.

The self-efficacy beliefs that shape residents’ orientation to learning and impact their attribution tendencies appear to play a significant role in how feedback is received (Archor, 2010; Teunissen, et al., 2009). Specifically, learners with a strong sense of self-efficacy are generally confident in their ability to meet new challenges, more frequently focus on learning as opposed to performance goals, and are more apt to assume personal responsibility when setbacks occur as opposed to blaming external forces (Bandura, 1986). It is reasonable to expect that residents with a stronger sense of self-efficacy would be more receptive to feedback. Those with confidence in their capabilities as learners are more likely to interpret feedback as a mechanism to support their growth rather than as criticism of their personal limitations.

Clinical teachers’ credibility has also been shown to heavily impact residents’ perceptions of feedback (Bing-You, Paterson, & Levine, 1997; Gosman, et al., 2005; Mann, et al., 2011; Sargeant, et al., 2011; Teunissen, et al., 2009; Watling, et al., 2008). The credibility of clinical teachers is enhanced for residents through sustained interaction in the clinical context. Specifically, residents appear to value feedback more highly when it is based on directly observed clinical performance and provided by individuals whom they believe know them well enough to make sound judgments about their clinical competence.
Qualities of the social learning environment also effect residents’ perceptions of feedback with the impact either facilitative or inhibitory (Mann, et al., 2011; Sargeant, et al., 2011; Teunissen, et al., 2009). For example, incongruence between espoused and enacted practices may cause residents to question the applicability of the feedback they receive and undermine its perceived value. In contrast, clinical contexts where learning is modeled as a life-long process and conceptions of feedback as a catalyst for learning are promoted tend to foster more positive perceptions for residents.

In addition to residents’ perceptions of feedback, their willingness to purposefully pursue it in the clinical learning environment has been studied (Kennedy, G, Baker, & Lingard, 2009; Tamuz, Giardina, Thomas, Menon, & Singh, 2011). Residents’ feedback seeking behaviour seems to be shaped, in part, by perceptions of potential threats to their professional credibility. Specifically, they appear to engage in impression management techniques in an effort to preserve professional credibility with their clinical teachers and have been found to more readily seek feedback from allied health professionals and peers.

Beyond residents’ perceptions of feedback and wish to preserve professional credibility, how they respond to it has also been linked to characteristics of the feedback they receive. For example, when feedback is provided (e.g., immediate versus delayed), the intention of the feedback (e.g., supporting transfer of knowledge versus skills development), the nature of the message (e.g., suggestions about how to improve versus confirmation), and whether or not residents are supported in using feedback information (e.g., goal setting) have all been shown to effect residents’ responses (Archor, 2010). Ultimately, facilitative feedback that includes information about how to improve, is response-specific, goal directed, and delivered in a timely manner is believed to best support clinical performance improvement.
Overall, researchers argue for more active involvement of residents in the feedback process as a means of supporting uptake (Bindal, et al., 2011; Cooke, et al., 2010; Regehr G, Eva K, Ginsburg S, Halwani Y, & R., 2011; Setna, et al., 2010). In the context of factors identified to impact residents’ perceptions of feedback, this could have beneficial effects. Such an approach would conceivably support residents’ adopting a learning orientation and emphasize responsibility for and ownership of their learning process. It could also foster more meaningful interactions between residents and clinical teachers, ultimately enhancing the credibility of clinical teachers in the eyes of those residents. Furthermore, formally embedding resident-driven feedback processes in the clinical learning context make the culture of learning explicit. However, the development of facilitative processes (e.g., assessment strategies, performance review protocols) guided by the characteristics of optimal feedback would be required to realize all of this potential, something not readily available in the PGME context.

**Meeting the Challenge of Assessment in Residency Education**

This review of stakeholder needs underlines the complex nature of assessment in the PGME context. Balancing the needs of all stakeholders is more than simply the selection of the right assessment tools. It requires embedding facilitative processes to support residents’ learning within the educational context by, for example, building in structured opportunities for learners to review and discuss progress with their clinical teachers. Furthermore, it depends upon establishing an institutional culture that recognizes and values the iterative nature of professional growth. Ultimately, assessment in residency education is an educational design challenge composed of multiple elements that interact in complex ways (Van der Vleuten & Schuwirth, 2005). Thinking about assessment as a system of interrelated parts and processes embedded
within an institutional culture, with the potential to influence each other in diverse ways, acknowledges the complexity of the task.

In summary, an assessment system congruent with the requirements of accrediting bodies that also meets the needs of educational leaders, clinical teachers and residents would address several major goals. Ideally, such a system would better bridge the formal curriculum of The CanMEDs Physician Competency Framework with the clinical practice setting to establish the focus of assessment based on clinical practice activities, and identify and/or develop tools to assess these in situ. It would also promote shared standards of performance among clinical teachers and with residents and support the systematic collection of performance data to serve as a basis for meaningful feedback, allow for the capture of patterns of performance over time, and ultimately inform ITER completion. Finally, it would promote the active involvement of residents in the assessment process and support clinical teachers’ provision of meaningful feedback in effective and efficient ways.

Reading this Thesis

This dissertation is presented in five chapters. In the Introduction, I described the context of PGME in relation to the challenge of assessment. The theoretical framework that guides this scholarly effort is introduced in Chapter 2. In Chapter 3 I make explicit the case study methodology used to explore, document, and analyze our process. Chapter 4 presents the findings of this inquiry. Chapter 5 provides a discussion of the results, areas for future research, limitations of this study and a conclusion.

There are a variety of significant organizations, departments, documents, forms, and instruments presented in this thesis. While these will always be introduced with their formal names, they are typically integrated into the daily language of those involved in PGME as
acronyms. All acronyms used in the dissertation are presented in alphabetical order in Appendix A.

Chapter One Reflections

In this work I include reflections gleaned during the dissertation writing process. These often take the form of purposefully elicited feedback provided by my research partners. Their insights are offered to illustrate the authenticity of my narrative about our experience. Dr. Connelly’s email reaction to an earlier draft of Chapter 1 is one example of this when he writes:

Very strong work. I really enjoyed reading it, and thought it really captured well the initial journey and the complexity of the situation. I had absolutely no issue with the accuracy of any of your statements from my perspective. It was not only nostalgic in thinking back to how we got started, but was great to see how the work could really contribute to both education and medicine. You really pulled that together well (EMAIL: Dr. C to me, 16/01/2012).

Beyond authenticating the narrative of our experience, Dr. Connelly’s reaction illustrates how our collaborative learning process continues to unfold. The reference to his emergent understanding of how this work simultaneously contributes to the disciplines of education and medicine is an example of this. (DISSERTATION REFLECTIONS: 08/06/2012)
CHAPTER TWO: THEORETICAL FRAMEWORK

The Role of Theoretical Frameworks in Research

There are differing opinions about the role of theoretical frameworks in research. Some authors speak to the interconnections among researchers’ epistemological stances, theoretical perspectives, methodological designs, and selections of methods (Crotty, 1998). Others place a focused emphasis on the relationship between theory and methods, arguing that theory dictates method selection by defining the researcher’s theoretical lens (Creswell, 1998). Anfara and Mertz (2006) define theoretical frameworks as “any empirical or quasi-empirical theory of social and/or psychological processes, at a variety of levels, (e.g., grand, mid-range, and explanatory), that can be applied to the understanding of phenomena.” (p. xxvii). In their opinion, theoretical frameworks are multi-functional, serving not only to direct the focus and process of inquiry, but scaffold understandings about what is discovered (Anfara & Mertz, 2006). Bordage (2009) takes a similar stance defining them as “ways of thinking about a problem or a study, or ways of representing how complex things work the way they do” (p. 313). He writes metaphorically about their role as lighthouses and magnifying glasses, illuminating only partial views of reality and magnifying only certain elements of a problem. According to Bordage, theoretical frameworks are the ‘backbone or soul’ of research endeavor, ultimately shaping “what you choose to do and how you interpret your outcomes and results” (2009, p. 318). Although authors often associate theoretical frameworks with singular theory, others speak of it in terms of “theoretical harmony” (B.D. Hodges & Kuper, 2012). Hodges and Kuper suggest research approaches that encompass multiple theories serve to “illuminate the plurality of ways in which the construct can be understood, taught, assessed, and researched.” (p. 7). These authors discuss theoretical harmony as the application of multiple theories that illuminate different aspects of a
singular construct. I suggest an additional manner in which theoretical harmony is achieved, particularly in this educational research. Specifically, thinking about theoretical harmony as the threading together of theories about the constructs of learning, assessment, and feedback creates a broader, more contextually accurate theoretical framework to anchor, guide, and interpret the research. For me, linking those theories with the selection of an approach to inquiry pushed the idea of theoretical harmony even further. This is how I interpret the role of theoretical frameworks in my research.

My theoretical framework serves three principle functions: (a) it situates my research within the broader research landscape and makes explicit both the linkages between theories in use and the assumptions that underpin them, (b) it is a cognitive tool I use to scaffold both my own and others’ understandings about how things fit and work together, and (c) it informs my design and data collection choices, and shapes my analyses and interpretations of results. My theoretical framework is my research map. It establishes my location in current literature, situates my thinking, and shapes my research journey.

As my research focuses on the creation of an assessment system and the emergence of an assessment culture in residency education, I draw on theories of self-regulated learning, formative assessment, and feedback to construct the conceptual model that frames this work. I begin with a review of theories that contribute to the theoretical framework. I then present my conceptual model of learning and assessment at the postgraduate level as it has emerged from the interplay between theory and action in this research.

**Self-regulated Learning Theory**

From a theoretical perspective, self-regulated learning skills could be considered fundamental to lifelong learning (LLL) as the means by which it is enacted (Segers, Dochy, &
A successful career in medicine is, by its nature, a commitment to LLL that involves the on-going enhancement of knowledge and skills to ensure the highest quality of patient care (Aspin, 2000; Field, 2001; Koper, 2005). Considered from the perspective of supporting the development of learning skills that will allow future physicians to meet the learning demands of professional practice, the self-regulated learning model holds promise.

Self-regulated learning (SRL) theory is centrally concerned with explaining, “how students become masters of their own learning process” (Zimmerman, 2008, p. 181). It is based on the analysis of cognitive processing, such as the monitoring and control of cognitive strategies (Heikkila & Lonks, 2006). Butler and Winne (1995) characterized SRL as “a deliberate, judgmental, adaptive process” (p. 246) and offer the following description:

In academic contexts, self-regulation is a style of engaging with tasks in which students exercise a suite of powerful skills: setting goals for upgrading knowledge; deliberating about strategies to select those that balance progress towards goals against unwanted costs; and as steps are taken and the task evolves, monitoring the accumulating effects of their engagement. As these events unfold seriatim, obstacles may be encountered. It may become necessary for self-regulating learners to adjust or even abandon initial goals, to manage motivation, and to adapt and occasionally invent tactics for making progress (p. 245).

As Butler and Winne’s description suggests, SRL is conceptualized as a dynamic, proactive, recursive process engaged in by individuals to enhance learning and performance. Learners use SRL skills to independently assess on-going learning needs, and effectively monitor and ultimately control cognitive processes to meet emergent learning demands (Zimmerman, 2008).
According to the SRL cognitive model of learning, learners engage in four phases of self-regulatory processes during successful learning. These phases include: (a) goal-setting; (b) managing cognitive strategies; (c) regulating motivation; and (d) monitoring progress (Bandura, 1986; Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008). The phases of the SRL cognitive model of learning are presented in Figure 1. These phases are not believed to be hierarchical, leaving open the possibility for multiple phases operating simultaneously (Pintrich, 2004). The overlap among them in my visual representation reflects this relationship.

Furthermore, these regulatory processes are commonly conceptualized to occur across a range of areas or domains including: (a) cognitive, (b) motivational/affective, (c) behavioral, and (d) contextual, to reflect the influence of social context on learning (Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008).

Although SRL processes are internal to the learner, theorists widely acknowledge the potential for external factors to both influence and support how these phases unfold (Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008). For example, learners’ interpretations of task

![Figure 1: Four phases of the Self-regulated Learning (SRL) cognitive model.](image-url)
demands help to shape personal goal-setting processes. Similarly, established standards for performance can support learners’ monitoring of their learning progress during self-assessment activities. Therefore, explicitly considering how these internal cognitive processes and external factors interface during the learning process provides important insights that can be used to guide assessment system design.

The monitoring phase of the SRL model is the internal cognitive assessment system that continually oversees performance in relation to learning goals. For example, a resident might want to be able to manage a preterm infant. According to the SRL model, once the resident has established that learning goal, she then plans how she intends to meet it and selects the learning strategies she will employ (e.g., reading around a case, discussing the case with peers) (Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008). Over time, the resident monitors her progress towards that goal (e.g., Are my learning strategies effective? Am I more confident about my ability to care for this infant? Should I adjust my approach in some way?) (Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008). She also manages her effort by regulating motivation (e.g., This is a worthy investment of time if I wish to specialize in neonatology) (Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008).

Some of the information the resident uses to determine her progress in relation to that learning goal comes from outside the individual in the form of established standards for performance and external feedback about the gap between current performance and those standards (Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2008). In this way, the monitoring phase of the SRL model links closely to assessment activities as it is through these activities that external feedback is generated. This connection is evident in theories of assessment and feedback.
Evaluation, Assessment, and Feedback

Scriven (1967) is credited as the first to articulate the relationship between evaluation and feedback with his introduction of the concept of formative evaluation in relation to program evaluation. Although Scriven conceptualized evaluation as a single process, he made a distinction in the purposes that the evaluation process might serve. Always in relation to program evaluation, Scriven differentiated between the evaluative purpose of informing program design and development and that of judging merit, worth, and significance of program processes and/or outcomes (Scriven, 1976). In the first instance, he considered the evaluation purpose to be formative while in the second summative. According to Scriven, regardless of purpose, the evaluation process began with the assessor making a summary judgment based on a comparison between the current and desired states. The distinction for him, were the consequences that followed the initial act of judgment. In Scriven’s view, formative evaluation happened when summary judgments were accompanied with information about how to close the gap between the two states (1967). Specifically, summative evaluation addressed the function of judgment making, while formative evaluation included the provision of information to inform improvement. In Scriven’s conceptualization of these differing functions of evaluation, summative evaluation was always a necessary precursor to formative evaluation because the generation of feedback, the distinguishing feature of formative evaluation, was based on conceptions of summary judgments (Taras, 2005). However, in translating the concept of formative evaluation into the learning context, Sadler (1989) proposed the distinction between summative and formative assessment was more than the availability of feedback. He argued that judgments made for summative and formative assessment purposes were qualitatively different because they were based on different theoretical assumptions about learning.
**Formative Assessment Theory**

Writing in the late 1980s, Sadler proposed a theory of formative assessment for complex learning situations and advanced a reason why feedback did not always lead to performance enhancement (1989). As defined by Sadler, complex learning situations involved “students in actively synthesizing and integrating ideas, concepts, movements or skills to produce extended responses in some form” (p. 124). He defined formative assessment as “concerned with how judgments about the quality of student responses (performances, pieces, or work) can be used to shape and improve student’s competence by short-circuiting the randomness and inefficiency of trial-and-error learning.” (Sadler, 1989, p. 120). He contrasted this with summative assessment that focused on “summing up or summarizing the achievement status of a student” (Sadler, p. 119). According to Sadler, the goal of formative assessment (FA) in complex learning was to scaffold learners’ capacities “to monitor the quality of their own work during actual production.” (Sadler, p. 119). He conceptualized FA as an active process engaged in by teachers and learners that culminated in the development of what he called learners’ “evaluative expertise” that would enable them to self-monitor in situ (Sadler, p. 143). From Sadler’s perspective, the development of this expertise was fundamental in complex learning because without it, learners would be forever dependent on teachers’ feedback to guide improvement. Central to his thesis was the idea of graduated learner autonomy through the development of this expertise that would allow them to recognize and act on performance gaps during production.

Sadler stipulated three conditions that learners had to satisfy in order to make the transition from the dependency on teacher-generated feedback to effectively self-monitoring their improvement (1989). First, learners had to grasps the desired standard of performance. Second, they had to make multicriterion comparisons between their current level and the aimed
for standard in situ. Finally, they had to strategically draw on a suite of strategies that would allow them to close the gap between the current and sought after standard. Sadler argued that learners’ transition to a state of effective self-monitoring was dependent upon the quality of teacher-generated feedback available to them during the learning process. However, based on his analysis of instructional practices in educational contexts at the time, the quality of teacher-generated feedback was inadequate for this purpose.

Sadler (1989) noted that the nature of feedback normally made available to learners focused primarily on test results (e.g., correct/incorrect). He attributed this narrow focus to behaviourist theories of learning that underpinned research on learning and subsequently influenced teacher education. The implications of this behaviourist philosophical frame were that knowledge was conceptualized as a directly transferable commodity, learners were passive recipients of that knowledge, and the learning process was by consequence predominantly one of memorization (Driscol, 2005).

According to Sadler (1989), behaviourist ideas about knowledge and learning were evident in the way instruction was organized in that learning was assumed to follow a predictable, sequential path. The role of assessment was to measure learners’ acquisitions of material by testing how much they could accurately recall. A task, Sadler admitted was well suited to the use of quizzes and progress tests. The nature of assessment was such that learners could either demonstrate they had acquired material with correct responses or they could not. Results were most often reported as numerical scores (e.g., number of correct responses) and knowledge of those results was feedback because it signaled to learners what material they had missed. However, Sadler questioned the utility of this type of feedback beyond a class of narrowly defined learning objectives.
In Sadler’s opinion, limiting feedback to knowledge of results posed a particular problem where complex learning was concerned (1989). In contrast to the behaviourist lockstep model of learning, development in complex learning situations was conceptualized as multidimensional and continuous. These important differences in how the learning process was understood rendered the underlying assumptions of behaviourist theories of learning inadequate to guide thinking about instruction and assessment in complex learning. Evidence of which would later emerge in the cognitive psychology research (Bransford, Brown, & Cocking, 2000).

Sadler (1989) proposed that understandings about complex learning drew heavily on the interpretive tradition as opposed to the objectivist one underpinning behaviourist conceptions of learning. This tradition aligns with sociocognitive and social constructivist theories of learning (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991; Shepard, 2000; Winne & Hadwin, 2010). The underlying assumptions associated with these theories are that: (a) knowledge is not directly transferable but constructed based on prior knowledge and experience; (b) that learners are active participants continually synthesizing and integrating new information and experiences with prior understandings; and (c) learning is a growth process involving the negotiation of meanings with others and constructing individual understandings shaped by those social negotiations (Driscol, 2005).

Instructional design based on constructivist conceptions about learning are not strictly focused on the organization and delivery of information as in behaviourist models of instruction (Sadler, 1989). In these complex learning situations, the emphasis is on providing opportunities to discuss and work with ideas, compare and contrast understandings, and arrive at shared meanings (Bransford, et al., 2000; Brown, et al., 1989; Bruner, 1990; Lave & Wenger, 1991). The goal of assessment is to provide opportunities for learners to share their evolving
understandings. Optimally, assessment tasks for achieving this goal are open-ended in the form of essays, projects, and/or performances as opposed to highly structured tests. The role of assessment is to collect evidence about learners’ multidimensional growth.

Assessing multidimensional growth entails teachers making qualitative judgments about learners’ responses as opposed to scoring items on tests (Sadler, 1989). Sadler defined a qualitative judgment as “one made directly by a person, the person’s brain being both the source and the instrument for the appraisal” (p. 124). He suggested teachers used implicit conceptions of quality gained through experience that he termed “guild knowledge” to formulate these judgments (p. 126). He argued that providing learners with feedback that focused only on summative information about their overall performance (e.g., knowledge of results) meant that the process of constructing those qualitative judgments remained hidden from view. According to his theory, this inhibited the development of learners’ evaluative expertise because they lacked access to the guild knowledge or professional expertise upon which those judgments were based.

Furthermore, Sadler (1989) posited that this situation could account for the inconsistent research findings about the impact of feedback on learning. Ultimately, he was suggesting that instructional situations where the impact of feedback was negligible were likely attributable to insufficient feedback information leading up to the testing situation. Consequently, learners had not transitioned to a state of independent self-monitoring. As a result, learners lacked the ability to monitor their performance in situ, which was reflected in the testing situation as poor performance. Sadler’s hypothesis was ultimately supported by later research on learning (Black & Wiliams, 1998; Bransford, et al., 2000; Brown, et al., 1989; Gibbs, 2006; Price & O'Donovan, 2006; Yorke, 2003).
For Sadler (1989), formative assessment in complex learning was a social process of collaborative interaction between teachers and learners. It involved negotiating the meaning of standards for performance, and sharing how and when they applied. It focused on collaboratively identifying gaps between learners’ current performance and agreed upon standards and developing a suite of strategies to address those gaps. Feedback played a key role in the process as the mechanism through which teachers shared their guild knowledge about standards and strategies for improvement, ultimately scaffolding learners’ capacity to independently assess and shape performance.

Essentially, Sadler (1989) positioned summative and formative assessment as different philosophical orientations to assessment. He spoke of the primary distinction between the two as that of purpose and effect. The purpose of summative assessment being to summarize achievement for reporting and certification, and formative assessment serving a supportive function fostering learners’ self-monitoring capacity. Although he explicitly recognized the duality of purpose he proposed, he offered no guidance about how these different forms of assessment would function together in practice. In fact, his philosophical argument in support of formative assessment underlined the incongruence of the two, suggesting he saw them functioning as different assessment tracks. In so doing, Sadler set the stage for an unproductive dichotomy between formative and summative assessment. This was unfortunate because had Sadler followed his line of reasoning about the need for a different orientation to the assessment of complex learning based on constructivist theoretical assumptions about learning, he might have re-envisioned summative assessment processes as well. For example, if the role of assessment in complex learning situations is to collect evidence about learners’ multidimensional growth, as was the premise of Sadler’s argument for formative assessment, then the strongest
evidence of growth would be a learner’s pattern of performance across multiple domains over time. This would entail valuing all assessment for the contribution it made to constructing a learner’s overall performance profile as opposed to designating a separate assessment track (e.g., summative) for summarizing learner achievement status.

Framed in this way, summative assessment would have become more than a reductionist act of combining multiple marks to derive a cumulative score. It could have become a process of formulating holistic judgments based on the qualitative analysis of emergent performance patterns that took into account a wide range of assessment activities including both structured (e.g., quizzes and tests) and more open-ended tasks (e.g., essays, projects, and/or performances).

Combining different philosophically oriented assessment activities at this micro level would be acceptable because, as Sadler acknowledged, assessments based on objectivist understandings about learning can be useful indicators of knowledge based objectives. It could be sensibly argued that some form of foundational knowledge base would be required in any complex learning situation. Therefore, though these indicators are insufficient in and of themselves to guide judgments about the quality of complex learning, they can provide valuable partial information.

Imagine a situation where learners are engaged in patient care. It is inconceivable that even after years of training, any one of them would have encountered and been assessed on every possible medical condition. Ideally, they will have managed a large sample of cases under the watchful eyes of clinical teachers. Over time, astute teachers would observe learners’ growing abilities to translate knowledge into practice in situations when novel cases present. Documented evidence of learners’ capacities to do so provides rich assessment information about how those learners perform in situ when faced with novel situations. Eventually, it may be acceptable to
integrate performance information with examination results. In effect, it would be possible to extrapolate that based on past performance (e.g., evidence of effective knowledge translation into the practice setting), correct responses to knowledge based exam items are one indicator of how learners would react if presented with a similarly novel situation in a practice setting. It is important to underline that aside from providing evidence of foundational knowledge, the value of examination results would be enhanced by learners having demonstrated their ability to transfer knowledge into practice. Ultimately, it is the contextual lens of clinical performance patterns that extends the value to examination results. Integrating assessment information in this manner compensates for the limitations of the objectivist underpinnings of an examination based assessment approach (e.g., decontextualize knowledge), while simultaneously harnessing its strengths (e.g., content coverage). This approach to making summative judgments about complex learning emphasizes the qualitative nature of the act. Multiple sources of assessment information must be meaningfully integrated to construct an understanding about the quality of learning a learner has achieved.

Positioning summative assessment as a “macro level” qualitative judgment making process might have avoided the unfortunate dichotomy that Sadler’s argument for formative assessment provoked. Such an argument would have allowed him to remain true to the overarching constructivist notion of learning he so strongly advocated as the basis for formative assessment. Unfortunately, in preserving the objectivist philosophical underpinnings of summative assessment, Sadler missed the opportunity to better align the summative function of assessment with those constructivist understandings. This missed opportunity had far reaching implications for assessment practice and unanticipated consequences for learning.
Empirical Support

In the 1990s, educational systems worked to accommodate a broadening range of purpose that assessment was required to fulfill (e.g., Gibbs, 2006). Along with the summative purposes, efforts to incorporate formative strategies were on the rise. In addition, assessment was fast becoming a mechanism for public accountability, where learner performance on national examinations was used as an indicator of institutional performance (Shepard, 2000). As the function of assessment expanded, tensions escalated.

Discussion in educational circles contrasted summative assessment for credentialing and accountability purposes with formative assessment as levers to support learning (Black & Wiliams, 1998). Often these discussions characterized summative assessment as an important and necessary institutional function, but one that excluded and potentially interfered with the role of assessment as a learning tool (Gipps, 1994). It was argued that assessment practices based on a measurement model of assessment (e.g., summative) undermined instructional goals and limited students orientation to and capacity for learning because it diverted students’ attention away from learning and towards achievement (Shepard, 2000).

In the late 1990s, Black and Wiliam (1998) published an extensive review of empirical evidence that identified formative strategies that positively impacted students’ learning outcomes, for example: (a) using questioning to elicit understanding; (b) providing focused feedback about how to improve; (c) sharing criteria of quality with students; and (d) implementing self and peer assessment strategies. Ultimately, Black and Wiliam’s work provided empirical support for Sadler’s theory of formative assessment for complex learning. They argued for an increased focus on what would later come to be known as “Assessment for Learning” (AFL) which was contrasted with the summative function “Assessment of Learning”
(Assessment Reform Group, 1999). Like Sadler’s theory, AFL strategies formally acknowledged the iterative nature of learning and emphasized the active involvement of learners in the process of assessment. In essence, AFL strategies spoke to the development of learners’ self-monitoring skills and considered the role of assessment as a mechanism of support.

As powerful as the contribution of empirical support for the formative purpose of assessment was, it also fueled the false dichotomy between formative and summative assessment set in motion by Sadler’s (1989) blending of philosophically divergent orientations to assessment. As a result, summative assessment retained its objectivist underpinnings. At the same time, its measurement-focused orientation was easily demonized for its negative impact on learning (Gipps, 1994; Taras, 2005).

**Sustainable Assessment**

Building on the work of Sadler (1989), Black and Wiliams (1998), and others, Boud (2000) proposed sustainable assessment as a lens for thinking about the long-term implications of assessment practice. He defined sustainable assessment as “assessment that meets the needs of the present without compromising the ability of students to meet their own future learning needs.” (p. 151). He argued that assessment should be framed as “an indispensable accompaniment to lifelong learning” because learners’ abilities to self-assess were fundamental to productive learning processes (p. 151). From his perspective, the ultimate goal of assessment activities should be preparing learners to actively engage in lifelong assessment. He saw this not as an add-on to the crowded agenda that assessment had to fulfill, but as a lens through which the value of assessment activities could be viewed and should be judged.

So where Sadler’s theory of formative assessment was situated within the immediate formal learning setting with a view to fostering learners’ self-monitoring skills, Boud (2000)
considered the complexities that a transition to the workplace would contribute. He emphasized the fact that learning and work were inextricably intertwined. Once in the workplace, professionals had to recognize emergent learning needs, set learning goals, develop criteria and set standards, gather feedback and evidence about progress, and make judgments about quality independent of a formal assessment structure (Boud & Falchikov, 2006). Furthermore, the process would be continually re-worked as novel situations were encountered and contextual cues shifted. The hallmark of lifelong assessment would be learners’ abilities to flexibly adapt to new environments, seeking out and cultivating forms of feedback embedded there (Boud, 2000).

This pushed beyond Sadler’s thinking in terms of gaining access to guild knowledge that would serve as the basis of self-monitoring and spoke to learners’ participation in the co-construction of that knowledge. Boud (2000) understood lifelong assessment to be a longitudinal developmental process purposefully supported by assessment activity at the postsecondary level. It began with fundamental skills development, progressed to learners’ shared ownership of the assessment process and ultimately led to their full participation in assessment, as would be required in professional practice.

**Translating Theory into Residency Education**

While self-regulated learning (SRL) theory speaks to how cognitive processes function within the learner, assessment theory offers insight about how assessment activities interact to support and impede those processes. It is important to recognize that these theoretical understandings emerged from observations and conditions that are found in most formal educational contexts. This learning context is significantly different from the clinical learning setting of residency education, however parallels can be drawn. For example, the SRL cognitive model of learning can be translated into residency education by reframing formal learning tasks
as residents’ patient care responsibilities and other professional acts (e.g., leading a meeting, presenting at journal club, supporting more junior learners). Similarly, while the purposes for assessment described above remain appropriate and can be adopted, the assessment activities intended to serve these purposes must be adapted to accommodate the unique contextual conditions that are prevalent in a workplace learning setting. The work of Nicol and Macfarlane-Dick (2006) provides a useful starting point for considering how this adoption/adaption framework can be operationalized.

Drawing on theories of SRL and FA, Nicol and Macfarlane-Dick (2006) proposed seven principles of good feedback practice intended to foster learners’ self-regulation at the post-secondary level. According to their synthesis of the research literature, good feedback practice entails: (a) clarifying what good performance is for learners (e.g., standards, exemplars); (b) providing structured opportunities for self-assessment, (c) delivering information that helps learners troubleshoot their own performance and self-correct (e.g., providing comments as opposed to grades), (d) supporting dialogue between learners and teachers about learning (e.g., discussions about strategies for improvement), (e) encouraging positive motivational beliefs and self-esteem (e.g., fostering incremental as opposed to fixed views of ability), (f) ensuring learners have the chance to act on feedback to close performance gaps (e.g., resubmitting an assignment), and (g) using information about learners’ performance to adjust teaching (Nicol & Macfarlane-Dick, 2006).

In effect, the seven principles of good feedback practice make explicit Sadler’s theory of formative assessment for complex learning. Although developed for use in formal, post-secondary educational contexts these principles can be adopted because they provide a useful framework for considering the role of assessment in supporting SRL in residency education. To
be effective however, these principles must be adapted to adequately reflect behaviours and interactions that are recognized as appropriate and congruent with residents’ contexts for learning.

Of particular significance to assessment design in residency education is the concept of sustainable assessment because it situates assessment as the preparatory efforts for workplace learning demands (Boud, 2000). For example, Boud and Falchikov (2006) wrote about the need to orient assessment in formal learning environments in ways that would equip learners for the unforeseeable demands of workplace learning. However, in contrast to post-secondary environments where learning demands of the workplace are often unforeseen, the career path in residency is set and the kinds of learning physicians do are, to a certain extent knowable. Specifically, residency education unfolds in the clinical practice setting and is situated at the very threshold of independent professional practice. This provides a significant advantage over the post-secondary context where specific future learning demands are frequently indeterminate.

In effect, understandings about the learning demands of practicing physicians provide the frame for designing assessment activities uniquely tailored to scaffold skills residents require to meet those future demands. Furthermore, the situatedness of residency learning in the clinical practice setting affords access to and practice with the contextual cues residents will rely on to inform adjustments in practice beyond training. Therefore, the point of departure for designing an autonomy-oriented assessment system for residency education is an understanding of how lifelong learning is enacted by practicing physicians.

**Physicians’ Enactment of Lifelong Learning in Practice**

A conceptual model of the physician’s learning cycle is presented in Figure 2. This model outlines the relationship among professional responsibilities, contextual cues, and SRL processes
for practicing physicians. According to the model, professional responsibilities produce contextual cues (e.g., patient outcomes, external benchmarks, interactions with colleagues, teaching evaluations). Physicians actively seek out these cues that constitute external feedback information. This external feedback is internalized through self-assessment processes by aggregating information and interpreting it in relation to implicit standards of performance physicians hold for themselves. In Sadler’s terms, self-assessment is the act of physicians formulating qualitative judgments about their current level of performance. Self-assessments that indicate that implicit standards have not been met signal a need for further learning (e.g., unanticipated medical complications). This triggers physicians’ SRL processes, establishing the baseline for monitoring learning progress and stimulating the physician to take action through goal setting (e.g., determining reasons for complications and ways to avoid these in the future).

Once learning goals are established, cognitive strategies are implemented to address those goals. These might involve reviewing the case with colleagues and revisiting practice guidelines. Throughout this process the physician monitors progress towards meeting the goals of determining why the complications occurred and seeking ways to avoid these in the future. Should selected strategies fail to provide the insights sought, the physician may adopt additional ones like referring to the research literature to inform revisions to current practice guidelines. In this way the physician actively manages cognitive strategies to optimize learning effort. As this
Figure 2: The Physician’s Learning Cycle
process unfolds the physician also regulates motivation to sustain effort in the learning quest by for example, focusing on the implications for future patients. Optimally, this process results in new learning that informs adjustments in practice. At this point the iterative cycle begins again, as the physician monitors contextual cues to assess the impact of these adjustments.

The example above portrays the physician’s learning cycle as a linear process to emphasize how contextual cues interact with SRL through self-assessment processes. However, highly self-regulating physicians would engage in multiple learning cycles simultaneously. In effect, physicians continuously monitor performance cues, formulate self-assessments, seek new understandings, integrate these into practice and monitor on-going impact in the unending quest to optimize professional performance and positively impact patient outcomes. In essence, this entire recursive process represents the learning cycle of professional practice in medicine.

The physician’s learning cycle can be understood to unfold in four key stages. The first stage of the cycle involves formulating qualitative judgments about current levels of performance. During the act of self-assessment the physician draws on cues collected from the clinical practice setting, aggregating that information and comparing it with implicitly held standards of performance. In the second stage SRL processes are engaged by using qualitative judgments about current performance to identify learning needs and establish a baseline against which learning progress will be monitored. SRL is enacted in the third stage, optimally resulting in new learning. Ideally, this new learning is integrated into practice in the fourth stage of the cycle as adjustments in practice. Ultimately, a competency-based, autonomy-oriented assessment system for residency education would address critical points in the physician’s learning cycle where residents would be anticipated to require support.
The Role of Assessment in Supporting Residents’ Learning

A conceptual model of the resident’s learning cycle is presented in Figure 3. This model provides an overview of the relationship among residents’ performances, assessments, and SRL processes. Essentially, the model builds on the physicians’ learning cycle by integrating the relevance of residents’ learning contexts and future learning demands and incorporating the principles of good feedback with a view to supporting SRL. The first stage of a resident’s learning cycle involves the resident’s self-assessment process that entails formulating qualitative judgments about current levels of performance and the effectiveness of their learning process. Viewed through the lens of formative assessment theory this poses several challenges for residents. As novices, residents possess a limited grasp of the contextual cues that merit consideration and are in the process of establishing implicit standards to use as benchmarks for their qualitative judgments about performance (Sadler, 1989). Therefore, unlike experienced physicians who may be more capable of independently formulating judgments about their performance, residents lack the tools they need to do so effectively. In addition, judging the effectiveness of their learning process may be difficult for residents. Unlike their physician counterparts who draw on well-established repertoires of strategies paired with knowledge about how and when these are best applied, residents are in the process of constructing this understanding as well. Their repertoires may be quite limited and the potential to mismatch strategies with desired learning goals quite high. Ultimately, this limits residents’ abilities to assess the effectiveness of their learning effort so they can use that information to direct future effort. Sadler (1989) talked about these challenges for learners in relation to teachers sharing guild knowledge about standards of quality and strategies for improvement that would support the development of learners’ evaluative expertise. The sharing of clinical teachers’ guild
knowledge with learners is initiated in the resident’s learning cycle through assessment activities that generate feedback information. According to the principles of good feedback practice this would involve clinical teachers making explicit the standards against which residents’ performances are judged and offering advice about enhancing future performance (Nicol & Macfarlane-Dick, 2006). Ideally, this advice would unfold as assessment conversations between residents and their clinical teachers. During these conversations the meaning of feedback would be negotiated. Residents’ qualitative judgments about their performance would be elicited and explored through the process of comparing residents’ perceived strengths and areas for improvement with those of their teachers. This process would help residents identify contextual cues that provide important performance information, bring their standards for performance in line with their teachers, and gather advice about what and how to do things differently next time.

Framed in this way, assessment conversations support residents’ growing abilities to monitor their performance in situ and emphasizes an iterative orientation to learning that in turn encourages positive motivational beliefs and self-esteem (Nicol & Macfarlane-Dick, 2006). This stage of the learning cycle functions to calibrate residents’ emergent standards for performance. Residents’ better-calibrated standards subsequently enhance the quality of judgments they can make about their performance and improves the overall accuracy of self-assessments.

Residents’ more accurate judgments about performance are advantageous going into the second stage of the learning cycle when SRL processes are engaged. As in the physician’s learning cycle, residents use these judgments to inform goal setting and establish a baseline for monitoring learning progress. Beyond this, clinical teachers’ guidance about future learning can stimulate residents’ reflections about the effectiveness of strategies employed in the past. This has significant implications during the third stage of the learning cycle when SRL processes are
Figure 3: The Resident’s Learning Cycle
enacted because it helps residents make informed decisions about the suitability of cognitive strategies given their new learning goals, thereby enhancing the management of these strategies. As residents’ SRL processes unfold, they monitor progress, manage cognitive strategies, and regulate motivation. Optimally, this cognitive effort results in new learning that residents use to inform adjustments to performance during the fourth stage of the learning cycle. Effectively, this final stage constitutes residents acting on the feedback they were originally provided to close performance gaps (Nicol & Macfarlane-Dick, 2006).

Overall, the resident’s learning cycle highlights key points during the learning process when residents require support and the kinds of learning support that clinical teachers need to provide. An autonomy-oriented assessment system would provide a framework that aligns with residents’ learning needs and assists clinical teachers in providing that support in a manner that aligns with residents’ SRL competence. Clinical teachers’ grasp of residents’ SRL competence would be gleaned during assessment conversations when residents’ interpretation of feedback, emergent standards for performance, and cognitive strategies are actively elicited.

In effect, assessment conversations serve a diagnostic function for clinical teachers by helping them understand where the learners are in their progress towards SRL competence. They can then use that information to adjust the kind of support they provide. For example, it may become evident to the clinical teacher that a resident has misinterpreted feedback in some way. Left unchecked this situation could lead the resident to make ineffective adjustments in future performance. Beyond that, these discussions present the opportunity for clinical teachers to explore how misinterpretations occur and share their interpretations, providing an alternative lens that might help residents view similar feedback differently in the future. Alternatively, a resident may appear overconfident in his/her ability, signaling to the clinical teacher a problem with that
resident’s emergent standards for performance. Making such problems explicit and examining the source of miss-calibration decreases the likelihood of the resident dismissing feedback that is contrary to self-perception. Another example might be a clinical teacher’s observation that a resident has difficulty translating knowledge into practice. The clinical teacher could suggest engaging in peer discussion around active cases as a strategy to facilitate the development of this skill. These examples demonstrate how clinical teachers could use information about learners’ SRL competence to adjust their teaching (Nicol & Macfarlane-Dick, 2006). Central to this approach is the idea of graduated learner autonomy where learners are actively supported in developing toward SRL competence. However, the implications of an autonomy-oriented assessment system reach beyond this micro level of assessment. Fundamentally, constructivist notions of learning as multidimensional growth would need to be embedded across the entire assessment system.

**Autonomy-oriented Assessment**

In an autonomy-oriented assessment system all assessment activities would be conceptualized as snapshots out of which rich learner profiles are gradually constructed and form the basis for macro level qualitative judgments about overall competence. As conceptualized, the construction of learner profiles is not merely an additive process of summing scores across assessment activities but an in-depth analysis of emergent performance patterns. The concept of emergent performance patterns is critical given the complexity of professional practice that is the focus of resident learning. For example, consider what might constitute evidence of a resident’s competence in the CanMEDs domain of communication. Ideally, such evidence would include multiple instances of effective communication with a variety of individuals (e.g., patients, allied health professionals, peers, and teachers), across a range of situations (e.g., explaining
complicated procedures, delivering bad news, case reporting, and charting), and in various contexts (e.g., ward, clinic, office, auditorium). From a qualitative perspective, the more robust this collection of evidence becomes, the better able we are to triangulate that evidence and enhance the trustworthiness of the conclusion we draw. In other words, the more refined the learner’s profile becomes, the more confidence we can have in our qualitative judgment about that resident’s competence in the domain of communication. A significant advantage of this approach is that it better represents the multidimensional growth that residents experience during training and respects variation in the way that multidimensional growth can unfold both across residents and over time.

Thinking about assessment as a constructivist/qualitative exercise in gathering and analyzing evidence to inform macro level judgments about competence is a significant departure from the way it is currently organized. It requires moving away from the practice of using results of a select few assessment tasks (e.g., summative OSCEs) as the basis for high-stakes decision-making about complex learning. Framed in this way, assessment could no longer be usefully conceptualized as separate formative and summative tracks because all assessment would be valued for the extent to which it contributed to the emergence of complex performance patterns and ultimately the construction of learners’ performance profiles.

Furthermore, residents would need to be fully involved at the macro level of formulating judgments about their overall performance just as they would be at the micro level through assessment conversation. Their involvement at the macro level is critical because it is at this level that assessment information is meaningfully aggregated. Ideally, this would unfold in periodic performance reviews when residents take responsibility for presenting evidence of learning progress, identifying learning needs, and articulating a professional learning plan for
meeting those needs. Practice doing this is central to developing the skills they will use to aggregate contextual cues, identify learning needs, and establish performance baselines during the self-assessment process beyond residency.

If our goal is to foster the development of SRL competence as preparation for participation in the self-regulated profession of medicine, then residents must assume ever-increasing responsibility and ownership of the assessment process up to and including the point of autonomy. We can not continue to reserve such autonomy until after a physician’s transition to independent practice because in doing so we ignore our ethical responsibility to ensure that graduating residents are adequately prepared and capable of assuming that autonomy.

Using assessment in this way better honors constructivist notions of learning, focuses residents’ attention on their multidimensional growth and casts assessment as a tool for supporting and monitoring that growth over time. From the perspective of preparing residents for the learning demands of independent practice, this is advantageous because this approach intentionally supports the development of learning strategies and assessment skills they will carry forward with them into practice and fosters a growth orientation to learning that professional practice in medicine demands.

Chapter Two Reflections

I wish to make clear that I did not enter the field of practice with this elaborate grasp of residency education, nor comprehensive vision of how assessment should operate to support learning for residents. These understandings emerged over time as products of the collaborative learning process this research has enabled. These ideas are artifacts of this research in and of themselves. Even as I write this dissertation and engage in reflection about the overall process to this point, I see things differently and continue to learn. Every time I share my writing with
members of the wider medical education community at Queen’s University, I am enriched for
the insights they share and humbled by how this work resonates with them. It is reactions like
that of Dr. Griffiths pointing at the resident learning cycle model stacked on top of my notebook
as she rose to leave a meeting and saying, “I like that” (FIELDNOTE: 30/09/11) and Dr. Flynn
upon seeing it, requesting that she share it with clinical educators at the RCPSC in the future
(FIELDNOTE: 01/05/2012) that reassures me that my interpretations are meaningful in our
community. I am also reminded that this work is the initial step in our thinking, learning, and
realizing the potential of assessment in residency education (DISSERTATION REFLECTIONS:
26/05/2012).
CHAPTER THREE: CASE STUDY METHODOLOGY

“More will be pursued than was volunteered. Less will be reported than was learned.”
(Stake, 1995, p 441)

Introduction

In this chapter I describe the case study design and methods used in this research. First, I provide an overview of case study methodology. Next, I describe the case study design, research participants, data collection, analysis, and presentation of results. Finally, I discuss issue of ethics and quality in qualitative research.

Case Study Methodology

Case study methodology is a qualitative inquiry design frame used to address descriptive (e.g., what) and/or explanatory (e.g., how and/or why) research questions within a naturalistic setting (Stake, 2005; Yin, 2006). In general, case studies can be used to: (a) construct thick description of complex interactions within a site to promote understanding or explanation of social phenomena, (b) confirm theoretical assumptions, and (c) generate theory (Eisenhardt, 2002; Yin, 2006). Often associated with a constructivist philosophical orientation, case study research aims to produce firsthand understandings of people and events frequently in collaboration with research participants (Baxter & Jack, 2008; Stake, 1995). According to Yin (2006) “compared to other methods, the strength of the case study method is its ability to examine, in-depth, a ‘case’ within its ‘real-life’ context” (p. 111). Therefore, case study methodology is a mode of situated inquiry, favouring uniqueness over generalizability. Thomas (2011) defines case studies as,

analyses of persons, events, decisions, periods, projects, policies, institutions, or other systems that are studied holistically by one or more methods. The case that is the subject
of the inquiry will be an instance of a class of phenomena that provides an analytical frame—an object—within which the study is conducted and which the case illuminates and explicates (p 513).

According to this definition, it is through the bounding of the case that the analytical frame is established. However, it is through the lens of the case that greater understanding about the wider analytical frame can be achieved.

Case study methodology employs a variety of research methods including both qualitative and quantitative strategies. Various sources are used to triangulate evidence with the goal of constructing robust case descriptions (Yin, 2006). The specific methods used (e.g., participant observation, interviews, artifacts) and analysis strategies employed in case studies are diverse and ultimately depend on individual case circumstances. However, the use of narrative to weave understandings about a case from diverse data sources is characteristic of this approach (Flyvbjerg, 2006).

One final aspect of case methodology that is important to highlight is its ability to contribute to theory development. This contribution can occur in two notable ways. First, through the iterative process which characterizes case study methodology, both internal (within case) and external (between case and larger discourse/theory). Second, as a population of cases within a given domain begins to accumulate, cross-case analysis can draw out generalizable trends to inform practices beyond local case contexts (Stake, 2005).

Case Study Design

Thomas (2011) proposes a typology of case studies that delineates what he calls five classificatory layers including: (a) Subject or case; (b) Object or analytic frame; (c) Purpose; (d) Approach; and (e) Process. According to Thomas (2011) his “typology encourages a clear
articulation of the distinctness and necessity of both subject and object; it encourages consideration of theoretical or illustrative approaches, methodological decisions, and decisions about process (p. 518).

The subject of this case study is changes in assessment thinking and practice in the Pediatric residency program at Queen’s University and whether those changes can eclipse departmental boundaries. The case has instrumental value in illuminating how an autonomy-oriented assessment system and culture can be cultivated in residency education. The analytic frame for the case was constructed based on a review of the research literature identifying weaknesses in assessment practices in residency education more generally. The approach was theoretical with the expressed intention of explaining how a shift in assessment culture might be fostered and achieved in Pediatrics. In order to expose shifts in the culture of assessment in Pediatrics, a diachronic or longitudinal process was adopted. Three watershed events that exemplified change were systematically selected from the mass of data collected over a three-year study and constitute the findings of this research.

Participants

In the context of prolonged engagement in case study research there is often a widening circle of participation over the life of the project (Baxter & Jack, 2008; Stake, 2005). In the following sections I describe three groups of people who contributed to this case study research including: (a) My research partners, (b) Our participants, and (c) My critical friends.

Research Partners

Case study research has been described as a collaborative learning process (Baxter & Jack, 2008; Stake, 2005). In this sense case study research unfolds in partnership “with” practitioners as co-creators of knowledge. My research “partners” in this project were the
members of the Pediatric sub-committee responsible for student assessment. Ultimately, this group defined the focus of this work through the frustrations they shared, questions they raised, and insights they offered. Our relationship evolved dramatically over the course of this study. At the beginning my interaction was almost exclusively with the chair of the committee, Dr. Connelly, as we developed funding proposals and the ethics application. Once funding was secured and the project was underway this expanded to include all members. Although committee membership changed over the life of the study, it always included a mix of three or four physicians along with one or two residents. Attendance at meetings fluctuated depending upon members’ availabilities and the nature of participation of individuals ranged from less active (e.g., contribution to group discussion) to full involvement (e.g., data collection, presenting at grand round/conferences).

Research Participants

I use the term “participants” to distinguish members of the Pediatric residency education community whose feedback was sought during development and piloting phases. This group included clinical teachers outside the membership of PAC and three first year residents, none of whom were members of our committee. Clinical teacher participants provided feedback on draft rubrics. First year residents participated in semi-structured individual interviews after completing two rotations in the NICU during which pilot rubrics were used.

Critical Friends

I call the third group of people involved in this research my critical friends. Critical friends, generally help researchers: (a) Deal with feelings of isolation triggered when moving between the collaborative case study context and solitary research activities (e.g., reflection and
writing); (b) Guard against bias that can arise from a researcher’s elaborate tacit knowledge of a case; and (c) Manage information overload due to the abundance of data generated in case study research (Foulger, 2010). I drew on an eclectic group of individuals to support me in managing these challenges and conceptualize them as divided into three major subgroups. I visualize these subgroups as a series of concentric circles with my Supervisor, Dr. Shulha and I located at the center. At the next level I include the two members of my dissertation committee, Drs. Klinger and Van Melle and my professional colleague Dr. Fostaty Young. At the perimeter I include members of Queen’s medical education community like Dr. Flynn (former Associate Dean PGME and current Vice Dean Education), and Drs. Griffiths and Schultz (Queen’s Department of Family Medicine).

**Primary.** As my primary critical friend and Supervisor, Dr. Shulha was instrumental in helping me carve out the portion of this case study that became the focus of my dissertation writing. We met frequently over several months, for periods of up to three hours to engage in what Foulger (2010) calls “External Conversation” (p. 144). Through these intense conversations Dr. Shulha pushed me to identify the watershed events reported in Chapter 4 and consider narrative as a medium for sharing the contextual richness of this case study.

**Secondary.** My secondary critical friends, Drs. Klinger and Van Melle approached my work from two very different orientations. Although our interactions were limited in comparison to the intensity of my relationship with Dr. Shulha, they were of significant impact. Through pointed questions during my proposal defense, Dr. Klinger pushed me to explore my role in the research and make explicit the beliefs and values that guided my interaction with others. Similarly, Dr. Van Melle challenged me to acknowledge how ownership of this research is
shared with my research partners. These insights served as launching pads for critical reflexivity about core values of this research process.

Dr. Fostaty Young, another of my secondary critical friends played a somewhat different role. She brought a familiarity with the postsecondary assessment literature along with an understanding of the learning context and culture of medical education. These understandings allowed her to critique my argument for a shift in our orientation to assessment in residency education. We met weekly in our professional capacities as assessment and evaluation specialists for the School of Medicine. During those meetings we frequently discussed topics related to my dissertation work and it was valuable and stimulating to hear her alternative perspectives.

Tertiary. Finally, tertiary level critical friends served as sounding boards for the ideas that emerged over the course of my writing and ensured those ideas were communicated in a manner that was accessible to an audience of physicians.

Data Collection

I employed a range of methods to gather qualitative data for this case study, including: (a) Document archiving (e.g., emails, minutes, and artifacts); (b) Active recording (e.g., field notes); (c) Journaling; and (d) Interviewing residents.

Document Archiving

A variety of strategies were used to archive documents generated in this case study. Minutes of meetings and artifacts (e.g., draft EPA rubrics) were stored in an online community space and organized in folders according to topic. The online systems tracks versions by date of upload and author. Only those with administrator privileges can reorganize or remove items from the online system. Dr. Connelly, myself, and our research assistant were the only individuals
who had administrator privileges for the online space. I also maintained an email folder with all
project related email communications.

**Active Recording**

I carry a notebook with me at all times to record meetings, observations, and short
reflections that occur during or between meetings. All entries are dated and a short description of
the nature of the interaction is recorded at the beginning of the entry (e.g., Faculty development
committee meeting). Notebooks are stored in chronological order in my home library.

**Journaling**

I also maintain an electronic journal on my laptop. Here I document reflections about on-
going projects and literature I am reviewing. These are generally more extensive reflections than
those recorded in my notebook. All entries are dated.

**Interviewing Residents**

Semi-structured individual thirty-minute interviews were conducted with 3 residents.
Questions focused on their experience of assessment since entering the Pediatric residency
education program and their experience with EPA Rubrics during rotations through the NICU. I
conducted all interviews at the convenience of the residents. All interviews were electronically
recorded and I noted significant insights in my notebook. Unfortunately, one audio-file was
damaged due to technical malfunction; however, I did have my own notes of the interaction for
reference. I transcribed the remaining two interviews verbatim. Interview transcripts and my
notes were analyzed through an iterative process of comparing and contrasting residents’
descriptions of their experience both in the program and specifically in the NICU.
Data Analysis and Selection

The wide assortment of data gathering methods employed over the course of this case study captured a rich history of group process and product. However, one of the challenges of implementing multiple data gathering strategies is the abundance of data that is generated. The researcher is then faced with the arduous task of deciding what to include when reporting.

This selection process can easily become overwhelming if not approached systematically (Foulger, 2010). As a first step in this process I constructed a project timeline in tabular format. I reviewed and cataloged the entire collection of data and artifacts accumulated over the two and a half year period from December 1, 2009 to July 7, 2012. I listed dates in the first column, events in the second, outcomes in the third, and data sources in the fourth (See Appendix B: Pediatric Data Catalog). It was a grueling but invaluable task that allowed me to engage in what Levin (2012) describes as a ‘scientific reflection process’ (p. 143). According to Levin, the “researcher must be able to take a step back from the involvement and critically analyze what he or she has participated in” to attain this depth of reflexivity (p. 143). The cataloging process served this function for me by providing the distance and opportunity for me to visualize the case holistically. Seeing it from this vantage point allowed me to distinguish watershed events that occurred over the life of the project more clearly and explore their relationship to the whole (Levin, 2012).

Presentation of Findings

Narrative inquiry (NI) is a method researchers use to convey the richness of case study research. According to Connelly and Clandinin (1990), “Narrative inquiry in the social sciences is a form of empirical narrative in which empirical data is central to the work” (p. 5). A variety of empirical data can be used for this purpose including for example, field notes, journal records,
and email messages. Ultimately, narrative inquirers build a “sense of the whole” out of these “concrete particularities of life that create powerful narrative tellings” (Connelly & Clandinin, 1990). In effect, NI involves the weaving together of theoretical understandings with stories of practical experience in a form of dialectic expressed in “narratives of experience” (Connelly & Clandinin, 1990). It is a process often undertaken in collaboration with those involved in the narrative. Standard practice in this form of inquiry is the sharing of drafts with those involved as a collaborative meaning making process and checks of the authenticity of emergent explanations.

Essentially, the purpose of NI is not to determine causal relationships but convey “explanations gleaned from the overall narrative” (Connelly & Clandinin, 1990). Since case study is a collaborative learning process, these explanations must be negotiated collectively among partners to gain what Kemmis and McTaggart call “a collaborative sense of legitimacy” (2008, p. 297). It is with this explicit purpose in mind that I use narratives of experience in the presentation of this case study research.

**Ethical Issues**

This research was reviewed and received ethics clearance from the Health Sciences Research Ethics Board at Queen’s University in March of 2010, and was renewed in March 2011 and March 2012 (See Appendix C: Ethics Approval Letter).

Participant ethics in case study research can be complicated. Where appropriate, standardized ethical protocols were followed, for example, seeking informed consent from residents who participated in interviews and released assessment data (See Appendix D: Resident Consent Form). Consent from others involved in this research was more complex. As an example, my research partners and critical friends. My approach to ensuring informed consent from research partners has been the sharing of drafts throughout the writing process. In this way
I honour the collaborative nature of this research and ensure the authenticity of my narrative of our experience from their perspectives (See Appendix E: Research Partner Consent Form). In instances where specific individuals have been referenced, this is done with permission from those individuals.

**Quality**

Numerous frameworks for judging research quality appear in the literature. These range from the conventional criteria grounded in the positivist/postpositivist tradition (e.g., internal validity, objectivity, replicability, and generalizability) to alternative, transgressive, postmodernist conceptualizations such as crystalline validity (Guba & Lincoln, 2005). The advent of alternative frameworks is based on the argument that conventional markers of quality are inadequate for supporting judgments about the quality of research grounded in traditions outside the positivist/postpositivist purview. Although there is general agreement about the need to make judgments about the quality of research, there is little consensus about criteria to guide those judgments (Bradbury-Huang, 2010; Freeman, Marrais, Preissle, Roulston, & Pierre, 2007; Greenwood & Levin, 2007; Guba & Lincoln, 2005; Herr & Anderson, 2005; Moss, et al., 2009; Reason & Bradbury, 2008). Scholars frequently argue that the relevance of criteria depends on its alignment with ontological and epistemological orientations of researchers (Creswell, 2009; Guba & Lincoln, 2005). However, multiple frameworks have been proposed for use even within similar research traditions, resulting in a mixture of terms to describe strikingly similar aspects of research (See for example, Guba and Lincoln’s concept of “educative authenticity” and Schwandt’s (1996) concept of “critical intelligence.”)

Given the current state of the field, the prospect of proposing a framework for judging the quality of one’s doctoral research is a daunting task. Having reviewed various sets of criteria for
judging research quality, explored arguments about what should guide selection (e.g., philosophical orientation, approach to inquiry), and considered the criteriology debate that argues against normative standards as a mechanism of control (see for example Schwandt, 1996), I settled with Tracy’s (2010) eight “Big-Tent” criteria. Tracy (2010) takes a practical approach to the conundrum of criteria for quality in qualitative research arguing they “help us learn, practice and perfect” (p. 838). She writes about the usefulness of rules and guidelines in supporting the development of expertise based on Dreyfus, Dreyfus, & Athanasiou’s novice/expert learning research and how these rules and guidelines ultimately “serve as helpful pedagogical launching pads” (1986, p. 838).

In presenting her eight criteria of quality in qualitative research, Tracy (2010) makes explicit her intentions to:

- provide a parsimonious pedagogical tool, promote respect from power keepers who often misunderstand and misevaluate qualitative work, develop a platform from which qualitative scholars can join together in unified voice when desired, and encourage dialogue and learning amongst qualitative methodologists from various paradigms (p. 389).

Ultimately, Tracy argues that if we avoid conflating means (e.g., methods: practices, skills, and crafts) with ends (e.g., common end goals of strong research) it is possible to honour the diversity of orientations to research and still agree on criteria for quality. Towards this end she proposes eight “Big-Tent” criteria for judging excellence in qualitative research including: (a) worthy topic; (b) rich rigor; (c) sincerity; (d) credibility; (e) resonance; (f) significant contribution; (g) ethics; and (h) meaningful coherence. In the sections that follow I introduce the
criteria, explain how each can be interpreted within the context of this research, and ultimately share how they were met.

In terms of the criterion of “worthy topic,” Tracy (2010) asserts that good quality research is relevant, timely, significant, interesting, or evocative. This research grew out of a grassroots initiative led by Pediatric educational leaders dissatisfied with the way they were assessing their residents, unsure about how to improve the situation, and eager for change. It is relevant in terms of the daily work of clinical teachers and residents. It is also timely as the recent Future of Medical Education in Canada – Postgraduate project report (2011) indicates that assessment in residency education is an important area in need of development. Finally, it is significant and interesting because it challenges us to re-conceptualize the role of assessment in residency education and professional learning more generally.

With regard to the criterion of “rich rigor” Tracy (2010) writes about requisite validity or the richness that grows out of the complexity of a researcher’s theoretical understandings brought to bear in the research process and an abundance of data that serves as the basis for rich descriptions and explanations. She also refers to face validity gained through a researcher’s prolonged engagement in the field and transparency about data gathering, selection, and analysis procedures.

Although I arrived in the field with theoretical understandings about learning and assessment as they apply in formal postsecondary learning settings, my prolonged engagement in the field helped me adapt and extend those understandings to residency education. Furthermore, in line with the emergent nature of case study research, I was conscious of the need to gather as much data as possible in multiple forms (e.g., field notes, minutes, emails, artifacts, electronic recordings of interviews and meetings). This abundance of data is reflected in the “slice of life
data” I weave into my case study narrative. I also shared my process for organizing and selecting the data for inclusion in this report, aware that these decisions have significant implications for the findings reported here. Together all of these experiences have enhanced my methodological competency and fostered my professional growth.

Tracy’s (2010) criterion of “sincerity” can be understood in part as the researcher’s efforts to make explicit the biases, goals, and foibles she brings to the research endeavor. Fundamentally, it is the “honesty and authenticity with one’s self, one’s research, and one’s audience” achieved through the process of self-reflexivity (p. 842). But it is also the degree of transparency about entry into the field of study, the challenges and unexpected turns experienced during the research process. I attempt to meet this criterion through my journaling over the course of this project and during the writing process with the sharing of insights, reflections and recollections placed at the end of chapters. This practice emerged as I wrote. It was a mechanism that helped me deal with the linearity of text as a medium. It allowed me to insert the insights that arose as a product of writing, sometimes days or months after having completed a section.

“Credibility” is the criterion Tracy (2010) uses to describe how readers’ confidence and trust are enhanced through the provision of thick description, the use of triangulation, and the sharing of member reflections. It is through the windows of the narratives of our experience that readers come to understand how my research partners and I engaged in our collaborative learning process. Similarly, I use the strategy of triangulation by interrelating theory with examples from our experience to promote deeper understandings and afford readers the opportunity to judge the credibility of the interpretations I propose. Finally, I share my research partners’ reactions to my narrative of our experience, not as verification but to illustrate how we continue to learn together even in reflecting about our process in retrospect.
Tracy (2010) uses the term “resonance” to describe a researcher’s ability to move an audience both with the way the writing is done and what is written about. She writes about the strategic use of emotion to draw the reader into the story, but she also describes the potential of vicarious experience to move readers to action. I struggled to incorporate these in my writing. Coming from a background in psychology, narrative is an unfamiliar, somewhat vulnerable genre for me. I frequently caught myself slipping back into the distanced third person perspective at some points during the writing process. It was out of that struggle that the hybrid style incorporating both formal academic and narrative genres emerged.

In terms of the criterion of “significant contribution”, Tracy (2010) writes about the questions a researcher might ask about her research for example, “Does the study extend knowledge?” “Improve practice?” “Generate on-going research?” Here she refers to the theoretical, heuristic, practical, and methodological significance of the work. I perceive the contribution of this work to be both theoretical and practical. From a theoretical perspective autonomy-oriented assessment in residency education extends our thinking about the role of assessment in the professional learning process. From a practical perspective shifting the assessment culture in residency education is a formidable task. Certainly our residents receive more and higher quality feedback than they did when we began this project, and take greater ownership of the assessment process, but we have yet to fully operationalize autonomy-oriented assessment. In terms of generating on-going research, this work raises important questions about how the shift in assessment practice towards patterns of performance and learner profiles will impact learning in residency and beyond.

Tracy (2010) defines the criterion of “ethics” in terms of procedural, situational, relational and exiting. The procedural level refers to issues related to institutional Ethical Review
Boards including such things as consent, confidentiality, and accuracy in reporting, among others. Situational ethics refers to the researcher’s responsibility to continually weigh potential costs (e.g., harm) against the benefits of research. Relational ethics underline the researcher’s responsibility to foster mutual respect, dignity, and connectedness with others involved in the research. Exiting ethics speaks to how the researcher leaves the field of study and how they represent that field once outside.

The criterion of ethics was satisfied in this research through the review and approval of the Health Sciences Ethics Board at Queen’s University and by securing consent from participants and research partners. But it is also reflected in the relationships I maintain with my research partners.

Tracy’s (2010) final criterion of “meaningful coherence” speaks to the manner in which the researcher’s process embodies the philosophical orientation s/he espouses, the manner of presentation promotes the underlying intentions of the work, and the skillfulness with which the research focus, reviewed literature, findings and conclusions are meaningfully interwoven. As case study research, this work is grounded in the constructivist philosophical orientation. It honours the emergent quality of our collaborative learning process thought explicit attention to the way our relationship evolved over time and extends to the manner in which member reflections have been meaningfully incorporated into the development of this report.

Furthermore, the underlying intention of this research was to provide insight into our process of re-conceptualizing assessment in residency education. However, the presentation of this research must also meet the requirements for a doctoral dissertation. Achieving the right balance between these competing demands is challenging. In this respect, Tracy (2010) writes about the need for authors to ensure that “their representation style matches the goals of the
project” (p. 848). The overall structure of this work was a purposeful effort to bridge the disciplinary boundaries between education and medicine. Ultimately, the aim of this work is to provoke cross-disciplinary conversations that will promote mutual understandings with the goal of expanding conceptions of the role of assessment in professional learning contexts more generally.
CHAPTER FOUR: FINDINGS

Introduction

Chapter 4 presents the findings of this case study that address the research question: How does a residency program cultivate an autonomy-oriented assessment culture? Results are organized around three watershed events that address the following three enabling question: (a) How do Pediatric educational leaders and an assessment specialist co-create an assessment system to capture residents’ performance in the clinical learning environment? (b) How is the culture of assessment within the Pediatric Department shifted as a result of evolving understandings about assessment? and (c) Can assessment innovation in Pediatrics eclipse departmental boundaries?

For the purpose of this case study I define a watershed event as a contextually bound moment when there was shared recognition that we had experienced an event in time in a chain of events when an important change occurred. I use three such events to anchor my examination of the process of change in assessment thinking and practice in Pediatric residency education and how that change is impacting the wider PGME community at Queen’s University. These watershed events emerged out of the analysis of data collected over approximately 3 years beginning in December 2009.

I identify the first watershed event as the emergence of EPA Rubrics. Insights derived from interviews with residents about their experiences using these in the NICU constitute the second watershed event. Refining the Rubric Descriptor Bank is the third watershed event. In the following sections I present each watershed event, describe how each emerged and explain their significance.
## Watershed Event 1: Emergence of the First EPA Rubric

On May 3rd 2010 Dr. Zielinski shared the Common Problems rubric presented in Figure 4 as an attachment to an email sent to Dr. Connelly. Dr. Zielinski had been mapping EPAs for the subspecialty of Respirology. In the body of his email to Dr. Connelly, he identified three

<table>
<thead>
<tr>
<th>Overall Approach</th>
<th>Incomplete</th>
<th>Partially Complete</th>
<th>Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>History and Physical Examination generally adequate for most patients</td>
<td>History is not well structured or overly time consuming</td>
<td>Aware and shows has approach to identify common and likely diagnosis</td>
<td></td>
</tr>
<tr>
<td>Inappropriate basic problem formulation</td>
<td>Able to manage common diagnosis/</td>
<td>Aware of less common diagnosis</td>
<td></td>
</tr>
<tr>
<td>Unable to initiate appropriate basic management in common causes of presentation</td>
<td>Not clearly shown to consider more complete Ddx in approach</td>
<td>Appropriate management plans that would be expected of a consultant pediatrician</td>
<td></td>
</tr>
<tr>
<td>History missing expected at a 1st year resident level*</td>
<td>History is overly broad and time consuming in its approach</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History is overly focused on one/few possible diagnosis</td>
<td>History is overly focused on one/few possible diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing aspects of physical examination for problem formulation</td>
<td>Does not look for “red flags”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Problem Formulation</td>
<td>Missing basic details expected at a 1st year level*</td>
<td>Appropriately concise and relevant</td>
<td></td>
</tr>
<tr>
<td>(This section is meant to evaluate the formation of the problem list based on the individual patient – ie the “wheezing” child in front of you – not the general approach to the wheezing child)</td>
<td>Does not look for “red flags”</td>
<td>Pertinent positive and negative points/aspects acquired</td>
<td></td>
</tr>
<tr>
<td>Does not consider relevant diagnosis for patient presentation</td>
<td>Misinterprets physical findings</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment/Investigation Formulation</td>
<td>Missing aspects of physical examination for problem formulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inappropriate plans (consistently wrong treatments, investigations, follow up)</td>
<td>Does not look for “red flags”</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Able to identify proper/likely diagnosis</td>
<td>Able to identify other possible diagnosis based on patient presentation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unable to identify other possible diagnosis based on patient presentation</td>
<td>Able to identify proper/likely diagnosis as well as appropriate differential diagnosis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Able to use information gathered to appropriately rank possible diagnosis based on likelihood</td>
<td>Able to use information gathered to appropriately rank possible diagnosis based on likelihood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate use of investigations</td>
<td>Appropriate use of treatments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriate follow up plans</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 4: Common Problems Rubric*
categories of EPAs and defined the content of each category. His list included: (a) Common Problems: Chronic/Recurrent Wheezing/Noisy Breathing Child, Chronic/Recurrent Coughing Child, Dyspnea, Snoring/OSA Child and Recurrent Pneumonias Child; (b) Common Diseases: Cystic Fibrosis, Asthma; and (c) Skills: Chest x-ray interpretation, Pulmonary Function interpretation. Dr. Zielinski explained that he had yet to consider communication skills and practice management issues but had begun to think about how the Common Problems category might be assessed. He sought Dr. Connelly’s opinion about the utility of his approach when he wrote, “See if you think this track of thinking is usable.” (EMAIL: Dr. Z to Dr. C, 03/05/2010).

In the preamble to the rubric Dr. Zielinski explained he envisioned using the “Overall Approach” section of the rubric as the basis for assessment. From his perspective the balance of the rubric would serve a feedback function for residents specifically, “to help identify deficiencies/strengths and to help advise where to improve.” (EMAIL ATTACHMENT: Dr. Z to Dr. C, 03/05/2010). He clarified that he anticipated assessments being based on case discussion rather than directly observed performance when he wrote, “the evaluation of the history will be based on how the resident communicates/organizes the details obtained from their history – not the actual communications skills with the families – that can be evaluated in a different format.” (EMAIL ATTACHMENT: Dr. Z to Dr. C, 03/05/2010). Further, Dr. Zielinski defined the levels of performance he used in his rubric as:

“Incomplete” – history/PE/management approach shows clear holes in knowledge and resident would not likely be able to manage the common diagnosis of the common problems appropriately. (As a staff you will repeat everything/most things after review).
“Partially Complete” – approach shows a better structure and knowledge base and can manage the basic problems (ie asthma for wheezing – or migraine for headache), but is at risk for missing less common diagnosis (ie immunodeficiency for wheezing – or brain tumor for headache). (As a staff you are relatively happy – but need to clarify/seek further information regarding history/PE etc frequently – looking for missing information which you suspect might be lurking – probably a level expected at a beginning of a subspecialty rotation for a 1st/2nd year resident)

“Complete” – The resident’s overall approach is relevant and concise (As a staff you may still need to clarify details – however looking for details at the level of a subspecialist or to help in extra learning or to pinpoint the final more complicated diagnosis – however the resident is usually on the right track). (EMAIL ATTACHMENT: Dr. Z to Dr. C, 03/05/2010).

Based on Dr. Zielinski defined levels of performance he was thinking about what resident performance would look like in practice and anticipating where he expected a resident to be upon entry into a subspecialty rotation (e.g., “probably a level expected at a beginning of a subspecialty rotation for a 1st/2nd year resident’). Furthermore, he provided a frame of reference for clinical teachers by describing the action they might take when faced with residents at various levels (e.g., “As a staff you will repeat everything/most things after review”).

Finally, Dr. Zielinski qualified his use of the term ‘appropriate’ in the rubric explaining, “I use the word “appropriate” in the assessment a lot – probably cause I’m not sure what that means in each case. There may be a different level of appropriate for history/PE/problem list as there is for management.” (EMAIL ATTACHMENT: Dr. Z to Dr. C, 03/05/2010).
How EPA Rubrics Emerged

Dr. Zielinski’s act of sharing his rubric for common problems in Respirology was a watershed event because it sparked a major shift in PAC members’ activities. To this point the group had focused on organizing the Pediatric residency curriculum into themes, identifying EPAs for each theme, and mapping those to CanMEDs Roles and enabling competencies defined in the RCPSC Objectives of Training for Pediatrics. This activity was based on a three-step curriculum design process outlined by Scheele et al. (2008). According to Scheele et al the three-step curriculum design process entailed: (a) Structuring the content of a profession using themes; (b) Selecting a limited number of EPAs for each theme; and (c) Matching EPAs to CanMEDs Roles using a variety of assessment tools.

Initial structuring of the Pediatrics curriculum occurred at the first EPA planning meeting held on April 1st 2010 in a conference room on the 10th floor of Kingston General Hospital. Six members of the committee attended including four faculty members and two residents. During the meeting the group discussed the concept of EPAs and exchanged ideas about how these could be defined in Pediatric medicine. At the time the only example of residency education organized around EPAs had been described for the specialty of Gynaecology (Scheele et al., 2008). The example in the literature was procedurally based and posed a challenge for committee members as they struggled to translate the concept into Pediatric medicine.

After some discussion, Dr. Connelly had the group attempt to identify some EPAs. He handed out pads of sticky notes in various colours and I secured an eight-foot strip of paper on one wall of the conference room. Dr. Connelly invited committee members to write EPAs on sticky notes and post them on the wall. The use of sticky notes allowed members the flexibility to shift individual EPAs as need be.
The room was a flurry of activity as members posted their EPA ideas. Once finished, the group sat down at the conference table facing the wall of EPAs. The concept map of EPAs is reproduced electronically in Figure 5. Although small adjustments to the positioning of clusters have been made in recreating the electronic version of the original map (e.g., ‘WARD’ moved from far left to lower left quadrant). Figure 5 is an exact replica of how committee members labeled and positioned individual EPAs, including for example, the links some members drew between major and minor concepts.

According to the EPA concept map, committee members conceptualized the content of Pediatric medicine as two distinct domains of practice (e.g., Inpatient and Outpatient). Headings for clusters and subsequent groupings of activities under each differed significantly. Some clusters focused exclusively on medical problems (e.g., Ill (ACUTE) Newborn), while others for example, “Practice Management” were an eclectic collection of activities ranging from “patient flow” to “consult requests.” The collaborative mapping exercise made individual committee members’ thinking about EPAs explicit and stimulated further group discussion about what could constitute an EPA in Pediatric medicine. Discussion about the applicability of EPAs to Pediatrics continued by email over the following week.

One theme that carried over into post-meeting email exchange was the difference committee members perceived between a surgically focused discipline (e.g., Gynaecology) and Pediatric medicine. In their opinion, surgical specialties had clearly definable, directly observable surgical procedures as the central focus of assessment. They contrasted this with the nature of professional activity in Pediatrics as they conceptualized it. Specifically, Dr. Zielinski noted that the ability to “manage discharges/prescriptions /complex follow up/ etc...and
Figure 5: Pediatric Assessment Committee EPA Concept Map
especially communicate are skills to pediatricians in some ways as important as being able to perform a caesarean section is to an obstetrician” (EMAIL: Dr. Z. to PAC, 01/04/10). He also underlined how the focus of assessment shifted depending upon the practice context (e.g., Inpatient/Outpatient). In this respect he contrasted the ward environment where it might be possible to directly observe residents’ examinations of children with the outpatient context of clinics where residents saw patients independently and subsequently reviewed those cases with their clinical teachers. According to Dr. Zielinski, clinical teachers in this context were “observing problem solving and knowledge base and using these opportunities to teach [management]” (EMAIL: Dr. Z. to PAC, 01/04/10).

At a fundamental level Dr. Zielinski was struggling to translate the procedurally based example offered in the literature to his complex understanding of the daily practice of Pediatrics. It was evident that he was searching for ways to logically group professional activities and weighing the pros and cons of different ways of doing that. He toyed with the idea of dividing between acute and chronic management as opposed to problem specific groupings. In doing so he referred to the example of asthma and seizures when he wrote, “you can maybe argue that there is more overlap in managing a status asthmaticus and status epilepticus than the status asthmaticus and chronic asthma” (EMAIL: Dr. Z to PAC, 01/04/2010).

In my response to Dr. Zielinski’s email I reminded committee members that we could define EPAs for Pediatrics as whatever we wanted, provided we had a rationale for doing so. I also tried to bring them back to the EPA literature by encouraging them to “focus on identifying ‘core activities’ that merit targeted assessment and can function as ‘proxies of competence’ within our context(s)” (EMAIL: me to PAC group, 02/04/2010). Furthermore, I cautioned against adopting a reductionist approach fearing they would “begin separating out elements of a
single ‘patient encounter’ too early” and potentially lose “the complexity of Pediatric resident performance we wish[ed] to capture” (EMAIL: me to PAC, 02/04/2010). Finally, I shared my reflections about the meeting and how I “wondered if the types of management responsibilities wouldn’t be associated with a number of EPAs and represented in the assessment tools we developed for those EPAs” (EMAIL: me to PAC, 02/04/2010).

A few days later another committee member posted an email to the group. Dr. Acker suggested that EPAs might be better suited to longitudinal assessments rather than rotation specific tools. She envisaged building the pediatric “learning portfolio and ‘longitudinal learning plan’ with EPAs that are cross-rotational” (EMAIL: Dr. A to PAC, 06/04/2010). She felt this could help avoid the reductionist approach I warned against and engage learners to “take some ownership and responsibility for making sure they are learning” (EMAIL: Dr. A to PAC, 06/04/2010). Overall it was clear that committee members were working hard to define EPAs in Pediatrics while at the same time thinking more broadly about a vision for an assessment system.

These discussions continued over the next several meetings as committee members grappled with how to define EPAs in the context of Pediatric medicine. Throughout the process I struggled to distill categories of professional activities and identify contextual constraints in relation to assessment from the fast-paced group discussion. I frequently emailed Dr. Connelly as follow-up to the meetings to clarify and extend my emergent understandings about Pediatric residency education. One example of this was our email exchange following the April 15th 2010 committee meeting wherein I wrote,

I know I keep harping on this but I need to know if there are “processes” that could function as categories. Forgetting for the moment the aspect of medical knowledge about certain presentations.....if that makes any sense at all. According to our discussion today I
have a list of 5 “types” of tasks so far: (a) Patient encounters (history & exam,...........);
(b) Presenting a patient; (c) Handover; (d) Chairing a meeting; and (e) Leading rounds.

(EMAIL: me to Dr. C, 15/04/2010).

In his email response Dr. Connelly suggested that documentation—charting and dictating consult letters be added to my list (EMAIL: Dr. C to me, 15/04/2010). However, getting at the contextual constraints across the various learning contexts (e.g., ward, clinics, etc.) was more challenging.

As a Neonatologist, Dr. Connelly’s practice was based almost exclusively in the NICU. For the most part, residents on rotations through the NICU functioned in a closed educational environment with the notable exception of being on called to attend high-risk deliveries (e.g., pre-term, prolonged labor). As such, Dr. Connelly’s perspective in regard to contextual constraints was coloured by his clinical scope of practice. These insights emerged from other sub-specialist committee members who raised concerns about the feasibility of conducting direct observations of residents within the context of clinics. I reiterated these issues to Dr. Connelly in my email when I wrote, “I also feel it is important to stay grounded in reality....by this I am referring to what Dave (on my left) was talking about in terms of what he could manageably assess during a rotation....at this point in time.” (EMAIL: me to Dr. C, 15/04/2010). I encouraged Dr. Connelly to heed concerns expressed by committee members when I wrote, “These are the types of limitations we will have to deal with and better we face those up front. I think we can get at that by asking what they can manageably assess NOW in the reality of practice.” (EMAIL: me to Dr. C, 15/04/2010).

Although I understood from committee members’ discussions that contextual constraints existed, I had no first hand knowledge about what those looked like in terms of the various practice setting in Pediatrics (e.g., Ward, Children’s Outpatient Center). Nor did I fully grasp the
types of professional activities undertaken by both clinical teachers and residents in those
different contexts. I expressed my concerns about this to a resident committee member who
graciously offered to take me on a guided tour. I was grateful for the opportunity.

We met at Kingston General Hospital (KGH) on April 26th 2010 at 9:00 o’clock in the
morning. Dr. Lysecki had just finished his call shift. Our first stop on our tour was to attend
Rounds for the Pediatric Ward held in the conference room on the 10th floor of KGH. During
Rounds staff coming on shift: (a) Review what has transpired during the night; (b) Report on
patients’ current status; and (c) Discuss management plans.

As we entered the conference room the attending, charge nurse, residents, and medical
students were gathered around the conference table. Dr. Lysecki and I sat side by side against the
wall, somewhat removed from the group. As the meeting progressed he used my notebook to
describe what was happening in real time. For example, as the charge nurse began to speak he
wrote that she was “filling in information that may have been passed on by the Nursing team”
(FIELDNOTE: 26/05/2010). He outlined similarities between the roles of Junior residents and
medical students when he wrote, “Jr’s + medical students do the same work, Jr’s are just
expected to be better.” He also defined differing scopes of responsibility between Juniors and
Seniors when he wrote, “Jr - needs to establish proficiency at presenting (underlined in the
original text) on rounds, Sr – needs to establish proficiency @ leading rounds” (FIELDNOTE:
26/05/2010). Along with these insights he shared his interpretations of what was being reported.
For example, when a first year resident reported that she had yet to perform a physical
examination on one of her patients he wrote, “She knows that physical examination is necessary
for his disease. She phrased it as being about trust with the pt, but really, she lacks trust in
herself. She’s asking for one of them to be present + show her how” (FIELDNOTE: 26/05/2010).
He elaborated further by explaining “Eliciting pain in a child is necessary + not a skill you are born with – Esp in children too young to understand. She needs to be taught.” (FIELDNOTE: 26/05/2010). He suggested that this would happen during ‘walk-around rounds’ when either the Senior or attending would, “walk her through it or observe (underlined in the original text) her w/feedback.” (FIELDNOTE: 26/05/2010).

Dr. Lysecki also made explicit reference to assessment opportunities within the context of rounds. For example, as the Senior discussed a patient’s management plan with a medical student he indicated that the attending had, “adequate opportunity here to directly observe (the Senior’s) ability to (1) manage the patient, (2) teach the medical student” (FIELDNOTE: 26/05/2010). Making reference to the procedurally focused Gynaecological EPA of ‘normal delivery’ provided by Sheele et al. (2008), he wrote, “This is as close to watching a resident deliver a baby as it gets!” (FIELDNOTE: 26/05/2010). As we listened to the final case presentation he underlined that overall we had seen opportunities to assess five of the seven CanMEDs Roles including: (a) Medical Expert (with questions); (b) Communicator; (c) Collaborator; (d) Manager; and (e) Health Advocate.

After “rounds” we toured of the Neonatal Intensive Care Unit (NICU). Dr. Lysecki explained the responsibilities of residents during a rotation there. He also showed me the ventilator used for newborns and briefly described what resuscitating a newborn would involve. We then walked over to Hotel Dieu Hospital (HDH) to visit the Children’s Outpatient Center (COPC).

The COPC clinic is a very busy place where children with chronic medical conditions are followed. Once again, we stood side by side against a wall and Dr. Lysecki explained what we were observing. He pointed out the Chief resident, explaining that he was a fourth year resident
currently studying for his certification exams. Although there was an attending on duty, the Chief resident directed the activity of more junior residents. He stood at the reception desk reviewing patient files. Periodically, a Junior resident would emerge from an examination room and approach the Chief resident. Dr. Lysecki explained they were discussing the patient the Junior resident had just examined. Sometimes the Chief resident would follow the Junior back into the examination room. Dr. Lysecki explained in such cases the Chief resident may be verifying the findings reported by the Junior, at others he might be performing a physical examination maneuver the Junior did not know how to perform or doing one the Junior had not thought to perform (DOCTORAL JOURNAL: 26/05/2010).

After having explained how things worked, Dr. Lysecki sought permission from a parent to observe the history and physical examination of her child. The parent granted consent and we stood off to one side of the examination room as the Junior resident performed a history and physical on the young girl. The Junior then excused herself from the room and reported her findings to the Chief resident. They both returned a few minutes later. The Chief resident introduced himself and asked a few additional questions, verified the physical findings and discussed these with the parent. Dr. Lysecki thanked the child and parent for allowing us to observe and we exited room, returning to the reception desk (DOCTORAL JOURNAL: 26/05/2010). By this time it was almost noon, Dr. Lysecki and I walked back towards KGH together. As we walked Dr. Lysecki reviewed the various assessment opportunities we had observed during our tour. It had been a very productive morning in terms of familiarizing me with the diversity of practice settings in Pediatric residency education. In effect, Dr. Lysecki helped me construct my frame of reference for discussions about contextual constraints that arose in our committee meetings.
The next major committee meeting took place on April 29, 2010, lasting almost two hours. Two faculty members, one resident, our prospective research assistant and I were in attendance. By the end of the meeting the group had: (a) agreed to focus EPA development within the context of subspecialty rotations; (b) assigned committee members to various rotations; and (c) defined the task of identifying no more than 10 EPAs for each rotation and linking these to rotation specific objectives. Members were encouraged to note additional EPAs that might function as “longitudinal EPAs.” A weekly meeting schedule was also established at this time.

Dr. Connelly emailed members outlining the plan of action for those unable to attend the meeting. He acknowledged challenges the group had faced, “Defining EPAs has been difficult, as we struggled between rotation vs. longitudinal, location vs. subspecialty etc” (EMAIL: Dr. C to PAC, 30/04/2010). He also shared the rationale for the action plan “So we thought we would start from the most concrete: let’s start by defining EPAs for rotations. And to be the most concrete, let’s start with subspecialty rotations.” (EMAIL: Dr. C to PAC, 30/04/2010). He listed subspecialty rotations assignments and reminded members to note prospective longitudinal EPAs as those emerged. Dr. Connelly cautiously acknowledged that attendance at weekly meetings would be fluid depending upon committee members’ availabilities.

Having already mapped EPAs for Neonatal/Perinatal Medicine rotations for PGY1, 2 and 3, Dr. Connelly attached these to his email as a model for how committee members might approach the mapping task. His matrix defining EPAs for PGY1s is presented in Figure 6. EPAs are listed down the far left column and CanMEDs Roles across the top. The corresponding numerical identifiers of CanMEDs enabling competencies appeared in the fields of the matrix.
Four days later Dr. Zelinski shared his approach to mapping EPAs for Respirology. Dr. Zelinski divided the respiratory content into categories (e.g., common problems, common diseases, and skills) and drafted a rubric for common problems. When he shared this with Dr. Connelly by email he wrote, “I had spent a chunk of my day working on this – however it looks very different than the mapping approach. I am not sure whether this would work or not” (EMAIL: Dr. Z to Dr. C, 04/05/10). Dr. Connelly forwarded that email to me and wrote, “I'd like your thoughts on this, and how we might use it” (EMAIL: Dr. C to me, 04/05/10).

I replied that Dr. Zelinski “has developed standards of performance! This is what I was talking about in the meeting when I asked what a PGY1, PGY2, etc would be expected to look like in practice” (EMAIL: me to Dr. C, 04/05/10). I had raised the idea of standards several times with the group but was struggling to help them make these explicit. Dr. Zelinski had found a way to do that with his rubric. Seeing the value, Dr. Connelly integrated his work with Dr. Zelinski’s rubric idea and shared it later that same morning, “Take a look at the attached document: this is my attempt to try and tie the grid I developed to Dave's work. As you can see below Dave, Laura applauded your setting standards of performance. I knew she'd make sense of this for me.” (EMAIL: Dr. C to Dr. Z and I, 04/05/10).
Figure 6: Neonatal/Perinatal Medicine PGY1 Rotation EPAs

Ultimately, this was a watershed event that shifted group activity in a significant way.

Although I had talked about standards for performance with the group, Dr. Zielinski had operationalized them. Once Dr. Connelly understood what the rubric was, he worked to integrate the approach with work he had been doing. Building on Dr. Zelinski’s work, Dr. Connelly integrated his mapping process with the rubric format and shared it with committee members by email explaining,
As you will recall, my first drafts had the EPAs down the left hand side, and then mapped those EPAs to existing enabling competencies that are listed in the RC Objectives of Training for Pediatrics. Dave Z. then sent me his work, which established some standards of performance for EPAs – focusing mainly on the Medical Expert role. We tried to then combine these into preliminary "rubrics" for a given EPA. I have attached the first draft of one for Respirology (which only has the Medical Expert blocks filled in) and two relatively simple Neonatal EPA rubrics. (EMAIL: Dr. C to PAC, 05/05/2010).

Furthermore, Dr. Connelly explained, that constructing rubrics forced him to better discriminate which competencies actually applied to a particular EPA. Specifically he wrote, “It also allowed me to drop some of the mapping elements (as they were really a bit of a stretch!)” (EMAIL: Dr. C to PAC, 05/05/2010). The rubric for “Management of Common Respiratory Presentations” is presented in Figure 7. Notably, in addition to incorporating numerical identifiers of applicable CanMEDs enabling competencies, Dr. Connelly inserted Dr. Zelinski’s list of presentations at the top of the rubric. The new EPA rubric format was the topic of discussion at the committee meeting on May 6, 2010. The group agreed to adopt the rubric format as a template for EPAs.
The Significance of EPA Rubrics for PAC

The emergence of the EPA Rubric was significant for two reasons: It mobilized group activity and became a platform for negotiating standards of performance among committee members.

**Mobilizing group activity.** With the advent of the EPA Rubric template came the prospect of conducting a pilot study, an idea I tabled in the committee meeting on May 27th 2010. Committee members were enthusiastic about conducting a pilot study and suggested limiting it to rubrics developed for the subspecialties of Neonatology, Cardiology, Respirology, and Hematology/Oncology (MEETING MINUTES: 27/05/2010). Furthermore, they proposed rubric developers act as leads in consulting with the clinical teacher for those subspecialties to
gather feedback about EPA rubrics and generate buy-in for the initiative. The deadline of June 10\textsuperscript{th} 2010 was set for the completion of draft rubrics for the four pilot rotations.

By the committee meeting of June 10\textsuperscript{th} 2010, EPA rubrics for designated pilot rotations were in various stages of development. Seven Neonatology and five Hematology /Oncology EPA rubrics had been drafted and uploaded into our online community space. Although only the initial common problems EPA rubric had been drafted for Respirology, a plan to work on ones for diseases and system skills had been established. In addition, Dr. Connelly had worked on mapping the Cardiology rotation. He was confident that with minor adjustments the three EPAs identified for Respirology would likely work for Cardiology (MEETING MINUTES: 10/06/2010). He had plans to meet with the clinical teacher in Cardiology to discuss this potential.

The emergence of the rubric template and planned pilot study served to mobilize committee members in a way the three-step curriculum design process proposed by Scheele et al. (2008) had not. In a nine-week period from June 10\textsuperscript{th} to August 16\textsuperscript{th} 2010 the group developed and refined EPA rubrics for the four pilot rotations and established an implementation plan. Ultimately, it was the tangible activity of developing EPA rubrics that engaged members to analyze the curriculum for subspecialty rotations and map EPAs.

**Negotiating standards of performance.** In effect, EPA rubrics became a platform for making standards of performance explicit among committee members. This was accomplished through a rubric review process initiated in the committee meeting held on the June 10\textsuperscript{th} 2010. Two things became immediately apparent from my review of the rubrics posted to our online community space prior to that meeting. First, committee members often described similar resident activity in slightly different ways. Second, many rubric descriptors provided little
information about what performance levels actually looked like in practice. With a view to establishing consensus among group members regarding descriptions of similar resident activity and enhancing the quality of information communicated in descriptors, I shared an edited version of the common problems rubric at the June 10th 2010 meeting.

During the committee meeting I led the group through a comprehensive review of the generic common problems rubric. I began by sharing my observation about the similarities I was seeing across rubrics, specifically regarding the types of resident activity that had enabled me to construct the generic common problems rubric. I suggested it would be strategic for us to leverage this by negotiating set categories of attributes associated with different CanMEDs Roles (e.g., history, physical, problem formulation, investigations, and management for Medical Expert; doctor/patient relationship, inter-professional, intra-professional, and documentation for Communicator) as these emerged over the course of our work.

The review exercise also fostered important discussion about what levels in a rubric actually represented. The group agreed that Level 2 was the “passable performance for a rotation” and Level 3 described “what we should all aspire to be” (MEETING MINUTES: 10/06/2010). Although these definitions would be revisited numerous times over the course of rubric development, the review exercise opened the door to negotiating standards of performance among the group. This was an important development because standards of performance have been characterized as knowledge possessed collectively by a community of practice (Lave & Wenger, 1991). Although agreement among community members as to what these standards are is often assumed, there is evidence of the need for communities to purposefully negotiate these in order to establish shared understanding (Wright, 1996). Ultimately, the rubric review process
supported this negotiation process by making explicit the standards of performance upheld by the group.

Furthermore, in reviewing individual descriptors I encouraged committee members to explain what they meant by ambiguous terms like “appropriate” and “effective”. We then worked together to craft wording that explicitly described hallmarks of resident performance they looked for in the practice setting. Overall, the rubric review process was so successful it became a regular agenda item at our weekly meetings.

**How do Pediatric educational leaders and an assessment specialist co-create an assessment system to capture residents’ performances in the clinical learning environment?**

According to the preceding analysis, the process of creating an assessment system to capture residents’ performance in the Pediatric clinical learning environment can be understood to unfold in 5 phases. The first phase required Pediatric educational leaders to collaboratively negotiate the meaning of EPAs in the context of Pediatric medicine and establish enabling constraints or concrete parameters (e.g., sub-specialty rotations) that served to bound the task of identifying EPAs and mapping the Pediatric curriculum. This phase began on April 1st and continued until April 29th 2010. The second phase entailed linking the conceptual task of curriculum mapping with the tangible activity of developing EPA rubrics. This phase was set in motion by the emergence of the first EPA rubric on May 3rd 2010 and continued until mid-August. The third phase involved a goal setting process (i.e., pilot study) and establishing strategies to secure buy-in from clinical teachers (e.g., rubric leads’ review processes). This phase began on May 27th 2010. The fourth phase began with the emergence of a formal rubric review process that facilitated a collaborative refining of the focus of assessment in the clinical learning environment and the negotiation of standards of performance among group members.
This phase was initiated on June 10th 2010 with the review of a generic rubric that combined multiple rubrics developed to this point. The fifth phase entailed developing an implementation plan that included, for example, negotiating form layout and format (e.g., rip sheets and binders), launch dates, and completion requirements (e.g., 1 per week). Discussion of this plan was initiated on June 24th 2010 and revisited in committee meetings on June 30th and July 29th 2010.

It is important to note that although the phases of development are presented sequentially above, there was overlap among these phases. This overlap is represented in Figure 8.

*Figure 8: Visual Representation of Development Phases*

What is also reflected in the graphic of our development process are the proportions of time allocated to different phases. Although only rough approximations, this visualization of our process underlines the large proportion of time dedicated to socially negotiating meaning and bounding the task at the outset of the project. However, the month spent doing that was dwarfed by the extensive effort invested in refining the focus of assessment and negotiating standards of performance among group members. This finding supports the assertion that although standards may be based on similarly held conceptions of quality upheld by a community of practice, they must be explicitly negotiated for shared understandings to be achieved (Lave & Wenger, 1991; Wright, 1996). Our experience underlines how time consuming this negotiation process can be.
Shifting an Assessment Culture

Watershed Event 2: Residents’ Response to EPA Rubrics in the NICU

In April 2011, I interviewed 3 first year residents about their experience of using EPA Rubrics in the NICU. According to residents, EPA Rubrics provided a useful framework for structuring and supporting the feedback process. Feedback generated with the assessment tools was highly valued and residents supported the idea of EPA Rubrics becoming a resident initiated assessment practice. These findings set in motion a fundamental shift in assessment culture in the Department of Pediatrics.

How Residents Experienced EPA Rubrics

Through the implementation of EPA Rubrics the Pediatric Assessment Committee sought to guide the focus of resident assessment in clinical contexts and ultimately scaffold clinical teachers’ provision of meaningful feedback. Interviewing residents about their experience with these assessment tools in the NICU was a watershed moment because it was an opportunity to see how well these goals were achieved. Towards this end, I contacted 4 first year residents who had completed rotations in the NICU during the 6-month period since EPA Rubrics had been available. One resident was completing an off-site rotation at the time and unable to participate in an interview. The remaining 3 agreed to interviews. I met with residents individually at their convenience (e.g., post-call, before academic sessions) for approximately 30 minutes each. Interviews were semi-structured with an opening question asking about residents’ general experiences of assessment in the program. I then produced copies of the NICU EPA Rubrics and asked if residents had encountered them during their NICU rotations.
Assessment experience. Residents described ITERs as the principle form of assessment in use and agreed they provided limited support for their learning. When asked about the kinds of assessments encountered since beginning the program one resident reported experiencing “little aside from ITERs” (INTERVIEW: R2, 12/04/2011). Another explained, “Not really that much. Certainly there are on-line evaluations done at the end of a rotation but they are not very helpful.” (INTERVIEW: R1, 08/04/2011).

Although residents seemed comfortable with the level of supervision they received in terms of patient safety, they valued immediate feedback on their clinical performance and wanted more. When asked if she felt supported in her learning one resident responded,

On some rotations.........no. I would say I feel supported in that if there is a problem with a patient or if you’re worried about them, the doc will be there to back you up or a Senior resident. Its not like I am concerned for anyone’s safety but I guess I don’t always feel like I am getting better at things because you are going through it doing a lot of it on your own. You can ask but its not like it’s a regular occurrence. (INTERVIEW: R1, 08/04/2011)

Another resident shared her experience of a case presentation to a clinical teacher on the Ward about a surgical consult for one of her patients. She talked about the teacher’s response to her presentation of the case. The clinical teacher had reminded her that, “they are consulting, you don’t have to take their word....we have to sign-off on those orders and make the decisions.” (INTERVIEW: R3, 13/04/2011). She explained that in retrospect she understood why the clinical teacher had responded that way and how that reaction had stimulated her reflection about the need to make her thinking explicit in case presentations. Specifically she explained,
I was just discussing the case and I guess I was just talking about management plans and I talked about surgery without saying... this is what they’ve suggested we should consider, I have, this is why I think we should go along with their suggestion...but I didn’t do that. (INTERVIEW: R3, 13/04/2011).

When asked if she had found the discussion useful she explained, “yes, oh yes, definitely...great kind of feedback at the moment, right then and there. Some people might get defensive about it but I don’t mind being told because I think its good learning” (INTERVIEW: R3, 13/04/2011). When asked if she received enough of these types of feedback opportunities she lamented that more would be better.

**EPAs.** When presented with EPA Rubrics, only 2 of the 3 residents reported having used them during their rotation in the NICU. The resident who had not encountered the rubrics on her rotations through the NICU was disappointed, remarking, “I would have really liked to have known about these.” (INTERVIEW: R2, 12/04/2011). Upon reviewing the documents she remarked on their potential as an advanced organizer to help residents prepare for procedures, “like the UVC/UVA replacement checklists – those are very helpful.”(INTERVIEW: R2, 12/04/2011).

The residents who had used EPA rubrics during their time in the NICU talked about the impact of rubrics in relation to feedback. As one resident explained, “It was nice to get feedback directly after you did something so it was fresh in my mind and not 3 months later when you often get the ITER and you are like...what did I do?”(INTERVIEW: R1, 08/04/2011) For this resident, the timing of feedback was important for promoting improvement, she valued being able to relate it directly to an activity. The other resident spoke about the powerful personal
impact of feedback on her self-efficacy beliefs and the catalytic potential it had to support further learning,

one on one teaching with the staff...it was amazing....that kind of bedside teaching...it gives me such confidence that what I do has been perceived by an expert and they have a couple of things to make it better. Obviously, it wouldn’t feel great if they ripped apart the whole thing, but generally they have a few suggestions and it gives you an awareness of where you are. I find for me, its confidence building. I’m where I should be....I’m doing well. I like that it gives you the opportunity to improve and be even better.

(INTERVIEW: R3, 13/04/2011).

Both residents spoke about how the use of EPA rubrics increased the specificity of feedback they received, “It was nice because it had a direct focus on some skill or something I just did as opposed to a general overview type comments.”(INTERVIEW: R1, 08/04/2011) They obviously valued the quality of feedback and kind of interaction with clinical teachers the use of rubrics promoted,

Somewhere can say ‘oh you’ve done a good job’ but it doesn’t mean anything...what you want someone to say is here is what you’ve done well, here’s what needs improvement... afterwards we sat down together and Dr. X filled out the evaluation in front of me and we discussed, “Oh that was excellent”.... “I’ll give you that there”, you got to speak.

(INTERVIEW: R3, 13/04/2011)

Finally, residents were positive about EPA Rubrics becoming a resident initiated process. One resident responded that she would, “definitely support the idea of residents taking responsibility for having 1 of each completed on each rotation in the NICU” (INTERVIEW: R2, 12/04/2011) The other talked about the need for them to be made a standard requirement as a
way of ensuring they were completed, “... if you have to do it that week, you know its like anything, if its concrete...the time is made.” (INTERVIEW: R3, 13/04/2011)

The Significance of Residents’ Experiences

Residents’ insights about their experience with EPA feedback rubric were significant for two reasons: They underlined the value of rubrics for supporting assessment and feedback processes and revealed residents’ readiness to adopt a resident-driven assessment model in the clinical environment. Together these insights set in motion a shift in assessment culture within the Department of Pediatrics.

Rubrics. According to residents, EPA Feedback Rubrics guided the focus of assessment and supported clinical teachers to provide meaningful feedback. As opposed to generic comments about having done a good job, residents reported having assessment conversations with clinical teachers about their performance. These interactions enhanced their self-efficacy beliefs, potentially making them more receptive to the feedback offered by their clinical teachers (Archor, 2010; Sargeant, Mann, van der Vleuten, & Metsemakers, 2008; Teunissen, et al., 2009; Watling, Driessen, van der Vleuten, & Lingard, 2012). The implementation of rubrics also appears to impact how residents used feedback. Residents reported that immediate, response-specific feedback that included information about how to improve allowed them to link it directly with their clinical activities (Archor, 2010; Sargeant, et al., 2011). Overall, rubrics seemed to support residents’ conceptions of feedback as a catalyst for learning and enhanced their perceptions about the clinical environment as a valuable learning space (Anderson, et al., 2010; Mann, et al., 2011; Sargeant, et al., 2011; Teunissen, et al., 2009).

Resident-driven assessment. Residents’ readiness to take ownership of and responsibility for driving the use of EPA Feedback Rubrics in the clinical environment was a
promising finding. Researchers argue this kind of active involvement supports residents’ uptake of feedback (Bindal, et al., 2011; Cooke, et al., 2010; Regehr G, et al., 2011; Setna, et al., 2010). Beyond supporting up-take, it is a particularly practical model in the clinical learning environment where supervision is often sequential. Ultimately, empowering residents with the responsibility to ensure rubrics are completed is a strategic approach to managing assessment in the busy clinical environment. However, this represented a significant shift in departmental assessment culture.

Communicating a vision of assessment. Departmental leadership adopted two strategies to support this shift in assessment culture. The concept of resident-driven assessment was introduced formally to residents during an academic half-day event in July 2011. I was asked by the Program Director to facilitate a one-hour session. I began the session by introducing myself and talking about how I think about learning and the role of assessment. I used my Resident’s Learning Cycle model from my dissertation work to do this. I had put together a handout of sample assessment tools with the model as the cover page so it was right in front of them as I described my thoughts about learning and assessment. I referred specifically to the model and walked them through the role of assessment, the function of feedback in calibrating their self-assessment skills and the relationship of those skills to SRL. I then emphasized the parallels between their learning at the PGME level and future professional practice. Throughout the discussion I underlined the critical aspect of them taking responsibility for and ownership of their learning.

Our discussion then turned to the assessment table presented in Figure 9. I constructed the table to identify the various assessment tools we were introducing and indicate CanMEDs roles each tool was designed to assess. We then reviewed the sample tools and briefly discussed
how these should be used, what residents’ responsibilities were for ensuring they were assessed and what their responsibility for summarizing performance data would be.

<table>
<thead>
<tr>
<th>Assessment Tools</th>
<th>CanMEDs Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct Observation:</strong></td>
<td>ME  CM  CL  MG  HA  SC  PR</td>
</tr>
<tr>
<td>Encounter Cards &amp; Summary form</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>EPA Rubrics (various)</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>OSCE</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>Observed History &amp; Physical</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td><strong>Other:</strong></td>
<td>ME  CM  CL  MG  HA  SC  PR</td>
</tr>
<tr>
<td>Allied Health Professional Assessment</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>Topic Presentation Rubric</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>Consult Note Assessment</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td><strong>Objective Testing:</strong></td>
<td>ME  CM  CL  MG  HA  SC  PR</td>
</tr>
<tr>
<td>AAP PREP</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>Short Answer Questions</td>
<td>X  X  X  X  X  X  X</td>
</tr>
<tr>
<td>American Board Examination</td>
<td>X  X  X  X  X  X  X</td>
</tr>
</tbody>
</table>

*Figure 9: July 2011 Assessment Table Handout*

It was an engaging discussion with high-level questions about things like validity and how we could be sure that the tools were valid. I explained that validity in this circumstance is not conceptualized as a property of a tool but rather as an argument that is constructed founded on the quality of inferences that can be made based on the data generated with the tool (Kane, 2001; Stobart, 2009). I explained that various aspects will need to be explored and that they would need to be involved in constructing that argument. I provided examples of the kinds of questions that would guide our examination of validity in this instance including: “Was the feedback generated with the tools useful to learners?” and “Did the patterns of performance captured with the tools predict future performance?” The level of discussion was fascinating and the residents were totally engaged and obviously comfortable asking questions throughout the session.
At the end of the session I handed out feedback sheets and asked residents to answer two questions. Specifically, “What did you find most interesting about the discussion today and why?” and “What would you like to learn more about in relation to assessment?” No identifying information was required and I explained these data would be used to prepare my report for the Program Director. Twelve feedback sheets were completed and I provide a summary based on my report to the Program Director (ACADEMIC HALF-DAY REPORT: Me to Program Director, 23/07/2011).

Two overarching themes emerged in response to the question about what residents found most interesting and why. The overview of the assessment system and how beneficial it was to understand and have the opportunity to talk about it was a dominant theme. Concepts related to self-regulated learning, mentioning ownership and responsibility for learning and the role of feedback in learning were also prominent. Themes that emerged in response to the second question about what residents would like to learn more about in relation to assessment included issues related to staff buy-in, including responsibility for the provision of feedback, what staff were learning, and how attitudes of both staff and residents could be changed. The concept of utility, specifically the meaningfulness of feedback generated with assessment tools and what staff members’ opinions were about the usefulness of these also emerged.

One month later I spoke with clinical faculty at an annual retreat in August of 2011 and introduced the concept of resident-driven assessment. We provided the same handout of all new assessment tools. Clinical teachers were advised about assessment requirements and encouraged to support residents in meeting them. Ultimately, this was framed as a professional responsibility for residents and clinical teachers.
Following these events the Program Director circulated regular email communications. These emails introduced new rubrics as they emerged, sought feedback to inform adjustments to the tools, and periodically reminded residents and clinical teachers about completion requirements. The Encounter Card Rubric for minor Emerge and clinics is one example of how residents and clinical faculty have become more actively involved in shaping the new assessment system and ultimately shifting the culture of assessment. After piloting the new rubric for several months the Program Director polled residents and clinical teachers by email about their satisfaction with the tool. Clinical teachers indicated they found the original descriptor for the developing level of the history attribute “Focuses too narrowly on one/few possible diagnoses” was too low. We subsequently revised that descriptor to “Focused and concise” as presented in Figure 10.

Eliciting and responding to feedback in this manner openly communicated to residents and clinical teachers that their opinions were valued and had a tangible impact. The assessment system was not merely something that was happening to them, but a change they could be actively involved in shaping. In effect, this open communication strategy became a platform for negotiating the standards for performance across the wider Pediatric community (Lave & Wenger, 1991). This was particularly important in light of evidence of inconsistent performance standards among clinical teachers.
Empowering the community. Although seeking and incorporating feedback from the Pediatric community was strategic, efforts to more fully involve residents and clinical teachers in assessment development are also being made. One example is the Senior Consult Clinic Rubric presented in Figure 11. In effect, development of assessment tools (e.g., EPA Feedback rubrics), originally the purview of PAC has dissipated out into the wider Pediatric community. This transition was facilitated by the legacy of committee activity; namely, my professional growth as an assessment specialist in postgraduate medical education and the Rubric Descriptor Bank.
Figure 11: Senior Consult Clinic Rubric

In the case of the Senior Consult Clinic Rubric, I met with a clinical teacher to establish the focus of assessment. This entailed discussing the types of activities a senior resident assumed during consult clinics and determining the scope of interaction residents have with clinical teachers in that particular learning environment. Based on this analysis I drafted a rubric. Most attributes for the rubric could be drawn directly from the Rubric Descriptor Bank: However, one important focus of assessment in the Consult Clinic learning environment is a senior resident’s ability to manage patient flow. Something that was not yet defined in our bank. The clinical teacher and I worked together to develop descriptors for this new attribute that was subsequently added to the bank. Similar collaborations are planned with residents to develop rubrics for handover and senior ward resident.

The Pediatric assessment system now includes a wide range of assessment tools. These are reflected in the table provided at the residents’ academic half-day session this year (see Figure 12.) What is notable in Figure 12 are the 6 new rubrics listed. This diversity allowed us to capture resident performance data across all seven CanMEDs Roles from a variety of individuals
(e.g., patients, allied health professionals, peers, and clinical teachers), across a range of situations (e.g., explaining complicated procedures, delivering bad news, charting), in various contexts (e.g., ward, clinic, office, classroom), and do so over time.

Although the assessment system is paper-based at present, MEdTech – our information technology group is working on a framework that will allow us to move all rubrics online. This will be especially advantageous for aggregating assessment data across rubrics to create ITER style reports. By ensuring ITERs are based on actual resident performance data as opposed to informal discussions with colleagues, our system will address a significant weakness that has been identified with the ITER system (Bullock, et al., 2011; Chou, et al., 2009; Dudek, et al., 2008; Norcini & Burch, 2007; Williams, et al., 2005).

Beyond the benefit of improving the quality of judgments about residents’ emerging competence, the move to an electronic framework for managing rubric assessment data will address a practical administrative burden. Automating the generation of ITER style reports will eliminate the need for clinical teachers to complete ITERs in the ONE45 system.

Of paramount importance, the electronic framework will better support residents’ ongoing access to assessment data so they can more easily monitor their own learning progress (Boud & Falchikov, 2006; Gibbs, 2006; Nicol & Macfarlane-Dick, 2006; Price, et al., 2010; Shute, 2008). This is critical because residents expressed frustration with managing the current paper-based system in my academic half-day session in July 2012. Ultimately, they want easy and continuous access to their assessment data so they can use it to inform their learning effort and seek additional support as needed.
### Assessment Tools

<table>
<thead>
<tr>
<th>Assessment Tools</th>
<th>CanMEDs Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct Observation:</td>
<td></td>
</tr>
<tr>
<td>Encounter/Ambulatory Cards &amp; Summary forms</td>
<td>X X X X X X X</td>
</tr>
<tr>
<td>EPA Rubrics - NICU &amp; Cardio/Resp*</td>
<td>X X X X X X X</td>
</tr>
<tr>
<td>Senior Consult Clinic Rubric (PGY4) NEW!</td>
<td>X X X X X</td>
</tr>
<tr>
<td>COPC Jr. Attending (PGY4) NEW!</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Inpatient Senior - Attending review (PGY4) NEW!</td>
<td>X X X X</td>
</tr>
<tr>
<td>OSCE</td>
<td>n/a</td>
</tr>
<tr>
<td>Observed History &amp; Physical</td>
<td>X X X X X X</td>
</tr>
<tr>
<td>Additional Assessment Tools:</td>
<td></td>
</tr>
<tr>
<td>Patient Feedback Cards (Clinics) NEW!</td>
<td>p.8</td>
</tr>
<tr>
<td>MSF Rubric RECENT!</td>
<td>X</td>
</tr>
<tr>
<td>Resident Teaching Feedback Rubric NEW!</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Topic Presentation Rubric</td>
<td>X X X X X</td>
</tr>
<tr>
<td>Consult/referral Note Assessment</td>
<td>p.12</td>
</tr>
<tr>
<td>Objective Testing:</td>
<td></td>
</tr>
<tr>
<td>AAP PREP</td>
<td>n/a</td>
</tr>
<tr>
<td>Short Answer Questions</td>
<td>n/a</td>
</tr>
<tr>
<td>American Board Examination</td>
<td>n/a</td>
</tr>
</tbody>
</table>

*EPA rubrics are currently under review

**Please Note** formatting of some assessment tools in this package were adjusted for printing

Figure 12: July 2012 Assessment Table Handout

**How does the culture of assessment within the Department of Pediatrics shift as a result of evolving understandings about assessment?**

Thinking about transformation in assessment practice as cultural change is not a new phenomena (Gipps, 1994; Segers, et al., 2003; Shepard, 2000). Authors write about the shift from psychometric measurement or testing orientations towards educational assessment.

Typically this cultural shift involves changes in: (a) conceptions of learning; (b) the focus and strategies of assessment; (c) the role of learners; and (d) how results are used (Gipps, 1994; Segers, et al., 2003; Shepard, 2000). According to the preceding analysis, the cultural shift in assessment is well underway in the Department of Pediatrics at Queen’s. Residents and clinical
teachers are embracing their shared responsibility for assessment. This is aided by assessment rubrics that help to focus assessment on important aspect of residents’ clinical activities and supports clinical teachers in providing timely, high quality, targeted feedback. Beyond that, community members are assuming development responsibilities and actively working to shape how assessment happens in the department. Our work towards developing an electronic framework to meaningfully aggregate assessment data for reporting purposes and to inform residents’ learning is further evidence of the significant shift in assessment culture. Furthermore, the fact that residents are concerned about when the electronic framework will be in place so they can continuously monitor their progress and direct their learning effort signals they conceptualize learning as on-going professional growth.

**Eclipsing Departmental Boundaries**

**Watershed Event 3: Refining the Rubric Descriptor Bank**

The Rubric Descriptor Bank is the product of almost three years of collaborative effort among Pediatric educational leaders, clinical teachers, residents, and an assessment specialist. The bank is a repository of descriptions of residents’ daily activities categorized by CanMEDs Roles (e.g., Medical Expert, Communicator, Collaborator). Multiple attributes have been defined for each CanMEDs Role and three levels of performance are described for each attribute. For example, some attributes associated with the Manager Role include: Prioritization, Leadership, Time management, Problem solving, and Patient flow. The three defined levels of performance for the attribute of patient flow include: (a) becomes overwhelmed quickly and falls behind; (b) sees all assigned cases – accumulating delays over time; (c) manages patient flow with minimal/justifiable delays. The Rubric Descriptor Bank holds more than 100 attributes and facilitates the efficient development of high quality rubrics that are used to assess resident
performance in the clinical learning environment (See Appendix F: The Rubric Descriptor Bank).

Although this resource is the cumulative effort of numerous individuals over an extended period of time, its emergence in its current form was a watershed event. On July 7th 2012 I spent fourteen hours refining rubric attributes and descriptors in our excel workbook. To this point we had hired two research assistants (RAs) over the course of the project to manage and organize the bank. These RAs were able to enter new attributes and associated descriptors and reformat the excel workbook. However, in spite of their advanced degrees in education, refining rubric attributes and descriptors was beyond their expertise. In effect, this task demanded both specialized expertise in assessment and an in-depth understanding of Pediatric residency education. It was a task I could not delegate: it was something I had to perform myself.

**How the Rubric Descriptor Bank Emerged**

Scoring rubrics are generally considered to be of two types, holistic and analytical (Fostaty-Young & Wilson, 2000; Jonsson & Svingby, 2007; McMillan, 2007). Holistic rubrics describe performance in terms of criteria grouped together to represent categories of performance (e.g., emerging, developing, achieving). As the label suggests, these types of rubrics are used to make holistic judgments about the quality of performance overall and are more commonly used in large-scale assessment contexts (Jonsson & Svingby, 2007). The difficulty with rubrics of this kind is the assumption that the range of learner performance will conform to multiple criteria associated with pre-determined categories (McMillan, 2007). However, this is often not the case in practice. Given the holistic nature of these instruments, raters frequently face selecting the most suitable category even when some criteria of the performance represented in that category was not adequately met. This limits the specificity of feedback available to learners. Although
learners may achieve a standard of performance most closely associated with a category, information about specific areas in need of improvement are not identified. In a large-scale assessment context this is not an issue since the purpose is predominantly summative. However, the lack of feedback information does present a problem in contexts where these instruments serve a formative function. Consequently, the specificity of analytical rubrics is considered a better fit for such circumstances.

Analytical rubrics commonly list criteria individually and provide either numerical rating scales (e.g., 1,2,3, etc.) or quantitative narrative descriptors (e.g. none, some, all) for each criterion (Fostaty-Young & Wilson, 2000; Jonsson & Svingby, 2007; McMillan, 2007). Although scoring each criterion separately is helpful in supporting learners in identifying personal areas of strengths and weaknesses, these instruments provide little guidance about how to improve because they fail to describe what different levels of performance look like. In effect, they lack important information about the standards of performance for criteria. This is problematic because it is exactly this type of information learners need to guide future learning effort. In formal learning settings this shortcoming is often dealt with by providing exemplars and rater training (Jonsson & Svingby, 2007).

So where holistic scoring rubrics provide an overview of general standards of performance, they offer learners insufficient individualized guidance to support improvement. On the other hand, analytical scoring rubrics identify individual areas of strength and weakness, but offer little guidance in terms of facilitating both learners’ and teachers’ understandings of performance standards.

Ultimately, in order to provide learners with quality feedback a rubric design that marries the strengths of both holistic and analytical scoring rubrics is ideal. This hybrid design describes
performance according to multiple criteria in each performance category and provides a
mechanism for identifying specific aspects of performance that fall outside pre-determined
categories. The advantage of this model is improved transparency (e.g., explicit standards of
performance) and increased feedback specificity (e.g., descriptive criteria by category). This
hybrid model emerged out of the work on EPA Feedback Rubrics.

As EPA Feedback Rubrics for Pediatric subspecialty rotation emerged in the summer of
2010 they were systemically entered into an Excel workbook. Originally, individual excel sheets
were created for each of the seven CanMEDs Roles with attributes from different rubrics
grouped together within each sheet (e.g., history, physical, problem formulation, investigations,
and management). A section of the initial Excel workbook is presented in Figure 13.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Rotation/EPA</th>
<th>Learning Objectives</th>
<th>Level 1 Descriptors</th>
<th>Level 2 Descriptors</th>
<th>Level 3 Descriptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Exam</td>
<td>Neo/Discharge</td>
<td>2.3</td>
<td>q Performs incomplete discharge examination</td>
<td>q Performs complete discharge examination</td>
<td>q Completes concise/relevant examination</td>
</tr>
<tr>
<td>Neo/Normal Exam</td>
<td>3.3</td>
<td></td>
<td>q Omita basic physical examination maneuvers</td>
<td>q Misses mild symptoms/signs</td>
<td>q Performs a concise and relevant examination, taking into account the special requirements of a newborn</td>
</tr>
<tr>
<td>Resp/Common Presentations</td>
<td>3.3</td>
<td></td>
<td>q Omita basic physical examination maneuvers</td>
<td>q Misses mild symptoms/signs</td>
<td>q Performs a concise and relevant examination, taking into account the special requirements of a newborn</td>
</tr>
<tr>
<td>Resp/Common Resp Problems</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiac/Common Problems</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Figure 13: First Excel Workbook Format*

This format facilitated cross-rubric comparisons of performance descriptors in these early stages
and helped identify CanMEDs Roles that were underrepresented (e.g., Health Advocate).
In the summer of 2011 major effort was invested to update the Excel workbook. Data entry had fallen behind rubric development so these needed to be entered. As well, many rubrics had gone through numerous revisions, some of which were not reflected in the workbook. The original Excel workbook design with CanMEDs Roles on separate sheets made data entry cumbersome because of the need to flip between sheets when entering a single rubric. Consequently, the seven sheets were collapsed and columns within the sheet were reordered (e.g., Rotation/EPA moved to the end). A column for CanMEDs Roles was added, and the Learning Objectives column was dropped (See Figure 14). Once all of the new rubrics were entered and revisions to existing ones made, there were a total of 178 attributes in the bank. Many of these had multiple descriptors associated with each level of performance. Over the following year descriptors were further refined based on feedback from users. Redundancies were gradually eliminated and new descriptors added.

<table>
<thead>
<tr>
<th>CanMEDs</th>
<th>Attribute</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
<th>Rotation</th>
<th>EPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical Expert</td>
<td>Physical Exam</td>
<td>Omits basic physical examination maneuvers.</td>
<td>Misses mild symptoms/signs. Recognizes aspects of Ps for problem formation. Misinterprets physical findings.</td>
<td>Completes concise/relevant examination.</td>
<td>Cardiology</td>
<td>Common Cardiac Presentation</td>
</tr>
<tr>
<td>Medical Expert</td>
<td>Physical Exam</td>
<td>Examination is not adequately focused on illness (e.g., too broad/narrow)</td>
<td>Performs complete exam. Limited interpretation of findings.</td>
<td>Focuses exam on disease. Considers systems that may play part in disease triggers/control. Holistic interpretation of findings.</td>
<td>Cardiology</td>
<td>Chronic Cardiac Problems</td>
</tr>
<tr>
<td>Medical Expert</td>
<td>Physical Exam</td>
<td>Omits basic physical examination maneuvers.</td>
<td>Misses mild symptoms/signs. Recognizes aspects of Ps for problem formation. Misinterprets physical findings.</td>
<td>Completes concise/relevant examination. Accurately interprets physical findings.</td>
<td>Respirology</td>
<td>Common Respiratory Presentation</td>
</tr>
<tr>
<td>Medical Expert</td>
<td>Physical Exam</td>
<td>Examination is not adequately focused on illness (e.g., too broad/narrow)</td>
<td>Performs complete exam. Limited interpretation of findings (findings of allergic rhinitis, etc. in asthma).</td>
<td>Focuses exam on disease control/progression. Ascertain possible reasons/barriers for poor control. Considers alternative diagnoses in difficult to control/treat patients.</td>
<td>Respirology</td>
<td>Chronic Respiratory Problems</td>
</tr>
<tr>
<td>Medical Expert</td>
<td>Physical Exam</td>
<td>Omits basic physical examination maneuvers.</td>
<td>Misses aspects of physical examination specific to the newborn. Misses mild symptoms/signs. Misinterprets physical findings.</td>
<td>Performs concise &amp; relevant examination, taking into account special requirements of a newborn. Recognizes signs and symptoms and integrates their relevance. Considers the significance of any identified abnormalities.</td>
<td>Neonatal</td>
<td>Newborn Examination</td>
</tr>
</tbody>
</table>

**Figure 14: Second Iteration of the Excel Workbook**

In its most recent iteration, the Bank is currently organized by CanMEDs Roles that appear in rows above attributes and descriptors in each of the seven categories (See Figure 15). A column to indicate Source has been added. The Source column lists various rubrics an attribute
is included on, as many are now used on multiple rubrics (e.g., Encounter Card Rubrics (EC Rubrics), Senior Consult Card (SCC)). Descriptors are now listed in individual rows and when multiple descriptors were associated with a single attribute, the attribute is numbered to reflect steps in or phases of a process (e.g., Physical1, Physical2). This approach was purposefully adopted to facilitate data aggregation at the attribute level once an electronic platform becomes available. Ultimately, this will enhance the specificity of feedback information that can be generated with the system by allowing us to drill down and examine for example, whether a resident is having difficulty identifying investigations, interpreting results, or both. Furthermore, aggregating performance data at the attribute level will allow us to construct patterns of performance across clinical learning setting and determine with a greater degree of confidence whether observed difficulty is attributable to a lack of knowledge or experience in a particular area (e.g., Neonatology) or a more general weakness in overall performance (e.g., clinical reasoning).

**Figure 15:** Current Rubric Descriptor Bank (limited version)
The Significance of the Refined Rubric Descriptor Bank

The refined Rubric Descriptor Bank is significant because high quality rubrics can be quickly generated by determining the focus of assessment and identifying attributes that best fit that purpose. Furthermore, just as committee members identified the transferability of attributes between subspecialty rotation rubrics early in the EPA development process, attributes also appear to transfer across postgraduate programs.

Facilitating rubric development. Developing high quality rubrics that explicitly communicate what different levels of performance look like in medical practice is a time consuming process. After determining the focus of assessment (e.g., attributes), shared standards of performance must be established to inform descriptor development. In effect, committee members engaged in both establishing the focus of assessment by identifying attributes for CanMEDs Roles and negotiating standards of performance for these during the EPA development process. Because of this sustained effort, much of the time consuming work of rubric development has been done. Although the current bank is not exhaustive, a solid foundation has been established upon which further work can build. At a fundamental level it is far easier to craft descriptors for new attributes now that language conventions have been established and performance standards for other attributes can be referenced.

Can assessment innovation in Pediatrics eclipse departmental boundaries?

Assessment innovation in Pediatrics is seeping across departmental boundaries in various ways and in multiple directions. One example of this was an initiative spearheaded by the former Department Head for Pediatrics. A practicing general surgeon, Dr. Jones was motivated to develop a Surgical Procedures rubric. She and I collaborated to draft the rubric. I then shared the draft with our Associate Dean of postgraduate medicine, Dr. Walker, also a
general surgeon and incorporated his feedback. We piloted a small sample of rubrics with General Surgery residents. A multi-site validity study of the Surgical Procedures rubric is currently being proposed as the focus of a research project for one of our third year surgery residents applying to Queen’s Clinical Investigators Program.

In another instance, the Pediatric Multisource Feedback (MSF) Rubric to support allied health professionals’ assessments of residents is being adopted by other programs. With minor revision, the MSF rubric was adapted for the Departments of Neurology and Family Medicine. The rubrics are currently being piloted across the three programs and are the focus of a collaborative validity study.

Queen’s Department of Family Medicine also recently received funding to purchase tablets so patients can complete feedback forms online. They have expressed interest in the newly developed Pediatric Patient Feedback Rubric for this purpose. More recently, the Program Director for Anesthesiology has reviewed the Rubric Descriptor Bank. He is considering moving to a rubric model as a solution for the variable standards of performance among clinical teachers reflected in daily assessments of Anesthesiology residents’ performances in the operating room.

The Resident Teaching Rubric is another outcome of the bank that has potential for both resident assessment and program evaluation purposes. This rubric is designed to capture residents’ bedside teaching of medical students and more junior residents. Aside from the obvious benefit for residents in terms of gathering feedback to inform improvements in teaching performance, this tool has the potential to generate large grained data about the overall quality of teaching in the clinical environment. Having access to this kind of information is particularly advantageous given accrediting bodies increased emphasis on practice-based learning and the push to enhance the quality learning in the clinical environment. Our Undergraduate Medical
Education group is especially interested in the potential of having access to aggregate data about quality of teaching in different clinical learning environments.

All of this assessment innovation at Queen’s School of Medicine can be traced back to the dedicated investment of members of the Pediatric Assessment Committee. However, it is an ever-widening circle of collaborators who push our innovation forward.
CHAPTER FIVE: DISCUSSION AND CONCLUSION

Discussion

The purpose of this case study was to examine how a program in residency education cultivates an autonomy-oriented assessment culture. This research focused on efforts in the Department of Pediatric at Queen’s University to make that shift. Specifically, I was interested in exploring how we initiated an assessment system to capture Pediatric residents’ performances in the clinical environment, how that system supported a cultural shift within that program, and whether assessment innovation grounded in Pediatrics could seep out into the wider PGME community at Queen’s University.

The central contribution my dissertation makes in response to this research purpose is that it is possible to shift the culture of assessment within a Pediatric residency program. That shift can be understood to unfold over a prolonged period of time through a process of mediating both social and regulatory requirements. In the following sections I describe what that process has looked like in the context of the Department of Pediatrics at Queen’s University and what our understandings about that process foreshadow.

Recognizing Need

The shift in Pediatric assessment culture began with educational leaders in the department identifying the need for change. In reviewing educational practices in preparation for the end of an accreditation cycle, the RPC identified assessment as an area in need of improvement. They struck an assessment sub-committee and named Dr. Connelly chair. Dr. Connelly was knowledgeable about assessment practice in the department, having previously held the position of Program Director.
One weakness in assessment Dr. Connelly wished to address as Chair of the assessment sub-committee was the ITER system. Based on his experience as a former Program Director he knew how difficult it was to get clinical teachers to complete ITERs in a timely manner, something that was not unique to Pediatrics (Bullock, et al., 2011). He was also very aware of the critical importance of ITER completion rates during accreditation reviews.

Based on his experience of completing ITERs organized around CanMEDs, he felt he understood why low ITER completion rates were a chronic issue department wide. In his view, ITERs had become endless checklists of items that were difficult to link with what he actually observed residents do in the clinical learning environment, a concern also voiced by medical education researchers (Govaerts, et al., 2007; Verma, et al., 2005). He suspected many of his departmental colleagues felt similarly. In effect, low ITER completion rates could be interpreted as a departmental expression of passive resistance in the face of assessment practice that was not aligned with the educational environment as perceived by clinical teachers.

Dr. Connelly encountered Entrustable Professional Activities (EPAs) at a journal club in September, 2009 (Scheele, et al., 2008; ten Cate & Scheele, 2007). He was immediately struck by the potential EPAs had to bridge the formal curriculum of CanMEDs with the clinical learning environment. Although he was not certain how to operationalize them in Pediatrics at that time, he recognized the potential they had to establish the focus of assessment on actual clinical activities.

Re-conceptualizing Assessment

In October of 2009, I was introduced to Dr. Connelly in a meeting with the Vice-Dean Education. We met again later that same week at a faculty development function on assessment. In our brief encounters we identified our common interest in the area of assessment and more
specifically, a shared enthusiasm for EPAs. Over the next few months we worked to conceptualize how EPAs might translate into Pediatric residency education, sought funding to support our initiative and ethics approval for this case study. Once funding was secured and ethics was approved we worked with members of the PAC to develop an assessment system based on EPAs. At a fundamental level our goal was to develop an assessment system to capture residents’ performances in the clinical learning environment.

**Realizing Change**

EPAs were the conceptual vehicle that allowed us to bundle multiple learning objectives in meaningful ways and develop assessment tools to assess these in situ. In effect, through the EPA Rubric development process, committee members established the focus of assessment in pediatric clinical learning settings and negotiated standards of performance in the form of EPA Rubrics (Lave & Wenger, 1991).

Through opportunities to pilot rubrics, the committee was able to ensure the new assessment system was supporting clinical teachers’ provision of meaningful feedback in effective and efficient ways (Nicol & Macfarlane-Dick, 2006). Residents’ positive responses to rubrics and readiness to assume responsibility for driving the assessment process changed the way assessment was conceptualized in the department. Previously, the ITER system symbolically placed clinical teachers in the seat of power with regard to assessment. Our new resident-driven assessment system is a symbol of shared ownership with residents.

**Engaging and Empowering Community**

Championing the shift in assessment culture was taken-up by a new Program Director and former PAC member, Dr. Acker. She provided opportunities for residents and clinical
teachers to negotiate the meaning of standards of performance and have input in adjusting those standards when so required. This was accomplished both formally (e.g., retreats and academic half-days) and informally (e.g., emails eliciting feedback on new rubrics, reminders about completion requirements). These activities empowered residents and clinical teachers, fostered feeling of shared ownership of the change, and engaged them in shaping the assessment system. The Rubric Descriptor Bank, a legacy of the PAC’s work with EPAs was instrumental in this regard. Rubric development has dissipated out into the wider Pediatric community, with clinical teachers and residents becoming actively involved in identifying important assessment opportunities (e.g., Senior Consult Clinics, Handover) and collaboratively developing rubrics that align with them.

Residents’ enthusiasm about the quality feedback they receive and their push for continuous access to assessment information, paired with clinical teachers’ completion rates are evidence of a shift in assessment culture in Pediatrics. Three years ago ITERs were the only form of assessment wherein residents’ clinical performance was documented. At that time, delays of a month or more between clinical performance and documented feedback were the norm, and the quality of feedback provided was highly variable. Today, residents actively seek out and collect a minimum of one rubric per week and frequently more. They also receive feedback about their clinical performance from other health professionals they work with in the clinical environment, patients they treat, and learners they teach.

Although progress has been great, we struggle to manage all of the assessment information we are generating. This is frustrating because it hinders our ability to fully leverage assessment information to support learning. Our current priority is to migrate to an electronic framework. This will allow us to systematically collect and aggregate rubric data, construct
patterns of resident clinical performance over time, and ultimately automate ITER style report
generation. Once this is in place we will be closer to our goal of meeting the requirements of our
regulatory body (RCPSC) and enhancing the quality of assessment information available to
support our residents’ learning.

MEdTech, the in-house technology group at Queen’s School of Medicine is currently
building a system to do this. There are plans to pilot the new system in January 2013. Once our
data management issues have been solved we will be looking to the next important phase of our
shift towards an autonomy-oriented assessment culture: namely, progress and promotion. Dr.
Acker, the Program Director and I have already begun discussing what that might look like.

Re-envisioning Progress and Promotion

Gains have been made in the area of progress and promotion to be sure, particularly with
respect to the scope of assessment information used to inform judgments about residents’ annual
progress. However, these remain an annual event as opposed to occurring more frequently,
which is particularly problematic for more junior residents. This must change.

There are two problems associated with the current structure. First, annual reviews are
too infrequent to support residents in developing their abilities to formulate qualitative judgments
about their learning progress. Residents need support learning how to analyze and triangulate
evidence of their learning, recognize emergent performance patterns, and construct their learner
profiles. Furthermore, they require guidance in identifying learning needs and developing
professional learning plans. Annual reviews simply will not provide the kind of sustained support
required to foster these activities.

Second, the current structure binds us to a time-based model of residency education with
the idea that promotion (e.g., PGY1, 2, 3, 4) happens on July 1st each year. This model provides
only one annual window of opportunity for a resident to be recognized to have attained sufficient competence to assume more professional responsibility. In effect, this model is based on behaviourist concepts of learning (Driscoll, 2005). The underlying assumptions are that all residents enter residency at the same point in development and progress through residency at the same rate. This model is completely at odds with current conceptions of competency-based medical education (Jason R. Frank, et al., 2010; Iobst, et al., 2010; Snell & Frank, 2010) and contemporary understandings about learning and growth (Bransford, et al., 2000; Brown, et al., 1989; Butler & Winne, 1995; Winne & Hadwin, 2010). It is contrary to the culture of autonomy-oriented assessment that we are working so hard to achieve.

Ideally, review processes would be framed in the Vygotskian notion of the Zone of Proximal Development (Vygotsky, 1978). Optimally, these processes would be established as a graduated resident-driven reporting structure whereby residents take ever-increasing responsibility for constructing their learner profiles and developing professional learning plans. As evidence of their competence grows they could present arguments for promotion with reference to their learner profiles.

This multi-year process could culminate in senior residents presenting formal arguments for their readiness to enter independent practice at departmental RPC meetings. This could look much like the process undertaken in a doctoral oral candidacy examination, except the focus of examination would be residents’ learning profiles. Through such a process residents and members of the RPC could engage in collaborative reflective processes both about residents’ readiness and program performance more generally. Obviously, these are only the beginning muses about how we can better align progress and promotion processes with our emerging
autonomy-oriented assessment culture. But they serve to underline the iterative process of this cultural shift.

Summary

Over a three year period, educational leaders in the Department of Pediatrics recognized the need to change their orientation to assessment, re-conceptualized and realized change in assessment practices, and engaged and empowered the community to support the shift in assessment culture. Having navigated this first phase of change we can reflect upon the process and identify key elements that have contributed to our success so far.

The first of these is leadership. This began at the departmental level in the RPC and was handed off to Dr. Connelly and his committee who then developed a conceptual map for change and ways to operationalize it in the form of EPA rubrics described as the first watershed event. However, it was Dr. Acker, the newly appointed Program Director and former member of PAC who brought innovation into the hands of the community and fostered ownership and up-take among residents and clinical teachers. The second key element is the Pediatric community, a group of engaged professionals open to seeing residency education happen differently. Specifically, the Pediatric community adopted a resident-driven assessment model described as the second watershed event. The third key element that pushed this process forward was the refining of the Rubric Descriptor Bank described as the third watershed event. This resource has become the corner stone of our assessment system and makes the standards of performance explicit for residents and clinical teachers. It is a living document that embodies the fundamental value of growth as it expands to accommodate an ever-increasing range of resident learning activity.
As we celebrate the progress made, we are cognizant of the need for further change, specifically in the area of progress and promotion. No doubt, re-envisioning those processes will raise possibilities about how the certification process might evolve, particularly with respect to the OSCE certification examination. Three questions that come immediately to mind are: Could learner profiles based on patterns of performance derived from triangulated assessment information prove more powerful evidence of physician competence than one very expensive, nationally administered, standardized examination? Second, what difference would a portion of funding currently allocated to OSCE certification examinations make in the quality of the clinical learning environment? Lastly, could that money be redirected to hire more clinical teachers to support high quality assessment in situ and formally recognize the significant contribution made by clinical teachers currently in place?

Ultimately, we are resigned to the fact that our progress is matched with new assessment challenges emerging on the horizon. Although assessment innovation in Pediatric residency education at Queen’s is exciting, in words attributed to the management philosopher Peter Drucker, “Culture eats strategy for breakfast” (Mickos, 2011). In essence and within the context of this research, this bold statement means that regardless of how valuable rubrics are as assessment tools to facilitate the effective provision of high quality feedback linked with standards of performance, without establishing a culture to support these practices – the strategy will most likely fail. Based on the theoretical framework outlined in this thesis, an autonomy-oriented assessment culture that recognizes and values the iterative nature of professional growth is crucial to both supporting current innovation and envisioning further development. Essentially, five theoretically grounded guiding principles point us back to the culture we wish to realize in practice. These five principles are the touchstones for this emergent culture and include:
1) Conceptualizing learning as a social, active process of negotiating meanings and constructing understanding based on prior knowledge and experience.

2) Focusing attention on multidimensional growth – using in this instance, the CanMEDs Roles as an overarching framework.

3) Moving away from a high-stakes orientation to assessment based on the false dichotomy between formative and summative assessment and embracing it as a tool for supporting and monitoring growth over time and across contexts.

4) Actively supporting learning strategy and assessment skill development in our learners...a concept well captured in the Scholar Role but something we need to be more purposeful support in practice.

5) Fostering a growth orientation to learning that emphasizes how this happens over time and embracing the concept of graduated autonomy by dismantling scaffolds and relinquishing control to learners gradually as they develop the skills to manage their own learning.

Fundamentally, medicine is complex and continually evolving, and requires a life long commitment to learning. We need to ensure our graduates are ready to assume this responsibility prior to entering independent practice. Autonomy-oriented assessment culture guided by the theoretically grounded principles outlined above hold promise for informing this cultural shift in residency education.
Implications

Case Study Research: A Vehicle for Change

The qualitative inquiry design frame of case study research is used to address descriptive (e.g., what) and/or explanatory (e.g., how and/or why) research questions within a naturalistic setting (Stake, 2005; Yin, 2006). Grounded in constructivist philosophy, case study research aims to co-construct understandings of people and events with research participants (Baxter & Jack, 2008; Stake). Ultimately, the strength of the case study method is the focus on examining a case in-depth within its context (Yin). These qualities make a case study approach particularly useful as a vehicle for documenting and supporting changes in assessment thinking and practices.

Furthermore, case study methodology employs a variety of research methods that produce various data sources that are triangulated with the goal of constructing robust case descriptions (Yin, 2006). Shifting assessment in residency education is a complex educational design challenge that unfolds over time and across multiple settings. Flexibility with respect to the selection and scope of methods and analytic strategies characteristic of case study methodology affords the responsiveness to assessment change that other approaches may not support.

Moreover, the use of narrative to weave understandings about a case from diverse data sources helps to make findings accessible (Flyvbjerg, 2006). Communicating the complexity of change to others can be difficult, particularly with respect to how findings might transfer across contexts. The richness of narrative approaches to reporting can function as a bridge to support this translational process. Ultimately, this can facilitate readers’ efforts to identify parallels between the contextual constraints and opportunities associated with this particular case and their own educational contexts.
One final aspect of case methodology that makes it a valuable vehicle to support change in assessment thinking and practices is its ability to contribute to theory development. First, the iterative quality of the case study approach aligns well with the emergent process of assessment change that is frequently unpredictable. It provides a structure for exploring and understanding that emergence over time in relation to theories that inform that change. Second, as the population of assessment change cases accumulate, cross-case analysis can be performed to draw out transferable trends that can inform practices beyond local case contexts (Stake, 2005).

**The Role of an Assessment Scientist**

Throughout this research process I looked to participatory forms of research to guide my thinking about my collaboration with members of the Pediatric community. One concept I found particularly valuable was that of researcher positionality. According to Herr and Anderson (2005), researcher positionality can be understood to unfold along a continuum. At one pole, the insider – for example a practitioner examining his/her own practice. The opposing pole being outsider initiated research, often associated with academic researchers entering a field of practice to conduct research. Herr and Anderson write about the various permutations between these poles and identify reciprocal collaboration as the ideal form of participatory research. They characterize this form as interdisciplinary partnerships – a concept that resonated with me in relation to my experience with members of the Pediatric community.

According to Herr & Anderson this ideal is rarely achieved in practice because of the time required to establish such high quality relationships. Fundamentally, it takes time for outsiders to become familiar enough with the research context to usefully contribute and for insiders to recognize the potential contribution an outsider can make. A second factor they identify as crucial to productive interdisciplinary partnerships is the outsider’s orientation.
Specifically, they underline the need to adopt a humble stance – recognizing that all participants bring important understandings to the research endeavor and fostering a mutual respect for that specialized expertise.

I describe my relationship with members of the Pediatric community as reciprocal collaboration that emerged over a prolonged period of time and continues to evolve. Fostering a culture of assessment change is challenging even when members of a community in which that change is being fostered are receptive to it.

At a fundamental level, an assessment scientist must be willing to invest the time and energy required to develop an in-depth understanding of the educational context, including the logistical constraints and opportunities that characterize that setting. This kind of understanding is not easily achieved without the active support of insiders. Establishing partnerships with individuals across the spectrum of stakeholders is crucial in this respect (e.g., clinical teachers and residents). Demonstrating a genuine interest in the diversity of perspectives these different groups provide, linking them with educational theory, and sharing those understandings to promote more complex interpretations of the educational context are an important contribution an outsider can make.

Beyond developing and promoting rich contextual understandings, an assessment scientist should assume a humble, flexible stance to foster the kinds of partnership that will optimize the knowledge and understandings that both parties bring to the change effort. The adoption of this orientation signals to partners a readiness to learn what is unique about a particular learning context in which change is being sought. This also helps make explicit the kind of adaptive process that translating educational theory into practice requires, ultimately leaving open the possibility of adjusting or even abandoning an approach should the need arise.
Such a stance functions to preserve the credibility of an assessment scientist with partners who may be unfamiliar with a more principle-centered approach to change.

As the partnership evolves and progress towards change is made it is important to identify and celebrate markers of progress. Just as we were able to identify watershed events in our progress towards an autonomy-oriented assessment culture in this project, assessment scientists and research partners should facilitate the labeling of such events and promote the celebration of these over the course of a project. Ultimately, shifting an assessment culture is a longitudinal process and recognizing progress can function to sustain the kind of effort and engagement such change requires.

**Research Boundaries**

There are limitations in every study and it is the responsibility of researchers to make these explicit. The first limitation of this research is a characteristic of the case study methodology. Specifically, this research is conducted on one program and is therefore not generalizable. However, the theoretically grounded guiding principles that supported the shift towards an autonomy-oriented assessment culture in the Department of Pediatrics could inform others’ efforts to cultivate a similar shift in culture.

A second limitation of this case study is the limited number of people involved even within the Department of Pediatrics. Although most stakeholder groups within the department are represented, including residents, clinical teachers, educational leaders, and departmental leadership, more involvement would have enriched the findings. There was also strong support for the initiative from higher-level leadership, including Dr. Walker, Associate Dean of PGME, Dr. Flynn, Vice-Dean Education and Dean Reznick, who supported our initiative by inviting Dr.
ten Cate, an international scholar in EPAs to present on the topic and spend time with PAC in particular. Regretfully, none of these leaders were interviewed for this research.

The third limitation of this research is the fact that the researcher and author of the narrative account of this change process is a focal member (i.e., Assessment Specialist) of Postgraduate Medical Education in the School of Medicine where the research was conducted. I endeavored to mitigate potential bias through the triangulation of data collection methods and participants.

**Opportunities for Further Research**

The opportunities for research in the area of autonomy-oriented assessment are many. Of immediate importance are validity studies to examine the trustworthiness of decisions about residents’ performances based on assessment rubrics. Exploring for example, the construct validity of rubric attributes for the CanMEDs competencies they are intended to represent and the extent to which rubrics can discriminate differential levels of performance (e.g., junior versus more senior residents) and detect change/growth over time. But also explore in more depth residents’ and clinical teachers’ perceptions of the quality and usefulness of feedback generated with these tools. Moving beyond Pediatrics, examining how other residency programs can adapt the Rubric Descriptor Bank will be important.

Further research focusing on how autonomy-oriented assessment culture is fostered and emerges is paramount given the limitations of this single case study. For example, how do other residency programs cultivate shifts in assessment culture? Examining how autonomy-oriented assessment culture permeates the wider medical education community is also relevant.

Longer-term research should examine the impact of shifts to an autonomy-oriented assessment culture on residents’ self-assessment abilities, self-regulated learning skills
development, and ultimately examine how this preparation impacts orientations to professional learning out in practice. One interesting line of inquiry would explore residents’ developmental trajectories in comparison to benchmarking based on training models with traditional summative-based assessment cultures. Specifically, do developmental trajectories of residents situated in autonomy-oriented assessment cultures differ from those in summative-based assessment cultures? Finally, exploring time to completion across these different assessment cultures would be interesting.

**Conclusion**

Re-conceptualizing assessment in residency education and cultivating a culture that recognizes and values the iterative nature of professional growth is a formidable task. It required us to find innovative solutions to complex assessment challenges and confront and shift long established traditions. This case study examined how the Department of Pediatrics at Queen’s University undertook that challenge through the lens of three watershed events.

The first watershed event was the development of an assessment system to capture residents’ performance in the Pediatric clinical learning environment. This was accomplished through the 5 phase process of: (a) negotiating the meaning of EPAs in Pediatrics and establishing enabling constraints, (b) operationalizing EPA in the form of rubrics, (c) goal setting and buy-in seeking, (d) collaboratively refining the focus of assessment and negotiating standards of performance, and (e) developing an implementation plan.

The second watershed event explored residents’ experiences with using EPA Rubrics in the NICU. This confirmed the value of the new assessment tools and fostered the adoption of a resident-driven assessment model in the department.
The third watershed event was the refining of the Rubric Descriptor Bank that is the product of three years of collaborative work. The bank has engaged Pediatric residents and clinical teachers to take ownership of and further shape their assessment system. It has fostered a multi-program validity study and spawned several other promising research and practice initiative. Effectively, the bank is moving Pediatric assessment innovation into the wider PGME community at Queen’s.

Together all of this work has established the foundation for a robust competency-based orientation to assessment and fostered the emergency of an autonomy-oriented assessment culture in the Department of Pediatrics.

Since identifying the need for innovation in assessment almost 3 years ago, the Department of Pediatrics at Queen’s University has worked tirelessly to realize change. The initiative could easily have lost momentum upon receiving our very successful accreditation results at the end of our cycle in October 2011. But that did not happen. In fact, the positive review and strong program leadership supported further growth. We continue our process from a position of strength, guided by the theoretically grounded principles that inform our emergent autonomy-oriented assessment culture.
REFERENCES


RCPSC (2006). General Information Concerning Accreditation of Residency Programs


ten Cate, & Scheele, F. (2007). Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Academic Medicine, 82(6), 542-547.


development. *Best Practice & Research Clinical Obstetrics & Gynaecology, 24*(6), 703-719.


### APPENDIX A

**List of Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>CanMEDs</td>
<td>The CanMEDs Physician Competency Framework</td>
</tr>
<tr>
<td>CaRMs</td>
<td>Canadian Resident Matching Service</td>
</tr>
<tr>
<td>CFPC</td>
<td>College of Family Physicians of Canada</td>
</tr>
<tr>
<td>CMQ</td>
<td>Collège des médecins du Québec</td>
</tr>
<tr>
<td>COPC</td>
<td>Children’s Outpatient Center</td>
</tr>
<tr>
<td>EPA</td>
<td>Entrustable Professional Activities</td>
</tr>
<tr>
<td>FA</td>
<td>Formative Assessment</td>
</tr>
<tr>
<td>HDH</td>
<td>Hotel Dieu Hospital</td>
</tr>
<tr>
<td>ITER</td>
<td>In-Training Evaluation Report</td>
</tr>
<tr>
<td>KGH</td>
<td>Kingston General Hospital</td>
</tr>
<tr>
<td>LLL</td>
<td>Life-long Learning</td>
</tr>
<tr>
<td>NI</td>
<td>Narrative Inquiry</td>
</tr>
<tr>
<td>NICU</td>
<td>Neonatal Intensive Care Unit</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective Structured Clinical Examination</td>
</tr>
<tr>
<td>PAC</td>
<td>Pediatric Assessment Committee</td>
</tr>
<tr>
<td>PGME</td>
<td>Postgraduate Medical Education</td>
</tr>
<tr>
<td>RCPSC</td>
<td>Royal College of Physicians and Surgeon of Canada</td>
</tr>
<tr>
<td>RPC</td>
<td>Residency Program Committee</td>
</tr>
<tr>
<td>SRL</td>
<td>Self-regulated Learning</td>
</tr>
<tr>
<td>UG</td>
<td>Undergraduate</td>
</tr>
</tbody>
</table>
## APPENDIX B

### Pediatric Data Catalog

<table>
<thead>
<tr>
<th>Date</th>
<th>Event</th>
<th>Outcome</th>
<th>Data source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct. 2009</td>
<td>Vice-Dean’s Meeting</td>
<td>Met Dr. C</td>
<td>Recollection</td>
</tr>
<tr>
<td>Oct 8 2009</td>
<td>Assessment Workshop</td>
<td>EPAs are discussed and Dr. C suggests meeting to discuss his assessment needs</td>
<td>Recollections</td>
</tr>
<tr>
<td>Nov 10 2009</td>
<td>Initial meeting with Dr. C</td>
<td>Plan to attend December Committee meeting</td>
<td>FN</td>
</tr>
<tr>
<td>Dec 1 2009</td>
<td>Pediatric Assessment Committee meeting</td>
<td>Introduction to group and presentation of components of an assessment system</td>
<td>Dr. C’s email to committee, My reflections about how naive I was – meeting with Lyn</td>
</tr>
<tr>
<td>Dec 16 2009</td>
<td>Report to RPC</td>
<td>Sub-committee progress report</td>
<td>Dr. C’s email to RPC</td>
</tr>
<tr>
<td>Jan 10 2010</td>
<td>SEAMO funding announcement</td>
<td>Decision to include a research component in EPA project</td>
<td>Email exchange with Dr. C regarding funding opportunities</td>
</tr>
<tr>
<td>Jan 19 2010</td>
<td>Meeting to discuss funding opportunity</td>
<td>Distribution of responsibilities &amp;</td>
<td>Dr. C’s email up-date to committee members</td>
</tr>
<tr>
<td>Jan 20- Feb 29 2010</td>
<td>Preparation of funding and ethics proposals</td>
<td>Ethics approval &amp; Funding</td>
<td>Email feedback from Dr. C on Proposals</td>
</tr>
<tr>
<td>Feb 29 2010</td>
<td>Submitted</td>
<td>Ethics approval &amp; Funding</td>
<td>N/A</td>
</tr>
<tr>
<td>Apr 1 2010</td>
<td>Committee meeting</td>
<td>Announce funding</td>
<td>Minutes</td>
</tr>
<tr>
<td>Apr 1-6 2010</td>
<td>Post-meeting email exchange</td>
<td>Revealed the struggles committee members were having in operationalizing EPAs for PEDs</td>
<td>Email between committee members</td>
</tr>
<tr>
<td>Apr 7 2010</td>
<td>Recreating product of committee meeting</td>
<td>Inspiration doc of EPA mapping exercise</td>
<td>Artifact: EPA Concept map</td>
</tr>
<tr>
<td>Apr 15 2010</td>
<td>Post committee meeting</td>
<td>List of 6 potential EPA activities</td>
<td>Email exchange with Dr. C</td>
</tr>
<tr>
<td>Apr 20 2010</td>
<td>Presentation at Pediatric Ground Rounds</td>
<td>Introduce concept of EPAs to Pediatric Faculty &amp; Residents</td>
<td>Minutes</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Outcome</td>
<td>Data source</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Apr 26 2010</td>
<td>Tour with resident committee member – Dr. L</td>
<td>Better understanding of educational landscape</td>
<td>FN recorded during ‘Rounds’ and recollections</td>
</tr>
<tr>
<td>Apr 29 2010</td>
<td>Committee meeting</td>
<td>Agree to define rotation specific EPAs</td>
<td>Minutes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Committee members tasked with defining EPAs for assigned rotations</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>&amp; mapping CanMEDs enabling competencies</td>
<td></td>
</tr>
<tr>
<td>Apr 29 2010</td>
<td>Post-meeting email</td>
<td>Dr. C shares mapping matrix</td>
<td>Artifact: NICU – PGY1, 2 &amp; 3</td>
</tr>
<tr>
<td>May 3 2010</td>
<td>First Rubric</td>
<td>Dr. Z defines levels of performance for common problems EPA</td>
<td>Email exchange between Drs Z &amp; C and me.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Artifact: Common Problems Rubric</td>
</tr>
<tr>
<td>May 5 2010</td>
<td>Revised Rubric</td>
<td>Dr. C incorporates mapping into rubric &amp; includes list of presentations/New Template shared</td>
<td>Email to Committee members</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Artifact: Revised Rubric Template</td>
</tr>
<tr>
<td>May 6 2010</td>
<td>Committee meeting</td>
<td>First EPA rubric is discussed, a working definition of levels of</td>
<td>FN: Dr. J’s definition of three levels of</td>
</tr>
<tr>
<td></td>
<td></td>
<td>performance is defined</td>
<td>performance</td>
</tr>
<tr>
<td>May 6 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifacts: NICU Discharge &amp; Exam rubric, Hem/Onc &amp; RESP lists of EPAs</td>
</tr>
<tr>
<td>May 18 2010</td>
<td>Ottawa Conference in Miami</td>
<td>Dinner with Dr. ten Cate: discussion of future of MedED &amp; our project.</td>
<td>FN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suggests changing label levels to: Full supervision, Partial Supervision,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Supervision on Demand</td>
<td></td>
</tr>
<tr>
<td>May 20 2010</td>
<td>Committee meeting</td>
<td>Dr. C distills 3 categories of presentations from RCPSC OofT-Chronic Outpatients, New Consults, Chronic Conditions</td>
<td>Minutes</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Outcome</td>
<td>Data source</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>May 27 2010</td>
<td>Committee meeting</td>
<td>Discuss piloting EPA rubrics &amp; getting buy-in from rotation supervisors</td>
<td>Minutes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rubric developer to act as leads in consulting with rotation supervisors.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Deadline of June 10th set for completion of Cardio, PGY1 Neonatal, Resp, Hem/Onc</td>
<td></td>
</tr>
<tr>
<td>May 27 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifact: Draft NICU resus rubric</td>
</tr>
<tr>
<td>Unsure of exact date</td>
<td>Dr. C’s meeting with Rotation supervisors (NICU)</td>
<td>Elicits suggested changes &amp; buy-in</td>
<td>Audio tape of meeting</td>
</tr>
<tr>
<td>Jun 3 2010</td>
<td>Committee meeting</td>
<td>Review progress: Hem/Onc EPAs need to be collapsed – focus on ‘core activities’ as opposed to diagnoses/what’s wrong Idea of performance patterns &amp; utility of EPAs for residents is discussed Need to consider how data gathered with rubrics will be aggregated.</td>
<td>Minutes</td>
</tr>
<tr>
<td>Jun 8 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifact: NICU Fluid M &amp; uncomplicated admission rubrics</td>
</tr>
<tr>
<td>Jun 10 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifact: NICU common problems, topic presentation; Revised Hem/Onc List of EPAs &amp; 6 rubrics; Cardio list of EPAs</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Outcome</td>
<td>Data source</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Jun 10 2010</td>
<td>Committee meeting</td>
<td>Review of generic rubric: discussed use of the word ‘miss’ in EPAs. Specifically what was meant by this: “didn’t do it” (an issue of performance), “looked at it: didn’t know it” (an issue of recognition), “looked at it: knew what it meant; but misinterpreted it (an issue of interpretation) Meaning of levels is revisited Dr. C’s recognition of transferability of descriptors across EPAs (e.g., collaborator). Guiding question in determining an EPA: “what can I do 100% of the time when I see them (residents) in the rotation?” Explored ideas about technology</td>
<td>Minutes Dr. C’s 2\textsuperscript{nd} report to RPC</td>
</tr>
<tr>
<td>Jun 10-17 2010</td>
<td>Generic Rubric review</td>
<td>Iterations of Generic Rubric</td>
<td>Artifacts: Generic V1, 2, 3, &amp; 4</td>
</tr>
<tr>
<td>Jun 16 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifact: Resp system skills rubric</td>
</tr>
<tr>
<td>Jun 17 2010</td>
<td>Committee meeting</td>
<td>Revisited levels: what distinguishes these and what this looks like in the context of CanMEDs Roles other then Med Expert.</td>
<td>Minutes</td>
</tr>
<tr>
<td>Jun 17 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifacts: Resp revised common problems &amp; common diseases; Cardio test/procedures &amp; common problems rubric</td>
</tr>
<tr>
<td>Jun 18 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifacts: NICU revised exam, Cardio revised test/procedures</td>
</tr>
<tr>
<td>Jun 24 2010</td>
<td>Committee meeting</td>
<td>Logistics for piloting: rip sheets</td>
<td>Minutes</td>
</tr>
<tr>
<td>Jun 24 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifacts: RESP revised system skills</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Outcome</td>
<td>Data source</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Jun 30 2010</td>
<td>Committee meeting</td>
<td>Launch date for NICU set for September 2010</td>
<td>Minutes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avoid the use of ‘Pilot’ use “Stages of development and Phases of implementation”</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Edits to Resp: procedural skills rubric reviewed.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Cardio review</strong> (Dr. J with Dr. F)- Acute &amp; chronic can be same as RESP – cardio test EPA added</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clarification about Level 3: competent general community pediatrician</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expectation of 1 EPA per week proposed</td>
<td></td>
</tr>
<tr>
<td>Jul 8 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifact: Cardio revised test/procedures &amp; common P</td>
</tr>
<tr>
<td>Jul 15 2010</td>
<td>Excel database Drafted</td>
<td>Descriptors from all rubrics are organized by CanMEDs (individual sheets)</td>
<td>Artifact: First Excel</td>
</tr>
<tr>
<td>Jul 18 2010</td>
<td>Upload to MedTECH</td>
<td></td>
<td>Artifact: Cardio revised test/procedures</td>
</tr>
<tr>
<td>Jul 29 2010</td>
<td>Committee meeting</td>
<td>Logistics &amp; formatting (Binders &amp; rip sheets)</td>
<td>Minutes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Professionalism: consult with Dr. A....Level 1: no, Level 2: yes, Level 3: facilitating others’</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Outcome</td>
<td>Data source</td>
</tr>
<tr>
<td>--------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sep 28 2010</td>
<td>7 - NICU EPA Rubrics launched</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sep 30 2010</td>
<td>2 NICU rubrics completed with resident</td>
<td></td>
<td>Artifacts: Common P, Fluid M</td>
</tr>
<tr>
<td>Oct 2 2010</td>
<td>1 NICU rubric completed resident</td>
<td></td>
<td>Artifact: Resus</td>
</tr>
<tr>
<td>Oct 8 2010</td>
<td>1 NICU rubric completed resident</td>
<td></td>
<td>Artifact: Exam</td>
</tr>
<tr>
<td>Oct 21 2010</td>
<td>1 NICU rubric completed resident</td>
<td></td>
<td>Artifact: Resus</td>
</tr>
<tr>
<td>Nov 10 2010</td>
<td>1 NICU rubric completed resident</td>
<td></td>
<td>Artifact: Fluid M</td>
</tr>
<tr>
<td>Nov 11 2010</td>
<td>3 NICU rubric completed resident</td>
<td></td>
<td>Artifacts: Uncomplicated admission, Discharge, Exam</td>
</tr>
<tr>
<td>Nov 29 2010</td>
<td>1 NICU rubric completed resident</td>
<td></td>
<td>Artifact: Common P</td>
</tr>
<tr>
<td>Nov 2010</td>
<td>2 – Ward EPA launched</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dec 2 2010</td>
<td>1 NICU rubric completed resident</td>
<td></td>
<td>Artifact: Common P</td>
</tr>
<tr>
<td>Jan 2011</td>
<td>3 – Cardio &amp; Resp EPAs launched</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apr 8, 12, &amp; 13 2011</td>
<td>Data collection: interviews with 3 NICU residents &amp; faculty</td>
<td>EPA feedback rubrics support feedback process &amp; residents’ readiness to assume more active responsibility for seeking feedback</td>
<td>Transcripts of interviews with Faculty &amp; 3-PGY1 Residents</td>
</tr>
<tr>
<td>Jul 22 2011</td>
<td>Resident Academic ½ day</td>
<td>Introduction of resident-driven assessment model. Feedback from residents.</td>
<td>Report prepared for Program Director</td>
</tr>
<tr>
<td>Jun 28 – Jul 7 2011</td>
<td>Excel workbook up-dated</td>
<td>All rubrics entered by research assistant</td>
<td></td>
</tr>
<tr>
<td>Aug 30 2011</td>
<td>Pediatric Program Retreat</td>
<td>Introduction of resident-driven assessment process</td>
<td>Meeting agenda</td>
</tr>
<tr>
<td>Sep 2 2011</td>
<td>Future of Medical Education Town Hall meeting</td>
<td>Associate Dean of UGME quote “having learners slows you down”</td>
<td>FN</td>
</tr>
<tr>
<td>Sep 30 2011</td>
<td>Meeting</td>
<td>Dr. G’s comments she likes resident learning cycle model</td>
<td>FN</td>
</tr>
<tr>
<td>Date</td>
<td>Event</td>
<td>Outcome</td>
<td>Data source</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
</tr>
<tr>
<td>Jan 16 2012</td>
<td>Dr. C’s response to Chapter 1</td>
<td>Quote about the authenticity of chapter</td>
<td>Email exchange</td>
</tr>
<tr>
<td>May 1 2012</td>
<td>Faculty development meeting</td>
<td>Dr. F’s positive feedback about learning cycle model</td>
<td>FN</td>
</tr>
<tr>
<td>Jun 8 2012</td>
<td>Reviewing</td>
<td>Reflections about Dr. C’s and my continuing collaborative process</td>
<td>DJ</td>
</tr>
<tr>
<td>Jun 19 2012</td>
<td>Pediatric Program Retreat</td>
<td>Sharing curriculum mapping process and assessment system developments</td>
<td>Summary of curriculum mapping process and assessment system</td>
</tr>
<tr>
<td>Jul 7 2012</td>
<td>Rubric descriptor bank up-dated and revised</td>
<td>Consolidated two and a half years of work</td>
<td>Artifact: Rubric descriptor bank</td>
</tr>
<tr>
<td>Jul 20 2012</td>
<td>Resident Academic ½ day</td>
<td>Review of resident-driven assessment system and gathering feedback from residents about system functioning and suggestions for improvement.</td>
<td>Report shared with residents and Program Director.</td>
</tr>
</tbody>
</table>
APPENDIX C

Ethics Clearance Letter

QUEEN'S UNIVERSITY HEALTH SCIENCES & AFFILIATED TEACHING HOSPITALS RESEARCH ETHICS BOARD

March 15, 2010

This Ethics Application was subject to:

☐ Full Board Review
☐ Meeting Date:  
☒ Expedited Review

Dr. Robert J. Connolly
Department of Pediatrics
Wakins 3
Kingston General Hospital

Dear Dr. Connolly,

Study Title: Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen's University
Co-Investigators: Ms. Laura April McEwen

I am writing to acknowledge receipt of your recent ethics submission. We have examined the protocol and consent form for your project (as stated above) and consider it to be ethically acceptable. This approval is valid for one year from the date of the Chair's signature below. This approval will be reported to the Research Ethics Board. Please attend carefully to the following list of ethics requirements you must fulfill over the course of your study:

› Reporting of Amendments: If there are any changes to your study (e.g. consent, protocol, study procedures, etc.), you must submit an amendment to the Research Ethics Board for approval. (see http://www.queensu.ca/vpr/reb.html).

› Reporting of Serious Adverse Events: Any unexpected serious adverse event occurring locally must be reported within 2 working days or earlier if required by the study sponsor. All other serious adverse events must be reported within 35 days after becoming aware of the information.

› Reporting of Complaints: Any complaints made by participants or persons acting on behalf of participants must be reported to the Research Ethics Board within 7 days of becoming aware of the complaint. Note: All documents supplied to participants must have the contact information for the Research Ethics Board.

› Annual Renewal: Prior to the expiration of your approval (which is one year from the date of the Chair's signature below), you will be reminded to submit your renewal form along with any new changes or amendments you wish to make to your study. If there have been no major changes to your protocol, your approval may be renewed for another year.

Yours sincerely,

Chair, Research Ethics Board

March 15, 2010

Study Code: PAED-253-10

Investigators please note that if your trial is registered by the sponsor, you must take responsibility to ensure that the registration information is accurate and complete.
APPENDIX D

Resident Consent Form

Re: Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University

Letter of Information: Resident

You are invited to participate in a project entitled "Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University". This letter provides you with background information to the project.

This project is being conducted as part of the Pediatric Assessment Committee’s evaluation of the implementation of our new assessment system in the Pediatric Residency Training Program. It is part of a larger initiative within the School of Medicine that is focused on improving the overall quality assessment. Given the significant changes in the assessment system the committee wishes to explore residents’ perceptions about how the system has impacted their learning. As such we are inviting residents to participate in an hour-long interview to review how the system is being implemented, what effect it has had on your learning and recommendations for improvement. Although all residents of the Pediatric Residency Training Program are invited to express their interest in participating in this project, the number of interviews that will be conducted is limited.

Participation in an interview will entail meeting with an interviewer. These encounters will be electronically recorded and later transcribed, however your identify will be kept confidential and your name and all identifying information will be removed from the final transcription. The results of this study will be coded in such a way that individual identity will not be attached in any way to the final data produced. There are no known risks affiliated with participation in this research. Confidentiality will be maintained throughout the study. This includes publication of the results. The data may be used for presentation as well as in similar research projects should the opportunity arise. Should such secondary uses of the data be required, confidentiality of participants will be maintained.

Your participation is completely voluntary and you are free to withdraw from the study without reason at any point. Should you choose to withdraw you may request removal of all or part of your data. Further, choosing not to participate or choosing to withdraw will have no adverse consequences to you.

By completing the attached consent form, you will be giving consent to your participation in this study and that you have understood the nature of your participation in this research. If you have any questions about this project, please contact the researchers Laura April McEwen at laura.mcewen@queensu.ca or Dr. R Connelly at connelr@KGH.KARI.NET or Dr. Albert Clark, Chair, Ethics Review Board: (613) 533-6081 or clarkaf@queensu.ca.

Sincerely,

Laura April McEwen, B.A., M.A., Ph.D. Candidate  Robert J. Connelly, MD, FRCPC
Assessment and Evaluation Consultant, Medical Director, NICU
Office of Health Sciences Education Kingston General Hospital

144
Re: Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University

Letter of Consent: Pediatric Resident

Thank you for agreeing to participate in this project entitled "Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University". This letter provides you with background information to the project.

This project is being conducted as part of the Pediatric Assessment Committee’s evaluation of the implementation of our new assessment system in the Pediatric Residency Training Program. It is part of a larger initiative within the School of Medicine that is focused on improving the overall quality assessment. Given the significant changes in the assessment system the committee wishes to examine how the system is being implemented, what effect it has had of learning and collect feedback about ways in which it can be improved to better meet your needs. As such we are inviting residents to participate in an hour-long interview to review how the system is being implemented, the effect it has had on your learning and recommendations for improvements. Given the important role of assessment in supporting resident learning the committee wishes to ensure that residents have input in shaping the way the new system is designed. As such we are inviting residents to participate in an interview to review prototype assessment tools and discuss assessment processes.

With your approval, this interview will be electronically recorded and later transcribed. Confidentiality will be maintained throughout the project in that numeric codes will be substituted for names in all transcripts.

Your signature below acknowledges that you have volunteered to participate as a subject in this investigation. Prior to providing consent please read the following statement regarding your participation in the study.

• This study explores resident’s experiences with the new assessment system.
• Involvement in this research will require NO additional work beyond participation in this one-hour focus group.
• Participation in this study is voluntary and may be terminated at any time by request.
• Participation in this project and/or withdrawal from this project will not adversely affect you in any way.
• This study will not involve any greater risks than those ordinarily occurring in daily life. It is not possible to identify all potential risks in any procedure, but that all responsible safeguards have been taken to minimize potential risks.
• The results if this research may be published or reported to government agencies, funding agencies, or scientific groups, but your name will not be associated in any way with any published results.
• The purposes and procedures of the study have been satisfactorily explained.
• I have read and kept a copy of the letter of information. I read this letter of consent prior to signing it.

Name: ___________________________ Signature: ___________________________ Date: ____________

In the event that you have any questions, please feel free to contact the researchers, Laura McEwen at laura.mcewen@queensu.ca or Dr. R Connelly at connellr@KGH.KARI.NET

In the event that you have any questions, concerns or complaints, please feel free to contact: Dr. Albert Clark, Chair, Ethics Review Board, (613) 533-6081 or clarkaf@queensu.ca

145
APPENDIX E

Research Partner Consent

Re: Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen's University

Letter of Information: Pediatric Assessment Committee Member

You are invited to participate in a project entitled "Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University." This letter provides you with background information to the project.

This project is being conducted as part of the Pediatric Student Assessment Committee’s evaluation of the development of an assessment system for Pediatric Residency Education. It is part of a larger initiative within the School of Medicine that is focused on improving the overall quality assessment.

As a member of the Pediatric Assessment Committee your participation will entail providing consent to use archival data including: (a) minutes of meeting, (b) artifacts generated by the group (e.g., assessment tools) and (c) email exchanges between committee members to describe the Pediatric Student Assessment Committee’s developmental process for the purpose of Laura April McEwen’s Dissteration research. As a active collaborator in this research project the opportunity to review in advance all publications of results will be afforded to you. Furthermore, your identity will be revealed in publications of results in order to formally acknowledge your significant contribution to this collaborative research effort unless you specifically request otherwise.

There are no known risks affiliated with participation in this research.

Your participation is completely voluntary and you are free to withdraw from the study without reason at any point. Should you choose to withdraw you may request removal of all or part of your data. Further, choosing not to participate or choosing to withdraw will have no adverse consequences to you.

By completing the attached consent form, you will be giving consent to your participation in this study and that you have understood the nature of your participation in this research. If you have any questions about this study, please contact the researcher Laura April McEwen at laura.mcewen@queensu.ca. If you have any concerns about your rights as a research participant please contact Dr. Albert Clark, Chair, Ethics Review Board: (613) 533-6081 or clarkaf@queensu.ca.

Sincerely,

Laura April McEwen, B.A., M.A., Ph.D. Candidate
Assessment and Evaluation Specialist,
Postgraduate Medical Education, Queen’s University
Re: Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University

Letter of Consent: Members of the Pediatric Student Assessment Committee

Thank you for agreeing to participate in this project entitled "Design and Evaluation of a Theory Based Assessment System in Pediatric Resident Training at Queen’s University". This letter provides you with background information to the project.

This project is being conducted as part of the Pediatric Student Assessment Committee’s evaluation of the development of an assessment system for the Pediatric Residency Education. It is part of a larger initiative within the School of Medicine that is focused on improving the overall quality assessment.

As a member of the Pediatric Student Assessment Committee your participation will entail providing consent to use archival data including: (a) minutes of meeting, (b) artifacts generated by the group (e.g., assessment tools) and (c) email exchanges between committee members to describe the Pediatric Student Assessment Committee’s developmental process for the purpose of Laura April McEwen’s Dissertation research. As a active collaborator in this research project the opportunity to review in advance all publications of results will be afforded to you. Furthermore, your identity will be revealed in publications of results in order to formally acknowledge your significant contribution to this collaborative research effort unless you specifically request otherwise.

Your signature below acknowledges that you have volunteered to participate as a research collaborator in this investigation. Prior to providing consent please read the following statement regarding your participation in the study.

- This study documents the efforts of Pediatric Student Assessment Committee to develop an assessment system for Pediatric Residency Education.
- Involvement in this research will require NO additional work beyond participation in the activities of Pediatric Student Assessment Committee.
- Participation in this study is voluntary and may be terminated at any time by request.
- Participation in this project and/or withdrawal from this project will not adversely affect you in any way.
- This study will not involve any greater risks than those ordinarily occurring in daily life. It is not possible to identify all potential risks in any procedure, but that all responsible safeguards have been taken to minimize potential risks.
- The results of this research may be published or reported to government agencies, funding agencies, or scientific groups, and your name will be associated with published results unless you specifically request otherwise.
- The purposes and procedures of the study have been satisfactorily explained.
- I have read and kept a copy of the letter of information. I read this letter of consent prior to signing it.

Name: ______________________________ Signature: __________________ Date: ______

If you have any questions about this study, please contact the researcher Laura April McEwen at laura.mcewen@queensu.ca. If you have any concerns about your rights as a research participant please contact Dr. Albert Clark, Chair, Ethics Review Board: (613) 533-6081 or clarkaf@queensu.ca.
# APPENDIX F

## Rubric Descriptor Bank (limited edition)

### Medical Expert

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>History-generic</td>
<td>EC Rubrics</td>
<td>Misses basic relevant info OR Overly detailed</td>
<td>Focused and concise</td>
<td>Identifies pertinent risk factors and acquires details, seeking corroborative info as required</td>
</tr>
<tr>
<td>Physical1</td>
<td>EC Rubrics</td>
<td>Omits basic Px maneuvers.</td>
<td>Performs basic Px maneuvers</td>
<td>Performs complete relevant Px</td>
</tr>
<tr>
<td>Physical2</td>
<td>EC Rubrics</td>
<td>Misinterprets physical findings.</td>
<td>Identifies findings relevant to problem formulation.</td>
<td>Identifies signs and symptoms and integrates their relevance.</td>
</tr>
<tr>
<td>Problem Formulation</td>
<td>EC Rubrics, SCC</td>
<td>Limited differential without prioritizing</td>
<td>Correct differential with prioritizing for common cases</td>
<td>Correct differential including plausible rarer items and prioritized for complex</td>
</tr>
<tr>
<td>Investigations1</td>
<td>EC Rubrics, SCC</td>
<td>Proposes irrelevant/incorrect investigations.</td>
<td>Identifies investigations, but use is indiscriminant.</td>
<td>Strategic use of investigations (e.g., justifiable cost/benefit)</td>
</tr>
<tr>
<td>Investigations2</td>
<td>EC Rubrics, SCC</td>
<td>Misinterprets results</td>
<td>Correctly interprets results</td>
<td>Results of investigations inform management (brings it all together, makes sense of all info)</td>
</tr>
<tr>
<td>Management1a</td>
<td>EC Rubrics, SCC</td>
<td>Proposes incorrect treatment OR Provides inadequate management plan</td>
<td>Manages uncomplicated/commonly encountered diagnoses.</td>
<td>Identifies &amp; manages treatment for complicated/less common diagnoses.</td>
</tr>
<tr>
<td>Management1b</td>
<td>EC Rubrics, SCC</td>
<td>Does not monitor response to treatment</td>
<td>Monitors response to treatment but unsure how to adjust.</td>
<td>Integrates results of follow-up investigations</td>
</tr>
</tbody>
</table>

### Communicator

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc/Patient 1a</td>
<td>Patient Feedback Rubric (PFR)</td>
<td>Dismissive, abrupt and/or rude</td>
<td>Listened to what I had to say</td>
<td>Asked questions about what I said</td>
</tr>
<tr>
<td>Doc/Patient 1b</td>
<td>PFR</td>
<td>Disregarded my beliefs and concerns</td>
<td>Respected my beliefs and concerns</td>
<td>Discussed my expectations</td>
</tr>
<tr>
<td>Doc/Patient 1c</td>
<td>PFR</td>
<td>Used language I did not understand</td>
<td>Used language I understood</td>
<td>Asked question to checked that I understood</td>
</tr>
<tr>
<td>Communicator</td>
<td>Doc/Patient 1d</td>
<td>PFR</td>
<td>Did not describe test procedures</td>
<td>Described test procedure</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td>-----</td>
<td>---------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Doc/Patient 1e</td>
<td>PFR</td>
<td>Did not explain test results</td>
<td>Explained test results</td>
<td>Checked that I understood what the results meant</td>
</tr>
<tr>
<td>Doc/Patient-PEDs</td>
<td>PFR</td>
<td>Did not connect with my child.</td>
<td>Tried to interact with my child in an appropriate way</td>
<td>Involved me to help my child feel more comfortable</td>
</tr>
<tr>
<td>Doc/Patient 1f</td>
<td>PFR</td>
<td>Was unclear about when/if we needed to return</td>
<td>Explained when we should return and who we should see</td>
<td>Checked that I understood how to arrange appointment</td>
</tr>
<tr>
<td>Doc/Patient 2a</td>
<td>Multisource Feedback Rubric (MSF)</td>
<td>Does not respond to patient’s needs for comfort and support</td>
<td>Responds to patient’s needs for comfort and support</td>
<td>Anticipates patient’s needs and plans accordingly</td>
</tr>
<tr>
<td>Doc/Patient 2b</td>
<td>MSF</td>
<td>Does not respond to family’s need for info and/or support</td>
<td>Discusses initial plan of care with family</td>
<td>Provides on-going info and support to family.</td>
</tr>
<tr>
<td>Doc/Patient 2c</td>
<td>MSF</td>
<td>Ignores rights and choices of patient when planning care</td>
<td>Rights and choices of patient are usually accommodated</td>
<td>Rights and choices of patient guide planning of care</td>
</tr>
<tr>
<td>Doc/Patient 2d</td>
<td>MSF</td>
<td>Overwhelmed by demanding interpersonal situations</td>
<td>Handles demanding interpersonal situation</td>
<td>Manages demanding interpersonal situations with compassion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intraprofessional1a</td>
<td>Resident Teaching Feedback Rubric (RTF)</td>
<td>Dismissive and/or abrupt</td>
<td>Listened to what I had to say</td>
<td>Used questions to probed my thinking</td>
</tr>
<tr>
<td>Intraprofessional1b</td>
<td>RTF</td>
<td>Appeared disinterested</td>
<td>Identified my strengths and weaknesses</td>
<td>Stimulated me to explore my strengths and weaknesses</td>
</tr>
<tr>
<td>Intraprofessional1c</td>
<td>RTF</td>
<td>Defensive</td>
<td>Responded positively to questions</td>
<td>Encouraged questions</td>
</tr>
<tr>
<td>Communicator</td>
<td>Attribute</td>
<td>Source</td>
<td>Needs Attention</td>
<td>Developing</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Case reporting</td>
<td>Sr. consult clinics (SCC)/COPC jr. attending/EC Rubrics</td>
<td>Omits pertinent info</td>
<td>Presents all pertinent info</td>
</tr>
<tr>
<td></td>
<td>Documentation1</td>
<td>MSF</td>
<td>Documentation is inaccurate/incomplete</td>
<td>Documentation may be unclear</td>
</tr>
<tr>
<td></td>
<td>Documentation2</td>
<td>MSF</td>
<td>Writing is illegible</td>
<td>Writing can be difficult to read</td>
</tr>
<tr>
<td></td>
<td>Documentation3-Charting</td>
<td>SCC</td>
<td>Inaccurate OR incomplete</td>
<td>Complete &amp; accurate with use of some uncommon abbreviations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collaborator</th>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intra/Inter1a</td>
<td>MSF/RTF</td>
<td>Ignored suggestions from team</td>
<td>Receptive to feedback from team</td>
<td>Actively sought input from team</td>
</tr>
<tr>
<td></td>
<td>Intra/Inter1b</td>
<td>MSF/RTF</td>
<td>Does not acknowledge others’ learning needs</td>
<td>Identifies others’ learning needs</td>
<td>Supports others’ learning</td>
</tr>
<tr>
<td></td>
<td>Intra/Inter1c</td>
<td>MSF/RTF</td>
<td>Ignored interpersonal conflict</td>
<td>Recognized interpersonal conflict</td>
<td>Managed interpersonal conflict</td>
</tr>
<tr>
<td></td>
<td>Interprofessional</td>
<td>RTF</td>
<td>Disrespectful to learners and colleagues</td>
<td>Respected other learners and colleagues</td>
<td>Encouraged others’ respect for learners and colleagues</td>
</tr>
<tr>
<td></td>
<td>Interprofessional-other</td>
<td>RTF</td>
<td>Does not accept assistance from others when needed.</td>
<td>Accepts assistance from others when needed.</td>
<td>Leverages the strength of all team members.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Manager</th>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Responsiveness</td>
<td>MSF</td>
<td>Requires multiple pages/responds slowly</td>
<td>Responds to pager/limited delay</td>
<td>Advises team in case of delays</td>
</tr>
<tr>
<td></td>
<td>Prioritization</td>
<td>MSF</td>
<td>Ignores urgent requests</td>
<td>Usually prioritizes urgent requests</td>
<td>Prioritizes urgent requests</td>
</tr>
<tr>
<td></td>
<td>Leadership</td>
<td>MSF</td>
<td>Appears overwhelmed in an emergency</td>
<td>Responds effectively in an emergency</td>
<td>Demonstrates leadership in an emergency</td>
</tr>
</tbody>
</table>
## Manager

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisory skills</td>
<td>COPC jr. Attending (COPC-jA)</td>
<td>Appears overwhelmed by supervisory responsibilities</td>
<td>Rarely overwhelmed by supervisory responsibilities</td>
<td>Manages supervisory responsibilities effectively</td>
</tr>
<tr>
<td>Leadership</td>
<td>Inpatient Sr. card</td>
<td>Overwhelms team members</td>
<td>Delegates workload strategically</td>
<td>Inspires confidence/ Supports excellence</td>
</tr>
<tr>
<td>Time management</td>
<td>MSF/Inpatient Sr.</td>
<td>Appears overwhelmed by workload</td>
<td>Rarely overwhelmed by workload</td>
<td>Manages workload effectively</td>
</tr>
<tr>
<td>Organization</td>
<td>Inpatient Sr. card</td>
<td>Appears disorganized/distracted</td>
<td>Organized and focused</td>
<td>Supports others’ organization and focus</td>
</tr>
<tr>
<td>Problem solving</td>
<td>MSF/COPC-jA</td>
<td>Uncertain/indecisive</td>
<td>Solves problems/makes decisions with minimal delay</td>
<td>Solves problems/makes decisions as they arise</td>
</tr>
<tr>
<td>Patient Flow</td>
<td>SCC</td>
<td>Became overwhelmed quickly and fell behind</td>
<td>Saw all assigned cases - accumulating delays over time</td>
<td>Managed patient flow with minimal/justifiable delays</td>
</tr>
<tr>
<td>ER consult</td>
<td>Ward EPA</td>
<td>Has difficulty prioritizing urgency of ER referral.</td>
<td>Evaluates urgency of ER referral and responds as warranted.</td>
<td>Communicates estimated response time to ER</td>
</tr>
<tr>
<td>Resource management</td>
<td></td>
<td>Does not consider resource costs</td>
<td>Considers resource costs but knowledge about cost effectiveness may be limited</td>
<td>Utilizes resources in a cost-effective manner</td>
</tr>
<tr>
<td>Technology use</td>
<td></td>
<td></td>
<td></td>
<td>Uses information technology to optimize patient care</td>
</tr>
</tbody>
</table>

## Health Advocate

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barriers to care-Gen</td>
<td>EPAs</td>
<td>Does not consider barriers to ongoing care</td>
<td>Identifies barriers to ongoing care</td>
<td>Addresses barriers to care with suitable plan</td>
</tr>
<tr>
<td>Determinants of health</td>
<td></td>
<td>Does not identify factors affecting patient health (bio/psychosocial/economic)</td>
<td>Identifies factors affecting patient health (bio/psychosocial/economic)</td>
<td>Focuses on harm reduction strategies</td>
</tr>
<tr>
<td>Access to testing?</td>
<td></td>
<td>Does not address delays in patient access to testing</td>
<td>Determines reasons for delays in patients' access to testing</td>
<td>Works to expedite patients' access to testing as required</td>
</tr>
<tr>
<td>Test results?</td>
<td></td>
<td>Does not follow-up on lab results</td>
<td>Contacts lab for results</td>
<td>Works to expedite access to lab results as required</td>
</tr>
<tr>
<td>Health Advocate</td>
<td>Attribute</td>
<td>Source</td>
<td>Needs Attention</td>
<td>Developing</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------</td>
<td>----------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Access to services</td>
<td></td>
<td>Does not consider organizations that provide additional patient services</td>
<td>Considers other organizations that provide patient services</td>
</tr>
<tr>
<td></td>
<td>Patient/Family Resources</td>
<td>EPAs</td>
<td>Does not consider family resources that might impact management plan.</td>
<td>Considers family resources when formulating management plan.</td>
</tr>
<tr>
<td></td>
<td>Patient education</td>
<td>Clinics-RESP/Diabetes, NICU-breastfeeding</td>
<td>Does not identify educational opportunities with patients related to determinants of health</td>
<td>Identified education opportunities with patient related to determinants of health</td>
</tr>
<tr>
<td></td>
<td>Patient Education-RESP</td>
<td>Resp EPAs</td>
<td>Demonstrates incorrect technique for use.</td>
<td>Demonstrates correct technique for use to family and child (when possible)</td>
</tr>
<tr>
<td></td>
<td>Patient Education-RESP</td>
<td>Resp EPAs</td>
<td>Does not explain correct technique for use to family and child.</td>
<td>Describes overall technique for use of device.</td>
</tr>
<tr>
<td></td>
<td>Community resources</td>
<td></td>
<td>Limited knowledge of community resources</td>
<td>Demonstrates knowledge of community resources but unsure how and when to use</td>
</tr>
<tr>
<td></td>
<td>Risk assessment-PEDs</td>
<td></td>
<td>Does not identify children at risk</td>
<td>Identifies children at risk</td>
</tr>
<tr>
<td></td>
<td>Access to benefits</td>
<td>Benefit cap/restricted medication, services</td>
<td>Identifies need to intervene on behalf of patient but unclear how to proceed</td>
<td>Drafts request to provider for adjustments in coverage but may not may lack detail</td>
</tr>
<tr>
<td>Attribute</td>
<td>Source</td>
<td>Needs Attention</td>
<td>Developing</td>
<td>Achieving</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Scholar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>SCC</td>
<td>Ignored feedback. Avoided feedback. Defensive about feedback</td>
<td>Accepted feedback.</td>
<td>Sought feedback</td>
</tr>
<tr>
<td>New learning</td>
<td>SCC</td>
<td>Does not research current trends in evidence-based practices.</td>
<td>Accesses evidence-based material and incorporates into practice.</td>
<td>Questions routine practices when in conflict with current evidence</td>
</tr>
<tr>
<td>Attribute</td>
<td>Source</td>
<td>Needs Attention</td>
<td>Developing</td>
<td>Achieving</td>
</tr>
<tr>
<td>Teaching-formal</td>
<td>Presentation PR</td>
<td>Selects topic that is irrelevant to some team members.</td>
<td>Select topic of interest to all team members.</td>
<td>Tailors topic presentation to address info needs of all</td>
</tr>
<tr>
<td>Teaching-formal</td>
<td>PR</td>
<td>Demonstrates inaccurate or incomplete knowledge of topic.</td>
<td>Demonstrates knowledge of key aspects of topics.</td>
<td>Presents thorough topic summary and important, advanced differentials.</td>
</tr>
<tr>
<td>Teaching-formal</td>
<td>PR</td>
<td>Does not actively engage audience</td>
<td>Uses standard teaching strategies (e.g., open-ended questions).</td>
<td>Using alternative teaching strategies (e.g., storytelling).</td>
</tr>
<tr>
<td>Teaching-formal-research</td>
<td>PR</td>
<td>Lacks thorough investigation of chosen topic.</td>
<td>Includes classic textbook teaching of topic.</td>
<td>Incorporates up to date evidence/ journal articles/guidelines</td>
</tr>
<tr>
<td>Use of Technology</td>
<td>PR</td>
<td>Visuals detract from message.</td>
<td>Visuals relate to message.</td>
<td>Visuals enhance message.</td>
</tr>
<tr>
<td>Teaching-surgical</td>
<td>SP-EPA</td>
<td>Missed opportunities to support others’ learning</td>
<td>Responded to less experienced learners’ questions</td>
<td>Guided less experienced learners to complete tasks</td>
</tr>
</tbody>
</table>
### Scholar

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bedside Teaching</td>
<td>RTF</td>
<td>Dismissed/ignored my learning objectives</td>
<td>Demonstrated knowledge of my learning objectives</td>
<td>Discussed my learning objectives and helped me plan how to meet them</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Demonstrated how to perform skills</td>
<td>Debriefed performance after demo</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Created opportunities for me to observe him/her</td>
<td>Encouraged/facilitated my observation of others</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Provided useful feedback during or immediately after directly observing me</td>
<td>Sought opportunities for me to apply feedback provided</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Disregarded my level of experience when teaching</td>
<td>Adjusted his/her teaching to my level of experience</td>
<td>Easily accommodated learners of varying levels of experience</td>
</tr>
<tr>
<td>Bedside Teaching</td>
<td>RTF</td>
<td>Demonstrated little interest in teaching</td>
<td>Demonstrated interest in teaching</td>
<td><strong>Demonstrated enthusiasm for teaching / Engaging teacher</strong></td>
</tr>
</tbody>
</table>

### Professional

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Source</th>
<th>Needs Attention</th>
<th>Developing</th>
<th>Achieving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conduct</td>
<td>COPC-jA</td>
<td>Exhibits dishonesty, lack of commitment, respect, compassion.</td>
<td>Acts with honesty, commitment, respect, and compassion.</td>
<td>Encourages others’ honesty, commitment, respect &amp; compassion.</td>
</tr>
<tr>
<td>Modeling</td>
<td>RTF</td>
<td>Engaged in unprofessional behaviors I will avoid as a doctor</td>
<td>Demonstrated some qualities of the kind of doctor I want to be</td>
<td>Served as a role model of the kind of doctor I want to be (e.g., knowledgeable, professional, empathetic)</td>
</tr>
<tr>
<td>Limitations - a</td>
<td>MSF</td>
<td>Does not seek assistance when required</td>
<td>Aware of personal limitations</td>
<td>Seeks consultation/supervision freely</td>
</tr>
<tr>
<td>Limitations - b</td>
<td>COPC-jA, Inpatient Sr. card</td>
<td>Does not seek assistance when difficulties arise.</td>
<td>Accepts assistance when offered</td>
<td>Seeks assistance</td>
</tr>
<tr>
<td>Limitations - c</td>
<td>SCC</td>
<td>Did not seek help when required</td>
<td>Aware of personal limitations</td>
<td>Sought help/consultation when required</td>
</tr>
<tr>
<td>Patient responsibility</td>
<td>MSF</td>
<td>Avoids responsibility for patient care</td>
<td>Aware of patient care responsibilities</td>
<td>Accepts responsibility for patient care</td>
</tr>
<tr>
<td>Patient interests</td>
<td>Ward EPA</td>
<td>Disregards best interests of patient</td>
<td>Demonstrates a growing awareness of ‘best interest’ of patient</td>
<td>Decision-making is focused on ‘best interest’ of patient</td>
</tr>
<tr>
<td>Professional Attribute</td>
<td>Source</td>
<td>Needs Attention</td>
<td>Developing</td>
<td>Achieving</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----------------</td>
<td>-------------------------------------------</td>
<td>------------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
<tr>
<td>Feedback-other</td>
<td>MSF</td>
<td>Disregards feedback</td>
<td>Accepts feedback</td>
<td>Seeks feedback</td>
</tr>
<tr>
<td>Privacy &amp; confidentiality</td>
<td>MSF/RTF</td>
<td>Disregards patient privacy &amp; confidentiality</td>
<td>Respects patient privacy &amp; confidentiality</td>
<td>Encourages others’ respects for patient privacy &amp; confidentiality</td>
</tr>
<tr>
<td>Consent</td>
<td>Ward EPA</td>
<td>Does not obtain consent or assent.</td>
<td>Acquires consent with potential areas of misunderstanding.</td>
<td>Obtains informed consent/assent.</td>
</tr>
<tr>
<td>Coping</td>
<td>MSF</td>
<td>Does not remain professional in stressful situations</td>
<td>Remains professional in stressful situations</td>
<td>Is professional in stressful situations and helps others cope.</td>
</tr>
<tr>
<td>Punctuality</td>
<td></td>
<td>Chronically late/absent</td>
<td>Rarely late</td>
<td>Punctual/Rarely absent</td>
</tr>
<tr>
<td>Disclosure</td>
<td>NICU-code/PCCU</td>
<td>Withdrawn and/or evasive when faced with unexpected events or procedure failure.</td>
<td>Remains calm but struggles to explain unexpected events or procedure failure.</td>
<td>Remains calm and honest in explanation of unexpected events or procedural failure.</td>
</tr>
</tbody>
</table>